

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO
CONHECIMENTO**

**CONTROLE DE AUTORIDADE PARA ENTIDADES PESSOAS EM
REPOSITÓRIOS INSTITUCIONAIS QUE USAM O DSPACE**

Anna Carolina Leite Cota

Belo Horizonte
2021

Anna Carolina Leite Cota

**CONTROLE DE AUTORIDADE PARA ENTIDADES PESSOAS EM
REPOSITÓRIOS INSTITUCIONAIS QUE USAM O DSPACE**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento, da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, como requisito para obtenção do título de Mestre em Gestão & Organização do Conhecimento.

Área de concentração: Ciência da Informação.

Linha de Pesquisa: Arquitetura e Organização do Conhecimento

Orientadora: Prof^a. Dr^a. Benildes Coura Moreira dos Santos Maculan

Coorientador: Prof. Dr. Max Cirino de Mattos

Belo Horizonte
2021

C843c

Cota, Anna Carolina Leite.

Controle de autoridade para entidades pessoas em repositórios institucionais que usam o DSPACE [recurso eletrônico] / Anna Carolina Leite Cota. - 2021.

1 recurso eletrônico (157 f. : il., color): pdf.

Orientadora: Benildes Coura Moreira dos Santos Maculan.
Coorientador: Max Cirino de Mattos.

Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 135-150.

Apêndices: f. 151-153.

Anexo: f. 154-157.

Exigências do sistema: Adobe Acrobat Reader.

1. Ciência da Informação – Teses. 2. Repositórios institucionais – Teses. 3. Autoridade – Teses. 4. Controle de vocabulário – Teses. I. Título. II. Maculan, Benildes Coura Moreira dos Santos. III. Mattos, Max Cirino. IV. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.4



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPG-GOC

FOLHA DE APROVAÇÃO

CONTROLE DE AUTORIDADE PARA ENTIDADES PESSOAS EM REPOSITÓRIOS INSTITUCIONAIS QUE USAM O DSPACE

ANNA CAROLINA LEITE COTA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Arquitetura e Organização do Conhecimento.

Aprovada em 27 de setembro de 2021, por videoconferência, pela banca constituída pelos membros:

Prof(a). Benildes Coura Moreira dos Santos Maculan (Orientadora)
ECI/UFMG

Prof(a). Max Cirino de Mattos (Coorientador)
Skema Business School

Prof(a). Milton Shintaku
IBICT

Prof(a). Cíntia de Azevedo Lourenço
ECI/UFMG

Prof(a). Célia da Consolação Dias
ECI/UFMG

Belo Horizonte, 27 de setembro de 2021.



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 27/09/2021, às 11:39, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Milton Shintaku, Usuário Externo**, em 27/09/2021, às 15:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Max Cirino de Mattos, Usuário Externo**, em 27/09/2021, às 15:39, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cintia de Azevedo Lourenco, Professora do Magistério Superior**, em 28/09/2021, às 09:48, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Celia da Consolação Dias, Coordenador(a) de curso de pós-graduação**, em 09/11/2021, às 15:31, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0982325** e o código CRC **A92E24E8**.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPG-GOC

ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA

ANNA CAROLINA LEITE COTA

Realizou-se, no dia 27 de setembro de 2021, às 09:00 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *CONTROLE DE AUTORIDADE PARA ENTIDADES PESSOAS EM REPOSITÓRIOS INSTITUCIONAIS QUE USAM O DSPACE*, apresentada por ANNA CAROLINA LEITE COTA, número de registro 2019663435, graduada no curso de BIBLIOTECONOMIA/DIURNO, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Benildes Coura Moreira dos Santos Maculan - ECI/UFMG (Orientadora), Prof(a). Max Cirino de Mattos - Skema Business School (Coorientador), Prof(a). Milton Shintaku - IBICT, Prof(a). Cíntia de Azevedo Lourenço - ECI/UFMG, Prof(a). Célia da Consolação Dias - ECI/UFMG.

A Comissão considerou a dissertação:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 27 de setembro de 2021.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 27/09/2021, às 11:39, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Milton Shintaku, Usuário Externo**, em 27/09/2021, às 13:47, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Max Cirino de Mattos, Usuário Externo**, em 27/09/2021, às 15:39, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cintia de Azevedo Lourenco, Professora do Magistério Superior**, em 28/09/2021, às 09:48, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Celia da Consolação Dias, Coordenador(a) de curso de pós-graduação**, em 09/11/2021, às 15:31, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0982203** e o código CRC **6F1EE0C3**.

AGRADECIMENTOS

A Deus, por ter me permitido chegar até aqui. Por me dar forças e energia para persistir e continuar. Sou grata por estar aqui.

À minha orientadora, Benildes Maculan, por ter me aturado durante toda a graduação e mais ainda agora no Mestrado. Agradeço de coração pela ajuda, paciência, colaboração, cumplicidade, dedicação e compreensão. E, acima de tudo, por não ter desistido de mim! Porque eu desisti, e muitas vezes! Sou grata por ter tido você ao meu lado todo esse tempo “turbulento”. E, espero que possamos continuar nossa caminhada juntas, mas agora só como amigas e colegas de profissão. Chega de academia, pelo menos, por enquanto. Obrigada por tudo e por tanto!

À minha amiga Paula, por me incentivar a trilhar os caminhos da academia. Sem você não teria nem tentado. Agradeço pelas palavras de incentivo, pelo seu otimismo invejável e fé inabalável e por acreditar que chegaríamos ao fim de mais essa batalha. É mais uma aventura para nossa conta! E estamos só começando. “Tamo junto, Marida!”.

Ao Max pelas preciosas contribuições.

Aos meus pais por me mostrarem o valor da educação e do conhecimento.

À minha mãe por me incentivar a seguir adiante mesmo na adversidade.

Aos meus irmãos e agregados, adoro vocês!

Ao Henrique, meu afilhado, no auge dos seus 3 anos, pelas frases de incentivo: “Força Dindinda, você consegue!”, “Dindinda é forte!”.

Aos meus amigos Bibliotecários, Júlio e Eneida, por fomentar em mim a chama da academia. Pois eis-me aqui! Grata por acreditarem em mim.

À Amanda Damasceno, Bibliotecária Clínica, que tive a honra de conhecer durante o programa de pós-graduação e com quem pude compartilhar angústias e experiências. Pela atenção e solidariedade com que me acolheu nos momentos de desespero durante a escrita deste trabalho.

À equipe da Biblioteca Guajajaras, Juliana, Jaqueline, Pollyanne, Lidiane, Everton, Mirian, Wadanu (*in-memoriam*), Leo Souza e Leo Brito e Show (Elaine), minha equipe para vida toda! Obrigada pela confiança!

Aos meus amigos Daniel, Tânia, Juuhhh, Vanessa, Marcus, Raquel, Mateus (Terapeuta), Matheus (Paulista) e Camila.

À Lélia pelo acolhimento e escuta no momento mais difícil.

Aos colegas de pós-graduação, pela troca de conhecimentos e experiências profissionais.

Aos professores Milton Shintaku, Cíntia Azevedo Lourenço, Célia da Consolação Dias que me honraram ao aceitarem o convite para compor a Banca de Defesa desta Dissertação.

A orientadora agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de Produtividade em Pesquisa, processo número 303650/2019-2.

Obrigada por acreditarem em mim!

Mas se depender de mim, eu vou até o fim

Até o fim

Não vim até aqui pra desistir agora

Até o fim, Engenheiros do Hawaii

RESUMO

O incremento das Tecnologias de Informação e Comunicação (TIC) acarretou alteração na estrutura da comunicação científica e alavancou iniciativas de acesso e compartilhamento de informação. Como exemplo, cita-se o Movimento *Open Access* (OA), resultado da *Budapest Open Access Initiative* (2002), que institucionalizou duas formas de acesso: via dourada e via verde. Nesse cenário, surgem os Repositórios Institucionais (RIs), utilizados como instrumento de apoio às universidades públicas brasileiras para facilitar a divulgação e o acesso à produção científica, tendo os bibliotecários e gestores a responsabilidade de os povoar. Esse povoamento, seja por autopovoamento, depósito mediado ou autodepósito, exige cuidado com a qualidade dos metadados, pois são eles que estruturam e oferecem suporte à descrição dos dados, permitindo que os registros sejam armazenados, organizados, preservados e recuperados. Dentre os diversos tipos de metadados, este estudo destaca os problemas relacionados ao controle de autoridade para entidade pessoa, que se referem a erros de digitação, registros duplicados, abreviações distintas e variações dos nomes, que impactam negativamente na recuperação da informação. Esses problemas têm estreita relação com a ambiguidade que envolve a determinação dos nomes dos autores. Considerando que o *DSpace* é o *software* mais adotado pelos RIS nacionais, este estudo propõe investigar como ocorre o controle de autoridade para entidades pessoas em repositórios institucionais que usam o *DSpace*, evidenciando os elementos computacionais nativos ou compatíveis (não nativos) com o *software*, contribuindo com as discussões acerca da desambiguação de nomes próprios. Os procedimentos metodológicos adotaram abordagem qualitativa, com objetivos exploratório e descritivo, tendo natureza básica, usando como método a revisão de literatura. Como instrumento de coleta de dados, adotou-se o questionário aberto, aplicado aos gestores dos repositórios por meio do *Google Forms*, e para a gestão de referências da revisão foi utilizado o *Zotero*. Os resultados demonstraram que a atividade de controle de autoridade para entidade pessoa é um processo complexo, pois envolve tratar o problema da desambiguação dos nomes próprios (variantes, abreviações e mudanças de nomes, erros ortográficos, pseudônimos, homônimos, entre outros). O controle de autoridade hoje extrapola as barreiras das bibliotecas físicas, exigindo um trabalho cooperativo e coordenado, envolvendo os diferentes atores no processo, tanto em nível local, regional, estadual e internacional, pois já existe tecnologia robusta o suficiente para lidar com metadados heterogêneos. Como contribuição, acredita-se que os esclarecimentos sobre os elementos computacionais nativos do *DSpace* podem auxiliar na implementação e manutenção dos RIs pelos seus gestores, e, dessa maneira, facilitar a gestão das informações sobre os pesquisadores ali publicadas, contribuindo para o aprimoramento da comunicação científica.

Palavras-chave: Repositórios Institucionais. Controle de autoridade para entidade pessoa. Ambiguidade e desambiguação. *DSpace*.

ABSTRACT

The increase in Information and Communications Technology (ICT) yields to a changing in the structure of scientific communication and enhances initiatives for accessing and sharing information. For instance, one can cite the Open Access Movement (OA), a result of Budapest Open Access Initiative (2002), which institutionalized two ways to access information: golden way and green way. In this scenario, Institutional Repositories (IRs) emerge as an instrument for supporting Brazilian public universities to easy dissemination and access of scientific production, and librarians and managers assumed the responsibility to populate them. This data population, whether by self-population, mediated deposit or self-deposit, requires care regarding quality of metadata, as it structures and provides support to data description, allowing the records to be stored, organized, preserved and retrieved. Among the distinct types of metadata, the present study highlights the problems regarding authority control for the entity “person”, which refer to typing mistakes, duplicate records, different abbreviations, and name variations, which negatively impact the information retrieval. These problems are closely related to the ambiguity encompassing the determination of authors’ names. Considering that DSpace is the most adopted software by Brazilian IRs, this study proposes investigate how authority control occurs for personal entities in institutional repositories that use DSpace, highlighting the computational elements native or compatible (non-native) with the software, contributing to discussions about the disambiguation of proper names. The methodological procedures adopted a qualitative approach, with exploratory and descriptive objectives, having a basic nature, using as method the literature review. As an instrument of data collection, it was adopted the open questionnaire, applied to the repository managers through Google Forms, and for the reference management of the review it was used Zotero. The results showed that the activity of authority control for person entity is a complex process because it involves dealing with the problem of disambiguation of proper names (variants, abbreviations and name changes, misspellings, pseudonyms, homonyms, among others). Authority control today goes beyond the barriers of physical libraries, requiring cooperative and coordinated work involving the different players in the process, at local, regional, state, and international levels, as there is already robust enough technology to deal with heterogeneous metadata. As a contribution, it is believed that the clarifications about the native computational elements of DSpace can help the implementation and maintenance of IRs by their managers, and, in this way, facilitate the management of information about the researchers published there, contributing to the improvement of scientific communication.

Keyword: Institutional Repositories. Authority control for person entity. Ambiguity and disambiguation. DSpace.

LISTA DE FIGURAS

| | |
|--|-----|
| Figura 1 – Formas de acesso e organização dos catálogos manuais externos..... | 64 |
| Figura 2 – Arquitetura técnica do <i>software DSpace</i> | 81 |
| Figura 3 – Modelo de informação do <i>DSpace</i> | 83 |
| Figura 4 – Exemplo de organização do RI da UFMG | 84 |
| Figura 5 – Exemplo de organização dos itens em coleções e comunidades | 85 |
| Figura 6 – Exemplo de chave de autoridade | 111 |
| Figura 7 – Autoridade controlada no <i>DSpace</i> | 112 |
| Figura 8 – Lista simples de tipos de documento | 114 |
| Figura 9 – Exemplo de busca no <i>DSpace</i> integrado com o ORCID..... | 116 |
| Figura 10 – Exemplo de interface do <i>DSpace</i> integrada com ORCID | 117 |
| Figura 11 – Campos do ORCID para identificação das autoridades | 118 |
| Figura 12 – Autoridade controlada pelo <i>DSpace</i> | 118 |
| Figura 13 – Exemplo de uso do <i>OpenRefine</i> | 120 |
| Figura 14 – Exemplo de banco de dados do <i>Access</i> | 122 |
| Figura 15 – Exemplo de uso do <i>Microsoft Access</i> | 123 |
| Figura 16 – Exemplo de organização de dados no <i>Excel</i> linhas e colunas | 124 |

LISTA DE QUADROS

| | |
|--|-----|
| Quadro 1 – Desafios relacionados aos metadados..... | 49 |
| Quadro 2 – Problemas encontrados no trabalho cooperativo do projeto AUTHOR | 70 |
| Quadro 3 – Características técnicas <i>DSpace</i> | 82 |
| Quadro 4 – Elementos <i>Dublin Core</i> | 86 |
| Quadro 5 – Pontos fortes e fracos do questionário | 89 |
| Quadro 6 – Vantagens e desvantagens das questões abertas..... | 90 |
| Quadro 7 – Identificação das Universidades Públicas Federais Minas Gerais | 93 |
| Quadro 8 – Universidades Públicas Federais Minas Gerais que utilizam <i>DSpace</i> | 94 |
| Quadro 9 – Campos de metadados <i>DSpace versus</i> ORCID..... | 119 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Busca preliminar..... | 27 |
| Tabela 2 – Resultado da busca por base de dados | 27 |
| Tabela 3 – Resultado da seleção inicial | 28 |

LISTA DE GRÁFICOS

| | |
|--|-----|
| Gráfico 1 – Questão 2: Como é realizado o controle de autoridade no Repositório Institucional da sua Universidade? | 97 |
| Gráfico 2 – Questão 3: Caso não realize o controle de autoridade, como justifica isso? | 98 |
| Gráfico 3 – Questão 4: O repositório da sua universidade possui política para cadastro e controle de autoridade? Se sim, por favor, descreva..... | 100 |
| Gráfico 4 – Questão 5: O repositório da sua universidade utiliza fonte externa para consulta de cadastro de autoridade. Se sim, por favor, descreva..... | 101 |

LISTA DE SIGLAS

| | |
|---------|---|
| ALA | <i>American Library Association</i> |
| BFM | <i>Bibliographic File Maintenance</i> |
| BN | Biblioteca Nacional |
| BOAI | <i>Budapest Open Access Initiative</i> |
| CATS | <i>Catalogue Support System</i> |
| CAU | Cabeçalho de Assunto Unificado |
| CBN | Controle Bibliográfico Nacional |
| CBU | Controle Bibliográfico Universal |
| CNPq | Conselho Nacional de Desenvolvimento Científico e Tecnológico |
| COEP | Comitê de Ética e Pesquisa |
| CONEP | Comissão Nacional de Ética em Pesquisa |
| DC | <i>Dublin Core</i> |
| DCMI | <i>Dublin Core Metadata Initiative</i> |
| DNB | <i>Deutsche Nationalbibliothek</i> |
| DPLA | <i>Digital Public Library of America</i> |
| Embrapa | Empresa Brasileira de Pesquisa Agropecuária |
| ETL | Extração, Transformação, e Carregamento |
| FECYT | <i>Fundación Española para la Ciencia y la Tecnología</i> |
| FIAB | Federação Internacional de Associação de Bibliotecas |
| FRAD | <i>Functional Requirements for Authority Data</i> |
| FRBR | <i>Functional Requirements for Bibliographic Records</i> |
| GARE | <i>Guidelines for Authority and Reference Entries</i> |
| GARR | <i>Guidelines for Authority Records and References</i> |
| HP | <i>Hewlett-Packard</i> |
| IBICT | Instituto Brasileiro de Informação em Ciência e Tecnologia |
| ID | Identificadores |
| IDEALS | <i>Digital Environment for Access to Learning and Scholarship</i> |
| IES | Instituições de Ensino Superior |
| IFLA | <i>International Federation of Library Associations</i> |
| IRALIS | <i>International Registry of Authors-Links to Identify Scientists</i> |
| ISADN | <i>International Standard Authority Data Number</i> |
| ISNI | <i>International Standard Name Identifier</i> |

| | |
|---------|--|
| ISTA | <i>Information Science & Technology</i> |
| JSPs | <i>JavaServer Pages</i> |
| KRIS | Kiwi Serviço de Informações de Pesquisa |
| LC | <i>Library of Congress</i> |
| LCNAF | <i>Library of Congress Name Authority File</i> |
| LCSH | <i>Library of Congress Subject Headings</i> |
| LISA | <i>Library and Information Science</i> |
| LISTA | <i>Information Science & Technology</i> |
| LSP | <i>Linked Systems Project</i> |
| MARC | <i>Machine Readable Cataloging</i> |
| MAT | Ferramenta de Análise de Metadados |
| METS | <i>Metadata Encoding & Transmission Standard</i> |
| MIT | <i>Massachusetts Institute of Technology</i> |
| NACO | <i>Name Authority Cooperative</i> |
| NCSA | <i>National Center for Supercomputing Applications</i> |
| NSDL | <i>National Science Digital Library</i> |
| OA | <i>Open Access</i> |
| OAI | <i>Open Archives Initiative</i> |
| OAI-PMH | <i>Open Access Initiative-Protocol for Metadata Harvesting</i> |
| OCLC | <i>Online Computer Library Center</i> |
| ORCID | <i>Open Researcher and Contributor ID</i> |
| OWL | <i>Ontology Web Language</i> |
| QDC | <i>Dublin Core Qualificado</i> |
| RDs | Repositórios Digitais |
| RIs | Repositórios Institucionais |
| RIS | <i>Research Information Systems</i> |
| RL | Revisão de Literatura |
| RLIN | <i>Research Libraries Information Network</i> |
| ROAR | <i>Register of Open Access Repositories</i> |
| SKOS | <i>Simple Knowledge Organization System</i> |
| SNI | <i>Standard Network Interconnect</i> |
| TCLE | Termo de Consentimento Livre e Esclarecido |
| TI | Tecnologia da Informação |
| TIC | Tecnologias de Informação e Comunicação |

| | |
|---------|---|
| UFMG | Universidade Federal de Minas Gerais |
| UH | Universidade de Houston |
| UNIMARC | <i>Universal Marc Format</i> |
| UNR | Universidade de Nevada |
| UTLAS | <i>University of Toronto Library Automation Systems</i> |
| VIAF | <i>Virtual International Authority File</i> |
| WLN | <i>Washington Library Network</i> |
| WNAF | <i>Western Name Authority File</i> |
| WoS | <i>Web of Science</i> |
| XML | <i>Extensible Markup Language</i> |

SUMÁRIO

| | | |
|------------|--|-----------|
| 1 | INTRODUÇÃO | 14 |
| 1.2 | Objetivos | 23 |
| 1.2.1 | Objetivo geral | 23 |
| 1.2.2 | Objetivos específicos | 23 |
| 2 | REVISÃO DE LITERATURA..... | 24 |
| 2.1 | Procedimentos da revisão de literatura (RL)..... | 24 |
| 2.2 | Descrição dos estudos da amostra da RL..... | 29 |
| 3 | FUNDAMENTOS TEÓRICO-CONCEITUAIS | 53 |
| 3.1 | A ambiguidade no âmbito dos nomes próprios..... | 54 |
| 3.2 | O controle de autoridade para entidades pessoas..... | 58 |
| 3.2.1 | Breve histórico sobre a catalogação e o catálogo | 59 |
| 3.2.1.1 | <i>O catálogo</i> | 62 |
| 3.2.2 | Controle de autoridade ao longo do tempo | 65 |
| 3.3 | Repositórios institucionais..... | 74 |
| 3.4 | <i>Software</i> DSpace..... | 80 |
| 4 | METODOLOGIA | 88 |
| 4.1 | Insumos tecnológicos | 88 |
| 4.2 | Instrumento de coleta de dados: questionário | 89 |
| 4.3 | Procedimentos metodológicos..... | 90 |
| 5 | RESULTADOS E DISCUSSÕES..... | 93 |
| 5.1 | Resultados da Primeira Etapa..... | 93 |
| 5.1.1 | Identificação das Universidades Públicas Federais existentes em Minas Gerais.. | 93 |
| 5.1.2 | Identificação das Universidades Públicas Federais mineiras que possuem RIs com a utilização do <i>software</i> DSpace..... | 94 |
| 5.1.3 | Identificação dos dados dos gestores dos RIs | 95 |

| | | |
|------------|---|------------|
| 5.1.4 | Construção do questionário, instrumento de coleta de dados para entrevistar os gestores dos RIs | 95 |
| 5.1.5 | Análise das respostas ao questionário aplicado aos gestores dos RIs | 96 |
| 5.1.5.1 | Breves considerações acerca da análise do questionário | 105 |
| 5.2 | Resultados da Segunda Etapa | 107 |
| 5.2.1 | Descrição da especificidade da versão mais atualizada do <i>software DSpace</i> | 107 |
| 5.2.2 | Mapeamento dos elementos computacionais (nativos e não nativos) de controle de autoridade no <i>DSpace</i> | 109 |
| 5.2.3 | Descrição dos elementos computacionais (nativos e não nativos) de controle de autoridade para a última versão do <i>software DSpace</i> | 110 |
| 5.2.3.1 | <i>Choice Management and Authority Control (nativo)</i> | 111 |
| 5.2.3.2 | <i>Authority Control of Metadata Values (nativo)</i> | 113 |
| 5.2.3.3 | <i>ORCID Integration (nativo)</i> | 115 |
| 5.2.3.4 | <i>Elementos computacionais não nativos do DSpace</i> | 119 |
| 5.2.3.4.1 | <i>OpenRefine</i> (não nativo) | 119 |
| 5.2.3.4.2 | <i>OpenOffice Calc</i> (não nativo) | 121 |
| 5.2.3.4.3 | <i>Microsoft Access</i> (não nativo) | 122 |
| 5.2.3.4.4 | <i>Microsoft Excel</i> (não nativo) | 123 |
| 5.2.3.4.5 | <i>Duplicate Detection Service</i> (não nativo) | 125 |
| 6 | CONSIDERAÇÕES FINAIS | 127 |
| 6.1 | Limitações da pesquisa | 130 |
| 6.2 | Estudos futuros | 132 |
| 6.3 | Contribuição da pesquisa | 132 |
| | REFERÊNCIAS | 135 |
| | APÊNDICE A – QUESTIONÁRIO | 151 |
| | APÊNDICE B – TCLE | 152 |
| | ANEXO A – PARECER CONSUBSTANCIADO | 154 |

1 INTRODUÇÃO

Desde o final do século XX, nota-se uma mudança na estrutura da comunicação científica e isso é reflexo do incremento advindo das Tecnologias de Informação e Comunicação (TIC). Essa mudança alavancou as iniciativas de acesso e compartilhamento de informação, dando origem, por exemplo, ao Movimento *Open Access* (OA). O OA é resultado da *Budapest Open Access Initiative* (BOAI, 2002), que institucionalizou duas formas de acesso livre: via dourada e via verde. Em linhas gerais, a primeira corresponde à publicação de artigos em periódicos de acesso livre, que não limitam o acesso ao conteúdo por meio de assinatura ou cobrança de taxa, e se sustentam pela cobrança, por exemplo, para a publicação. A segunda tem como principal característica o autoarquivamento (*self-archiving*), em que os próprios autores submetem os documentos em portais, periódicos ou repositórios (BAPTISTA *et al.*, 2007).

Os Repositórios Digitais (RDs) tiveram origem como um instrumento de apoio e suporte à via verde, ou de comunicação de acesso livre, pois têm como objetivo gerenciar a informação científica. Segundo Leite (2009), os RDs podem ser de três tipos: os institucionais, que divulgam a produção científica e intelectual de uma instituição; os temáticos ou disciplinares, que são restritos a áreas específicas de conhecimento; e os de teses e dissertações, que divulgam os resultados das pesquisas na pós-graduação. Neste estudo, têm-se como ponto central os RDs do tipo institucional, ou seja, os Repositórios Institucionais (RIs).

Os RIs podem ser definidos como uma biblioteca digital, já que seu conteúdo está inteiramente *on-line*, e têm como objetivo “guardar, preservar e garantir livre acesso, via internet, à produção científica no âmbito de uma dada instituição” (SAYÃO *et al.*, 2009, p. 9). Na contemporaneidade, Crow (2002) relata que se observa uma tendência dos RIs em unificar dois tipos de repositórios: o institucional e o de teses e dissertações. Isso vem ocorrendo porque os RIs respondem a questões estratégicas enfrentadas pelas instituições universitárias: facilitam o acesso à produção da universidade; dão um retorno sobre as pesquisas financiadas com o uso de bolsas de agências de fomento; aumentam a concorrência e reduzem o impacto do monopólio dos periódicos, a visibilidade e relevância da instituição; e tornam tangíveis a produção e a qualidade do que é produzido (CROW, 2002).

No Brasil, outro fator que impulsiona os RIs é o fato de as universidades públicas serem desafiadas a mostrarem para a sociedade os resultados do financiamento público que é

empregado nas suas atividades científicas (SERRA; ROLIM; BASTOS, 2018). Como uma das respostas a esse desafio, as iniciativas de RIs rompem os limites tradicionais de acesso à informação acadêmica, ampliam a acessibilidade, visibilidade e impacto das publicações, assim como melhoram o crescimento da pesquisa (NEMATI-ANARAKI; TAVASSOLI-FARAH, 2018), buscando dar maior visibilidade à produção institucional.

A acessibilidade informacional, que é condição para a garantia do acesso à informação no RI, é dependente da qualidade na entrada de dados no sistema, atividade denominada na Biblioteconomia como processamento técnico. Esse processo se norteia por fatores que “podem ser compreendidos como um conjunto de atividades realizadas em um setor específico da biblioteca que se encarrega da análise e representação temática (indexação) e descritiva (catalogação)” (FERNANDES; SOUZA, 2017, p. 13), com o objetivo de organizar os documentos e sua futura recuperação, e [que] “são norteados por uma vasta gama de códigos, padrões e regras” (FERNANDES; SOUZA, 2017).

No processamento técnico, um documento¹ é representado por seus elementos temáticos e descritivos, que são os pontos de acesso para responderem às buscas dos usuários. Os elementos temáticos são determinados a partir dos “conceitos expressos em um documento [...]”. Esta operação pressupõe um conhecimento do assunto do documento e uma definição precisa do nível de informação a ser preservado de forma a responder às necessidades dos usuários” (GUINCHAT; MENO, 1994, p. 31). Os elementos descritivos se constituem na determinação dos atributos referenciais dos documentos, tais como título, autor, editora, entre outros, que têm o papel de “individualizar os itens, de forma que não sejam confundidos entre si; reunir itens por suas semelhanças, estabelecendo relações entre si, e, finalmente, permitir a localização de um item específico” (MEY, 1995, p. 38-39). E, por fim, os pontos de acesso podem ser definidos como “um nome, termo, código, etc., por meio do qual os dados bibliográficos ou de

¹ Ortega (2016) discorre sobre o conceito de documento trazendo, entre outros autores, as contribuições de Paul Otlet (quaisquer objetos abordados informacionalmente; são matéria e conteúdo), Suzanne Briet (requer procedimentos de significação), Sagredo e Izquierdo (questão da utilidade), Robert Escarpit (objeto informacional concreto) e Jean Meyriat (quaisquer objetos com a função de transmitir uma mensagem). Partindo desses estudos, a autora elabora uma definição abrangente para documento bibliográfico: “documento se define por uma instância física (seu suporte material) e uma instância simbólica (informativa, contedística), sendo que esta última é sempre uma atribuição, ou seja, não é inerente ao objeto. A dimensão informativa do documento não é prévia, é construída no momento da interpretação. Se o documento é objeto construído, ou seja, não existe a priori, ou in natura, o documento bibliográfico também não é objeto pré-existente, mas depende de ações de significação que o tornem como tal” (ORTEGA, 2016, p. 56).

autoridade são procurados e identificados” (*International Federation of Library Associations – IFLA*, 2009, p. 9).

Todas as atividades realizadas no processamento técnico demandam controle. No âmbito do controle de autoridades, o termo “controle” se refere à garantia da univocidade entre os nomes utilizados para representar cada entidade e a entidade em si. Segundo Assumpção e Santos (2012, p. 6), “uma entidade [deve] ser identificada por apenas um ponto de acesso autorizado e um ponto de acesso autorizado [deve] identificar somente uma entidade”, sendo a atividade orientada por políticas (de indexação e catalogação) que possam manter tal consistência. Para tanto, adotam-se os códigos de catalogação, as linguagens de indexação e os sistemas de classificação, que têm como função dar apoio às atividades no processamento técnico. Desse processo, como produto ou instrumento de consulta, são gerados os catálogos, que orientam as buscas do usuário.

Desde a criação dos catálogos, nota-se uma preocupação com o controle de autoridade, quando são criadas listas com vocabulários controlados de assunto e de autoridade (nomes de pessoas, famílias e entidades coletivas). Esse controle tem como finalidade fazer com que a informação disponibilizada seja padronizada e uniformizada e auxilie o usuário a encontrar o que busca de maneira mais precisa. Neste estudo, trata-se do controle de autoridades em RIs.

O controle de autoridade para pessoas, famílias e entidades coletivas “requer a reunião de todos os nomes ou forma do nome pelo qual uma entidade é conhecida” (ASSUMPCÃO; SANTOS, 2012, p. 11). O “controle de autoridade obriga o uso de sistemas limitados e padronizados para a representação da informação e impõe uniformidade ao passo que rejeita divergência e diversidade” (MARAIS, 2004, p. 61, tradução da autora). Esse controle se refere, ainda, à “criação de registros de autoridade para cabeçalhos estabelecidos, no relacionamento dos registros de autoridade com os registros bibliográficos e na manutenção e avaliação do sistema de autoridade” (ROWLEY; FARROW, 2000, p. 369-370, tradução da autora).

Dentre os distintos tipos de controle de autoridade, no contexto desta pesquisa será tratado o controle de autoridade para entidades de nomes de pessoas, ou, de modo mais abreviado: controle de autoridade para entidade pessoa, com foco nas questões que envolvem a gestão de RIs.

1.1 Motivação, Problema e Justificativas

A motivação para se desenvolver esta pesquisa se deve a interesse profissional, à formação em Biblioteconomia e à atuação como bibliotecária há mais de 10 anos. Nesta atividade, lida-se com as questões que envolvem a busca e recuperação da informação em RIs pelos públicos atendidos, que é uma constante preocupação.

No desenvolvimento desta pesquisa, numa exploração inicial foram acessados alguns RIs de universidades federais do estado de Minas Gerais e de outros estados. Em uma análise na lista de autoridades para entidade pessoa, foram encontrados registros de autoridade com erros de digitação, estando um mesmo autor com entradas distintas ou duplicadas, falta de padrão ao abreviar o prenome (ora maiúsculas, ora minúsculas), entre outros equívocos. Essa ambiguidade na determinação do autor, que prejudica o controle de autoridade, inspirou a definição do escopo de investigação.

Veiga, Machado e Alves (2013) destacam que, desde a criação do ArXiv (*open archives*), em 1991, primeiro arquivo eletrônico de *preprints*, “pesquisadores passaram a criar arquivos eletrônicos de *preprints* e *posprints* como alternativa para publicação direta de seus trabalhos em texto completo” (MARCONDES; SAYÃO, 2002, p. 43). E, com a popularização da internet, tem-se um incremento de bases, periódicos e repositórios, que incentivam o autoarquivamento dos documentos produzidos, sobretudo no contexto científico, concentrado nas instituições (LEITE, 2007). Esse fato agilizou o processo de tornar pública e mais acessíveis todas as publicações acadêmicas das instituições, já que, em geral, os próprios autores/pesquisadores lançam no sistema a sua produção.

Salo (2009) destaca que uma dificuldade para o sucesso dos RIs é que eles foram concebidos como sistemas de depósito de conteúdo por autoarquivamento, sem qualquer controle de autoridade ou de qualidade no preenchimento dos metadados. Reforçando essa ideia, Assis (2013, p. 216) alerta que é necessário refletir sobre a “implantação de uma política de sucesso de autoarquivamento”. A autora afirma que essa não é uma prática muito adotada pelas instituições em geral, mas que elas “devem disponibilizar materiais de apoio e treinamento para que os pesquisadores se sintam confortáveis com o sistema” (ASSIS, 2013, p. 217). Esse fato evidencia um despreparo para o efetivo controle de autoridades, o que tem prejudicado a gestão

das coleções da produção institucional e causado impacto negativo na recuperação da informação.

Especificamente, discorrendo sobre a busca por nome de autores em RIs, Xia (2006) afirma que esse é um desafio à parte devido, sobretudo, aos problemas de controle de autoridade para entidades pessoas, consequência das diversas variações possíveis para um mesmo nome, que não só reduzem a precisão dos resultados, como também tornam a análise de citações pouco confiável. Essa dispersão referencial resulta em ambiguidade na determinação da autoria da produção científica em RIs, e se torna o problema central para a recuperação de informações sobre os autores.

Salo (2009) acrescenta que o depósito em RI mediado por bibliotecários ou outros profissionais da informação parece ser o método mais viável para se manter a qualidade da entrada de dados em RIs. Para a autora, essa conduta deveria ter melhorado a situação do controle de autoridades, mas que, na prática, a qualidade dos metadados ainda está longe do ideal. Afirma ainda, que mesmo os catalogadores interessados no tema parecem não perceber que o controle de autoridade para entidade pessoa deve ser entendido como um ponto chave dos RIs. Conforme alerta a autora, outros elementos impactam negativamente nessa qualidade, tais como o *software* do RI e o *design* dos serviços, porém ela afirma que é relevante estudar as questões em torno do controle de autoridade para entidades pessoas.

Salo (2009) afirma que se um usuário quer buscar no RI a produção acadêmica pelo nome de um único autor, terá dificuldades na recuperação desses itens, pois nem sempre o nome dele estará ligado a toda a sua produção e, talvez, nem as variações do mesmo nome estarão ligadas a uma única pessoa. Smalheiser e Torvik (2009, p. 288) afirmam que enquanto as bibliotecas tradicionais, em geral, fazem um esforço para a identificação manual e local dos nomes dos autores, “essa estratégia falha com o surgimento maciço de bibliotecas digitais” (SALO, 2009, p. 288).

Outros problemas salientados por Salo (2009) no controle de autoridade para entidades pessoas se referem aos nomes homônimos. Já que os RIs são, em geral, multidisciplinares, isso aumenta a possibilidade de ocorrência de duas pessoas com nomes idênticos e, também, de acontecer alteração dos nomes dos autores durante as suas carreiras acadêmicas. Com isso, fica caracterizada a ambiguidade na determinação dos autores.

Xia (2006) destaca três problemas básicos de ambiguidade, a partir de variações de nomes próprios, que afetam diretamente o controle de autoridade para entidade pessoa: a) formas diferenciadas: nome completo e abreviado; b) homônimos; c) mudança de nome ao longo do tempo: casamento, divórcio entre outros.

Rosado e Dias (2020, p. 4) afirmam que a “desambiguação torna os pontos de acesso únicos e inconfundíveis, impedindo que mais de um deles tenham informações idênticas, mesmo havendo semelhanças entre eles”, e, no caso de homônimos, “esses podem ser diferenciados com informações complementares do tipo: datas, obras, cargos, entre outras”. Percebe-se, assim, que o problema da ambiguidade na identificação dos autores continua ocorrendo há mais de dez anos, pelo menos desde o estudo de Salo, em 2009.

Matias (2015, p. 22) relata que o “povoamento [moderado] de um repositório institucional exige dos profissionais envolvidos a recuperação das publicações”, o que demanda bastante trabalho, e acrescenta que muitos RIs “coletam metadados de bases como a *Web of Science* (WoS) e a SciELO” (MATIAS, 2015, p. 23). Contudo a autora observa uma limitação nessa prática, uma vez que “a WoS indexa somente os periódicos mundialmente mais citados em suas respectivas áreas, e que a SciELO reúne apenas periódicos de acesso aberto da América Latina e do Caribe” (MATIAS, 2015, p. 23). Outra limitação dessas duas bases se refere ao fato de que,

na WoS, a identificação de autoria é realizada a partir dos nomes adotados nos trabalhos por meio de algoritmos de aprendizagem de máquina e clusterização (CRL, 2015). Na SciELO, essa identificação depende do uso consistente do nome do autor, obedecendo sempre a um mesmo formato em todas as publicações (MATIAS, 2015, p. 23).

Matias (2015) ressalta que, no Brasil, o portal Plataforma Lattes, mantido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) “é uma fonte promissora de metadados”, pois o portal

congrega dados sobre todos os trabalhos de vários tipos de pesquisadores, os quais têm incentivo institucional e profissional para mantê-los atualizados no Lattes, independente da revista ou editora responsável pela publicação. Além disso, cada Currículo Lattes é atrelado de forma unívoca à identidade civil (CPF) de seu portador, que concorda com um termo responsabilizando-se legalmente pelas informações ali contidas. Isso significa que os metadados estão agrupados por autor já em sua fonte original, ao contrário do que ocorre em outras bases (MATIAS, 2015, p. 23).

Apesar de ser considerada uma fonte promissora, Matias (2015, p. 23) destaca que “a falta de uma sistemática para a coleta, tratamento e importação dos metadados dificultam o aproveitamento do potencial da Plataforma Lattes”. O problema de controle que, no contexto

desta pesquisa, denomina-se de ambiguidade, é observado em diversas bases devido, principalmente, “às diferentes maneiras que os distintos autores e coautores cadastram uma mesma publicação em seus currículos” (MUGNAINI *et al.*, 2012, p. 265).

Mugnaini *et al.* (2012) citam que, internacionalmente, o projeto *Open Researcher and Contributor ID* (ORCID), lançado oficialmente em meados de 2010, é a iniciativa que parece mais promissora para o controle de autoridades para entidades pessoas. O projeto ORCID tem em vista a solução de problemas de ambiguidade para os nomes dos autores de todo o mundo, com a criação de um identificador único e universal para cada entidade pessoa, que seja claro, permanente e unívoco. Contudo, ele não parece ser, ainda, a solução cabal, pois verifica-se a ocorrência dos mesmos problemas no controle de autoridade para entidades pessoas, tais como a duplicação de registros, erros de digitação, erros básicos de registro dos sobrenomes, incorreções estas que podem ser encontradas relatadas na literatura (PÉREZ-ALCAZAR *et al.*, 2011; CARDOSO; MIRANDA, 2020).

Desde o protótipo Xanadu, que pretendia ser um “sistema de bases de dados onde os escritores podiam publicar diretamente os seus textos, vinculando-os a outros documentos, constituindo, dessa forma, uma biblioteca eletrônica universal *on-line* de documentos interconectados” (SAYÃO, 2008, p. 8), muitos têm sido os desafios para os gestores na implantação e manutenção de RIs. No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) promoveu acesso às tecnologias e recursos para a implementação de RIs, conforme relatos de Costa, Kuramoto e Leite (2013). Babini e Machin-Mastromatteo (2015, p. 478) destacam que para que os RIs se fortaleçam é necessário promover boas práticas, treinar os interessados, investir em infraestrutura e estabelecer políticas institucionais, de maneira que elas não fossem apenas recomendações, mas “transformadas em mandatos, exigindo que o depósito ou auto-arquivamento fosse um requisito a ser considerado no momento da avaliação acadêmica, da promoção e da posse”. E a formalização de políticas em RIs também é reforçada por Costa e Leite (2017), sobretudo no que diz respeito às questões de Acesso Aberto em cada país.

Segundo informações obtidas no *site* do IBICT (2012, *on-line*²), o crescimento do número de RIs no Brasil “foi acelerado pelo projeto IBICT-FINEP/PCAL/XBDB, que possibilitou a

² Informações disponíveis em: <https://ibict.br/informacao-para-a-pesquisa/repositorios-digitais#historico>, onde consta que a última modificação ocorreu no dia 08 set. 2020.

implantação de [...] repositórios institucionais em diversas universidades e instituições de pesquisa”. Segundo Hélio Kuramoto, coordenador de Projetos Especiais do IBICT em 2005, foi recomendado o uso do *software DSpace* - um *Open Sources* - para a criação de RIs, tendo em vista convergir todas as informações sobre a produção científica das universidades “para a Biblioteca Digital Brasileira, que será o grande repositório nacional da informação e do conhecimento científico nacional” (IBICT, 2005, *on-line*³). O *Register of Open Access Repositories* (ROAR) atualmente⁴ (maio de 2021) registra 4552 repositórios no mundo inteiro, e, desse total, quase 50% (2249, em números absolutos) são iniciativas implementadas com o *DSpace*. Em outra fonte, o *site OpenDoar* (2021)⁵, encontra-se a informação da existência no mundo de mais de 5721 repositórios e que, cerca de 39%, utilizam o *software DSpace*. Essa fonte relata que, no Brasil, há 151 repositórios digitais e, dentre estes, 83% utilizam o *DSpace*. Esses dados confirmam que existe um uso massivo desse *software* na gestão de RDs de todos os tipos.

De acordo com Ribeiro Segundo *et al.* (2013, p. 6), desde a versão 1.6, o recurso de controle de autoridade está presente no *DSpace* e, na versão 3.1, já foi possível fazer o controle no campo autor, que é “mantido no padrão *Dublin Core* Qualificado (QDC) pelo campo de metadado ‘dc.contributor.author’, onde: dc é o esquema de metadados; contributor, o elemento; e author, o qualificador”, cujos dados podem ser vinculados à base *Library of Congress Names Authority File - LC Names*. O autor também esclarece que “essa facilidade pode ser aplicada a qualquer campo no *DSpace*, pois é sempre possível adicionar novos descritores que atendam às especificações do repositório, e vinculá-los a uma determinada base de controle de autoridade” (RIBEIRO SEGUNDO *et al.*, 2013, p. 6).

Desde a versão 3.1, novas versões já foram implementadas, e, conforme relata Martins, Silva e Siqueira (2018, p. 56), o “*DSpace* encontra-se, em 2018, na versão 6.x”, e a versão mais recente é a 6.3 (2020), sendo que, em agosto de 2021 foi lançada a versão 7, após fase de testes e *feedback*. Espera-se, a partir de análise da literatura, da documentação do *DSpace* e em listas de discussões sobre o *software*, que existam diferentes elementos computacionais, na forma de aplicativos nativos e também compatíveis (não nativos) com ele, que são desenvolvidos por

³ Informação obtida a partir de entrevista com o pesquisador Hélio Kuramoto, coordenador de Projetos Especiais do IBICT em 2005, disponível em: <https://ibict.br/sala-de-imprensa/noticias/item/788-ibict-estimula-adocao-dos-open-archives-no-brasil>; publicada em 23 set. 2005

⁴ Informação retirada de: <http://roar.eprints.org/>, obtida em maio de 2021

⁵ Disponível em: <https://v2.sherpa.ac.uk/opensoar/>

pesquisadores de diferentes instituições, inclusive para apoiar a atividade de controle de autoridade.

Parte-se do pressuposto de que a implementação dos elementos computacionais nativos e também dos compatíveis (não nativos) com o *DSpace* é difícil para profissionais não habilitados a resolver os possíveis problemas técnicos envolvidos, já que eles exigem um grau avançado de conhecimento em informática.

Com base no exposto, sobre o problema e o pressuposto de pesquisa, esta pesquisa propõe responder à seguinte questão norteadora: como ocorre o controle de autoridade para entidades pessoas em repositórios institucionais que usam o *DSpace*, e quais são os elementos computacionais nativos ou compatíveis (não nativos) com o *software* que podem contribuir para esse controle pelos gestores?

Como contribuição, busca-se trazer luz à questão proposta, por entender que a disponibilidade de um documento em um RI não necessariamente melhora o acesso, nem mesmo consegue medir o impacto da produção dos autores, caso não exista uma entrada de dados com qualidade, que ofereça um serviço eficiente e abrangente. Ou seja, é preciso que a informação disponível no RI passe pelo processamento técnico de maneira sistemática, consistente e com qualidade de dados, para só então ser considerada útil ou acessível (LAWRENCE, 2001).

Acredita-se que o mapeamento dos elementos computacionais disponíveis para o *software DSpace* (nativos ou não nativos) pode contribuir no controle de autoridades, uma vez que os erros na atividade de controle de autoridade para entidade pessoa podem ser mitigados pela implementação dos recursos já existentes para essa finalidade. Ao compilar esses aplicativos, espera-se difundir a sua existência entre os profissionais bibliotecários e gestores de RIs.

Espera-se contribuir também para os pesquisadores, uma vez que a falta de uniformidade na determinação da forma correta de seus nomes na entrada (povoamento) no banco de dados tem sérias consequências para o cálculo real da produção científica dos próprios autores, para a visibilidade destes e de seus centros de pesquisa, em nível nacional e internacional. Essa falta de controle também dificulta a recuperação das publicações e das citações recebidas pelos autores, além de outras dificuldades, tal como a não correspondência dos resultados da pesquisa com os critérios estabelecidos pelo usuário.

1.2 Objetivos

Este estudo propõe atender ao objetivo geral e aos objetivos específicos aqui descritos.

1.2.1 Objetivo geral

Investigar como ocorre o controle de autoridade para entidades pessoas em repositórios institucionais que usam o *DSpace*, evidenciando os elementos computacionais nativos ou compatíveis (não nativos) com o *software*, contribuindo com as discussões acerca da desambiguação de nomes próprios.

1.2.2 Objetivos específicos

- a) Explorar o fenômeno da ambiguidade no âmbito da determinação de nomes próprios, tendo em vista compreender as questões para o controle de autoridade para entidade pessoa.
- b) Analisar e descrever as questões que envolvem o controle de autoridade, com foco no controle para entidade pessoa, trazendo elementos históricos para o entendimento da atividade nos RIs.
- c) Entender como o controle de autoridade tem sido realizado em RIs, sobretudo aqueles que usam o *DSpace*, a partir de uma revisão da literatura sobre o tema.
- d) Descrever informações sobre elementos computacionais nativos e compatíveis com o *software DSpace* disponíveis na sua documentação e na literatura, buscando aqueles que podem auxiliar no controle de autoridade para entidade pessoa.
- e) Verificar de que modo e se ocorre o uso de elementos computacionais para o controle de autoridade em RIs que utilizam o *DSpace*, buscando entender o controle de autoridade realizado para entidade pessoa.

2 REVISÃO DE LITERATURA

Este Capítulo está dividido em duas seções, sendo que, a primeira destaca os procedimentos adotados para a revisão de literatura e a segunda apresenta a descrição dos resultados em conformidade com as categorias.

2.1 Procedimentos da revisão de literatura (RL)

Esta seção descreve os procedimentos utilizados na revisão de literatura (RL) que, em síntese, norteou-se pelo seguinte planejamento: a) Determinação do objetivo da RL; b) Definição das fontes de informação a serem consultadas; c) Escolha dos descritores e *strings* de pesquisa; d) Definição dos critérios de inclusão e exclusão; e) Fase de Execução; f) Tratamento dos dados; g) Categorias de análise dos documentos da amostra; que serão detalhadamente descritas a seguir.

a) Determinação do objetivo da RL

A RL tem como objetivo: buscar informações sobre a implantação de RIs; boas práticas; desafios, soluções para a correção de metadados em lote; assim como propostas de elementos computacionais que possam auxiliar na gestão dos RIs, em especial daqueles que usam o *software DSpace*, para a melhoria do controle de autoridade e da qualidade dos metadados para entidade autor.

b) Definição das fontes de informação a serem consultadas

Como fonte de coleta de dados, optou-se por realizar as buscas em bases específicas no Portal Capes (*login* Universidade Federal de Minas Gerais – UFMG). No Portal Capes, foram consideradas as bases: *Proquest*, que inclui a base referencial com resumo *Library and Information Science* (LISA); *Ebsco*, que inclui a base referencial com resumo *Information Science & Technology* (ISTA); a base referencial com resumo e texto completo *Library, Information Science & Technology* (LISTA); as bases de texto completo, *Taylor & Francis Online* e *Wiley Online Library*, e as bases referenciais com resumo *Scopus* e *Web of Science*.

c) Escolha dos descritores e *strings* de pesquisa

Inicialmente, foram elencados os seguintes descritores para a realização da busca: “controle de autoridade”; “trabalho de autoridade”; “registros de autoridade”; repositórios; “repositórios institucionais” e *DSpace*. Também foram incluídos os seus correspondentes em inglês, respectivamente, “*authority control*”, “*authority work*”, “*authority records*”, *repositories*, “*institutional Repositories*” e *DSpace*. A tradução dos descritores para o inglês facilitará a busca de documentos nos demais idiomas, já que, em bases científicas, o *abstract* e os descritores em inglês são obrigatórios para todas as publicações, independentemente do idioma da publicação.

Os descritores foram combinados formando as seguintes *strings* de pesquisa:

(“authority control” OR “authority work” OR “authority records”) AND (repositories OR repository OR “institutional Repositories” OR “Institutional-repository”);
(“authority control” OR “authority work” OR “authority records”) AND DSpace;
(“authority control” OR “authority work” OR “authority records”)

(“controle de autoridade” OR “trabalho de autoridade” OR “dados de autoridade”) AND
(repositórios OR repositório OR “repositórios institucionais” OR “repositório
institucional”); (“controle de autoridade” OR “trabalho de autoridade” OR “dados de
autoridade”) AND DSpace; (“controle de autoridade” OR “trabalho de autoridade” OR
“dados de autoridade”)

d) Definição dos critérios de inclusão e exclusão

Para a filtragem e seleção de trabalhos a serem analisados nesta pesquisa, foram estabelecidos os seguintes critérios de inclusão e exclusão:

Critérios de Inclusão de documentos:

- publicados a partir de 2002 (origem do *DSpace*), até 31 de dezembro de 2019;
- publicados em inglês, português e espanhol;
- Tipo de documento artigo, por entendermos que esse tipo contempla os relatos de pesquisas concluídas ou em andamento.

Critérios de Exclusão de documentos:

- que não estejam disponíveis para acesso livre em texto completo;

- duplicados;
- que não atendam ao objetivo da revisão.

Optou-se pela busca avançada em todas as bases, no entanto, algumas das bases não permitiam, nessa busca inicial, aplicar o filtro de idioma, são elas: ISTA, *Taylor & Francis Online*, *Wiley Online Library*, *Science Direct*, *Scopus* e *Web of Science*. Para não excluí-las como fonte de busca, o filtro de idioma foi aplicado após a exibição dos resultados da busca.

e) Fase de Execução

A seleção inicial foi realizada a partir da aplicação dos critérios de inclusão e exclusão estabelecidos na fase de Definição dos critérios de inclusão e exclusão.

Na busca inicial, considerando as *strings* definidas e os critérios de inclusão e exclusão, observou-se um resultado com alta revocação e baixa precisão. Um volume final de documentos que excedia 50.000 itens.

Sendo assim, para diminuir a revocação, optou-se por simplificar a *string* para: (“*authority control*” OR “*authorities control*”) AND (“*institutional repositories*” OR “*Institutional repository*”)

A busca foi realizada novamente considerando a nova *string* e o resultado é apresentado na Tabela 1.

Tabela 1 – Busca preliminar

| Base de dados | "authority control" AND ("institutional repositories" OR "Institutional repository") | ("authority control" OR "authorities control") AND ("institutional repositories" OR "Institutional repository") | "authority control" AND "institutional repositories" OR "Institutional repository" |
|-------------------------|--|---|--|
| LISA | 74 | 74 | 1419 |
| ISTA | 7 | 9 | 313 |
| LISTA | 10 | 11 | 850 |
| Taylor & Francis Online | 136 | 136 | 1183 |
| Wiley Online Library | 2 | 2 | 157 |
| Scopus | 42 | 42 | 42 |
| Web of Science | 2 | 2 | 449 |
| Total | 273 | 276 | 4413 |

Fonte: Resultado da pesquisa, 2020. Nota: Busca realizada em 6 de abril de 2020.

A Tabela 1 apresenta três buscas distintas: a primeira e a segunda colunas destacam os resultados encontrados, realizando a busca pelo termo “*authority control*”, somente no singular e no singular e plural. Apesar de se considerar que a busca pelo termo no singular deveria retornar também os resultados para o termo no plural, observou-se uma diferença no resultado. Assim, optou-se por trabalhar com os resultados da segunda coluna.

A segunda e terceira colunas destacam a diferença dos resultados ao serem incluídas as expressões ligadas pelo operador *OR*, com e sem o uso dos parênteses. Novamente, notou-se uma alta revocação quando o parênteses foi ignorado na expressão de busca.

Ao final, obteve-se o resultado apresentado na Tabela 2.

Tabela 2 – Resultado da busca por base de dados

| Base de Dados | Número de documentos recuperados |
|-------------------------|---|
| LISA | 74 |
| ISTA | 9 |
| LISTA | 11 |
| Taylor & Francis Online | 136 |
| Wiley Online Library | 2 |
| Scopus | 42 |
| Web of Science | 2 |
| Total | 276 |

Fonte: Dados da pesquisa, 2020.

Ao final da busca, obteve-se um total de 276 documentos, sendo mais da metade dos documentos recuperados oriundos das bases *Taylor & Francis Online* e LISA, com 49% e 27%, respectivamente.

f) Tratamento dos dados

Na fase de Tratamento dos dados, o *software* de produtividade escolhido para gerenciar as citações foi o Zotero, onde ficaram armazenados os documentos coletados na execução da estratégia de busca.

Durante a busca nas bases de dados, os artigos selecionados foram extraídos e exportados nas extensões *Research Information Systems* (RIS). Dentro do Zotero, os documentos foram separados nas pastas correspondentes às bases das quais foram extraídos e foram descartados 44 documentos duplicados.

Em seguida, procedeu-se à leitura dos resumos, e, quando estes não foram considerados suficientes para a seleção, a leitura diagonal foi realizada. Nessa análise, foram descartados 154 documentos, uma vez que não atenderam critérios de inclusão e exclusão. Ao final, restaram para análise 78 artigos, conforme mostra a Tabela 3.

Tabela 3 – Resultado da seleção inicial

| Base de Dados | Número de documentos selecionados |
|--------------------------|-----------------------------------|
| LISA | 19 |
| ISTA | 0 |
| LISTA | 3 |
| Taylor & Francis Online* | 46 |
| Wiley Online Library | 2 |
| Scopus | 8 |
| Web of Science | 0 |
| Total | 78 |

Fonte: Dados da pesquisa, 2020.

Após a leitura dos 78 artigos, 45 foram descartados por fugirem à principal questão da RL. Após nova análise nos 33 documentos restantes, foram descartados oito documentos, que não atenderam a todos os critérios de inclusão e exclusão. Então, para a amostra final foram selecionados 25 documentos para análise completa.

g) Categorias de análise dos documentos da amostra

As categorias a serem analisadas estão alinhadas ao objetivo definido para esta revisão, ou seja, buscar informações sobre:

1. implantação de RIs;
2. boas práticas, desafios e soluções para a correção de metadados em lote;
3. elementos computacionais que possam auxiliar na gestão dos RIs para a melhoria do controle de autoridade e da qualidade dos metadados para entidade autor, principalmente aquelas com aplicação no *DSpace*.

2.2 Descrição dos estudos da amostra da RL

Esta seção apresenta os resultados da análise dos artigos selecionados para a amostra, num total final de 25 documentos. A descrição do conteúdo dos documentos da amostra trabalhada está apresentada em ordem crescente de ano de publicação.

O artigo *Personal name identification in the practice of digital repositories* (Artigo 1), publicado em 2006 por Jingfeng Xia, se propôs a analisar algumas características de aplicações de repositório existentes no que se refere à identificação dos nomes dos autores de artigos científicos, particularmente em repositórios digitais. O estudo destaca que “em uma situação ideal, os nomes pessoais são tão capazes de identificar objetos digitais de maneira única como são capazes de identificar pessoas” (XIA, 2006, p. 256).

O autor categoriza as três principais variações de nome: 1) nome completo e abreviado, escrita do nome aparece de forma inconsistente nas publicações; 2) pessoas diferentes que compartilham o mesmo nome ou semelhantes; e, por último, 3) uma mesma pessoa com nomes diferentes, o que acontece por mudança no estado cível, por exemplo. Porém, a quantidade de variações possíveis para um único nome afeta diretamente a qualidade da recuperação da informação, já que diminui a precisão dos dados recuperados. Além disso, segundo o autor, a não determinação correta do autor dificulta a análise de citações, e, nesse ponto, dificulta a avaliação da qualidade e impacto da pesquisa, pois torna pouco confiável os índices de produção dos pesquisadores.

O autor propõe que, ao fazer uma proposta de um RI, o ideal é que os metadados sejam revistos para que sejam incluídos campos que permitam desambiguar os autores. Para os RIs já existentes, segundo Xia (2006), deve-se avaliar a inclusão de novos metadados, mesmo que isso signifique algum nível de risco. Para o autor, o risco é inferior ao benefício, ele destaca a importância do envolvimento da equipe de Tecnologia da Informação (TI) no processo de implementação e manutenção de RIs.

Ele propõe, ainda, um conjunto de elementos de metadados para desambiguação: uso de identificadores compostos que combinem nome do autor, data de publicação e afiliação do autor; solicitar aos autores, no momento do depósito do trabalho, a indicação de todas as variantes possíveis do seu nome.

Em suas conclusões, o autor afirma que a falta de controle da quantidade de variações possíveis para um único nome afeta diretamente a qualidade da recuperação da informação, já que isso diminui a precisão dos dados recuperados. Além disso, ele afirma que dificulta a análise de citações, e, nesse ponto, atrapalha a avaliação da qualidade da pesquisa, pois tornam pouco confiáveis os índices de produção dos pesquisadores.

O artigo *Innkeeper at the Roach Motel* (Artigo 2), publicado em 2008 por Dorothea Salo, bibliotecária e professora da Escola de Informação da Universidade de Wisconsin, em Madison, é, segundo registro do *Google Scholar*, o seu artigo mais citado, com 214 citações (fev./2020). O artigo discute a viabilidade dos projetos de repositórios institucionais, levantando algumas questões que parecem inviabilizar o sucesso deles, analisando três *softwares*: EPrints, lançado 2000, *DSpace*, lançado em 2002, e Fedora, lançado em 2003.

A autora destaca que docentes e pesquisadores não se sentiram atraídos pela tarefa de povoamento dos repositórios, seja em função de não enxergarem os ganhos em visibilidade de suas produções, uma vez que “não ajudam o pesquisador a alcançar estabilidade, promoção ou financiamento” (*on-line*), seja pela frustração no uso do *software* para o autoarquivamento, que, muitas vezes, não é intuitivo, tendo problemas de usabilidade.

Dentre os problemas na gestão de RIs, a autora elenca os seguintes: 1) os *softwares* não oferecem ferramentas de revisão pelos pares; 2) os *softwares* não possuem mecanismos que permitem *feedback* dos pares; 3) recursos financeiros insuficientes; 4) recursos humanos insuficientes; 5) recursos materiais insuficientes; 6) falta de apoio da gestão institucional (alta

direção); 7) falta de apoio local, da equipe de bibliotecários; 8) o *software DSpace*, por exemplo, não pode criar um *link* a partir dos metadados do *Dublin Core*; 9) pouca aceitação do repositório institucional entre os bibliotecários de bibliotecas acadêmicas; 10) falta de suporte técnico por parte dos desenvolvedores de *softwares* livres, o que leva à dependência de listas de discussão específicas de ferramentas; 11) os três *softwares* de código aberto analisados (EPrints, Fedora e *DSpace*) oferecem dificuldades de instalação e manutenção, demandas caras de *hardware*, problemas de customização e desenvolvimento, e mecanismos insuficientes para ajustes entre o *software*, *websites* e serviços já existentes nas bibliotecas; 12) a maior parte do povoamento nos repositórios é mediada por terceiros, e não pelos próprios docentes, ou seja, por bibliotecários, por pessoal de suporte ou por pessoal de TI; 13) o *DSpace* não tem um *workflow* de trabalho que possa permitir um povoamento colaborativo; 14) para ser capaz de promover a visibilidade e medir o impacto da produção do docente e da instituição, o *software* deve ser capaz de fazer relatórios estatísticos, ou seja, contar acessos por autor, por coleção, por item e por arquivo; fornecer informações sobre os acessos ao longo de períodos de tempo definidos; rastrear os *links* de referência e os locais de leitura; permitir participar de conversas *on-line* sobre um dado trabalho; filtrar os acessos ociosos; 15) o *software DSpace* é escrito em Java, em vez de usar uma linguagem mais simples, tal como *Web-scripting* (como PHP ou *Ruby*), o que facilitaria as pequenas modificações pela comunidade de usuários gestores e bibliotecários, ou seja, por aqueles que não possuem conhecimento especializado; 16) o *software DSpace*, mesmo sendo de código aberto, não foi originalmente projetado com um API *plugin*, o que limita a modificação e a expansão de funcionalidades do pacote por terceiros, o que implica que as “mudanças no código requerem reavaliação demorada, correção e, possivelmente, modificação a cada nova versão do *software*” (*on-line*), o que torna a preocupação com os retrabalhos genuína; 17) as modificações feitas por desenvolvedores locais quase sempre não são compartilhadas pelos implementadores de repositório; 18) os erros de povoamento no *DSpace* deverão ser sempre corrigidos pelos administradores do sistema, uma vez que depois que o depositante arquivou o item, não poderá alterá-lo ou a seus metadados, assim como não poderá retirá-lo ou apagá-lo da coleção, sendo que não há uma política clara sobre quais campos de metadados são necessários preencher; 19) poucos bibliotecários são treinados e eficientes desenvolvedores de *softwares*; 20) muitas opções e funcionalidades do *software* ficam escondidas nos arquivos de configuração no servidor do sistema, geralmente controlados pelo pessoal de TI, ficando indisponíveis para os gestores dos repositórios pela falta de uma interface de acesso; 21) no *software DSpace*, a mudança no projeto visual padrão dos repositórios requer

a colocação de arquivos no servidor e depois o reinício do *software* no sistema servidor subjacente, muitas vezes não podendo ser realizado pelo gestor do repositório, exigindo a interferência do pessoal de TI.

Como conclusão, a autora afirma que os problemas não são insolúveis, mas que exigem reconsiderações acerca das missões, estabelecimento de objetivos e metas, definição de expectativas de desempenho, definição de responsabilidades e de curadoria de dados, alocação de tempo e de apoio administrativo, assim como todos os meios necessários para levar adiante os projetos determinados, ainda na fase de planejamento dos repositórios. A autora também alerta para a necessidade de a formação em biblioteconomia incluir conteúdos sobre acesso aberto e RIs no currículo, bem como fortalecer a capacidade dos alunos de escolherem e criticarem *softwares* de forma útil.

Um ano mais tarde, o artigo *Name authority control in institutional repositories* (Artigo 3), publicado em 2009, também por Dorothea Salo apresenta críticas à forma como os RIs foram criados, sem qualquer tipo de controle e o preço que vem sendo pago, por causa disso, ao longo dos anos. Já naquela época, a autora apontava a necessidade de se criarem metadados de qualidade, principalmente os relacionados ao controle de autoridade. A autora atribui a ausência de metadados de qualidade às formas de povoamento dos RIs: autoarquivamento e autopovoamento em lote. O primeiro, por ter como responsável pela submissão os próprios autores, e, o segundo, pela dificuldade de se preverem todas as variações possíveis de nomes espalhados em diferentes fontes.

Assim, a autora identifica como obstáculo para resolver os problemas de controle de autoridade: normas, *software* e fontes de informação para nomes. Em relação às normas, Salo (2009) entende que há ausência de regras para o controle de autoridade nos esquemas de metadados disponíveis para RI, mais especificamente o *Dublin Core*, que não prevê alternativas para o registro de autoridade que não seja autorizado, e não oferece um campo para descrever as variantes de um nome ou afiliação institucional, o que facilitaria a desambiguação dos nomes.

O *DSpace* é um dos *softwares* apontado pela autora como mais um obstáculo para o controle de autoridade, pois, além de adotar o *Dublin Core* como padrão de metadados, o programa é engessado por não facilitar a correção dos metadados em lote. Ademais, o *software* demanda o trabalho de um profissional especializado para tal tarefa. E, por último, as fontes de informação para nomes são outro obstáculo importante, pois os *softwares*, mesmo que permitam a consulta

a arquivos de autoridade externos, em geral, esses arquivos não atendem à demanda, já que os autores nele descritos se limitam, quase que exclusivamente, a autores de livros e não contemplam as autoridades locais. E mesmo que as instituições criem listas locais de autoridades, o fato de cada local controlar a sua própria lista de autoridade já se torna um novo e diferente problema. E, ainda assim, não seria uma solução, pois o *Dublin Core* não tem um campo que permita a descrição do vínculo institucional para a autoridade.

Os fundamentos sobre os quais o protótipo desenvolvido pelo Projeto Nomes do JISC foi baseado são: os Requisitos Funcionais para Dados de Autoridade (FRAD), interoperabilidade, reutilização de dados e intervenção manual mínima podem apontar o caminho para a evolução nesse campo de estudo. Segundo a autora, nessa área temática, a tendência de interfaces com a capacidade de permitir buscas e navegação multifacetadas levou ao desenvolvimento de um protótipo no qual, com base na representação do *Library of Congress Subject Headings* (LCSH) em *Simple Knowledge Organization System* (SKOS), uma ontologia foi projetada para auxiliar os usuários na escolha do termo correto, relacionando, automaticamente, as palavras-chave de entrada com os cabeçalhos de assunto apropriados no sistema (Papadakis *et al.*, 2009). Esta parece ser uma opção viável entre aquelas que visam alcançar a interoperabilidade sintática e semântica.

O artigo *Defining Best Practices in Electronic Thesis and Dissertation Metadata* (Artigo 4), publicado em 2009 pela bibliotecária Rebecca L. Lubas teve como objetivo analisar as melhores práticas para a criação de metadados de teses e dissertações, com o uso do *software DSpace*. O estudo foi realizado na Universidade do Novo México, que estava no processo de implantação do RI. A autora destaca quatro melhores práticas:

1) dois registros de metadados são melhores que um: as bibliotecas já possuem equipes de bibliotecários treinados para a criação de registros em *Machine Readable Cataloging MARC/AACR2*, e que a inclusão das teses e dissertações no catálogo tradicional, além do RI, ampliaria a capacidade de recuperação dos registros, já que esses catálogos tradicionais, normalmente, integram catálogos coletivos nacionais. Além disso, os metadados básicos fornecidos pelo autor, ao serem descritos da forma tradicional, deveriam receber padrões e controles de vocabulários, que são negligenciados fora da biblioteca tradicional. Ademais, a integração com catálogos coletivos deveria permitir à “biblioteca executar trabalhos de autoridade, normalização e desambiguação de nomes de autores” (p. 259);

2) melhorar os metadados gerados pelo autor, uso de listas pré-definidas de nomes de departamentos e professores: no *DSpace*, isso pode ser resolvido com a inclusão de listas pré-definidas, ao invés do preenchimento livre do campo. Essas soluções fornecem metadados consistentes e buscas mais precisas. Sugere, ainda, que os autores recebam um treinamento, para conscientizá-los sobre problemas básicos como o uso de maiúsculas e minúsculas e também é uma oportunidade de incentivá-los a fornecerem mais dados como palavras-chaves e informações adicionais sobre os nomes;

3) aprimorar os metadados do *Dublin Core*: propõe que os bibliotecários catalogadores utilizem sua expertise em MARC/AACR2 também ao descreverem os registros no *DSpace* com a utilização do DC qualificado (*dc.subject.lcsh*). Sugere ainda que os dados fornecidos pelo autor sejam mantidos (exceção para erros de grafia) e aprimorados;

4) investir em treinamento, pois, apesar de simples, o *Dublin Core* exige treinamento, principalmente, porque sua ausência pode acarretar confusão na descrição dos elementos. Além disso, é importante estar atento às tendências de comportamento dos usuários, das ferramentas de pesquisa e dos métodos de coleta por trás deles.

O artigo *Experiences in Deploying Metadata Analysis Tools for Institutional Repositories* (Artigo 5), publicado em 2009 por David M. Nichols, Gordon W. Paynter, Chu-Hsiang Chan, David Bainbridge, Dana McKay, Michael B. Twidale e Ann Blandford apresentam um estudo comparativo de duas ferramentas de análise de metadados para repositórios: a Ferramenta de Análise de Metadados (MAT), da Universidade de Waikato, e a ferramenta Kiwi Serviço de Informações de Pesquisa (KRIS), da Biblioteca Nacional da Nova Zelândia.

O KRIS foi criado como um serviço de descoberta a pesquisa em RIs da Nova Zelândia, que coleta dados a partir do OAI-PMH e armazena três tipos de dados: metadados *Dublin Core*, metadados internos e metadados administrativos do Repositório. Este último é o responsável por descrever a qualidade dos metadados coletados “cada registro possui zero ou mais campos de metadados administrativos [...] e cada um deles identifica um erro de metadados específico [...] uma condição que falha explicitamente em atender a um requisito das diretrizes de metadados, como um campo obrigatório ausente” (*on-line*). Essas informações levantadas ficam à disposição para consulta e edição e os dados podem ser manipulados de forma a criar outras ferramentas. De forma automática, os erros gerados podem ser localizados pelo gestor sob demanda, que pode realizar uma seleção de tipo de erro por coleção, tipo de erro ou conjunto

de erros. O gestor pode ainda se cadastrar no RSS e receber avisos e notificações dos erros. Além disso, o KRIS pode fornecer estatísticas sobre a qualidade dos registros em diferentes instituições.

O MAT foi projetado para uso do *Greenstone*, apesar de ter sido pensado para esse *software*, foi projetado em ambiente Web para não limitar seu uso apenas aos usuários do *Greenstone*. Essa ferramenta, após análise, gera tabelas de visualização com informações sobre valores duplicados (de acordo com o campo de metadados selecionado); com uso incorreto de espaços e pontuação (que leva a valores duplicados); informações de campos vazios, entre outras possibilidades. O uso dessas ferramentas pode permitir ao gestor do RI, a partir dos relatórios de erro, definir quais os metadados merecem sua atenção e impactam diretamente sobre o uso do RI. Atualmente, os RIs recebem registros de metadados importados e a ferramenta forneceria informações sobre os erros de incompatibilidade entre campos de metadados do *Dublin Core* ou de outros esquemas.

Mesmo que essas ferramentas não sejam aplicáveis ao contexto desta pesquisa, o artigo foi incluído por apresentar a importância de ferramentas de análise de metadados para repositórios institucionais que produzem visualizações e auxiliam os gestores a conhecerem seus metadados para, assim, traçarem a melhor estratégia de gestão/correção.

No artigo *Re-Discovering Repository Architecture: Adding Discovery as a Key Service* (Artigo 6), publicado em 2009 por Peter Sefton, o autor sugere “que deveríamos trabalhar arduamente para ampliar a visão do repositório, voltando a algo no espírito da definição de Lynch como um conjunto de serviços” (p. 85), o que não se resume apenas ao *software* instalado.

Ele propõe a adoção de mais de um aplicativo para auxiliar na gestão dos RIs, que possa ser complementado pelas funcionalidades do *software*, que auxiliem na gestão, usando, por exemplo, um grande índice ou sistema de descoberta. Ele indica a criação de índices independentes, inteligentes e flexíveis, que possam ser utilizados independentemente do *software* utilizado. Para o autor, a identificação de pessoas, com identificadores confiáveis e persistentes, é importante, e, se nas bibliotecas isso em geral é feito por listas de autoridades, com a indicação de uma forma canônica para o nome, nos RIs isso também deveria ser feito. Como nas bibliotecas tradicionais, essa é a parte mais cara da catalogação, conforme afirma o autor, pois também nos RIs há um alto custo no controle dos dados de autoridade para nomes.

Ademais, o autor afirma que nem todo *software* de gestão de RIs possui funcionalidade que permita gerenciar as IDs de todas as pessoas do repositório, associando um nome às suas diferentes formas. Assim, muitas vezes a gestão dos nomes é feita à mão.

O artigo *Author name disambiguation* (Artigo 7), publicado em 2009 por Smalheiser e Torvik, descreve o processo de desambiguação de nomes e os desafios que o envolvem: um indivíduo pode publicar sob vários nomes e isso inclui (variantes ortográficas, mudança de nome - casamento, conversão religiosa, mudança de sexo e pseudônimo); indivíduos com o mesmo nome (homônimos); metadados incompletos ou inexistentes; documentos com autoria múltipla ou institucionais. Os autores ressaltam a importância do processo para: pesquisadores em busca de parceiros e suas produções; agência de fomento para rastrear possíveis conflitos de interesse e para rastrear a produção de pesquisadores e pesquisas patrocinadas; mapear domínios do conhecimento e fazer análise de citações. Eles apresentam duas soluções de desambiguação: criar um registro único por autor, com ID, que agrupariam as publicações, e o processo manual.

A primeira, atualmente, pode ser comparada ao ORCID, no entanto, os autores apresentam como desvantagens contar com o autor para a atualização constante desses dados de maneira voluntária e ativa, com precisão e periodicidade por longos anos; se muitos autores têm apenas um único artigo publicado, e, assim, não participariam de uma iniciativa como essa; além disso, não haveria a garantia de que a instituição responsável por esses registros os manteria ao longo do tempo e que se responsabilizaria com todas as questões envolvidas no processo.

A segunda, desambiguação manual, seria contar com um processo que existe nas bibliotecas tradicionais, com o uso de catálogos já conhecidos tais como o mantido pela *Library of Congress* e a *Online Computer Library Center (OCLC)*, *Virtual International Authority File (VIAF)*. Para os autores, o processo manual tem como desvantagens: processo difícil, incerto e inviável mesmo em pequena escala. Os autores mencionam iniciativas de processo de desambiguação com algoritmos e acreditam que soluções com uso de modelos de aprendizagem pela máquina são alternativas plausíveis para o processo.

No artigo *Dublin Core, DSpace, and a Brief Analysis of Three University Repositories* (Artigo 8), publicado em 2010, Mary Kurtz faz uma análise comparativa de três repositórios que utilizam *DSpace*, de acordo com os princípios de qualidade dos metadados: completude, consistência e precisão. E avaliou os campos mais utilizados na busca: título, autor, assunto e resumo.

A autora conclui que os campos gerados de forma automática têm mais consistência e contribuem para a integridade dos dados. Porém, alguns campos ainda dependem do preenchimento não automático, pelos autores, por exemplo. Segundo a autora, é nesse ponto que está a maioria dos problemas, com erros na entrada dos dados, ausência de controle de autoridade e confusão sobre o entendimento dos campos de metadados que resultam em baixa precisão dos metadados.

Ela destaca que, das três universidades pesquisadas, a que obteve melhores resultados para a qualidade no campo assunto tinha revisão dos metadados revista por um bibliotecário antes do carregamento no repositório. Assim, a autora sugere que, por enquanto, talvez essa última opção seja a melhor alternativa para uma qualidade aceitável dos metadados nos repositórios.

No artigo *ACCEDA, repositorio institucional de la ULPGC: mejoras y desarrollos propios* (Artigo 9), publicado em 2011, Maria Pilar López Díez e Alejandro Melián Vinuesa relatam o processo de implantação do RI ACCEDA, usando o *software DSpace*.

Os autores informam que, em 2010, houve a mudança para a versão 1.6.2 do *DSpace*, por causa de novas funcionalidades oferecidas, especialmente o controle da autoridade. Eles destacam que esse controle é essencial para o RI,

pois ajuda a estabelecer um formulário único para os pontos de acesso de autor, garantindo a consistência e permitindo a recuperação de todas as obras do mesmo autor”, e, assim, “otimiza consideravelmente a recuperação de informações e também ajuda a evitar a duplicação e o ruído nas buscas (LÓPEZ DÍEZ; MELIÁN VINUESA, 2011, p. 131).

Uma vez que o controle de autoridade na versão 1.6.2 é baseado na lista de autoridades da *Library of Congress* (LC – Biblioteca do Congresso), os autores relatam que tiveram que encontrar uma solução mais adequada às especificidades do RI Acceda. Depois de várias tentativas,

e para não forçar o usuário a escolher uma forma particular de escrever seu nome, foi decidido utilizar o próprio banco de dados de autoridades da Acceda, baseado no uso de um controle AJAX (3), que mostra uma lista das autoridades existentes a partir da introdução do quarto caractere na caixa de busca (LÓPEZ DÍEZ; MELIÁN VINUESA, 2011, p. 131-133).

Assim, no momento da submissão, ao digitar o quarto caractere do nome, o sistema do *DSpace* retorna à lista de autoridades existentes para seleção. Além disso, deixam em destaque instruções para os autores de como preencherem os dados conforme as orientações da *Fundación Española para la Ciencia y la Tecnología* (FECYT), *International Registry of*

Authors-Links to Identify Scientists (IRALIS) e ResearcherID. Segundo os autores, “o sistema é aberto, já que cada vez que um documento com uma nova autoridade é arquivado, ele se torna parte do banco de dados”, e a “criação do controle de autoridade não só foi desenvolvida no sistema de autoarquivamento, mas também implementada na ferramenta de povoamento delegado [realizada pela biblioteca central]” (p. 132).

O artigo *A Study Of Authority Control in Spanish University Repositories* (Artigo 10), publicado em 2012 por Almuzara, Alvite Díez e Bravo, se propôs a estudar os repositórios espanhóis e avaliar o controle de autoridade realizado. Em uma análise em 26 RIs, as autoras concluíram certa desigualdade no controle de autoridade, a maioria não possui vocabulários controlados ou formas de se normalizar os registros neles depositados.

As autoras destacam que o padrão de intercâmbio OAI-PMH exige que os metadados sejam providos em *Dublin Core*, que não “incorpora mecanismos para controle de autoridade e não oferece a facilidade de distinguir entradas controladas ou de harmonizar variantes de um determinado ponto de acesso” (p. 97). Elas citam algumas iniciativas para o controle de autoridades, a saber: o Arquivo de Autoridade Internacional Virtual (VIAF); *People Australia*; ISNI *International Standard Name Identifier* (o esboço do ISO 27729), I (Identificadores Institucionais do NISO) e ORCID (*Open Researcher and Contributor ID*) para projetos específicos de controle de autores; e iniciativas para nomes corporativos, tais como: o Projeto pioneiro de Nomes, financiado pela JISC no *United Kingdom*, e o *Cooperative "Identities Hub"* da OCLC. Elas afirmam que a maior parte dessas iniciativas é baseada em cooperação entre bibliotecas ou instituições, tendo em vista “reduzir custos, reduzir esforços, evitar a duplicação e aumentar a utilidade dos serviços” (p. 97), e ressaltam que o trabalho cooperativo é essencial para o sucesso de projetos de controle de autoridade.

As autoras destacam que a versão 1.6.0 do *DSpace* já trouxe um módulo específico para o controle de autoridade, dando “a opção de escolher uma única entrada ou valor entre várias que são possíveis” (p. 97), apoiando a inserção de metadados corretos e “a interoperabilidade com outras aplicações fora do programa, tais como o banco de dados do pessoal da instituição e similares” (ALMUZARA; ALVITE DÍEZ; BRAVO, 2012, p. 97).

Sobre o controle de pontos de acesso para autores pessoais, as autoras destacam que a “longa tradição da biblioteconomia em padronizar tais detalhes de autoridade parece estar refletida nas respostas fornecidas em relação à implementação deste indicador” (p. 98), pois 23 RIs adotam

modelos de autoridade pré-definidos. Elas perceberam que os arquivos com maior incidência de padronização em suas entradas são de membros da própria comunidade acadêmica, para os quais as informações, em sua maioria, são obtidas nos sistemas administrativos da instituição, homogeneizadas a partir dos pontos de acesso no catálogo da biblioteca da universidade (muitas vezes o conteúdo do catálogo foi integralmente levado para o RI).

Ao final, as autoras concluem que a tarefa de gerir autoridades em RIs demanda a criação de políticas eficazes e uma gestão cooperativa que auxilie na criação dos registros de autoridade e “nos processos de catalogação, criação de metadados e recuperação da informação” (p. 101). Elas também apresentam, como “caminho para a evolução neste campo de estudo” (p. 101), o protótipo do projeto *Names of JISC*, desenvolvido em colaboração com a Biblioteca Britânica, que se baseia no FRAD, em “interoperabilidade, reutilização de dados e intervenção manual mínima” (ALMUZARA; ALVITE DÍEZ; BRAVO, 2012, p.101).

O artigo *Evolving Landscape in Name Authority Control* (Artigo 11), publicado em 2013 por Jinfang Niu, descreve seis tipos de abordagem que auxiliam as bibliotecas na desambiguação de nomes, por exemplo, obras de um autor que usa vários nomes e/ou várias formas de um nome. O autor propõe que as bibliotecas utilizem identificadores (ID) para vincular registros bibliográficos e registros de autoridade, tendo em vista facilitar a atualização de cabeçalhos autorizados. Isso se justifica porque os títulos de nomes autorizados não são estáveis, por vários motivos, entre eles: alteração das regras de controle de autoridade, alteração ou variantes de nomes. Assim, mesmo que o cabeçalho da autoridade sofra alguma alteração, o identificador permanecerá o mesmo, sem a necessidade de atualização dos registros vinculados à autoridade. O autor, no entanto, compreende que esse identificador seria local, e que, se compartilhado, precisaria ser substituído pelo cabeçalho de autoridade. O autor destaca a iniciativa da Federação Internacional de Associações e Instituições de Bibliotecas (IFLA), de 1979, que propôs o estabelecimento de um Número Internacional de Dados de Autoridade Padrão (ISADN). Porém o projeto não foi continuado por dificuldade de implantação.

Niu aponta a iniciativa do *Virtual International Authority File* (VIAF), que agrega arquivos de autoridade nacional e regional em um banco de dados central e atribui um ID, para, em seguida, mesclar esses registros para um mesmo autor, atribuindo um ID exclusivo a cada registro de autoridade. No entanto, essa alternativa também apresenta problemas, pois, se houver imprecisão nos registros de autoridade, cada um receberá um ID. Caso o problema seja

detectado, os IDs originais poderão ser desabilitados e um novo é criado. Se essa substituição acontecer, o ID atribuído ao registro pelo RI será quebrado. Ainda assim, o autor defende o uso de identificadores para aplicações, por exemplo, na Web Semântica.

Outra iniciativa citada pelo autor é o *International Standard Name Identifier* (ISNI), que atribui um ISNI para cada responsabilidade/identidade do autor. Ademais, cita o *Open Researcher and Contributor ID* (ORCID), que atribui IDs para autores, sendo que o código ORCID pode ser incorporado às publicações do autor. O autor enumera outras iniciativas que variam em abrangência e propósito: *OCLC Cooperative Identity Hub*; *Scopus Author Identifier*; *Project Names*; Sistema de informações de pesquisa atual na Noruega (CRIStin); Comitê Conjunto de Sistemas de Informação Identificador de pesquisador (JISC); e o Lattes no Brasil.

Ele levanta a questão de que de nada adianta ter várias ferramentas para o controle de autoridade com atribuição de ID, se uma mesma autoridade puder ter mais de um ID. Para evitar tal problema, o autor propõe um sistema de ID global, que comporte e incorpore dados dos outros sistemas, seja interoperável e aceite todos os padrões de metadados. Reconhece, porém, que esse grande sistema global não será suportado por um único banco de dados centralizado. Assim, o sistema global deve ser vinculado aos bancos de dados de autoridade locais. Segundo o autor, isso ampliaria a cobertura do controle de autoridade para uso em outros ambientes que não apenas nas bibliotecas, repositórios e arquivos.

No artigo *Developing an institutional repository: experiences at the library, Faculty of Medicine, University of Jaffna* (Artigo 12), publicado em 2014, Murugathas e Balasooriya relatam a experiência na implantação do RI com o *software DSpace* na Faculdade de Medicina, Universidade de Jaffna. Como desafios, destacam-se: desconhecimento do pessoal da biblioteca no que se refere às Tecnologias de Informação e Comunicação (TIC), isso inclui linguagem de programação, instalação e customização e conhecimento do sistema operacional.

Segundo os autores, se a equipe da biblioteca não possuir esses conhecimentos, será preciso contratar um profissional capacitado, principalmente, para a fase inicial de instalação. Os autores também reforçam a parceria que deve existir entre a biblioteca e a equipe de tecnologia da instituição.

Ademais, afirmam que a política de submissão e de direitos autorais também são imprescindíveis para a implantação do RI. Contar com o apoio da comunidade acadêmica para

o depósito e autoarquivamento são apontados como desafios. E os autores destacam que é papel do pessoal da biblioteca convencer a comunidade acadêmica sobre a importância do depósito e do quanto isso pode contribuir para a visibilidade da pesquisa e reputação dos autores, além de oferecer serviços de mediação que facilitem a vida do pesquisador. Apontam ainda a importância da padronização de dados, como nome de autor e instituição e a pertinência de se ter um arquivo de autoridade local para membros institucionais. Murugathas e Balasooriya (2014) ainda indicam a possibilidade de uso de *link* para a página pessoal dos autores ou para o sistema acadêmico no repositório.

Por fim, reforçam que um RI sustentável precisa de apoio institucional a longo prazo, recursos financeiros e pessoais, e a carga de trabalho deve ser incorporada às rotinas de trabalho da biblioteca. Ademais, segundo os autores, a oferta de treinamentos deve ser contínua.

No artigo *Transferring Cataloging Legacies into Descriptive Metadata Creation in Digital Projects: Catalogers' Perspective* (Artigo 13), publicado em 2014, Junli Diao e Mirtha A. Hernández descrevem o papel dos catalogadores no contexto digital e de que forma o legado da catalogação tradicional pode contribuir com o controle de qualidade e de autoridade na criação de metadados em projetos digitais. Os autores entendem que os catalogadores devem se apropriar da criação de metadados e acreditam que essa é uma extensão natural das atividades já desenvolvidas na biblioteca. Eles indicam que a qualidade dos metadados utiliza os mesmos critérios de avaliação da criação de registros bibliográficos: precisão, consistência e integridade, e a padronização da catalogação deve ser aplicada no contexto digital. Da mesma forma que na biblioteca tradicional, em contexto digital a catalogação continua sendo um processo intelectual.

Para os autores, os bibliotecários estão em vantagem se comparados a outros profissionais, pois já conhecem documentações complexas, tais como os códigos de catalogação, e isso “dará aos catalogadores a vantagem de interpretar e executar as diretrizes de aplicativos de metadados para obter precisão, consistência e integridade em projetos digitais” (p. 4). Sendo assim, os catalogadores devem, “progressivamente, assumir um papel de liderança, incorporando as habilidades dos padrões e práticas tradicionais de catalogação na criação de registros de metadados em projetos digitais” (p. 4).

No que se refere ao controle de qualidade, sugerem que os catalogadores incrementem os metadados descritivos e que utilizem listas de autoridade externas ou locais, já que nem todas

as entidades serão localizadas em listas externas. Ressaltam que a lista de autoridade local é uma ferramenta útil e que pode complementar o trabalho de catalogação e, assim, fortalecer o processo de recuperação pelo usuário.

Finalmente, os autores propõem que os catalogadores sejam criativos. Isso porque no ambiente digital existe uma diversidade de materiais a serem catalogados e o catalogador deve ser criativo para criar registros ricos e completos. Além disso, os materiais são exclusivos e únicos, sem muitos recursos de consultas a catálogos coletivos ou dados compartilhados, por exemplo. Nesse sentido, a catalogação deve ser criativa, na tentativa de antecipar a demanda dos usuários, que não podem ser vistos como consumidores passivos e precisam ser convidados a participarem, por exemplo, da revisão dos registros.

O artigo *La visualización de autores en un repositorio institucional a través del enfoque Model Driven con WebRatio* (Artigo 14), publicado em 2016, por Jose Texier, Marisa R. De Giusti, Gonzalo L. Villarreal e Ariel J. Lira propõe uma aplicação *Web* para gerenciar os dados de autoridade. A aplicação foi implantada no RI da *Universidad Nacional de La Plata* (SEDICI) com a utilização do *software* comercial WebRatio, que incorpora a metodologia *Model-Driven*.

Os autores defendem que esse tipo de aplicação é vantajosa, pois uma mesma aplicação pode atender as várias partes envolvidas na gestão do repositório, independente do sistema que utilizam (biblioteca, sistema acadêmico e departamento pessoal, por exemplo). Segundo os autores, a “entidade abstrata autora no SEDICI, são os metadados dos recursos que permitem identificar o criador intelectual do recurso, seja ele uma pessoa ou uma organização” (p. 70), e o autor “é considerado uma entidade abstrata no SEDICI porque tem suas próprias informações descritivas” (TEXIER *et al.*, 2016, p. 70).

Texier *et al.* (2016) afirmam que há diversas formas para a padronização de autoridades, e, assim, o controle da entidade “autor” deve ser completo (incluindo nomes e outras informações), independente do *software* de gestão adotado pelo RI, que deve ter funcionalidade que permita o controle desses dados. Sendo assim, o RI SEDICI utiliza o *software* DSpace, mas a proposta dos autores - o WebRatio - pode ser executada em qualquer outra plataforma.

O artigo *High-Quality Metadata and Repository Staffing: Perceptions of United States-Based OpenDOAR Participants* (Artigo 15), publicado em 2016, por Heather Moulaison Sandy e Felicity Dykas apresenta uma pesquisa com os gestores de RIs dos EUA, que têm os seus dados

coletados pelo *OpenDoar*. A intenção dos autores foi compreenderem se esses gestores colocaram em prática os padrões de qualidade de metadados para promoverem a interoperabilidade dos dados do RI.

Para os autores, avaliar a qualidade de metadados é problemático, “apesar da criação de indicadores ou *benchmarks* [...] devido à subjetividade da noção e à fluidez da aplicação” (*on-line*), pois os gestores e usuários de RIs somente “são capazes de identificar a qualidade para seus próprios propósitos” (*on-line*). Segundo os autores, os respondentes indicaram que “estão tomando medidas para alcançar a qualidade dos metadados, inclusive empregando profissionais da informação e outros especialistas para criar e revisar metadados e para selecionar documentação e melhores práticas às quais os repositórios devem aderir” (*on-line*). E ao não adotarem políticas e boas práticas para criação dos metadados, acabam sobrecarregando o sistema de busca do repositório.

Ao final do estudo, os pesquisadores apontaram que a qualidade dos metadados está diretamente relacionada a uma equipe “adequada e apropriada” para o RI, pois é uma necessidade crítica. Os resultados do estudo indicaram como obstáculos à criação de metadados de qualidade: equipe insuficiente; limitações de tempo da equipe, disponível para a tarefa; falta de habilidades que impedem que a equipe se envolva mais diretamente no suporte à criação dos metadados; e, apesar de os gestores serem bibliotecários, alegaram não fazerem uso de padrões ou de melhores práticas.

Os autores acreditam que esses problemas podem ser sintetizados por falta de experiência, tempo e treinamento (educação continuada) da equipe. Para eles, uma equipe eficaz e eficiente, equipada com os recursos corretos, normas que promovam a interoperabilidade e sistemas não cooperativos, poderiam suprir os problemas com ferramentas de gestão de RIs imperfeitas.

O artigo *Metadata for Electronic Theses and Dissertations: A Survey of Institutional Repositories* (Artigo 16), publicado em 2016, por Tom Steele e Nicole Sump-Crethar, apresenta um estudo realizado em RIs nos EUA, para conhecer as políticas de metadados e de controle de qualidade que são utilizados, principalmente, para descrever documentos do tipo teses e dissertações. Destaca-se o resultado mais relevante para a pesquisa: 59% das bibliotecas pesquisadas responderam que os metadados eram revistos por catalogadores ou outra equipe, independente de quem foi o responsável pela criação dos metadados, mas essa porcentagem cresce para 74% se forem considerados somente os RIs que usam o autodepósito.

A maioria dos pesquisados, representados por 88% do total, compreendem que a qualidade dos metadados nos RIs é importante, sendo que dessa porcentagem, 29%, entendem a qualidade como importante e 59% a consideram extremamente importante. No entanto, 61% dos respondentes alegaram não fazerem uso de qualquer vocabulário controlado para descreverem o campo destinado ao ponto de acesso de assunto, que é preenchido com as palavras-chave atribuídas pelo autor. A maior parte dos pesquisados alega que se tivesse mais tempo e recurso faria mais pelo controle de qualidade.

Os autores concluem que parece haver um consenso acerca dos metadados a serem adotados em 90% dos RIs, quais sejam: título, resumo, tipo de publicação, autor e assunto. O metadado de autor foi adotado em quase 100% dos RIs, exceto em dois: um que usa "contribuidor.autor", e outro que usa "contribuidor.conselheiro" (para conselheiro), situação que é provável ocorrer em outros RIs, segundo os autores. E as plataformas de *software* para RI que se destacaram foram o *Digital Commons* e o *DSpace*.

No artigo *Achieving and Maintaining Metadata Quality: Toward a Sustainable Workflow for the IDEALS Institutional Repository* (Artigo 17), publicado em 2017, Ayla Stein, Kelly J. Applegate e Seth Robbins descrevem os fluxos de trabalho criados no RI denominado *Digital Environment for Access to Learning and Scholarship* (IDEALS), da Universidade de Illinois em Urbana-Champaign, tendo em vista melhorar a qualidade dos metadados antes do carregamento de lotes de dados, evitando-se novos erros de metadados, assim como limpar os metadados antigos. Para tanto, utilizaram uma aplicação de *software* de código aberto *OpenRefine*, em lugar de revisarem e corrigirem metadados dentro da própria interface do RI. O RI da IDEALS possui metadados mistos, criados por humanos, autoarquivamento ou autodepósito e o autopovoamento (povoamento automático).

Os autores destacam que, mesmo a Universidade possuindo políticas de metadados, vocabulários controlados e um guia de boas práticas, o RI ainda possui erros nos registros. Os autores propõem que os metadados sejam revistos e limpos pelo menos uma vez por semestre, uma vez que o RI usa o *software DSpace*, que possui uma função para isso, o *MetadataValueUpdate*. Mesmo sendo considerada limitada, os autores afirmam que ela pode auxiliar no processo de correção e limpeza. Para o caso em questão, o projeto usa, em função do volume de dados, o *OpenRefine*. Outra ferramenta descrita pelos autores e adotada no projeto, foi a *OpenOffice Calc*, utilizada para limpar dados do título, para, por exemplo,

padronização entre maiúsculas e minúsculas. Os autores destacam que “a grande maioria dos erros foi introduzida quando os lotes de materiais foram carregados devido à falta de procedimentos abrangentes de revisão de garantia de qualidade” (p. 644), ou seja, falta de políticas de povoamento. Segundo os autores, o uso dessas ferramentas traz benefícios, com a melhora significativa da qualidade e consistência dos metadados.

Os autores concluem que a ausência de ferramentas úteis de gestão de metadados inerentes aos *softwares* dos RIs, obriga o uso de aplicações complementares e isso dificulta a rotina do fluxo de trabalho nos RIs. Ao final, eles concluem que é preciso uma revisão sistemática e regular dos metadados, a elaboração e implementação de um cronograma de limpeza de dados no RI, que devem estar registrados nas políticas oficiais do IDEALS.

No artigo *Evaluation of Spanish institutional repositories based on criteria related to technology, procedures, content, marketing and personnel* (Artigo 18), publicado em 2018, Rocio Serrano-Vicente, Remedios Melero e Ernest Abadal propuseram uma análise baseada em cinco categorias (tecnologia, procedimentos, conteúdo, marketing e pessoal) para avaliarem os RIs espanhóis, sendo que os resultados para a categoria "conteúdo" não serão apresentados, por se tratar de metadados de assunto.

No quesito tecnologia, os resultados indicaram que o *DSpace* é o *software* mais utilizado e 65% dos RIs importavam e exportavam dados para outros sistemas da universidade. E 41% ofereciam aos usuários *Altmetrics*, recurso que permite medir o impacto do objeto digital através das redes sociais, possibilitando ao usuário do RI compartilhar os documentos no *Facebook* e *Twitter*, por exemplo. Segundo os autores, as métricas, ao serem introduzidas no RI, incentivam a comunidade acadêmica a depositar seus trabalhos e conhecer o impacto de suas produções. Além disso, as métricas são ferramentas úteis aos bibliotecários para analisarem o RI.

Em relação aos procedimentos, destaca-se, entre outros resultados, a existência de manuais que descrevem como “inserir, validar, vincular” documentos nos repositórios. Isso inclui tutoriais em vídeo, ferramentas de suporte para depósito, FAQ e instruções sobre como inserir dados bibliográficos. Além disso, a pesquisa revelou que, em 83% dos RIs pesquisados, são os bibliotecários que verificam os registros antes da publicação definitiva no repositório. A atribuição de um ID exclusivo para autores não era uma alternativa adotada entre os

pesquisados, e 25% dos RIs usam o ORCID; apesar de entenderem a importância, ainda estão trabalhando na implantação.

A categoria *marketing* analisa de que forma os RIs são promovidos interna e externamente. Segundo os autores, internamente, o destaque da promoção cabe aos treinamentos, documentos de suporte, redes sociais e páginas de ajuda, além do apoio das autoridades acadêmicas e do depósito mediado tendo os bibliotecários como parceiros nessa tarefa. Essa mediação é realizada para que não se tome tempo dos autores e os desmotive a realizarem o depósito. Os autores ressaltam que algumas instituições adotam estratégias de incentivo aos autores, seja com recursos financeiros, seja vinculando o depósito no RI ao progresso da carreira. Destacam também a importância de envio aos autores de relatórios estatísticos de *downloads* e visitas aos documentos depositados por eles. A maioria dos RIs pesquisados não adota essa estratégia, alegando que as estatísticas são públicas e podem ser consultadas pelos autores. No entanto, acreditam que essas métricas devem ser tratadas no nível do autor, primeiro, para convencê-los da importância para o RI e para a visibilidade da pesquisa deles, e, segundo, que esses relatórios podem ser ferramenta para a concessão de prêmios de reconhecimento.

A categoria “equipe destaca a importância da parceria entre bibliotecários e a equipe de tecnologia da instituição. E, na maioria dos casos, a instituição não tem uma equipe dedicada em tempo integral ao RI.

O artigo *Extending Name Authority Work beyond the Cataloging Department: A Case Study at the University of Nevada, Reno Libraries* (Artigo 19), publicado em 2018 pelas bibliotecárias Dana M. Miller e Amy Jo Hunsaker, apresentam a proposta de criação de um fluxo de trabalho como ferramenta para estender o trabalho de autoridade para além dos catalogadores, os não-MARC, sobremaneira, em coleções especiais, arquivos e coleções digitais. Segundo as autoras, “é importante para aqueles que criam metadados consultar e usar os bancos de dados cooperativos disponíveis, como o *Connexion* da OCLC e o LCNAF (*Library of Congress Name Authority File*)” (p. 136). Assim, tanto os catalogadores profissionais quanto os pesquisadores (não-MARC) devem ser motivados a participarem do processo, para que a “busca e recuperação desses materiais ocorram com precisão e qualidade”, e, “para garantir isso, é vital aderir aos padrões das autoridades nacionais de nomes” (MILLER; HUNSAKER, 2018, p. 136).

As autoras destacam que as coleções especiais e as digitais vêm sendo

compartilhadas com públicos mais amplos, por meio de consórcios de arquivos regionais, tais como o Arquivo Online da Califórnia e o *Archives West*, e agregadores digitais, tal como a Biblioteca Pública Digital da América (DPLA), tornando-se responsável de cada instituição fazer com que seus materiais, e os criadores desses materiais, sejam descobertos em nível nacional e internacional (MILLER; HUNSAKER, 2018, p. 136-137).

Elas afirmam que as instituições individuais devem distinguir os seus autores particulares e compartilhar o produto em bancos de dados já existentes, a exemplo do LCNAF, que coletivamente compila, mantém e distribui informações sobre autores, podendo reduzir a carga de trabalho e evitar retrabalhos dos criadores de metadados locais.

As autoras expõem um estudo de caso, realizado na Universidade de Nevada (UNR), no Departamento de Metadados e Catalogação de Reno, da implementação de um fluxo de trabalho inclusivo, para catalogadores, pesquisadores e estudantes, para a criação de metadados para o campo de autores. Essa ferramenta permite “coletar e rastrear os nomes propostos e seus pontos de dados numéricos associados, que são, então, encaminhados a um catalogador centralizado e designado para criar registros de autoridade de nomes” (p. 147), minimizando os riscos de duplicação de esforços e de erros de povoamento. As autoras concluem que, ao se criarem registros de autoridade locais, para serem compartilhados com outras instituições, é importante que sigam os padrões internacionais para registros. E a criação dos registros pode ser cooperada e envolver outras áreas da instituição, não somente os catalogadores, desde que se crie um fluxo de trabalho e um formulário de informações sobre autores de preenchimento mais acessível, que não fiquem limitados aos jargões dos catalogadores, sobretudo sobre os elementos MARC e aplicação do RDA.

No artigo *A Case Study of ETD Metadata Remediation at the University of Houston Libraries* (Artigo 20), publicado em 2019, Santi Thompson, Xiping Liu, Albert Duran e Anne M. Washington relatam o processo de correção de metadados em lote, de registros de teses e dissertações no Repositório da Universidade de Houston, que utiliza o *DSpace* como *software* de gestão. Primeiramente, foram exportados os metadados do banco de dados do *DSpace* para um arquivo CSV, que foi analisado via *Excel*. A partir da análise do arquivo, foi possível traçar as estratégias e escolher as ferramentas para a execução dos trabalhos, que foi dividido em três etapas: comunicação, documentação e correção.

Na primeira etapa, escolheram o sistema *Basecamp*, uma plataforma de gerenciamento de projetos, facilitando os fluxos de trabalho para a atribuição de tarefas, documentar e

acompanhar decisões e registrar as atas das reuniões e a comunicação entre os envolvidos, que forneceram informações importantes sobre nomes de pessoas e departamentos, por exemplo.

A etapa de documentação foi imprescindível para que o trabalho pudesse ser replicado e ter continuidade, como referência futura. Para tanto, foi utilizado a PMwiki, uma plataforma de publicação *wiki* de código aberto, que permitiu “coletar respostas do pessoal das faculdades e departamentos e arquivar informações do projeto” (*on-line*), com capturas de tela e descrições de instruções fáceis de entender.

A etapa de correção, etapa final do processo, envolveu todas as edições necessárias nos metadados para que atingissem a qualidade necessária. Para isso, adotaram o *Microsoft Access*, para executar funções mais complexas, com a consolidação de colunas e células, e o *OpenRefine* com um recurso para a padronização de nomes de autor, orientador, departamento e faculdade, usando facetas, filtros e *clusters*. Para a padronização desses campos, os autores fizeram uso de vocabulários controlados, localmente, para dirimirem os erros no momento do depósito.

Os autores destacam que, para grande volume de dados, talvez as ferramentas mencionadas não sejam as ideais. E sugerem o uso de *scripts* no *DSpace*, que seria a opção automatizada para corrigir erros no depósito.

O artigo *Investigation of challenges in academic institutional repositories* (Artigo 21), publicado em 2019 por Joo, Hofman e Kim, destaca os principais desafios enfrentados para a gestão de RIs a partir de uma pesquisa realizada em bibliotecas americanas entre dezembro de 2016 e janeiro de 2017. Com uma amostra de 64 respondentes, que estavam ligados, em sua maioria, diretamente à gestão do RI da instituição. Como resultados, os autores destacam seis dimensões como desafios: dados, metadados, tecnologia, usuários, questão ética e administração. Em ordem de importância, os metadados foram apontados como o segundo maior desafio entre as dimensões do estudo. E, entre os 37 desafios apontados, no conjunto de todas as dimensões, foi o que recebeu maior peso.

O Quadro 1, adaptado de Joo, Hofman e Kim (2019, p. 534), destaca os principais tópicos relacionados ao desafio com os metadados, enfrentados pelos participantes.

Quadro 1 – Desafios relacionados aos metadados

| | |
|---|---|
| Recursos insuficientes | Não há recursos suficientes (por exemplo, indexadores ou orçamentos humanos) para executar a criação de metadados para itens digitais |
| Inconsistência na entrada de metadados | É difícil manter a qualidade consistente dos metadados em diferentes coleções |
| Interoperabilidade de metadados | É difícil criar esquemas de metadados para serem compartilhados entre diferentes sistemas |
| Falta de vocabulários controlados | Existe uma falta de vocabulários controlados disponíveis para indexar para determinados assuntos |
| Indexadores de recrutamento | É difícil recrutar indexadores humanos de qualidade para criação de metadados |
| Indexadores de treinamento | É difícil treinar indexadores humanos para criação de metadados |
| Dados incompletos | Alguns campos-chave de metadados não estão completamente documentados para itens digitais (informações ausentes para os principais campos de metadados) |
| Esquema de metadados existente insuficiente | Esquemas de metadados existentes (por exemplo, <i>Dublin Core</i>) não são suficientes para serem usados em repositórios institucionais |
| Poucos metadados específicos do domínio | Poucos esquemas de metadados estão disponíveis para representar domínios específicos |

Fonte: Adaptado de JOO; HOFMAN; KIM, 2019, p. 534, tradução da autora.

No artigo *Assessing Author Identifiers: Preparing for a Linked Data Approach to Name Authority Control in an Institutional Repository Context* (Artigo 22), publicado em 2019, Moira Downey destaca a dificuldade de os RIs utilizarem vocabulários de autoridade já estabelecidos como ISNI, pois as autoridades em RIs são, em geral, locais. Os autores afirmam que realizar o controle local é insustentável para instituições carentes de recursos, sendo assim, o RI perde a capacidade de oferecer informações precisas sobre a produção, além do impacto na qualidade da recuperação.

A partir dessas questões, a autora descreve um estudo realizado na Universidade de Duke que teve como objetivo avaliar as ferramentas de autoridade que fornecem URIs de autores. Os dados de autoria foram exportados para o *OpenRefine* e, a partir dele, foram comparados os registros de autoridade internos com os serviços LCNAF, VIAF e ORCID.

Como resultado, o ORCID se destaca por maior número de nomes correspondentes. Foi observado também a ausência de estudantes entre as correspondências com o ORCID. A autora observa que essa questão pode ser facilmente resolvida com a incorporação de políticas que orientem esse grupo acadêmico a utilizar o ORCID e incluí-lo como parte do fluxo de depósito. No entanto, o ORCID apresenta algumas desvantagens: necessidade de fontes de dados adicionais para desambiguar, já que alguns autores não fornecem informações necessárias para tal: vínculo institucional, por exemplo. Ademais, o autor pode optar por excluir o seu perfil.

Assim, até que o ORCID se consolide, a autora afirma que outras práticas complementares precisam ser adotadas pelas instituições.

O artigo *Planting Cedar: An Open Source Linked Data Vocabulary Manager at the University of Houston Libraries* (Artigo 23), publicado em 2019 por Andrew Weidner, Anne M. Washington e Xiping Liu apresentam o processo de desenvolvimento do Projeto Cedar (aplicativo), de migração de metadados e de implementação de fluxo de trabalho para povoar o novo ecossistema digital de dados interligados na *Web*, que engloba todas as bibliotecas da Universidade de Houston (UH), inclusive o seu RI.

Segundo os autores, “a Unidade de Metadados analisa os nomes usados no Repositório Institucional UH a fim de normalizar e criar autoridades locais para o corpo docente da UH” (*on-line*). Eles destacam a necessidade de manutenção de dados de autoridade locais, já que os autores locais não são localizados nos catálogos tradicionais. No entanto, como afirmam os autores, a criação das listas locais só é válida desde que nasçam com o propósito de compartilhamento, seja com outras iniciativas locais ou catálogos nacionais. A criação das autoridades faz parte do fluxo de trabalho do projeto, porém, eles não puderam ser importados para o novo sistema em lote devido à falta de padronização nos dados. Então, foi feita a inserção manual dos nomes.

Segundo os autores, também faz parte do fluxo de trabalho a verificação em duas fontes externas para as autoridades: o Arquivo de Autoridade de Nome da Biblioteca do Congresso (LCNAF), também conhecido como Arquivo de Autoridade NACO (Programa Cooperativo da Autoridade de Nome), e o Arquivo de Autoridade Internacional Virtual (VIAF). O fluxo de trabalho também envolve limpeza contínua de metadados ETD (tese e dissertação). A cada novo lote de dados, é feita a “padronização dos nomes dos assessores e membros do comitê, e os nomes dos departamentos” (*on-line*). Para isso, eles exportam os dados como um arquivo N-Triples ou RDF/XML e depois carregados no *OpenRefine*, através da extensão RDF, e, em seguida, são reconciliados, assegurando a consistência dentro e entre as coleções do RI. A ideia futura é mudar do MARC para BIBFRAME, estabelecendo “identificadores locais únicos, que podem ser inseridos no campo \$0, para adicionar a capacidade de ligação para nomes locais” (*on-line*).

O artigo *Local Name Authorities and Linked Data: the Landscape* (Artigo 24), publicado em 2019, por Anna R. Craft traz uma visão geral de diversos projetos de autoridades locais que

utilizam a abordagem dos dados vinculados, apresentando as boas práticas e os desafios. Segundo a autora, os principais desafios se referem à equipe, que tem restrições de tempo e habilidades, com base na experiência, para se dedicar ao trabalho de autoridade. Ela afirma que associar os nomes aos dados vinculados potencializa que sejam utilizáveis em um contexto mais amplo.

A autora destaca importantes pontos vantajosos do controle de autoridades, a saber:

facilita a busca de nomes conhecidos, fornece previsibilidade na busca e consistência em registros, organiza e colhe informações sobre autores, fornece eficiências para criadores de metadados, facilita a manutenção de dados ao longo do tempo, reduz erros em conjuntos de dados de nomes, diferencia nomes semelhantes e fornece contexto para nomes (CRAFT, 2019, *on-line*).

A autora destaca um relatório que foi publicado em 2018, pelo *Forum National Strategy for Shareable Local Name Authorities National*, que, além de fornecer informações e os resultados do projeto, destaca “detalhes técnicos, uma revisão dos agregadores de autoridades de nome existentes e estudos de caso de fluxos de trabalho de criadores de dados e agregadores de dados, alguns dos quais estão usando protocolos de dados vinculados e outros que não são” (*on-line*).

A autora conclui sugerindo que, ao embarcar nesse processo, as instituições criem fluxos de trabalho e melhores práticas do W3C, assim como boas práticas de publicação de dados vinculados, aproveitando a expertise em práticas e princípios de controle de autoridades dos bibliotecários que trabalham na catalogação e em áreas relacionadas.

O artigo *Western Name Authority File: A Pilot Regional Name Authority Project* (Artigo 25), publicado em 2019 por Anna Neatrour e Jeremy Myntti, apresenta o projeto *Western Name Authority File* (WNAF), numa iniciativa da Biblioteca J. Willard Marriott da Universidade de Utah e outras. Segundo os autores, o projeto tem como objetivo “a criação de um arquivo de autoridade regional de nomes pessoais e órgãos corporativos existentes em registros de metadados de coleta digital em várias instituições em todo o Oeste dos Estados Unidos” (*on-line*), em um trabalho cooperativo, que à época contava com 60 mil nomes. O processo de criação englobou a revisão de modelos de dados, coleta de metadados e limpeza e reconciliação de dados duplicados.

Os autores destacam que algumas tarefas de limpeza e aprimoramento de metadados, assim como a reconciliação e desenvolvimento de listas de autoridades, podem ser automatizadas,

mas que devem ser considerados nos projetos desse tipo o árduo trabalho manual associado à construção das listas, regionalmente. Eles alertam que

algumas das tarefas manuais que levaram mais tempo do que o esperado incluem a necessidade de padronizar metadados produzidos em uma variedade de instituições e sistemas parceiros, o trabalho de duplicação, pesquisar correspondências, formatar dados para testes de sistemas e as demandas de controle de qualidade de metadados (NEATROUR; MYNTTI, 2019, *on-line*).

Assim, no fluxo de trabalho cooperativo, devem-se incluir os extensos requisitos de tempo para a revisão manual, segundo os autores.

Para os autores, é preciso “desenvolver melhores métodos para conciliar dados existentes [...] e simplificar o processo para adicionar novos nomes ou vincular novas coleções aos nomes existentes” (*on-line*). Os autores também destacam que elementos tais como “datas de nascimento e ocupação são importantes para a desambiguação de nomes, além da presença de informações de localização associadas a esses nomes” (*on-line*).

3 FUNDAMENTOS TEÓRICO-CONCEITUAIS

Desde os primórdios, as bibliotecas se preocuparam com o controle bibliográfico, e instrumentos foram criados para facilitar essa tarefa, tais como sistemas de classificação bibliográficos, vocabulários controlados e catálogos, todos eles com o objetivo de organizar a informação para permitir o acesso. Com as novas tecnologias, a biblioteca teve que se readaptar e seu acervo, que antes era limitado, quase exclusivamente a documentos físicos, precisou incluir documentos digitais de naturezas e formatos diversos.

Então, enquanto a informação produzida estava restrita às limitações de uma biblioteca física, era concebível imaginar um Controle Bibliográfico Universal (CBU) como o proposto pela *International Federation of Library Associations* (IFLA), em 1970. No entanto, a tecnologia e a multiplicidade de suportes e formatos são agora um desafio ainda maior, já que além do controle é necessário armazenar, organizar e preservar todo esse volume de documentos para que seja acessado pelo usuário.

Com as novas tecnologias, o ambiente digital dá origem às bibliotecas digitais com diferentes denominações, e, entre elas, estão os Repositórios Institucionais (RIs), que buscam consolidar em um único ambiente a produção acadêmica da instituição. Nesse ambiente, ressalta-se a preocupação com os metadados, e a qualidade deles se torna essencial para uma recuperação com acurácia e precisão. Em especial, no que diz respeito ao controle de autoridade para entidade pessoa, se nas bibliotecas sempre houve o investimento em listas de autoridades, muitas vezes de maneira local, de rede ou regional, nos RIs já se torna essencial uma solução mais global.

Tendo como norte essas questões, este Capítulo apresenta fundamentos teóricos e conceituais para a sua compreensão e para as análises dos resultados, estando dividido em quatro seções. A primeira seção fornece uma descrição sobre a ambiguidade que envolve os nomes próprios, demonstrando a complexidade do processo de desambiguação para as variações de um mesmo nome. Depois, a seção dois destaca o foco deste estudo, que é o controle de autoridade para entidades pessoas, trazendo um breve histórico sobre a catalogação e o catálogo, assim como um relato sobre o controle de autoridades ao longo do tempo. Em seguida, a seção três apresenta os Repositórios Institucionais (RIs), destacando a importância da qualidade dos metadados e das implementações de políticas de gestão. Por fim, na seção quatro descreve-se o *software*

DSpace, que, conforme dados da ROAR, em maio de 2021, no mundo quase 50% das iniciativas de RIs são com esse *software*. Ademais, ele foi recomendado pelo IBICT para a implementação de RIs no Brasil, e, segundo o *OpenDoar* (2021), dos 151 RIs brasileiros, 83% utilizam o *DSpace*.

3.1 A ambiguidade no âmbito dos nomes próprios

No âmbito acadêmico, é essencial atribuir a cada produção científica a sua autoria, ou seja, o pesquisador que a escreveu, possibilitando verificar todas as obras publicadas por ele. Segundo Smalheiser e Torvik, (2009, p. 287)

a desambiguação do nome do autor compreende quatro desafios distintos: Primeiro, um único indivíduo pode publicar sob vários nomes - isto inclui a) variantes ortográficas e de grafia; b) erros ortográficos; c) mudanças de nome ao longo do tempo, como pode ocorrer com o casamento, conversão religiosa ou reatribuição de gênero; d) o uso de pseudônimos. Em segundo lugar, muitos indivíduos diferentes têm o mesmo nome - na verdade, nomes comuns podem compreender vários milhares de indivíduos. Em terceiro lugar, os metadados necessários são muitas vezes incompletos ou faltam completamente - por exemplo, alguns editores e bancos de dados bibliográficos não registraram os primeiros nomes dos autores, suas localizações geográficas ou informações de identificação, como seus graus ou suas posições. Em quarto lugar, uma porcentagem crescente de artigos acadêmicos não são apenas de autoria múltipla, mas representam esforços multidisciplinares e multi-institucionais. Em tais casos, a desambiguação de alguns dos autores não ajuda necessariamente a designar os autores restantes.

Para os autores, enquanto as bibliotecas tradicionais, em geral, fazem um esforço para a identificação manual e local dos nomes dos autores, “essa estratégia falha com o surgimento maciço de bibliotecas digitais” (SMALHEISER; TORVIK, 2009, p. 288). Eles ainda acrescentam que

o surgimento de grandes bancos de dados bibliográficos desafiou as análises de mineração de dados a compreender as características de grande escala dos dados como um todo, e a extrair, remontar e sintetizar as informações brutas para criar um conhecimento inteiramente novo (SMALHEISER; TORVIK, 2009, p. 289).

Assim, em ambiente digital é preciso que se encontrem formas de fazer a desambiguação dos nomes próprios, uma vez que “conhecer indivíduos (não apenas nomes de autores) é crucial

para estabelecer novos recursos como redes de citação, redes de colaboração e perfis de autores” (SMALHEISER; TORVIK, 2009, p. 289).

A Onomástica é a área do conhecimento que estuda os nomes próprios: quando se trata de pessoas, é o objeto de estudo da Antroponomástica, e, quando se trata de lugares, é o objeto de estudo da Toponomástica, e, no “âmbito internacional, os estudos onomásticos contam com uma larga tradição”, sendo que, “em alguns países, como a Inglaterra, a Suécia, a Itália e a Hungria, os nomes próprios começaram a ser estudados cientificamente no começo do século XIX” (SEIDE, 2021, p. 50). Isquierdo (2020, p.10) chama a atenção para o fato de que o estudo dos nomes próprios “implica considerações sobre particularidades que os envolvem, como questões históricas e ideológicas”. Nesse cenário, evidencia-se a ambiguidade, pois, segundo a autora, existem processos de renomeações, quando há substituição, acréscimo e/ou redução de palavras nos nomes próprios, uma vez que “o nome próprio, de pessoa ou de lugar, registra e perpetua crenças, valores, procedências de grupos sociais e, por extensão, da sociedade, em diferentes momentos de sua história com suas ideologias, devoções, motivações e também seus modismos e valores” (ISQUERDO, 2020, p.10).

Seide (2021, p. 50) destaca que, no Brasil, os estudos dos antropônimos (nomes próprios de pessoas) “se tornaram mais conhecidos a partir dos estudos de Dauzat”, já em meados do século XX, e que o linguista Mansur Rosário Farâni Guérios é pioneiro nesse campo de estudos, “cujo dicionário etimológico de nomes e sobrenomes (1981) é usado, ainda hoje, por vários pesquisadores da área” (SEIDE, 2021, p. 51). Mariani (2014, p. 133) afirma que

há várias possibilidades de pesquisa para se falar sobre o nome próprio. No âmbito dos estudos da linguagem, que remonta aos da lógica, da gramática, da retórica e da filosofia, o nome próprio deve ser situado a partir de estudos sobre o funcionamento geral dos nomes, mais especificamente em relação à questão da referência. Os nomes próprios de pessoas, em princípio, funcionariam de modo diferenciado e particularizado relativamente aos nomes comuns, uma vez que supõem, como referente, um objeto único.

Para a autora, o nome próprio “implica seu caráter referencial, produzindo um efeito de identidade por sua repetição. O nome próprio funciona como uma referência para o sujeito, uma vez que o sujeito é designado e se designa a partir desse nome que lhe é dado ao nascer” (MARIANI, 2014, p. 134). Também Seabra e Isquierdo (2018) afirmam que, na Onomástica, os nomes de pessoas têm uma função referencial. Mariani (2014) acrescenta que um nome próprio

é referencial porque não comunica um significado, “razão pela qual não se traduz um nome próprio, ou seja, não há significado intrínseco, embora o nome próprio traga uma significação para aquele que o porta”, e, assim, tem como função uma identificação pessoal, jurídica, familiar, na cultura onde interage e na sociedade, como um todo (MARIANI, 2014, p. 134).

Seide (2021) destaca que há diferentes tipos de nomes, tais “como os sobrenomes, os apelidos, os pseudônimos e os nomes civis (nomes oficiais formados por um ou mais nomes seguidos de um ou mais sobrenomes)” (SEIDE, 2021, p. 52). Sobre os nomes, tem-se que

o prenome, ou primeiro nome, é o antropônimo que antecede o sobrenome. Geralmente distingue o indivíduo dentro de grupos sociais de sua intimidade. Também é denominado nome de batismo por todos os dicionários [...]. Pode ser simples e composto. No primeiro caso, apresenta apenas um item lexical antroponímico (José) e, no segundo, mais de um (José Maria). De acordo com a legislação brasileira, o prenome, junto com o sobrenome, constitui o nome pessoal garantido por lei [...]. O prenome é utilizado oficialmente pelo portador durante toda a sua vida, mas o art. 1º da Lei Nº 9708, de 18 de novembro de 2008, possibilita sua substituição (AMARAL, 2001, p. 70).

Sendo assim, a possibilidade de mudança no nome de uma pessoa é outro fator que pode causar ambiguidade quando é preciso referenciá-la. Amaral e Seide (2020, p. 58) citam como exemplo uma conversa entre dois falantes, na qual o nome de Elis Regina é citado, e os dois a “identificam [como] um referente único, que é a cantora brasileira Elis Regina Carvalho Costa (1945-1982)”. Segundo os autores, isso somente é possível quando os dois falantes compartilham dessa informação, que faz parte do repertório lexical deles, pois o nome, em si, não carrega qualquer traço semântico que possa identificar uma classe de indivíduos com esse mesmo nome - Elis Regina.

No entanto, é possível que existam outras pessoas que possam ter o mesmo nome: Elis Regina. Esse fenômeno é conhecido como o de pessoas homônimas, que são aquelas pessoas que possuem um nome idêntico ao de outra pessoa, questão que dificulta a comunicação na sociedade, e, em especial, no ambiente acadêmico. Esse é um problema tão sério que foi instituída a Lei Federal nº 11.971/2009, que trata da obrigação de os cartórios, serviços extrajudiciais e distribuidores judiciais brasileiros fazerem constar nas certidões o maior número possível de elementos de identificação, para se evitarem os transtornos causados pela homonímia nos nomes das pessoas. Em relação ao controle de autoridade para entidades

peessoas, como já mencionado no Capítulo de Introdução, Salo (2009) afirma que, sobretudo em RIs multidisciplinares, a ocorrência de nomes idênticos é recorrente.

Pode-se considerar que os nomes do registro de nascimento são dispositivos administrativos e políticos para a identidade da pessoa e para o apontamento da experiência vivida (ALMEIDA, 2017). Para o autor, os registros de identificação das pessoas são mecanismos por meio dos quais governos supervisionam e mantêm controle das atividades que interessam ao Estado, e também as instituições usam esses mecanismos com semelhante objetivo. Em contrapartida, deve-se entender que esses mesmos mecanismos dão às pessoas acesso a direitos de toda natureza – política, jurídica e social.

No ambiente da academia, pode-se acrescentar que, por meio desses mecanismos, os profissionais garantem os seus direitos acadêmicos, que são aqueles que dão visibilidade e reconhecimento ao conjunto de suas atividades e produção técnico-científica.

Silveira (2013, p. 34) destaca que

na contemporaneidade, o autor pode ser visto por vários olhares. No caso de autores científicos, a publicação de artigos que revelam sua pesquisa é parte integrante de seu trabalho, sendo necessária para pleitear bolsas de fomento junto às agências. Portanto, a publicação para os cientistas se caracteriza como um "meio" para divulgar suas pesquisas, captar mais recursos e para o reconhecimento de seu trabalho.

Percebe-se, então, que o sentido do trabalho concretizado na academia ocorre no campo do reconhecimento e expressa o retorno do investimento que é realizado no trabalho pelos indivíduos. Assim, apropriando-se da ideia de Frege (1978), que estabelece uma relação entre referência e nome próprio, pode-se conceber que na comunicação científica, cada nome próprio deve se referir a apenas um objeto - nesse caso, a uma autoria -, que também deve ser apreendido coletivamente, para que não ocorra ambiguidade na computação de sua produção.

Discorrendo sobre a questão da referência, Searle (1996, p. 251), em seu artigo intitulado *Proper Names*, ressalta que um nome próprio deve ser sempre associado a um conjunto de descrições, necessárias e suficientes, que identificam aquele que é denominado por ele, sendo o nome um designador indireto, pois, “nomes (e sentenças) têm essencialmente um sentido, e apenas acidentalmente uma referência”. Para o autor, estando o nome próprio ligado às

descrições, ele tem como função ser a sua referência intencional e unívoca, uma vez que esse nome comunica a identidade de um indivíduo, sem que seja necessário recuperar os elementos descritos sobre esse mesmo indivíduo. O que o autor discute é exatamente o ponto que importa nesta pesquisa, que não é a referência dos nomes em si e sim a intenção de se referir a uma pessoa, e, no caso deste estudo, é a intenção que se tem de se referir aos autores em suas publicações, dando-lhes crédito por isso.

Nessa questão da autoria, também importam os direitos autorais, regulada no Brasil pela Lei nº 9.610 de 19 de fev. de 1998, que define o termo “autor” como

peessoa física criadora de obra literária, artística ou científica [...] é titular de direitos de autor quem adapta, traduz, arranja ou orquestra obra caída no domínio público, não podendo opor-se a outra adaptação, arranjo, orquestração ou tradução, salvo se for cópia da sua (BRASIL, 1998, *on-line*).

Os homônimos, as mudanças de nome, as grafias diferenciadas, as diversificadas abreviações que se podem estabelecer para um mesmo nome, as informações incompletas ou mesmo equivocadas sobre o registro de uma pessoa, são fatores que podem levar à ocorrência de ambiguidades na determinação da autoria dos documentos (BARBASTEFANO, 2015). Esses fatores são alguns dos elementos que dificultam o controle de autoridade para entidade pessoa, tópico que será discutido na próxima seção.

3.2 O controle de autoridade para entidades pessoas

Segundo a IFLA (2017, *on-line*, tradução da autora), o controle de autoridade, também denominado como controle de ponto de acesso, “refere-se à normalização dos pontos de acesso controlados (cabeçalhos) e à provisão de pontos de acesso alternativos e relacionados”.

Para garantir a consistência, é criado um registro de autoridade para cada título autorizado (ponto de acesso autorizado) para um nome próprio ou um assunto, etc. Um registro de autoridade é feito quando um cabeçalho é estabelecido, ou seja, autorizado para uso como entrada principal (título preferido e, se apropriado, o ponto de acesso autorizado para o criador), uma entrada adicional, ou entrada de assunto, pela primeira vez, durante a catalogação de um item da biblioteca. O controle de autoridade é o processo que é aplicado tanto às partes descritivas como às partes de análise do assunto da catalogação. Ele garante a consistência e correção dos nomes e cabeçalhos de assunto inseridos na descrição bibliográfica (IFLA, 2017, *on-line*, tradução da autora).

Nessa perspectiva, as funções básicas do controle de autoridade são para apoiar: a) a busca agrupando as obras de um mesmo autor pelo seu nome ou as obras anônimas pelos seus títulos; b) a identificação dos nomes de pessoas e organizações ou títulos de obras e expressões, adotando a desambiguação; c) a identificação e o relacionamento de recursos por assunto (IFLA, 2017). Assim,

o controle de autoridade é um processo que organiza informações bibliográficas em catálogos de bibliotecas, utilizando uma grafia única e distinta de um nome (ponto de acesso ou cabeçalho) ou de um assunto para cada tópico. Registro de autoridade é um registro que dá a forma autorizada (a forma selecionada para um título) de um nome pessoal, nome empresarial, sobrenome, nome de família, nome de lugar, título uniforme ou preferido, título de série, assunto, etc. no catálogo da biblioteca ou no arquivo de registros bibliográficos, e são listados em um arquivo de autoridade contendo títulos de itens da biblioteca (IFLA, 2017, *on-line*, tradução da autora).

Como se pode perceber, o trabalho de controle de autoridade envolve não somente o registro da documentação da forma preferida do cabeçalho, de suas variantes, de outros dados sobre a autoridade em questão e *links* para os registros de autoridade relacionados a ele, como também a manutenção e atualização contínua dos registros de autoridade e dos cabeçalhos em registros bibliográficos associados (TILLET, 1989).

Em especial, o controle de autoridades pessoa segue esse entendimento, visto que

o controle de autoridade rígido dos nomes de pessoas em uma base de dados ou biblioteca digital, com indicação dos nomes variantes, ou equivalentes, e reconhecimento daqueles semelhantes, por exemplo, contribui para oferecer mais eficiência nos processos de representação e de recuperação de informações. Estes controles ajudam a evitar inconsistências dos pontos de acesso e possibilitam aos usuários resultados mais satisfatórios e precisos no processo de busca por informação (ROSADO; DIAS, 2020, p. 3).

Essas inconsistências podem ser causadas por diferentes fatores, conforme já discutido na seção 3.1, anterior, sobre ambiguidades na representação de nomes próprios.

3.2.1 Breve histórico sobre a catalogação e o catálogo

A catalogação pode ser analisada em um sentido amplo, como uma atividade bibliográfica que tem como produtos os registros bibliográficos, incluindo a representação descritiva e a

representação temática (MEY; SILVEIRA, 2010). Nesse sentido, as autoras entendem a catalogação

como o estudo, a preparação e a organização de mensagens, com base em registros do conhecimento, reais ou ciberespaciais, existentes ou passíveis de inclusão em um ou vários acervos, visando a criar conteúdos comunicativos que permitam a interseção entre as mensagens contidas nestes registros do conhecimento e as mensagens internas dos usuários (MEY; SILVEIRA, 2010, p. 126).

As autoras afirmam que para atender a esses elementos, a catalogação deve ser realizada a partir de um conjunto de características essenciais, que são: integridade, clareza, precisão, lógica e consistência, que assim se explicam:

- integridade significa fidelidade, honestidade na representação, transmitindo informações passíveis de verificação;
- clareza significa que a mensagem deve ser compreensível aos usuários;
- precisão significa que cada uma das informações só pode representar um único conceito, sem dubiedades ou dúvidas;
- lógica significa que as informações devem ser organizadas de modo lógico;
- consistência significa que a mesma solução deve ser sempre usada para informações semelhantes.

Tais características relacionam-se às funções maiores de compreensão dos produtos catalográficos por parte do usuário e de auxílio ao mesmo. Para atingir essas funções, a catalogação utiliza-se de linguagem catalográfica, que possui semântica e sintaxe próprias (MEY; SILVEIRA, 2010, p. 128).

As autoras afirmam que esse conjunto de características irão individualizar os recursos bibliográficos, oferecer pontos de acesso, assim como permitir ao usuário entender os produtos catalográficos e se apropriar das informações que são comunicadas sobre os documentos (MEY; SILVEIRA, 2010).

Os pontos de acesso “são a parte pela qual os usuários podem acessar a representação de um recurso bibliográfico” (MEY; SILVEIRA, 2009, p. 95). Para a Declaração dos Princípios Internacionais de Catalogação (IFLA, 2016, p. 12) eles podem ser:

um nome, termo, código e entre outros pelo qual os dados bibliográficos ou de autoridade são pesquisados e identificados” e tem como objetivos “fornecer uma recuperação confiável de dados bibliográficos e de autoridade e seus recursos bibliográficos e; reunir e limitar os resultados da pesquisa.

Para atingir esses objetivos, os pontos de acesso devem ser normalizados e terem controle rígido, evitando-se incoerência e ambiguidade da informação fornecida. Nesse caso, a criação dos pontos de acesso deve ser feita com base em regras, políticas e instrumentos que auxiliem na escolha dos pontos de acesso ideais para o contexto dos usuários (MEY; SILVEIRA, 2009; IFLA, 2016).

Os pontos de acesso podem ser de dados bibliográficos e de autoridade. O primeiro se refere à descrição do recurso com a extração das informações nele contidas de forma a individualizá-lo entre os demais e pode incluir os pontos de acesso “autorizados para as obras e expressões (controlados) existentes no recurso, o título da manifestação (geralmente sem controle) e os pontos de acesso autorizados para os criadores das obras” (IFLA, 2016, p. 8).

Herrero Pascual (1999) sugere que o arquivo com dados de autoridades é composto por uma lista de nomes, que podem ser de pessoas, entidade, congressos, famílias e lugares geográficos; lista de títulos uniformes que podem ser individuais ou coletivos; e lista de assuntos que podem ser palavras-chave ou descritores. E cada um deles se configura como ponto de acesso.

Os dados de autoridade, foco desta pesquisa, estabelecem registros normalizados para nomes de pessoas, entidades, títulos uniformes ou assuntos que podem ser o ponto de acesso principal “[...] que tem como objetivo facilitar a identificação e a recuperação de documentos de forma a evitarem confusões na identificação que podem ser homônimos, sinônimos ou todas as variações possíveis de uma pessoa, entidade, obra ou conceito” (HERRERO PASCUAL, 1999, p, 121, tradução da autora).

Os dados de autoridade representam os pontos de acesso controlados e outras informações que as instituições utilizam para colocar obras de uma pessoa específica, família ou entidade empresarial, ou as várias edições de um título.

O controle de autoridade requer a reunião de todos os nomes ou formas do nome pelo qual uma entidade é conhecida, com o objetivo de fazer com que o usuário, ao utilizar em sua busca qualquer um desses nomes ou formas do nome, encontre um determinado recurso informacional ou todos os recursos associados à entidade por ele desejada (ASSUMPCÃO; SANTOS, 2012, p. 11).

Silveira (2013) traça uma trajetória das tentativas de controle de autoridade para pessoas na atividade da catalogação, trazendo uma série de elementos para entender o tratamento dado aos autores, destacando:

a) a identificação dos autores ficou recorrente somente a partir da Idade Média, quando teve início a identificação da autoria para incorporar aos catálogos; a autora relata que isso pode ter ocorrido pela censura às obras e autores naquela época, havendo a necessidade de se identificarem os autores em casos de possíveis punições;

b) no livro de St. Augustine - Trinity (De Trinitate) - publicado no século VIII, está uma das primeiras listas com dados de título e autor de que se tem notícia, ainda que ela possa ser considerada mais um inventário;

c) os catálogos melhor estruturados aparecem no século IX, e a Biblioteca de Reichenau (Alemanha) "compilou vários catálogos entre os anos 822 e 842", sendo que "o primeiro deles mantinha as obras de um mesmo autor reunidas" (SILVEIRA, 2013, p. 28); e o mosteiro St. Requier (França) compilou, em 831, um catálogo que fazia entrada para o autor, ainda que sem uma ordem evidente;

d) houve a organização das obras de um mesmo autor, separadas por assuntos, trabalho que foi realizado pelos frades agostinianos de York, em 1372, na Inglaterra;

e) houve a inclusão dos nomes dos tradutores no catálogo do convento de Bretton (Yorkshire), em 1558;

f) na França do século XVII, tratando a biblioteca como uma instituição pública e universal, Gabriel Naudé orientou a produção de catálogos em duas seções: de assunto e de autor, considerando que as bibliotecas deveriam estar abertas a disponibilizar todos e quaisquer tipos de autores e assuntos.

Independente se os elementos são bibliográficos ou de autoridade, os usuários têm acesso a esses dados utilizando catálogos, que são o produto da catalogação, tópico tratado na próxima subseção.

3.2.1.1 O catálogo

Um catálogo é uma compilação organizada de metadados bibliográficos que representa o acervo de uma determinada instituição e/ou recursos acessíveis em um determinado local. (JOURNEY; TAYLOR; MILLER, 2015, p. 24). Mey (1995, p. 9) os define como “canal de comunicação estruturado, que veicula mensagens contidas nos itens, e sobre os itens, de um ou vários acervos, apresentando-os sob forma codificada e organizada, agrupadas por semelhanças, aos usuários desse(s) acervo(s)”.

Nesse sentido, o catálogo representa a ferramenta de acesso do usuário ao acervo de uma biblioteca⁶.

A Declaração dos Princípios Internacionais de Catalogação (IFLA, 2016) propõe que o catálogo deve ser um instrumento que permita ao usuário: encontrar, identificar, selecionar, adquirir ou obter, e navegar. Encontrar um documento a partir de uma busca por atributos ou relações, identificar a partir dela a correspondência entre a busca e seu resultado e, assim, selecionar o recurso que lhe convém a partir da sua necessidade de questão inicial para então adquiri-lo ou localizá-lo para empréstimo, compra. Durante esse processo, ao usuário deve ser possível navegar pelo catálogo ou fora dele por meio de seus registros e relações.

Para atingir essas funções, Joudrey, Taylor e Miller (2015, p. 25-27, tradução da autora) destacam que o catálogo deve:

- “ser flexível e atualizado”, sendo a biblioteca um organismo vivo e em constante mutação o catálogo e os metadados nele contidos precisam ser atualizados (adição/edição, substituição ou exclusão) para espelhar a coleção da biblioteca;
- “apoiar as funções internacionalmente aceitas de encontrar, identificar, selecionar, adquirir ou obter, e navegar”, isso inclui metadados de fácil recuperação e controle de autoridade e certa exaustividade na descrição dos registros de forma a facilitar a busca e recuperação pelos usuários;
- “ser fácil de pesquisar, com instruções claras e com explicações sobre a abrangência de uma pesquisa”, recursos de busca integrada, booleana, autocompletar, busca por campos autor, título, palavras-chave ou tipos de documento;
- “ter uma exibição de resultados de pesquisa clara, compreensível e organizada”, devem permitir ao usuário explorar os resultados por data, relevância, autor, título;
- “ser acessível a todos os utilizadores”, precisa considerar usuários com expertise, inexperientes ou com algum tipo de deficiência, recursos de usabilidade e acessibilidade devem estar disponíveis (contraste, zoom);
- “ser preparado e mantido economicamente”, se é demasiado caro e custoso é inviável de ser mantido.

Considerando-se as funções e características dos catálogos, compreende-se a importância dele para que o usuário chegue aos documentos depositados em uma biblioteca e garanta a eficiência da recuperação da informação.

Os catálogos existem desde o século XIX, em diferentes formatos: livro, cartão, fichas e, mais atualmente, *on-line* ou automatizado. Inicialmente, organizavam-se os acervos em catálogos de livros, mas, por serem pouco flexíveis e muito caros, foram substituídos pelas fichas ou cartões.

⁶ No contexto desta pesquisa, o termo “biblioteca” é usado para representar todo o conjunto de informações armazenadas em um local, sejam elas bibliotecas físicas, virtuais ou digitais, repositórios, bases de dados de periódicos, arquivos e museus.

Nesse formato, cada registro bibliográfico é reproduzido em uma ficha (7,5 x 12,5) e cada ponto de acesso do registro tem uma ficha associada, autor e assunto, por exemplo (JOURNEY; TAYLOR; MILLER, 2015).

Mey e Silveira (2009) classificam os catálogos como manuais e automatizados. Os manuais se dividem em internos e externos. Os internos destinados aos bibliotecários e equipe da biblioteca e devem refletir os externos, que são destinados ao público.

Os catálogos manuais externos podem ser acessados e organizados de diferentes formas (FIGURA 1).

Figura 1 – Formas de acesso e organização dos catálogos manuais externos



Fonte: Adaptado pela autora, a partir de MEY; SILVEIRA, 2009.

Em catálogos manuais em fichas em acesso, cria-se uma ficha para cada forma de acesso do registro, para permitir que o registro fosse encontrado sob quaisquer dos pontos de acesso. Havia uma ficha/ponto de acesso para cada autor/responsabilidade/entidade associada a obra e suas variantes, bem como pontos de acesso ou ficha para títulos, variantes do título e assunto (JOURNEY; TAYLOR; MILLER, 2015).

Essas fichas podem ser organizadas alfabeticamente como em um dicionário: nome, título e assunto são combinados, palavra por palavra, em uma sequência alfabética; ou nome, título e assunto são organizados separadamente, em ordem alfabética, cada uma formando um catálogo-dividido, catálogo de autores, títulos e assuntos. Na organização sistemática, os assuntos ou registros são organizados pelo número de classificação adotado na biblioteca. Atualmente, essa forma de organização está em desuso, principalmente no ambiente *on-line* (JOURNEY; TAYLOR; MILLER, 2015; MEY; SILVEIRA, 2009).

Os catálogos internos que auxiliam equipe e bibliotecários nas atividades de catalogação compreendem os seguintes tipos: “catálogo de identidade (ou de autoridade de nomes); catálogos de assuntos (ou de autoridade de assuntos); catálogo de número de classificação; catálogos de títulos (títulos de séries e títulos uniformes); catálogo decisório; catálogo de registro, catálogo oficial e catálogo topográfico.” (MEY; SILVEIRA, 2009, p. 194).

3.2.2 Controle de autoridade ao longo do tempo

A qualquer época, o controle de autoridade exige a gestão da qualidade da informação disponibilizada, a partir de manuais de procedimentos e políticas institucionais para a atividade de catalogação, tendo em vista ser executada de maneira mais eficiente e com qualidade (ESPÍNDOLA; PEREIRA, 2017). As autoras destacam que esse tipo de política institucional tem como objetivos:

proporcionar melhorias na qualidade dos produtos finais da catalogação; assegurar que as decisões referentes ao processo de catalogação sejam uniformes, quer ao longo dos anos de funcionamento da UI [Unidade de Informação], quer pelos diferentes catalogadores responsáveis pela catalogação na instituição; aumentar a eficácia e a eficiência no processo de catalogação; documentar as mudanças ocorridas no processo de catalogação como forma de facilitar futuros processos de conversão e migração de dados, e proporcionar redução dos custos relacionados ao processo de catalogação (SILVA, 2009 citado por ESPÍNDOLA; PEREIRA, 2017, p. 74).

Com os catálogos em fichas, o registo de autoridade,

continha a forma autorizada para um cabeçalho e, noutras entradas, registravam-se as formas variantes autorizadas (como entradas remissivas “Veja também”) e as formas rejeitadas (como entradas remissivas “Veja”). O ficheiro de autoridade praticamente só existia em bibliotecas de grande porte, sendo em muitos casos o próprio ficheiro bibliográfico a fonte de informação sobre a(s) forma(s) autorizada(s) e as formas rejeitadas (CAMPOS, 2003, p. 2-3).

Para a autora, com a adoção dos Princípios de Catalogação de Paris em 1961, em especial, das formas consistentes nos cabeçalhos, contribuíram com o controle de autoridade e permitiram às bibliotecas “evitar informação inconsistente para o utilizador e facilitar o trabalho do catalogador”, pois “os registos de autoridade continham notas sobre as fontes utilizadas para estabelecer a forma autorizada ou para explicar a identidade de determinada pessoa de forma a poder distingui-la de outras com nome semelhante” (CAMPOS, 2003, p. 3). Ela destaca que,

no contexto automatizado, essa “tradição [...] passou a ser uma peça fundamental na validação dos pontos de acesso” (CAMPOS, 2003, p. 3).

O controle de autoridade sofreu um impacto com o processo de automação das bibliotecas. Esse processo incrementou o trabalho de cooperação de registros bibliográficos e introduziu os catálogos *on-line*. Algumas iniciativas ou tentativas de controle de autoridade e compartilhamento de registros bibliográficos podem ser destacadas, a saber: *University of Toronto Library Automation Systems* (UTLAS), Controle Bibliográfico Universal (CBU), *Online Computer Library Center* (OCLC), *Research Libraries Information Network* (RLIN), *Washington Library Network* (WLN), *Name Authority Cooperative* (NACO), *Linked Systems Project* (LSP), *Universal Marc Format* (UNIMARC), *Guidelines for Authority and Reference Entries* (GARE), AUTHOR, *Virtual International Authority File* (VIAF), *Open Researcher and Contributor ID* (ORCID) e *International Standard Name Identifier* (ISNI), além de outras iniciativas.

A UTLAS foi desenvolvida, inicialmente em 1967, para ser um sistema automatizado para a universidade, e, em 1971, foi expandido para fornecer serviços para as bibliotecas do mundo. Em 1978, foi criado o *Catalogue Support System* (CATS), um serviço *on-line* criado para facilitar a catalogação eletrônica e oferecer suporte a outras bibliotecas. Mas esse sistema não possuía controle de autoridade (CLACK, 1990, p. 12). Em 1984 foi vendido para a *Thompson International* (UNIVERSITY OF TORONTO LIBRARIES, 2018).

O projeto de Controle Bibliográfico Universal (CBU), desenvolvido pela UNESCO e IFLA em 1970, tem início com o interesse de estabelecer um sistema mundial para o controle e intercâmbio de informações bibliográficas, com os dados de todas as publicações do mundo. Para tanto, seria necessário um trabalho colaborativo internacional, com a criação de uma rede de bibliotecas nacionais (BNs), que cobririam as atividades bibliotecárias federativas, pressupondo que cada país poderia identificar e registrar as publicações e seus autores nacionais, a partir de um Controle Bibliográfico Nacional (CBN). O conceito por trás desse projeto tem origem desde Conrad Gesner (bibliógrafo, 1516-1565), em 1545, que compilou a *Bibliotheca Universalis*, e, depois, em 1892 com Paul Otlet e Henri La Fontaine, que estavam interessados em organizar uma biblioteca universal. Também é preciso destacar o trabalho da Federação Internacional de Associação de Bibliotecas (FIAB), que desde 1954 desenvolvia um trabalho semelhante ao do projeto CBU. O papel das BNs e dos CBNs tem estreita relação com

o controle de autoridade, impactando positivamente a recuperação de informações, apoiando a proteção do patrimônio cultural e da produção intelectual de cada país, contribuindo para a promoção internacional desses produtos (SOUZA; ORTEGA, 2014; CLACK, 1990).

A OCLC foi criada em 1971 com o objetivo de criar um sistema de catalogação compartilhada, inicialmente para as Bibliotecas de Ohio, mas, atualmente, é considerada uma rede global cooperativa de bibliotecas. Ela é uma rede sem fins lucrativos e com bibliotecas membros espalhadas por todo o mundo e responsável pelo *WorldCat*, considerado o maior catálogo em linha do mundo, com dados de mais de 70 mil bibliotecas (OCLC, 2021a; DIAS, 1977). No *WorldCat*, é atribuído um número de controle (OCN: Número de controle OCLC) a cada novo registro bibliográfico, que vinculam os registros do *WorldCat* aos registros do sistema da biblioteca local, conferindo uma chave de referência comum entre eles. A OCLC administra o Arquivo de Autoridade Internacional Virtual (VIAF), que é um arquivo de autoridade de nome internacional supervisionado por um Conselho e por representantes de diversas instituições, cujos números VIAF são usados como identificadores padrões, por exemplo, na *Wikipedia* (OCLC, 2021b).

O RLIN foi fundado por três universidades – Columbia, Harvard e Yale – e a Biblioteca Pública de Nova York, em 1975, para resolver problemas de bibliotecas que não eram atendidas pelo OCLC. (OCLC, 2021c; FALK, 1989). Possuía um subsistema de autoridades de nome da *Library of Congress* e da Biblioteca Pública de Nova York, que permitia aos membros buscarem por “nomes pessoais e corporativos, títulos uniformes, assuntos atuais, nomes geográficos e combinações de nome-título” (CLACK, 1990, p. 12). Em 2006, o RLIN foi incorporado à OCLC (OCLC, 2021c).

O WLN foi idealizado pela Biblioteca Estadual de Washington no ano de 1975, e ficou operacional em 1977, para atender as bibliotecas do estado e noroeste do Pacífico, e tinha como objetivo criar um catálogo e fichas que pudessem ser compartilhadas entre bibliotecas acadêmicas e públicas (MONTE-MOR, 1983). O destaque da base WLN foi o controle de autoridade “obtido através da manipulação *on-line* do título uniforme do autor e registros de autoridade de assunto em um arquivo de vocabulário (VOe) anexado aos registros bibliográficos” (CLACK, 1990, p. 13).

O NACO teve início em 1977, e atualmente é denominado *Name Authority Cooperative Program*, e é a fonte predominante para nomes nas bibliotecas dos Estados Unidos. É também

conhecido como *Library of Congress Name Authority File* (LCNAF) (CRAFT, 2019). Foi idealizado pela *Library of Congress* para ser um sistema de autoridade nacional, que oferecesse suporte a todas as bibliotecas nacionais, e, assim, diminuísse o retrabalho em projetos locais. Para isso, as bibliotecas foram selecionadas para fazerem parte da rede e contribuírem na criação dos registros de autoridade para nomes pessoais, corporativos e jurisdicionais, títulos uniformes e séries para o Arquivo de Autoridade LC/NACO. Desde a sua criação, o controle de autoridade foi uma premissa básica, que se mantém até hoje. Os atuais mais de 8 milhões de registros são gratuitamente acessíveis *on-line* através de uma interface *Web* fornecida pela Biblioteca do Congresso e estão disponíveis como dados vinculados (CRAFT, 2019). A adesão está aberta a qualquer instituição (CRAFT, 2019), porém, as instituições cooperantes NACO precisam seguir sete parâmetros fundamentais, a saber: *Basic documentation needed; New name authority record contributions; Changes to existing name authority records; Cancellation of name authority records; Bibliographic File Maintenance (BFM), Authority (NACO) Normalization, Literary Author Numbers (053s)* (CLACK, 1990; LIBRARY OF CONGRESS, 2021), e “também exige que as instituições cumpram metas anuais relacionadas à criação ou edição de registros de autoridade” (CRAFT, 2019, *on-line*).

O LSP foi uma iniciativa cooperativa da década de 80, pensada para criar um *link* de comunicação entre os diferentes sistemas de bibliotecas. O objetivo era contribuir com o compartilhamento das diferentes redes cooperativas fazendo com que uma rede compartilhasse não só entre os seus membros, mas também com outras redes cooperativas. Esperava-se que essa cooperação entre diferentes redes diminuiria os custos das redes, bem como o retrabalho na duplicação de registros. Como resultado, foi desenvolvido o *Standard Network Interconnect* (SNI), uma ferramenta que cria *links* entre registros, possibilitando a comunicação entre diferentes redes cooperativas de bibliotecas (CLACK, 1990).

Publicado em 1977, o UNIMARC foi adaptado em 1985 pela *International Federation of Library Associations and Institutions*. Foi criado para ser um padrão internacional do formato MARC, que permite o intercâmbio de dados bibliográficos, em formatos legíveis por máquina, entre sistemas de gerenciamento de bibliotecas e agências catalogadoras. Também surgiu com o propósito de dirimir as diferentes versões do MARC, que foram desenvolvidas respeitando práticas de catalogação nacionais que não eram compatíveis entre si. Desde a sua criação, ele vem sendo atualizado e adaptado para os diferentes tipos de formatos. Em 1981, foi criado o UNIMARC para controle de autoridades: o UNIMARC *Authorities*. Antes da sua criação, ao

criar um registro bibliográfico as agências inseriam o nome do autor para cada um dos registros. Com o UNIMARC *Authorities* foi necessário criar uma única forma oficial do nome para compor o arquivo de autoridades e cada entidade receberia um número único de controle (CLACK, 1990; IFLA, 1999).

Em 1984, têm-se os GARE, que tinham preocupação com o controle das formas dos nomes e desenvolveram metodologias para a produção de registros uniformes que pudessem ser permutados entre diferentes instituições (CAMPOS, 2003). Porém, segundo a autora, os problemas, tais como as diferenças de idiomas e de transliteração de um nome, tornaram a iniciativa insustentável. Em 2001, na revisão do GARE, que deu origem aos *Guidelines for Authority Records and References* (GARR), houve uma mudança de filosofia, e assim, o “critério de ‘uniformidade’ consagrado nos Princípios de Paris, 1961, dá lugar à adoção de um ou mais cabeçalhos ‘autorizados’, relacionados com a mesma entidade”, buscando-se assim, compilar todas as formas autorizadas em outros idiomas, incluindo os registros dessas autoridades como remissivas (CAMPOS, 2003, p. 11).

Campos (2003, p. 9) relata o projeto europeu AUTHOR, que esteve ativo entre janeiro de 1995 e dezembro de 1997, que teve como objetivo buscar a “troca internacional e a reutilização de registros de autoridade para autores, pessoa física e colectividades”. Nesse projeto, foi desenvolvido um protótipo acessível via *Web*, utilizando o protocolo Z39.50, quando se converteu “para o formato UNIMARC [...] uma amostra de registros de autoridade relativos às mesmas letras (e aos mesmos apelidos ou nomes de colectividades)” (CAMPOS, 2003, p. 9), retirados dos ficheiros de autoridade das bibliotecas participantes, representantes de cinco países (Bélgica, Reino Unido, França, Espanha e Portugal). Foram encontrados vários problemas durante o projeto, que dificultavam o trabalho cooperativo, tais como os apresentados no Quadro 2.

Quadro 2 – Problemas encontrados no trabalho cooperativo do projeto AUTHOR

Cada biblioteca tinha sua língua.

Regras de catalogação, formato e sistema automatizado diferentes.

Práticas e regras diferentes para criar os registros de autoridade: algumas indicavam a nacionalidade do autor, e outras não; algumas bibliotecas não faziam entradas de autoridade para grupos eventuais, como as conferências, e os mencionavam em nota.

Diferenças resultantes de usos linguísticos diversos, sobretudo em autores clássicos ou personalidades com formas vernáculas consagradas pelo uso.

Esquemas diversos de transliteração e romanização.

Formatos MARC diferentes que às vezes não tinham os elementos suficientes para uma conversão correta, tal como os que não previam o código de nacionalidade do autor.

Diferenças na grafia da mesma língua quando utilizada por países diferentes.

Fonte: CAMPOS, 2003.

O VIAF, lançado em 2003, tem como objetivo “reduzir o custo e aumentar a utilidade dos arquivos de autoridade da biblioteca, combinando e vinculando arquivos de autoridade amplamente usados e disponibilizando essas informações *on-line*” (OCLC, 2021d, *on-line*). Para atingir esse objetivo, foi criado um consórcio internacional de cooperação com mais de 40 organizações, de mais de 30 países, dentre os quais destacam-se: *Library of Congress* (LC), a *Deutsche Nationalbibliothek* (DNB) e o *Online Computer Library Center* (OCLC); esta última é, atualmente, a responsável pela manutenção do VIAF (OCLC, 2021d, *on-line*). Além disso, a combinação e a vinculação dos arquivos de autoridade são feitas por meio do *Linked Open Data*.

Em 2010, foi lançado o ORCID voltado para pesquisadores e cientistas na criação de uma identificação digital única, persistente, confiável e inequívoca, para dirimir as ambiguidades dos nomes pessoais (tais como: iniciais, nomes próprios completos, sobrenomes diferentes), atribuindo aos autores um código numérico e exclusivo. O ORCID permite aos autores fazerem a vinculação automática de seu identificador a instituições, afiliações e produções associadas, a atividades de pesquisa (conjunto de dados, patentes, materiais didáticos, publicações e subsídios, por exemplo), e tem se tornado uma ferramenta eficiente no acompanhamento da produção científica de docentes e pesquisadores (ORCID, 2021). Ainda que o ORCID possa ser adotado nas iniciativas de controle de autoridade para nomes de pessoa, ele tem como limitações se ater a autores “vinculados a instituições de pesquisa e ensino, [...]. Assim, não atenderá a todas as autoridades”, e, também, por não permitir “pleno aproveitamento dos dados

presentes no registro ao não exportá-los na totalidade para o modelo RDF” (SILVA; SERRA; SANTAREM SEGUNDO, 2017, p. 201). Atualmente, o ORCID tem como membros nacionais a Universidade de São Paulo (USP), Universidade Estadual de Campinas (Unicamp), Universidade Estadual Paulista "Júlio de Mesquita Filho" e Universidade Federal de Uberlândia (ORCID, 2021).

O ISNI, publicado como padrão válido em 2012, assim como o ORCID, foi criado com o objetivo de resolver os problemas que envolvem a ambiguidade de nomes. Para tanto, segue as diretrizes da ISO para a atribuição de identificadores persistentes para pesquisadores, inventores, escritores, artistas, criadores visuais, produtores, editores e outros, portanto mais amplo que o ORCID (ISNI, 2021). Por princípio, o ISNI atribui um número exclusivo a uma identidade pessoa, ligando-o a todos os outros possíveis signos que possam estar relacionados a ela (tais como pseudônimo, marca, variação ortográfica, etc.). Baseado em diversas fontes de dados, tal como o próprio VIAF, é hoje uma ferramenta do *Linked Data* e da *Web Semântica* (ISNI, 2021). Apesar de ser uma agência independente, ela segue os princípios teóricos e práticos do controle de autoridade em bibliotecas e conta com colaboradores como *The European Library*, *Bibliothèque Nationale de France* e a OCLC (ISNI, 2021). Para Angjeli, MacEwan e Boulet (2014, p. 19, tradução da autora), “VIAF e ISNI renovam os princípios do CBU e abrem uma nova era para a consolidação colaborativa e sem precedentes de identidades em escala internacional”.

Além das iniciativas anteriormente destacadas, cabe mencionar o papel das Bibliotecas Nacionais (BN) e Associações de Bibliotecas e Centros de Catalogação Nacionais empenhados em cumprir os Princípios de Catalogação propostos pela UNESCO em 1961. Nos Estados Unidos é a Biblioteca do Congresso (*Library of Congress* - LC) que cuida do controle das autoridades do país, e os dados são utilizados por outras bibliotecas no mundo inteiro (GRINGS, 2015).

No Brasil, a primeira iniciativa registrada de uma tentativa de controle foi em 1873, com o responsável pela seção de Manuscritos da BN, que elaborou um dos primeiros catálogos da Instituição. Mas somente em 1945, com a adoção do “Código da *American Library Association* (ALA) que foram criados os catálogos de autoridades de nomes, para auxiliarem na uniformização dos pontos de entrada de autores pessoais e coletivos” (GRINGS, 2015, p. 145). Na década de 80, a BN se incorporou à rede Bibliodata/Calco, projeto de automação das

bibliotecas da Fundação Getúlio Vargas que pelo interesse de outras instituições se tornou uma rede nacional de cooperação. Nesse mesmo ano, a BN adotou também o AACR2 e implantou o Cabeçalho de Assunto Unificado (CAU) e, assim, se consolidou como instituição nacional de referência para cabeçalhos de assunto (GRINGS, 2015, p. 146). Segundo a autora, em 1997 a BN migrou o banco de dados de autoridades e assuntos para o *software* OrtoDocs, compatível com o Marc (iniciando com o USMarc), e, no ano seguinte, já havia 26 mil registros de autoridades pessoais. Em 1999, a BN criou o sistema InterDocs, baseado na *Web*, porém, as bases de autoridades e terminologia somente foram disponibilizadas publicamente algum tempo depois (GRINGS, 2015). De acordo com a autora, esses dados ficaram nesse sistema até 2013, quando os dados foram migrados para o *software* Sophia, estando disponíveis numa busca unificada na aba Autoridades, em formato fichas e em formato Marc, que continham, na época, 280 mil entradas padronizadas de autoridades e hoje é formado por “nomes de pessoas, entidades coletivas, eventos associados à autoria de obras e assuntos” (BN, 2021, *on-line*).

Outra iniciativa a destacar, trata-se do projeto MARC (*Machine Readable Cataloging*), lançado em 1965 pela *Library of Congress* (LC) que, em 1976, publicou o *Authorities: a MARC Format* para fazer o controle de autoridade em catálogos digitais (ASSUMPCÃO; SANTOS; ZAFALON, 2017). Segundo os autores, desde então a LC tem publicado atualizações desse formato, que é denominado atualmente de Formato MARC 21 para Dados de Autoridade (MARC 21 *Format for Authority Data*). Assim, o MARC foi criado para representar registros bibliográficos em um formato padrão legível por máquina, e hoje contempla cinco formatos: MARC 21, dados bibliográficos, dados de autoridade, dados de acervos, dados de classificação e informações da comunidade. Em específico, o MARC para Dados de Autoridade foi criado “para fornecer informações sobre as formas autorizadas de nomes, assuntos e subdivisões de assuntos a serem usados na construção de pontos de acesso em registros MARC” (LIBRARY OF CONGRESS, 2009, *on-line*). Nesse formato, os cabeçalhos para pessoas correspondem ao campo 100 e se subdividem em três tipos: Primeiro indicador: designa o tipo de entrada do nome pessoal (0 – Nome na ordem direta; 1 – Sobrenome; 3 – Nome de família); Segundo indicador: indefinido (#). No subcampo "c" do Campo 100 é possível acrescentar títulos ou outras palavras associadas ao nome, e, no subcampo "d", é possível acrescentar datas associadas ao nome. Já no Campo 368, é possível indicar qualquer outro atributo da pessoa ou entidade coletiva, e, no subcampo “d”, pode-se inserir o título da pessoa. Outros exemplos são: Campo 375, subcampo "a": gênero; Campo 370 (subcampo "a": local de nascimento; subcampo "b": local de falecimento); Campo 371 (subcampo "a": endereço; subcampo "e": código postal;

subcampo "m": endereço de correio eletrônico); Campo 372 (subcampo "a": campo de atividade); Campo 374 (subcampo "a": ocupação); Campo 678 (subcampo "a": dado biográfico ou histórico); entre outros. “Acompanhando os desenvolvimentos realizados nos Estados Unidos [...], outros formatos para dados de autoridade surgiram, tais como o UNIMARC/Authorities: *Universal Format for Authorities*, cuja primeira edição é de 1991” (ASSUMPÇÃO; SANTOS; ZAFALON, 2017, p. 24).

Como iniciativas mais atuais, têm-se os Requisitos Funcionais para Registros Bibliográficos (FRBR – *Functional Requirements for Bibliographic Records*), que é o primeiro modelo conceitual (baseado em entidade-relacionamento - E-R) elaborado pela *International Federation of Library Associations and Institutions* (IFLA) para a catalogação (IFLA, 1998). Ele apresenta três elementos (entidades, atributos e relacionamentos), sendo que o elemento identidade possui dez itens, divididos em três grupos: grupo 1: obra, expressão, manifestação e item; grupo 2: pessoa e entidade coletiva; grupo 3: conceito, objeto, evento e lugar (IFLA, 1998). Juntamente ao FRBR, foi desenvolvido um modelo de entidade-relacionamento para pessoas, conceitos e eventos, denominado de Requisitos Funcionais para Dados de Autoridade (FRAD). O FRAD não restringe as informações para pessoas a datas de nascimento, uma vez que inclui uma ampla lista de atributos para nomes pessoais, incluindo: quaisquer datas associadas; títulos; gênero (biológico e social); lugar de nascimento e falecimento; país (nascimento, residência, etc.); filiações; endereço (residencial, comercial); idioma; atividades (profissional ou não); profissão/ocupação; biografia/história, e outras informações associadas à pessoa (SILVA; PAVÃO; GABRIEL JÚNIOR, 2019).

Os atributos do FRAD são exemplificados na RDA, o mais novo código de catalogação, que adota a abordagem teórica dos modelos conceituais FRBR e FRAD, com pretensões de substituir o AACR2 (SOUZA; BEZERRA; COSTA, 2016). Os autores indicam que alguns dos atributos FRAD são elementos centrais no controle de autoridade pessoas, tal como a data de nascimento, que é usada para desambiguar um nome pessoal, por exemplo. Ademais, destacam que é possível codificar os dados RDA em formatos e esquemas já consolidados, tais como o MARC 21 e o *Dublin Core*. Acredita-se que isso pode contribuir para a interoperabilidade entre sistemas, possibilitando o reuso de informação já estruturada.

Por fim, destaca-se que

o surgimento da Web no início da década de 1990, sem dúvida, significou uma revolução no modo de compartilhar informações. Os catálogos digitais, com seus dados bibliográficos e de autoridade, passaram a ser disponibilizados na Web e, apesar do novo ambiente tecnológico, os instrumentos de descrição, regras de catalogação e formatos para o intercâmbio de dados, principalmente, mantiveram-se em certa medida imutáveis, contribuindo, assim, para uma reprodução na Web do mesmo modelo de dados utilizado nos catálogos digitais anteriores a ela (ASSUMPTÃO; SANTOS; ZAFALON, 2017, p. 31).

Segundo os autores, no contexto brasileiro atual se discute sobre o uso dos princípios da *Web Semântica* e das práticas de *Linked Data* para enfrentar os desafios do controle de autoridade, numa tentativa de deixar para trás as limitações tecnológicas existentes no passado.

3.3 Repositórios institucionais

Os Repositórios Institucionais (RIs) são constituídos por toda a produção científica de uma instituição, pois,

um repositório institucional baseado na universidade é um conjunto de serviços que uma universidade oferece aos membros de sua comunidade para a gestão e divulgação de materiais digitais criados pela instituição e pelos membros da comunidade. É, essencialmente, um compromisso organizacional com a administração desses materiais digitais, incluindo a preservação a longo prazo, quando apropriado, bem como organização e acesso ou distribuição (LYNCH, 2003, p. 2).

E, mais especificamente, pode-se afirmar que a criação de um RI pressupõe acesso aberto, sendo que

todo repositório institucional de acesso aberto pode ser considerado um tipo de biblioteca digital. Entretanto, nem toda biblioteca digital pode ser considerada um repositório institucional. Um repositório institucional de acesso aberto constitui um serviço de informação científica - em ambiente digital e interoperável - dedicado ao gerenciamento da produção científica e/ou acadêmica de uma instituição (universidades ou institutos de pesquisa). Contempla a reunião, armazenamento, organização, preservação, recuperação e, sobretudo, a ampla disseminação da informação científica produzida na instituição (LEITE *et al.*, 2012, p. 7).

Nota-se que a criação e manutenção de um RI é uma responsabilidade a ser abraçada pela direção institucional, caso contrário, dificilmente obterá sucesso. Historicamente, os RIs nasceram em pleno movimento mundial do acesso aberto à informação científica, sendo considerados uma inovação no sistema de gestão da comunicação científica existente até então (LEITE *et al.*, 2012).

O modelo de origem da comunicação científica “atendeu principalmente aos interesses de editores e publicadores que visavam ao lucro crescente proporcionado pelas altas taxas das assinaturas de seus periódicos”, gerando o que foi “chamada ‘crise dos periódicos’ em meados das décadas de 80 e 90”. Isso também acentuou “o dilema ético envolvido no limitado acesso ao conhecimento científico que este modelo de comunicação científica proporcionava” (MATIAS, 2015, p. 27).

No início do século XXI, esse paradigma de publicação, até então consolidado, dos altos preços de assinatura dos periódicos e da restrição de acesso do conteúdo apenas aos assinantes, é fortemente questionado. Nesse contexto, se iniciam os movimentos de acesso livre.

O movimento de acesso livre à informação eclodiu como uma reação ao domínio das editoras no processo editorial de periódicos científicos, portanto, sobre os autores, e consequências como o alto custo das assinaturas, que atingiu bibliotecas e usuários em geral, limitados na possibilidade de serem informados sobre o conhecimento gerado e publicado em periódicos (PINHEIRO, 2014, p. 156).

Destaca-se a limitação ao acesso ao público em geral como ponto crucial para justificar o acesso livre à informação, pois a pesquisa, em muitos países, é financiada por órgãos públicos. A restrição de acesso não justificaria os investimentos em pesquisa realizados por esses órgãos, pois não estariam disponíveis para o benefício da sociedade em geral. “Um dos marcos deste movimento é a criação, em 1991, do primeiro repositório digital de *pre-prints*, no laboratório de física de Los Alamos, Novo México, EUA, coordenado pelo físico Paul Ginsparg” (MATIAS, 2015, p. 27). Para a autora, os repositórios digitais são também denominados *e-prints*, e, assim, os RIs, segundo Johnson (2002, tradução da autora), são um importante modelo alternativo de comunicação acadêmica em repositórios digitais, que busca quebrar o monopólio das editoras e aumentar a consciência sobre a produção intelectual de uma instituição.

Matias (2015, p. 30) destaca que a “Iniciativa de Arquivos Abertos e do Movimento pelo Livre Acesso à Informação Científica propõe o livre acesso à informação”, e a sua principal recomendação é a “construção de repositórios institucionais e a implantação de interoperabilidade entre estes repositórios por meio de princípios e padrões estabelecidos” em comum. De acordo com a autora, é, “a partir dessa iniciativa que surgem os principais *softwares* para a construção e manutenção de repositórios no mundo” (MATIAS, 2015, p. 30). Marcondes e Sayão (2009, p. 24) afirmam que a

[...] a criação de repositórios institucionais compreende um grande número de atividades que ensejam aspectos políticos, legais, educacionais, culturais e alguns componentes técnicos importantes. O encaminhamento correto desses vários aspectos e de suas interrelações é que vai determinar o perfil do repositório e a sua aproximação aos objetivos fixados pela organização e, por fim, o sucesso do empreendimento.

Apoiando essa posição, Viana, Márdero Arellano e Shintaku (2005, p. 3) indicam que

estabelecer um repositório institucional indica que a biblioteca está mudando seu papel de custódia para contribuir ativamente na mudança do modelo de comunicação científica. As bibliotecas mantêm a responsabilidade de gerenciar e arquivar material impresso. Mas, à medida que o volume de material para pesquisa de acesso aberto em formato digital cresce, o papel e o valor das coleções impressas declinam proporcionalmente.

Dessa maneira, segundo Leite (2009, p. 19), os RIs “provêm os mecanismos que aumentam tanto a eficácia da preservação da produção intelectual de pesquisadores e instituições acadêmicas, quanto a visibilidade de ambos”. Ademais, os “repositórios institucionais acadêmicos tornam-se facilitadores do acesso à produção intelectual das IES” na medida em que “reúnem, em um único local, a produção acadêmica das unidades e permitem a busca e o acesso por meio de uma base de dados ou *software*” (SHINTAKU; ROBREDO; BAPTISTA, 2011, p. 312).

Com a popularização dos RIs, foi preciso padronizar as suas políticas de criação, possibilitando a interoperabilidade entre os distintos RIs (MATIAS, 2015). A autora relata que a *Open Archives Initiative* (OAI) teve início em 1999, com o “objetivo principal de implantar uma política de interoperabilidade, dentre outros princípios, para a construção de repositórios digitais de acesso aberto à informação”, e, logo em seguida, foi criado o protocolo OAI-PMH (*The Open Archives Initiative Protocol for Metadata Harvesting*), que faz a “coleta mútua de metadados [...] [e] pode ser adaptado aos mais diversos padrões de metadados” (MATIAS, 2015, p. 38). E há grande variedade de padrões de metadados, tais como os “*Dublin Core*, RIS, DIDL, ISO, MARC 21, METS, MODS, MTD-BR, BibTeX, dentre outros. O mais utilizado mundialmente, no âmbito dos repositórios digitais abertos, é o *Dublin Core*” (MATIAS, 2015, p. 39). Segundo a autora, os metadados são considerados a “ferramenta fundamental para a consolidação da interoperabilidade entre os repositórios” (MATIAS, 2015, p. 38).

Dessa maneira, destaca-se como boa prática o uso dos metadados como um recurso para facilitar a recuperação dos documentos em RIs, e, para tanto, eles precisam ter qualidade. Currier (2004) relata que a qualidade dos metadados vai além das especificações e padrões de

metadados, da qualidade dos perfis de aplicativo ou até mesmo dos vocabulários ou taxonomias que são utilizados para descrever os recursos. Para o autor, deve haver também uma preocupação com criação dos conteúdos dos campos que descrevem os recursos dos repositórios, ou seja, “os campos especificados devem ser preenchidos com dados reais” (2004, *on-line*, tradução da autora).

Salo (2009) ressalta que os repositórios foram concebidos sem qualquer preocupação com a qualidade dos metadados ou com as práticas recomendadas de fluxos de trabalho para a gestão dos mesmos.

Claramente a qualidade dos metadados tem um papel importante a desempenhar na realização dos objetivos de repositórios de objetos de aprendizagem e arquivos e-Print, e muito esforço já foi feito para desenvolver abordagens padronizadas para a estrutura de metadados, mas até agora as questões em torno da criação de metadados de boa qualidade dentro dessa estrutura receberam surpreendentemente pouca atenção (BARTON; CURRIER; HEY, 2003, *on-line*, tradução da autora).

Os repositórios ou bibliotecas digitais oferecem serviços que dependem da existência e da forma como são descritos os registros disponibilizados, pois eles representam as propriedades do recurso em um esquema padrão de metadados. Sandy e Dykas (2016, p. 101, tradução da autora) afirmam que “os metadados oferecem suporte ao uso do repositório digital, permitindo que os objetos digitais sejam organizados, mantidos, recuperados e usados posteriormente”.

Assim como em bibliotecas tradicionais, a representação do recurso permitirá ao usuário localizá-lo, através dos pontos de acesso (autor, título, assunto). Sendo assim, se os metadados, gerados nesse contexto mais amplo, não representam corretamente um recurso, este ficará invisível e, com isso, afetará diretamente a recuperação de itens no repositório. De acordo com Zeng e Qin (2008 citado por YASSER, 2011, tradução da autora), os registros de metadados mal criados resultam em baixa recuperação e limitam a acessibilidade às coleções, exercendo, finalmente, um impacto negativo na adoção e uso contínuos de uma biblioteca digital.

Para que a troca de registros se torne tão eficiente quanto no compartilhamento das coleções bibliográficas tradicionais, a descrição dos metadados deve ser precisa e consistente, ou seja, devem ser metadados de qualidade. Park (2009, p. 215, tradução da autora) afirma que “a qualidade dos metadados reflete o grau em que esses desempenham as principais funções bibliográficas de descoberta, uso, procedência, ocorrência, autenticidade e administração”. Nessa perspectiva, observa-se a importância de recursos como VIAF, ISNI e ORCID para

auxiliarem no processo de padronização, precisão e consistência dos dados de autoridade em RIs.

Outra boa prática a ser observada para a criação e manutenção de RIs é a elaboração de uma política de informação, que “são políticas que irão regulamentar o funcionamento do RI, seu escopo/conteúdo, propósito, modos de depósito/submissão de documentos, direito do autor, dentre outras importantes questões”, que perpassam “pela política de povoamento do mesmo” (MATIAS, 2015, p. 41). A autora relata que há duas maneiras básicas de se povoar um RI: “realizadas/monitoradas por equipe capacitada responsável por esta atividade na instituição, ou o autoarquivamento, realizado pelo próprio autor”, e “ambas as modalidades possuem limitações” (MATIAS, 2015, p. 42).

Discorrendo sobre o povoamento por autodepósito, Kuramoto (2011) afirma que:

o povoamento de um repositório institucional (RI) é fruto de uma série de fatores, que se inicia pela sua gestão e continua com a adoção de uma política adequada e capaz de induzir os pesquisadores a auto-depositar a sua produção científica. Dentre esses fatores, encontra-se um programa de marketing para sensibilizar e induzir os pesquisadores a auto-depositar a sua produção científica (KURAMOTO, 2011, *on-line*).

Dessa maneira, a política de povoamento vai além de requisitos técnicos, pois, para ser eficaz deve,

prever ações referentes aos itens expostos: tipologia documental; depósito voluntário ou compulsório; permissão/embargo de documentos; direito autoral; divulgação e marketing do sistema de informação, com vistas a pleitear um RI de ampla aceitação e povoado sistematicamente pela comunidade acadêmica a que se destina, servindo como aporte para a divulgação da produção científica da Universidade, de forma a agregar valor à instituição e a todos que contribuirão para sua divulgação, através da disponibilização de trabalhos no RI (LEY, 2013, p. 95).

Para atingir toda a complexidade do processo de povoamento é preciso treinamentos dos pesquisadores que farão o autodepósito ou autoarquivamento, e também dos bibliotecários e gestores de RIs.

Já o depósito monitorado ou “mediado é a preferência dos pesquisadores, apesar dos repositórios serem construídos para o autoarquivamento” (COOKE, 2007, *on-line*). Segundo destaca a autora,

como o depósito mediado é o mais usual e os bibliotecários são os mais requisitados para realizar a tarefa, é importante que haja investimentos na

biblioteca para garantir que os bibliotecários tenham condições de assumir o depósito de toda a produção da universidade em que atuam (FREITAS, 2015, p. 121).

Assim, é preciso investir no treinamento dos bibliotecários, assim como na formação continuada deles.

Outra forma relevante é o autopovoamento ou o povoamento automático, com carregamento em lote, que inclui a integração do *software* gestor com ferramentas externas (*plugins*), assim como o rastreamento de dados a partir de fontes externas, tal como a do Currículo Lattes, que estão em acesso aberto, conforme evidenciado no Capítulo de Revisão de Literatura.

No caso do *software DSpace*, Carr (2008) esclarece que

tanto o EPrints quanto o *DSpace* permitem o carregamento de lotes, mas, mais especificamente, ambos suportam o novo protocolo SWORD⁷ [Simple Web-Service Offering Repository Deposit] para fazer depósitos automáticos em repositórios. Nós (os desenvolvedores do SWORD) esperamos muito poder trabalhar com os repositórios disciplinares já estabelecidos para permitir a alimentação automática de depósitos dos Repositórios Institucionais para os Repositórios Disciplinares e vice-versa (CARR⁸, 2008, tradução da autora).

Existem diferentes tecnologias e metodologias de povoamento automático relatadas na literatura, tais como Averkamp e Lee (2009), Walsh (2011), Assumpção (2013) e Masar (2014). Este último, fornece procedimentos para o carregamento automático de um RI, com dados obtidos dos registros da *Scopus* e da *Web of Science*, discorrendo sobre os problemas encontrados e quais as soluções adotadas no projeto. Especificamente sobre a identificação de pesquisadores nessa coleta automática de registros, tem-se o exemplo do RI da UNESP que se associou ao ORCID (jul./2015) para apoio ao controle de autoria (VIDOTTI; BASTOS; ROMANETTO, 2017).

Sobre o povoamento automático, Kuramoto ressalta que se o movimento do acesso livre for ignorado e se não se “desenvolver qualquer ação no sentido de criar repositórios ou provedores de serviços para colheita dos metadados de publicações ou repositórios nacionais e internacionais significa continuar dependente das publicações científicas comerciais” (KURAMOTO, 2006, p. 101).

⁷ O protocolo SWORD está disponível no DSpace desde a versão 1.5.

⁸ Disponível em: <https://www.southampton.ac.uk/~harnad/Hypermail/Amsci/7207.html>

No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), em sua cartilha Boas Práticas para a Construção de Repositórios Institucionais da Produção Científica, estabelecem que os RIs “lidam com a produção científica de uma determinada instituição” (LEITE *et al.*, 2012, p. 7), e,

constitui um serviço de informação científica – em ambiente digital e interoperável – dedicado ao gerenciamento da produção científica e/ou acadêmica de uma instituição (universidades ou institutos de pesquisa). Contempla a reunião, armazenamento, organização, preservação, recuperação e, sobretudo, a ampla disseminação da informação científica produzida na instituição (LEITE *et al.*, 2012, p. 7).

O IBICT é o principal responsável por disseminar o Movimento do Acesso Aberto no Brasil, com o projeto IBICT-FINEP/PCAL/XBDB, lançado em 2009, que viabilizou a implantação de 40 repositórios no país, por meio do fornecimento de *kits* tecnológicos, treinamento e suporte durante o processo de implantação dos repositórios. Além disso, em 2004, foi o órgão responsável por customizar, traduzir e distribuir o *software DSpace* (IBICT, 2019).

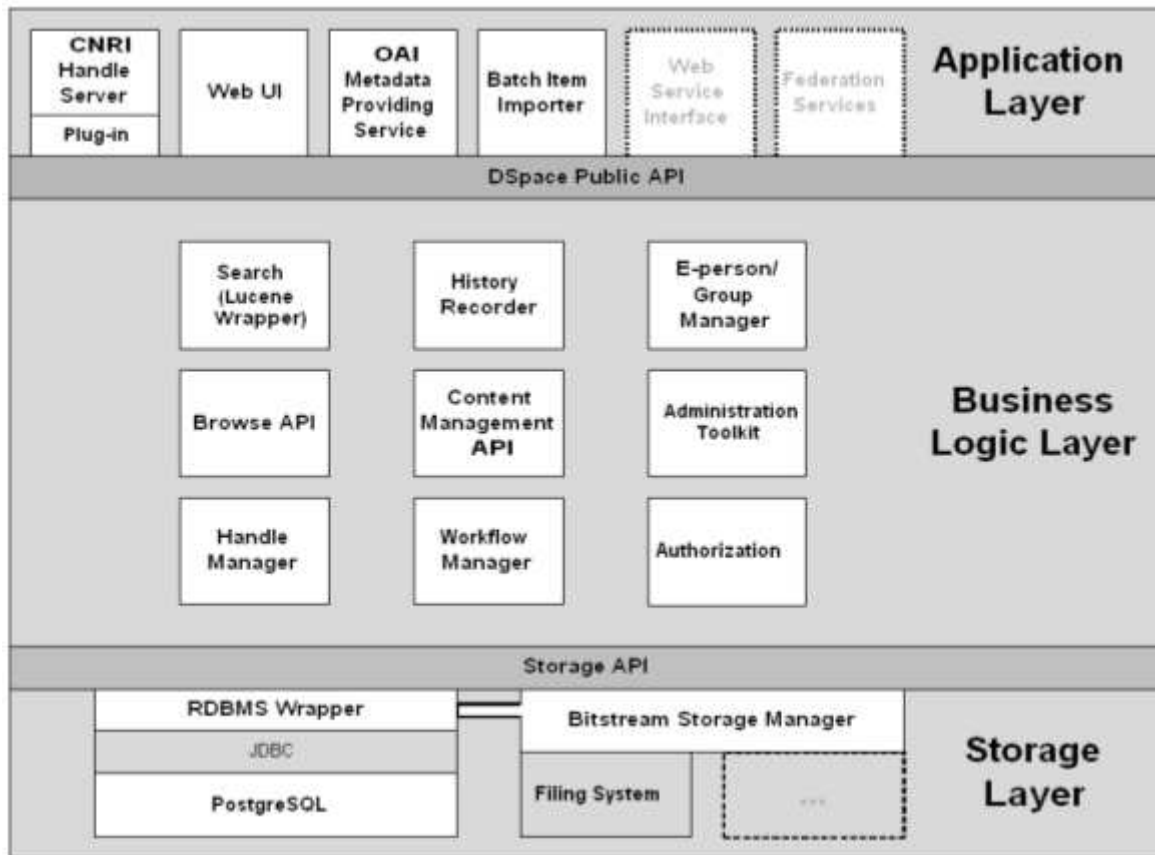
3.4 *Software DSpace*

Desenvolvido pelas bibliotecas do *Massachusetts Institute of Technology* (MIT) em conjunto com a *Hewlett-Packard* (HP), o *software DSpace* foi disponibilizado livremente em novembro de 2002. É um produto de código aberto nos termos da *BSDOpen Source License*, que permite ao usuário adaptar e expandir o *software* (DSPACE, 2020).

O *DSpace* é um conjunto de aplicativos *Web* e programas utilitários Java cooperantes, tais como *Apache Maven*, *Apache Ant*, *PostgreSQL* ou *Oracle* (sistema de gerenciamento de banco de dados objeto-relacional), *Tomcat* (um *Container Servlet* que serve para interpretar aplicações escritas em *Java* para *Web*). Esses aplicativos podem ser executados nos seguintes sistemas operacionais: *Linux*, *Windows* e *Mac OSX* (DSPACE, 2020).

Texier *et al.* (2016, p. 70) afirmam que os “recursos dentro do contexto de RI são objetos físicos ou digitais que são descritos com base na enumeração de um conjunto de dados específicos (metadados) que os distinguem de outros objetos”. Sendo assim, os *softwares* de gestão precisam apresentar um conjunto de características técnicas.

A Figura 2 mostra a arquitetura técnica do *DSpace*.

Figura 2 – Arquitetura técnica do *software DSpace*

Fonte: SMITH *et al.*, 2003, *on-line*.

Smith *et al.* (2003, *on-line*) explica que a arquitetura técnica do *DSpace* possui três camadas (armazenamento, negócios e aplicativos), “cada uma com uma API documentada para permitir personalização e aprimoramento futuros”. Segundo o autor,

a camada de armazenamento é implementada usando o sistema de arquivos, conforme gerenciado pelas tabelas de banco de dados postgresSQL. A camada de negócios é onde reside a funcionalidade específica do *DSpace*, incluindo o fluxo de trabalho, gerenciamento de conteúdo, administração e módulos de pesquisa e navegação. Cada módulo tem uma API para permitir que os adotantes do *DSpace* substituam ou melhorem essa função conforme desejado. Finalmente, a camada de aplicativo cobre as interfaces do sistema: a interface Web UI e o carregador em lote, em particular, mas também o suporte ao OAI e o servidor Handle para resolver identificadores persistentes aos itens *DSpace*. Esta é a camada que chamará grande atenção em versões futuras, pois adicionamos serviços Web para novos recursos (por exemplo, para apoiar a interoperação com outros sistemas), e definimos os serviços da Federação em toda a gama de instituições que adotam o *DSpace* (SMITH *et al.* 2003, *on-line*, tradução da autora).

Conforme destacou o autor sobre a camada de aplicativo, no âmbito do controle de autoria, essa parece ser a camada que mais tem relevância.

O Quadro 3 destaca as características técnicas do *DSpace*.

Quadro 3 – Características técnicas *DSpace*

Interface *Web*;

Fluxos de trabalho integrados, uso *workflow*, podendo um administrador aceitar ou não a submissão de um documento no repositório;

Mecanismo de pesquisa embutido, que permite a pesquisa e navegação em facetadas, o texto completo e os metadados dos registros também são pesquisáveis;

Armazena qualquer tipo de arquivo: DOC, PDF, XLS, PPT, JPEG, MPEG, TIFF;

Metadados baseado no *Qualified Dublin Core* (QDC), é interoperável com MARC e MODS;

Plugins, o *DSpace* tem embutido um conjunto de plug-ins (importação, exportação e edição de metadados em lote), além disso, existem outros que podem ser adquiridos;

Segurança, tem um sistema próprio de autenticação, mas pode ser integrado a outros sistemas;

Permissões, permite delegar acessos diferenciados para diferentes funções administrativas, bem como controlar as permissões de leitura ou edição no site, por comunidade, coleção, item ou arquivo;

Backup, todo o conteúdo pode ser exportado em arquivos *Archival Information Packages* (AIP), que podem ser usados para restaurar o sistema ou parte dele;

Protocolos: está em conformidade com os protocolos OAI-PMH, SWORD, OpenAIRE eDriver, padrões de interoperabilidade em repositórios;

REST, fornece Interface *Application Program* (API) RESTful que utiliza requisições HTTP para extrair, inserir, postar e deletar dados que permitem ao usuário conectar e interagir com aplicações na nuvem;

Banco de dados configurável, permite escolher entre os bancos de dados PostgreSQL ou Oracle;

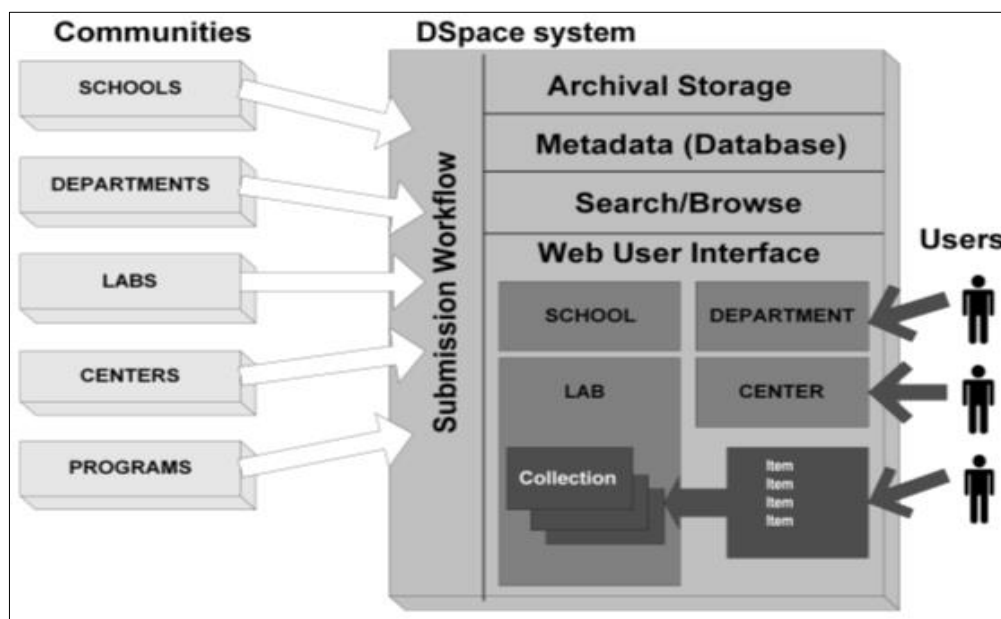
Armazenamento de arquivos configurável, arquivos podem ser armazenados em um servidor local ou em nuvem, por exemplo, Amazon;

Integridade de dados, ao realizar upload do arquivo o *DSpace* cria uma espécie de log que permite futuramente checar a integridade do arquivo;

Idiomas, disponível em mais de vinte idiomas

Fonte: DSPACE, 2020.

O modelo de informação do *DSpace* pode ser visto na Figura 3.

Figura 3 – Modelo de informação do *DSpace*

Fonte: SMITH *et al.*, 2003, *on-line*.

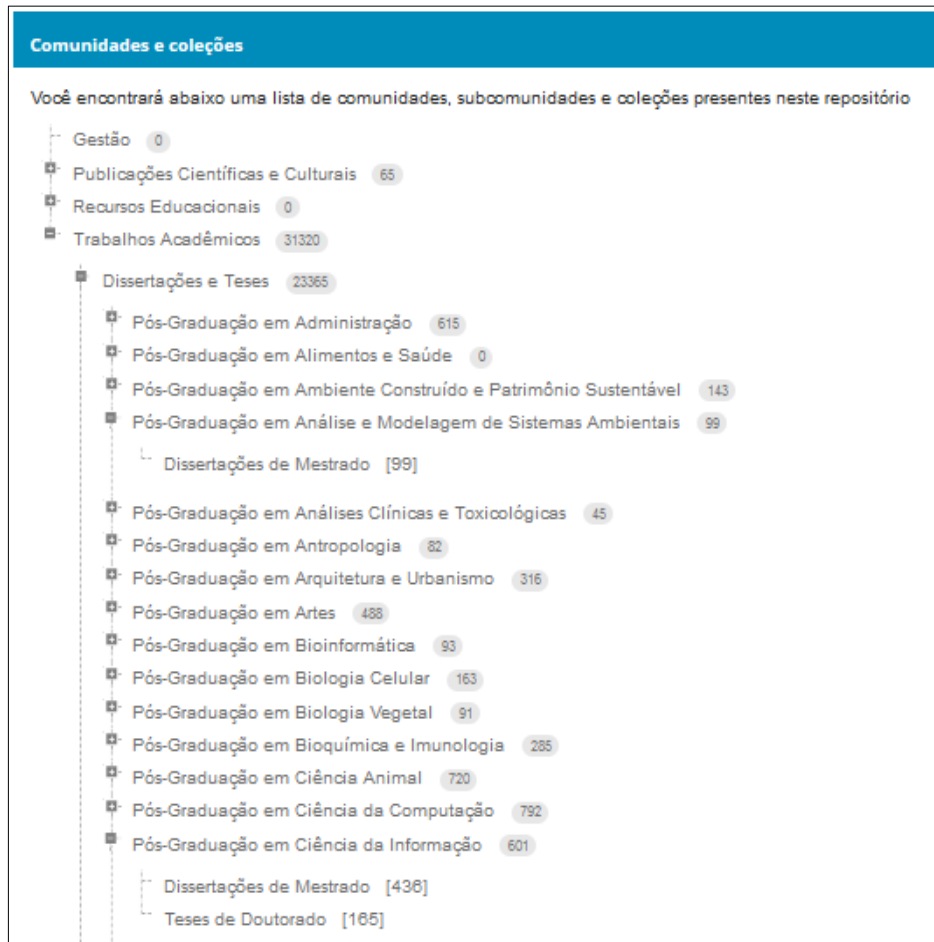
Segundo Smith *et al.* (2003, *on-line*), o “*DSpace* foi projetado para facilitar a participação dos depositantes”, sendo que o seu modelo de organização da informação “é construído em torno da ideia de ‘Comunidades’ organizacionais — subdivisões naturais de uma instituição que têm necessidades distintas de gerenciamento de informações”, pois cada Comunidade pode “adaptar o sistema para atender às suas necessidades particulares e gerenciar o processo de submissão em si”.

Destaca-se que a interface de navegação e busca do *software DSpace* se organiza de forma hierárquica:

- Comunidades: estrutura de mais alto nível e pode se dividir em subcomunidades. Em repositórios, normalmente representam as faculdades, institutos e cursos de pós-graduação. As comunidades não agrupam os documentos, apenas subcomunidades e coleções;
- Coleções: agrupam documentos semelhantes ou relacionados. Está sempre vinculada a uma comunidade ou subcomunidade. Em repositórios institucionais, é comum observar coleções por tipo de documento: dissertações, teses, artigos, por exemplo;
- Item: contêm os objetos digitais, conjunto de dados que compõem um item, os arquivos e os metadados que os descrevem. (SHINTAKU; MEIRELLES, 2010)

Na Figura 4, tem-se um exemplo de organização do Repositório Institucional da Universidade Federal de Minas Gerais (UFMG) em comunidade, subcomunidades e coleções, com seus respectivos itens.

Figura 4 – Exemplo de organização do RI da UFMG



Fonte: UFMG, 2020, *on-line*.

Observa-se, na Figura 4, que, após o nome de cada coleção, há a indicação do número de documentos que está disponibilizado nela. Ao clicar sobre a coleção ‘Dissertações de Mestrado’ da Comunidade ‘Pós-graduação em Ciência da Informação’, por exemplo, tem-se acesso aos itens que compõem a coleção (FIGURA 5).

Figura 5 – Exemplo de organização dos itens em coleções e comunidades

The screenshot shows a web interface for a digital repository. At the top, there is a blue navigation bar labeled 'Navegar'. Below it, there are several filter buttons: 'Data do documento', 'Autores', 'Orientadores', 'Título', 'Assunto', 'CNPq', 'Departamento', 'Programas', 'Tipo do Documento', and 'Tipo de Acesso'. Below the filters, there is a section for email notifications with a text box and an 'Assinar' button, and three RSS feed icons. The main content area displays a list of items from a collection, ordered by date of deposit in descending order (1 of 436 items). A 'Próximo >' link is visible. The list is presented in a table with the following columns: 'Data do documento', 'Título', 'Autor(es)', and 'Tipo'.

| Data do documento | Título | Autor(es) | Tipo |
|-------------------|--|--------------------------------|-------------|
| 7-Jun-2019 | Entrando em ação, movendo a cena: práticas informacionais nos ambientes do aplicativo Telegram | André Gustavo Fonseca da Silva | Dissertação |
| 5-Jul-2019 | Divulgação científica e cidadania nas páginas da revista Minas Faz Ciência Infantil | Manoel de Oliveira Silva | Dissertação |
| 29-Out-2018 | Inteligência competitiva: evolução, organização e tendências a partir de uma análise bibliométrica e cienciométrica | Lorena Stephanie de Camargo | Dissertação |
| 20-Mai-2019 | As relações de poder no discurso dos dispositivos legais e o desenvolvimento das coleções em bibliotecas públicas | Priscilla Pereira Gonçalves | Dissertação |
| 28-Jun-2019 | Entre o folclore e o patrimônio: um estudo de caso sobre o Centro de Arte Popular-Cemig | Amanda Dabêss de Carvalho | Dissertação |
| 5-Jun-2019 | Na teia do conhecimento: a busca de informação no varejo de moda de Belo Horizonte e os entraves enfrentados pelos seus gestores | Valéria Ayres Magalhães | Dissertação |

Fonte: UFMG, 2020, *on-line*.

Essa estrutura, além de facilitar a organização dos documentos no repositório, também facilita a busca dos itens nele depositados.

Como especificado no Quadro 3, o *DSpace* tem uma interface *Web* que permite a adoção de metadados em arquivos *eXtensible Markup Language* (XML).

Segundo Shintaku e Meirelles (2010, p. 59), o *DSpace* aceita a inclusão de diferentes esquemas de metadados, “com o objetivo de cobrir todos os aspectos importantes de uma informação em um suporte, para uma determinada finalidade, cria-se esquemas de metadados que contenham campos mais apropriados”, e que eles “devem ter apenas dois níveis, como o esquema de metadados *Dublin Core* qualificado, que possui no primeiro nível o elemento e no segundo, o qualificador”. Os autores afirmam que o “cadastramento de um novo esquema de metadados é simples no que se refere ao processo, porém, complexo quanto ao conhecimento necessário para fazê-lo com plenitude” (SHINTAKU; MEIRELLES, 2010, p. 59). Contudo, destacam que o “esquema de metadados padrão do *DSpace* é o *Dublin Core*, utilizado nos principais procedimentos que envolvem a apresentação e alimentação do repositório” (SHINTAKU; MEIRELLES, 2010, p. 59).

O *Dublin Core* tem origem de um trabalho internacional de cooperação entre OCLC e o *National Center for Supercomputing Applications* (NCSA), que teve como objetivo desenvolver um conjunto de elementos especializados para descrever fontes e recursos da *Web*

para facilitar a busca e a recuperação da informação (*Dublin Core Metadata Initiative* – DCMI, 2020).

O *Dublin Core* tem dois níveis de padrões: o simples, composto pelos 15 elementos, e o qualificado, proposto em 2000 para descreverem os elementos de maneira mais precisa e detalhada. Os qualificadores se enquadram em dois grupos: *element refinement* e *encoding scheme*. O primeiro atribui ao elemento mais especificidade, detalhando-o melhor. O segundo atribui ao elemento codificadores que lhe atribuem valor com uso de vocabulário controlado ou sistemas de classificação, por exemplo (DCMI, 2001).

O Quadro 4 descreve os 15 elementos do *Dublin Core*.

Quadro 4 – Elementos *Dublin Core*

| Elemento | Definição |
|--------------------|---|
| <i>title</i> | nome conhecido do recurso |
| <i>subject</i> | assunto do conteúdo do recurso (o assunto será representado usando palavras-chave) |
| <i>description</i> | descrição do recurso (ex.: resumo, sumário, notas) |
| <i>language</i> | idioma do recurso |
| <i>source</i> | referência a um recurso a partir do qual o recurso descrito foi derivado |
| <i>relation</i> | referência a recursos relacionados (ex.: coleção a que faz parte, livro a partir do qual o filme foi adaptado) |
| <i>coverage</i> | tópico espacial (nome de lugar ou coordenadas geográficas), temporal (período, data) ou jurisdição do recurso (abrangência de uma lei) |
| <i>creator</i> | entidade (pessoa, organização ou serviço) primariamente responsável pela confecção do recurso (ex. autor) |
| <i>contributor</i> | entidade (pessoa, organização ou serviço) responsável por ter realizado contribuições ao recurso (ex.: ilustrador, tradutor, orientador, diretor) |
| <i>publisher</i> | entidade responsável por tornar o recurso disponível (editora) |
| <i>rights</i> | informações sobre os direitos do recurso (Copyright, Creative Commons) |
| <i>date</i> | data associada a um evento do ciclo de vida do recurso |
| <i>type</i> | natureza ou gênero do recurso (texto, imagem, tese, dissertação, livro) |
| <i>format</i> | formato de um arquivo, meio físico ou dimensão de um recurso (html, ps, doc, mp3, wave, jpeg, mpeg, duração de 30 min, tamanho 300MB) |
| <i>identifier</i> | uma referência não ambígua do recurso em um dado contexto (URI, ISSN, ISBN, ...) |

Fonte: ROCHA, [200-?]

Cada campo possui um qualificador capaz de refinar o elemento principal. O campo *data*, que pressupõe ser a data de publicação, com o uso do qualificador pode atribuir ao registro um valor mais apropriado, como data de submissão, modificação, criação. No campo *relação*, tem-se a opção de atribuir o tipo de relação específica, parte de, versão de, é referenciado por. A inclusão dos qualificadores aperfeiçoa a semântica e facilita a identificação e recuperação dos registros.

O *DSpace* é um *software* totalmente hospedado (gestão total do hospedeiro), que pode ser instalado em qualquer sistema operacional (Mac OS, Windows ou Linux), sendo configurável ou customizável, atendendo a especificidades institucionais (personalização de metadados, idioma padrão, entre outros). Para a sua programação, adota a linguagem Java, que dá suporte a UTF-8 (8-bit *Unicode Transformation Format*), que é um método de codificação de caracteres Unicode (codificação binária; traduz um determinado caractere Unicode em uma sequência de binários) (JUVILER, 2020).

Dada as especificações técnicas do *DSpace* e suas características, é importante que qualquer ferramenta de correção de dados em lote ou de uso de *scripts* para correção automática de dados, leve tais elementos em consideração. Há algumas ferramentas que podem apoiar a tarefa de limpeza de metadados em RIs que usam *DSpace*, e um dos exemplos é o *OpenRefine*, de código aberto, gratuito e intuitivo, que permite a “exportação dos dados em um formato tabular, ou seja, seus metadados como um arquivo *Excel*, valor separado por tabulação [TSV], ou valor separado por vírgula [CSV]” (GALSON, 2017, *on-line*). Dessa maneira, segundo o autor, ao exibí-los como uma planilha, é possível visualizar os valores constituintes de qualquer campo em uma caixa de facetas, que, quando ordenadas em ordem alfabética permitirão encontrar os valores inconsistentes.

4 METODOLOGIA

Segundo Creswell (2014), um projeto de pesquisa contempla três componentes: os pressupostos filosóficos, a abordagem de pesquisa e, por último, os procedimentos e métodos de pesquisa. O presente estudo tem o pragmatismo como concepção filosófica, pois tem o intuito de aliar a teoria com a prática do controle de autoridade, em especial, para as entidades pessoas, tendo como motivação as situações vivenciadas em suas atividades profissionais como bibliotecária.

Seguindo a proposta de classificação de Gil (2008), esta pesquisa se caracteriza como pesquisa básica, na medida em que foi criado conhecimento e uma melhor compreensão sobre o controle para entidade pessoas em RIs. Quanto aos objetivos é exploratória, pois envolve familiarizar o pesquisador com o problema para torná-lo explícito; descritiva, uma vez que irá identificar as características de uma população com técnicas de coletas de dados padronizados, e qualitativa, pois, com base em seus objetivos, propõe uma análise interpretativa dos dados coletados para uma melhor compreensão do problema de pesquisa.

4.1 Insumos tecnológicos

Nesta seção, estão descritas as seguintes ferramentas de trabalho que foram necessárias para o desenvolvimento desta pesquisa: 1) o gerenciador de referências, Zotero e o, 2) componente do *Google Drive*, o *Google Forms*.

1) Zotero

O processo de busca de documentos para a composição do *corpus* desta pesquisa envolveu uma grande quantidade de dados, recuperando um alto volume de artigos, resumos e literatura cinzenta (não convencional e semipublicada), que precisou ser tratado e analisado, e para isso foi necessário recorrer a programas de gestão de bibliografias.

Para esta pesquisa, optou-se pelo Zotero, principalmente porque o gerenciador é *software* livre (gratuito) e de código aberto (*open source*), e a ferramenta foi tema da disciplina *Tools for research productivity* cursadas durante o Mestrado da proponente, o que facilitou a manipulação dos dados.

Lançado em 2006, pela Universidade de George Mason, teve origem como uma extensão no navegador *Mozilla Firefox*, empregado para armazenar, gerenciar, citar referências e gerar bibliografias de maneira automática. Atualmente, pode ser utilizado também como extensão em outros navegadores (*Chrome*, *Safari*), assim como na sua versão *desktop*, a *Standalone*, sendo compatível com *MS Word* e o *LibreOffice* (MENDES *et al.*, 2019).

2) *Google Forms*

O *Google Forms* ou Formulário é um aplicativo de gestão de dados de pesquisas que compõe o pacote *Google Drive*. O aplicativo permite criar, *on-line*, instrumentos de coleta de dados personalizados, tais como questionários e entrevistas. Esse aplicativo foi escolhido por ser gratuito, pela possibilidade de envio *on-line* e pela facilidade de compilação dos dados. Ademais, a ferramenta permite configurar, para cada pergunta, a obrigatoriedade ou não da resposta para as questões propostas, impedindo que o questionário seja retornado ao pesquisador em branco, por exemplo.

4.2 Instrumento de coleta de dados: questionário

Um questionário é “constituído por uma série de perguntas, que devem ser respondidas” (MARCONI; LAKATOS, 2003, p. 201). Gil (2008, p. 121) define o questionário “como a técnica de investigação composta por um número mais ou menos elevado de questões apresentadas por escrito às pessoas, tendo por objetivo o conhecimento de opiniões, crenças, sentimentos, interesses, expectativas, situações vivenciadas, etc.”. De modo geral, os questionários têm pontos fortes e fracos, como demonstrado no Quadro 5.

Quadro 5 – Pontos fortes e fracos do questionário

| Técnica de Coleta | Pontos Fortes | Pontos Fracos |
|-------------------|---|--|
| Questionário | <ul style="list-style-type: none"> • garante anonimato; • questões em geral, objetivas, de fácil pontuação; • padronização das questões garante a uniformidade; • flexibilidade de tempo para dar respostas; • fácil converter os dados em arquivos legíveis por máquina; • custo razoável e aceitável. | <ul style="list-style-type: none"> • baixa taxa de respostas; • sem comprovação de respostas; • dificuldade em pontuar questões abertas; • possibilidade de respostas influenciadas; • limitada às pessoas aptas à leitura; • sujeito a ambiguidade e polarização. |

Fonte: adaptado de RIBEIRO, 2008.

O questionário permite a elaboração de questões de forma sistemática, que auxiliam o pesquisador na coleta de informações e tem como objetivo compreender a opinião ou o conhecimento dos participantes da pesquisa sobre um determinado tema. Eles podem ser compostos por questões fechadas, com respostas pré-definidas, ou abertas, em que o participante constrói suas próprias respostas (SEVERINO, 2007).

O questionário adotado nesta pesquisa é do tipo questões abertas, conforme exposto no Apêndice A. Nesse tipo de questionário não há a sugestão de respostas prontas, que são dadas pelos respondentes, e, a análise das respostas não é estatística, pois busca trazer informações novas. Este tipo de questionário tem vantagens e desvantagens, como mostra o Quadro 6.

Quadro 6 – Vantagens e desvantagens das questões abertas

| Técnica de Coleta | Vantagens | Desvantagens |
|----------------------------------|---|---|
| Questionário de Questões Abertas | <ul style="list-style-type: none"> • estimulam cooperação; • admitem melhor avaliação das atitudes nas análises das questões; • evidenciam pontos para além daqueles conseguidos com as questões fechadas; • influenciam menos os respondentes do que as perguntas fechadas; • permitem comentários, explicações e esclarecimentos mais significativos; • permitem maior abrangência nas respostas. | <ul style="list-style-type: none"> • permite parcialidade do entrevistador, já que não há um padrão claro de respostas possíveis; • são difíceis de codificar; • possibilitam interpretação subjetiva; • podem ser prejudicadas pelas dificuldades e preguiça de redação dos respondentes; • permitem que os respondentes divagam ou fujam do assunto nas respostas; • são mais onerosas e demoradas para serem analisadas. |

Fonte: adaptado de MATTAR, 1994.

Analisando as vantagens das questões abertas, no item que trata sobre a possibilidade de maior abrangência nas respostas, destaca-se que, dependendo do tipo de conteúdo das perguntas, é possível verificar distintos elementos, tais como: fatos, crenças quanto a fatos, sentimentos, crenças quanto a sentimentos, padrões de ação e padrões de comportamento, estejam esses elementos no presente ou no passado.

4.3 Procedimentos metodológicos

Para melhor sistematização, os procedimentos metodológicos foram divididos em duas etapas: 1) Estudo com gestores: aplicação de questionário; 2) Mapeamento e descrição dos recursos de controle de autoridade para o *software DSpace* (nativos ou não), cujos detalhamentos estão descritos a seguir.

PRIMEIRA ETAPA: Estudo com gestores de RIs: aplicação de questionário

Para melhor compreensão de como o controle de autoridade tem sido realizado nos RIs das universidades que utilizam o *DSpace*, buscando evidenciar se utilizam ou não os elementos computacionais disponíveis para o *software* (nativos ou não), e a razão para o não uso (quando for o caso), propôs-se a elaboração de um instrumento de coleta de dados do tipo questionário. Para este estudo, fez-se um recorte nas Universidades Públicas Federais existentes em Minas Gerais, tendo em vista que a coleta dos dados ocorreu em meio à pandemia do COVID-19, estando as instituições em regime de tele trabalho, e apenas oferecendo os seus serviços básicos.

Esta etapa inclui as seguintes cinco fases:

1. Identificação das Universidades Públicas Federais existentes em Minas Gerais: busca pelas universidades que atendam ao escopo estabelecido no e-MEC.
2. Identificação das Universidades Públicas Federais mineiras que possuem RIs com a utilização do *software DSpace*: dentre as universidades identificadas na Primeira Fase, foram selecionadas aquelas que possuem RIs, e, dentre elas, aqueles que utilizam o *software DSpace*.
3. Identificação dos dados dos gestores dos RIs: levantamento dos gestores responsáveis pelos RIs, e dos dados de contato para a aplicação de um questionário.
4. Construção do instrumento de coleta de dados - o questionário - para entrevistar os gestores dos RIs: nesta fase é desenvolvido o instrumento de coleta de dados, para ser aplicado aos gestores dos RIs.
5. Análise das respostas ao questionário aplicado aos gestores dos RIs: tabulação e análise das respostas, com foco no escopo da pesquisa, tendo em vista realizar inferências.

SEGUNDA ETAPA: Mapeamento e descrição dos elementos computacionais de controle de autoridade para o *software DSpace* (nativos ou não)

Esta etapa compreendeu três fases, conforme descrito a seguir.

1. Descrição da versão mais atualizada do *software DSpace*: esta fase tem o intuito de estabelecer as características atuais do *software*, evidenciando as possíveis compatibilidades com os aplicativos não nativos.
2. Mapeamento dos elementos computacionais (nativos ou não) de controle de autoridade para o *software DSpace*: levantamento dos elementos computacionais para identificação daqueles que têm funções relacionadas ao controle de autoridade.
3. Descrição dos elementos computacionais (nativos ou não) de controle de autoridade: nesta fase se descrevem os elementos para a versão atualizada do *software DSpace*, a partir do Manual *DSpace* e dos documentos recuperados na fase dois, de mapeamento.

5 RESULTADOS E DISCUSSÕES

Neste Capítulo, os resultados obtidos nas duas etapas dos procedimentos metodológicos estão descritos em cada uma de suas respectivas fases.

5.1 Resultados da Primeira Etapa

5.1.1 Identificação das Universidades Públicas Federais existentes em Minas Gerais

O escopo de estudo desta pesquisa são as Universidades Públicas Federais do Estado de Minas Gerais, que estão identificadas no site e-MEC, Cadastro Nacional de Cursos e Instituições de Educação Superior, que reúne os dados das Instituições de Ensino Superior (IES) do país.

No site mencionado, a busca foi realizada na opção avançada, com seleção dos filtros: UF, categoria administrativa, e organização acadêmica. Foram selecionados, respectivamente, Minas Gerais, pública federal, universidade.

Como resultado, obteve-se um número total de 11 Universidades Federais existentes no Estado de Minas Gerais, conforme detalhamento no Quadro 7.

Quadro 7 – Identificação das Universidades Públicas Federais Minas Gerais

| Itens | Universidade | Sigla |
|-------|--|--------|
| 1 | Universidade Federal de Alfenas | UNIFAL |
| 2 | Universidade Federal de Itajubá | UNIFEI |
| 3 | Universidade Federal de Juiz De Fora | UFJF |
| 4 | Universidade Federal de Lavras | UFLA |
| 5 | Universidade Federal de Minas Gerais | UFMG |
| 6 | Universidade Federal de Ouro Preto | UFOP |
| 7 | Universidade Federal de São João Del-Rei | UFSJ |
| 8 | Universidade Federal de Uberlândia | UFU |
| 9 | Universidade Federal de Viçosa | UFV |
| 10 | Universidade Federal dos Vales do Jequitinhonha e Mucuri | UFVJM |
| 11 | Universidade Federal do Triângulo Mineiro | UFTM |

Fonte: Dados da pesquisa, 2020.

5.1.2 Identificação das Universidades Públicas Federais mineiras que possuem RIs com a utilização do *software DSpace*

Para a concretização desta fase, foi realizada uma visita aos sites oficiais de todas as Universidades mineiras identificadas na Primeira Fase desta Segunda Etapa. Isso foi feito, mais especificamente, para mapear a existência ou não de RI, e, por conseguinte, a utilização ou não do *software DSpace*, assim como a versão do *software* que está implantada no RI.

A verificação e reconhecimento, com clareza, da utilização do *software* e sua versão foi realizada com o acesso à página exclusiva do RI, quando se optou pela exibição do seu código fonte. Essa estratégia permitiu identificar as informações de maneira clara e confiável, sendo que os dados foram coletados em 20 de dezembro de 2019.

Ao final da busca nos sites, verificou-se, dentre as 11 universidades Públicas Federais existentes em MG, a existência de 10 universidades que possuem RI, fazendo a utilização do *software DSpace*, conforme mostra o Quadro 8.

Quadro 8 – Universidades Públicas Federais Minas Gerais que utilizam *DSpace*

| Item | Sigla | Tem RI | Utiliza DSpace | Versão |
|------|--------|--------|----------------|--------|
| 1 | UNIFAL | Sim | Sim | 4.2 |
| 2 | UNIFEI | Sim | Sim | 5.3 |
| 3 | UFJF | Sim | Sim | 6.3 |
| 4 | UFLA | Sim | Sim | 6.1 |
| 5 | UFMG | Sim | Sim | 5.7 |
| 6 | UFOP | Sim | Sim | 5.4 |
| 7 | UFSJ | Não | Não | - |
| 8 | UFU | Sim | Sim | 6.3 |
| 9 | UFV | Sim | Sim | 6.3 |
| 10 | UFVJM | Sim | Sim | 5.2 |
| 11 | UFTM | Sim | Sim | 4.2 |

Fonte: Dados da pesquisa. Nota: Coletados em 20 de dezembro de 2019.

A Universidade Federal de São João Del-Rei (UFSJ), apesar de atender aos requisitos iniciais para compor a amostra, foi excluída por não possuir um RI na época da coleta de dados

(dez./2019). No site oficial da Instituição encontra-se a informação de que o RI está em fase de implementação, mas, até a data de coleta de dados, isso ainda não havia ocorrido. Sendo assim, o conjunto da amostra ficou composto pelas dez universidades (QUADRO 8).

5.1.3 Identificação dos dados dos gestores dos RIs

Nesta fase, foram extraídos, também no site oficial das Universidades Públicas Federais de Minas Gerais (MG), os dados para contato com os gestores dos repositórios. Quando o site não disponibiliza o contato do gestor, optou-se por entrar em contato por telefone com a Biblioteca Universitária ou setorial para assim identificar o gestor do RI e seus dados de contato.

A proposta inicial era ligar para os números de telefone que constavam nas páginas institucionais dos RIs, para coletar o endereço de *e-mail* dos gestores, de forma mais precisa e confiável, assim como para uma conversa inicial sobre a pesquisa, e sobre a possibilidade de envio por *e-mail* de um questionário para coleta de informações sobre o processo de gestão e controle das autoridades no RI da universidade. Após inúmeras tentativas frustradas, provavelmente pelo trabalho remoto devido à pandemia Covid-19, decidiu-se por registrar os dados dos *e-mails* cadastrados nas páginas oficiais dos RIs.

5.1.4 Construção do questionário, instrumento de coleta de dados para entrevistar os gestores dos RIs

Nesta Quarta Fase, foi criado um questionário do tipo semiestruturado, composto por questões abertas, conforme exposto no Apêndice A, para ser enviado aos gestores dos RIs (conforme seção 5.1.3). Cabe destacar, que se optou por perguntas abertas para evitar o direcionamento das respostas. A intenção foi deixar o gestor livre para relatar a rotina do controle de autoridade nos RIs, assim como expor as possíveis angústias que esse trabalho pode envolver e identificar algum grau de desconhecimento dos gestores sobre as funcionalidades oferecidas pelo *software DSpace*.

Para acompanhar o questionário e resguardar o sigilo e confidencialidade da pesquisa, foi elaborado o Termo de Consentimento Livre e Esclarecido (TCLE, conforme APÊNDICE B), que é o protocolo para obtenção de consentimento da pesquisa com ética. Ele traz esclarecimentos (tais como responsabilidades, objetivos, justificativa, metodologia, possíveis e

previsíveis riscos, constrangimentos e desconfortos, benefícios, garantia de sigilo de dados e de outras garantias do participante), para que o participante possa decidir sobre a sua participação ou não na pesquisa. O projeto de pesquisa foi submetido ao Comitê de Ética e Pesquisa da UFMG (COEP/UFMG), em 28 de fev. 2020, sob o CAAE 29769320.8.0000.5149, atendendo à necessidade de avaliação pelo Sistema CEP/CONEP (Comitê de Ética em Pesquisa; Comissão Nacional de Ética em Pesquisa), conforme a Resolução CNS nº 510/2016. O projeto obteve aprovação, conforme Parecer Nº 4.048.320, emitido em 26 de maio de 2020 (ANEXO A). No formulário, o respondente só teve acesso ao questionário após ler e aceitar as condições previstas no TCLE.

O questionário é constituído por nove perguntas abertas, sendo que as duas primeiras são de identificação da Instituição e do gestor do RI. Neste último caso, opcional, é para o caso de o gestor ter interesse em receber os resultados da pesquisa, como previsto no TCLE. As outras sete perguntas são voltadas para o processo de inserção (povoamento) e gestão dos registros de autoridade no RI.

Antes do envio do questionário para a amostra mencionada, foi aplicado um pré-teste. A aplicação possibilitou identificar e sanar problemas de inteligibilidade das questões e problemas nos procedimentos metodológicos. Dessa maneira, os problemas foram sanados e o questionário aplicado.

5.1.5 Análise das respostas ao questionário aplicado aos gestores dos RIs

Conforme estabelecido na seção 5.1.2, a amostra compreendeu dez instituições que possuem RIs, utilizando o *software DSpace*. Destaca-se que a coleta dos dados se deu no auge da pandemia do COVID-19, período em que as instituições adotaram o trabalho remoto como alternativa para a manutenção de serviços básicos. Assim, o *link* do *Google Forms*, com as nove perguntas abertas (como descrito na seção 5.1.4) do questionário, com informações sobre a pesquisa e sobre a participação dos respondentes, foi enviado para o *e-mail* dos gestores, cadastrados nas páginas oficiais dos RIs (conforme seção 5.1.3), no dia 19 de junho de 2020.

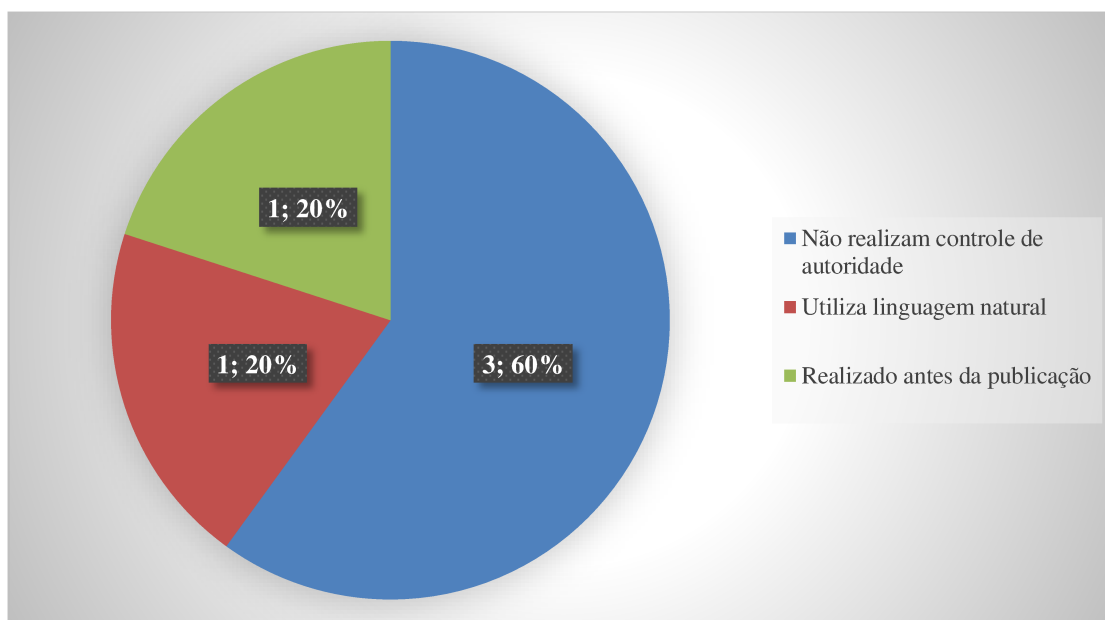
A taxa de retorno via formulário foi de 50% (cinquenta por cento) dos RIs da amostra. Ou seja, foram recebidas 5 respostas completas. Houve uma sexta resposta, enviada por *e-mail*, que foi excluída, pois, além de ser entregue fora do formulário, o participante não respondeu a todas as perguntas do questionário.

A seguir, apresentam-se os dados e a transcrição de alguns trechos das respostas, para efeito de análise e discussão. Os cinco respondentes foram identificados pelos códigos E1, E2, E3, E4 e E5, atribuídos de modo aleatório, mantendo-se o sigilo e confidencialidade, conforme previsto na Resolução CNS nº 510/2016. A exposição, análises e discussões serão feitas em separado, para cada uma das sete questões do questionário - da questão 2 à questão 8 -, visto que duas questões: a questão 1, que trata da identificação da instituição, e a questão 9, que solicita um endereço de *e-mail* para receber os resultados da pesquisa, serão mantidas em sigilo, por motivos éticos, atendendo à Resolução 516/2016.

QUESTÃO 2: Como é realizado o controle de autoridade no Repositório Institucional da sua Universidade?

Nesta questão 2, 3 (ou 60%) respondentes disseram não realizarem controle de autoridade; 1 (ou 20%) respondente disse que utiliza linguagem natural, e 1 (ou 20%) respondente disse que o controle de autoridade pessoas é realizado antes da publicação, provavelmente, por quem deposita a produção acadêmica.

Gráfico 1 – Questão 2: Como é realizado o controle de autoridade no Repositório Institucional da sua Universidade?



Fonte: Dados da pesquisa, 2020.

Este resultado evidencia o que foi discutido neste estudo, por autores como Salo (2009) e Xia (2006), ainda válidos na contemporaneidade. O que o primeiro destaca é que a concepção dos RIs foi feita sem atenção ao controle de autoridade, e, por sua vez, Xia (2006) ressalta que a

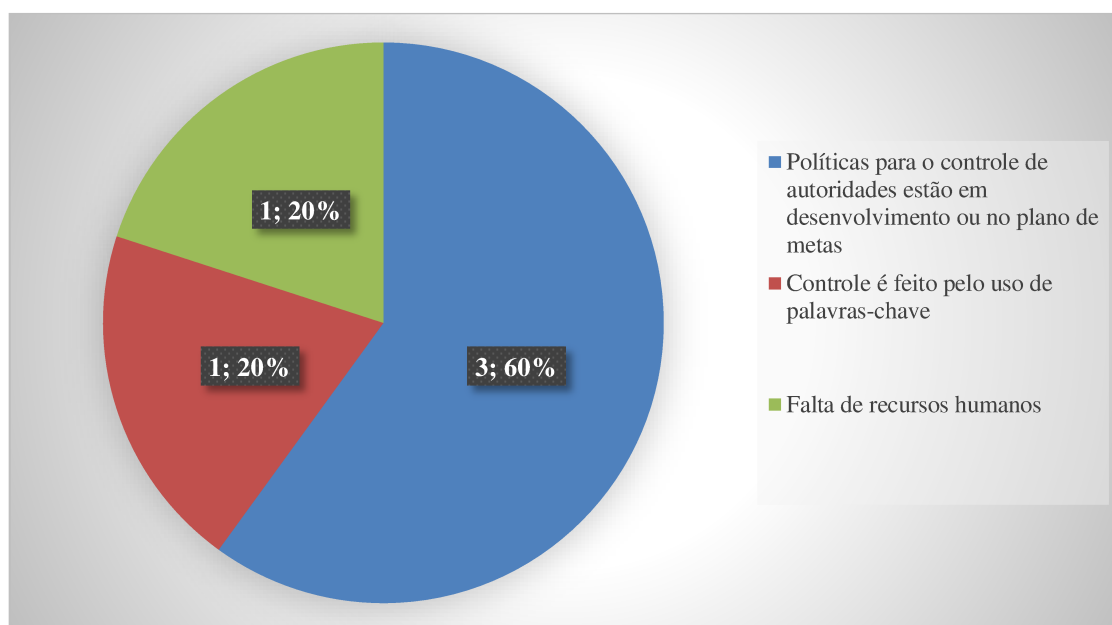
falta de um controle de autoridade nos RIs torna a busca e recuperação de informações mais difícil, sobretudo, quando se trata da busca por nomes de autores.

QUESTÃO 3: Caso não realize o controle de autoridade, como justifica isso?

Nesta questão 3, 3 (ou 60%) dos respondentes justificam afirmando que as políticas de povoamento para o controle de autoridades estão em desenvolvimento ou no plano de metas; 1 (ou 20%) dos respondentes justifica dizendo que o controle é feito pelo uso de palavras-chave; e 1 (ou 20%) dos respondentes justifica com a falta de recursos humanos, pois,

o repositório também é alimentado por autoarquivamento, o que dificulta muito o controle das autoridades. Até o momento não possuímos pessoal suficiente para que possamos acompanhar de maneira que possamos controlar as autoridades inseridas no sistema (E1, 2020).

Gráfico 2 – Questão 3: Caso não realize o controle de autoridade, como justifica isso?



Fonte: Dados da pesquisa, 2020.

De fato, a literatura trabalhada evidencia um despreparo para o efetivo controle de autoridades, e mostra que o autoarquivamento (*self-archiving*) foi uma estratégia criada para o povoamento rápido dos RIs, mas que isso foi feito sem a elaboração de políticas institucionais para orientarem os pesquisadores nessa tarefa (SALO, 2009; ASSIS, 2013). Outros problemas, também já apontados no referencial teórico, referem-se a fatores tais como recursos financeiros, materiais e humanos insuficientes para a árdua tarefa de gestão de RIs (SALO, 2008).

Sobre o uso das palavras-chave, que são palavras livres atribuídas em linguagem natural, no controle de autoridades do RI, foi entendido que o respondente se referia ao controle de

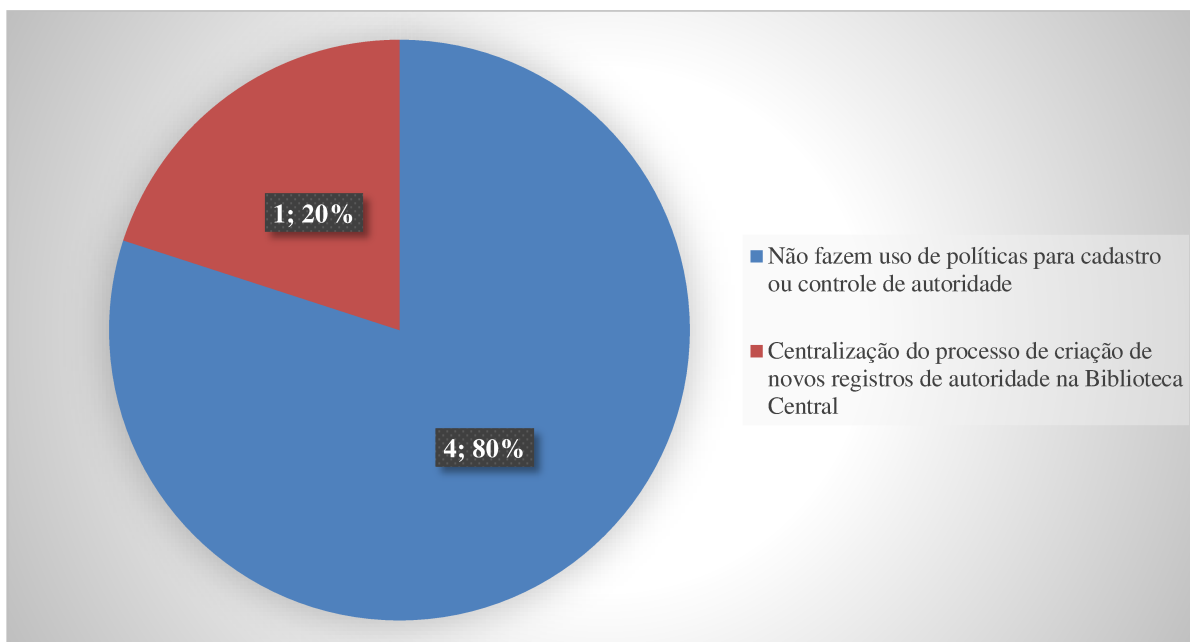
autoridades para assunto, elemento que não faz parte do escopo deste estudo. Contudo, considera-se não ser essa uma prática eficaz, quando aplicada de modo isolado. Mas, muitas vezes, é o que ocorre, como mostraram os resultados de Steele e Sump-Crethar (2016), na revisão de literatura, (Artigo 16). O controle dos assuntos em RIs, numa visão ideal, deveria utilizar um vocabulário controlado de termos/assuntos, para manter a consistência na recuperação de informações, e também permitir a inclusão de palavras-chave livres, aquelas fornecidas pelos próprios autores em suas publicações (ZAVALINA, 2010). Com o uso simultâneo das duas estratégias, esperam-se melhores resultados na busca do usuário (LOPES, 2002). Entretanto, a pesquisa de Steele e Sump-Crethar (2016) aponta que 61% dos repositórios investigados não adotam vocabulário controlado no campo de assunto. Assim, as respostas a esta questão confirmam a literatura sobre o controle de autoridade, a exemplo dos estudos tanto de Steele e Sump-Crethar (2016) quanto de Sandy e Dykas (2016), que apontam que, apesar do envolvimento dos bibliotecários na gestão dos RIs e do reconhecimento deles sobre a importância do controle de autoridade para a consistência e qualidade dos dados, o controle ainda é um desafio importante.

QUESTÃO 4: O repositório da sua universidade possui política para cadastro e controle de autoridade? Se sim, por favor, descreva.

Nesta questão 4, 4 respondentes (ou 80%) disseram que não fazem uso de políticas para cadastro ou controle de autoridade. O último respondente (ou 20%) afirma que

foi adotada a centralização do processo de criação de novos registros de autoridade na Biblioteca Central [...]. Mediante a necessidade de inserção de novos registros de autoridade pelas bibliotecas setoriais, estas solicitam a criação pela Central, através de formulário. [...] foram definidos os campos a serem utilizados para cada tipo de registros de autoridade, as fontes de informação para consulta dos termos, das remissivas, datas associadas, biografia e notas explicativas dos termos. Desta forma, tem sido possível melhorar a consistência e precisão dos pontos de acesso (E2, 2020).

Gráfico 3 – Questão 4: O repositório da sua universidade possui política para cadastro e controle de autoridade? Se sim, por favor, descreva.



Fonte: Dados da pesquisa, 2020.

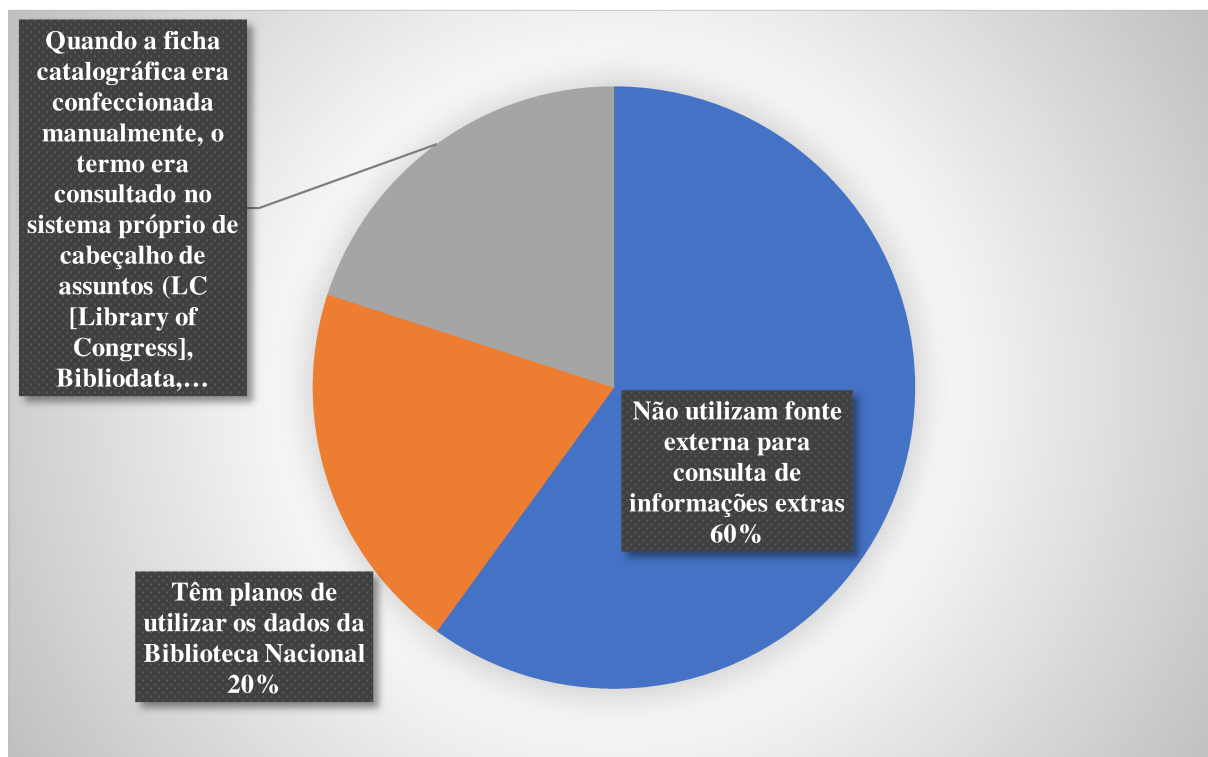
Não foi surpresa apenas um dos gestores ter respondido que havia uma política de povoamento, visto que a literatura evidencia que os erros de povoamento no *DSpace* muitas vezes se devem à inexistência de uma política clara que determine quais campos de metadados devem ser preenchidos e quais são os critérios e parâmetros para tal preenchimento (SALO, 2008). Acredita-se que, realmente, parece haver falta de apoio local para a criação de políticas de povoamento e para o treinamento de pesquisadores na atividade de autoarquivamento, conforme alertam Stein, Applegate e Robbins (2017) e Downey (2019). É importante que todos os responsáveis pela implantação e gestão dos RIs se atentem para o fato de que é preciso refletir e elaborar políticas de povoamento, materiais de apoio e de treinamento para o sucesso das iniciativas (ASSIS, 2013). Ademais, ao não adotarem políticas e boas práticas para criação dos metadados, os gestores acabam sobrecarregando o sistema de busca do repositório, atribuindo a ele a responsabilidade pelo mau desempenho (SANDY; DYKAS, 2016).

QUESTÃO 5: O repositório da sua universidade utiliza fonte externa para consulta de cadastro de autoridade? Se sim, por favor, descreva.

Nesta questão 5, 3 (ou 60%) dos respondentes disseram que não utilizam fonte externa para consulta de informações extras; 1 (ou 20%) dos respondentes disseram que tem planos de utilizar os dados da Biblioteca Nacional; e 1 (ou 20%) dos respondentes disse que “quando a ficha catalográfica era confeccionada manualmente, o termo era consultado no sistema próprio

de cabeçalho de assuntos (LC [*Library of Congress*], Bibliodata, Biblioteca Nacional)” (E3, 2020).

Gráfico 4 – Questão 5: O repositório da sua universidade utiliza fonte externa para consulta de cadastro de autoridade. Se sim, por favor, descreva



Fonte: Dados da pesquisa, 2020.

A decisão por não consultar fontes externas para o cadastro de autoridade vai ao encontro do que se vê relatado na literatura. Isso pode ser visto em Salo (2009; XIA, 2006; DOWNEY, 2019; CRAFT, 2019), que afirma que a consulta de nomes dos autores em catálogos de autoridade tradicionais como *Library of Congress* e VIAF poderia ser pouco viável, já que as autorias dos RIs são inéditas, na maioria das situações. O autor sugere arquivos de autoridade de nomes locais, que podem estar vinculados ao sistema acadêmico, e que poderiam ser compartilhados por diferentes repositórios. Craft (2019), Weidner, Washington e Liu (2019), assim como Neatrou e Myntti (2019), compartilham da mesma opinião, desde que esses arquivos sejam criados com o propósito de compartilhamentos locais ou nacionais. Almuzara, Alvite Díez e Bravo (2012, p. 106, tradução da autora)

reforçam a necessidade de estabelecer políticas eficazes para a gestão de autoridades neste tipo de coleção digital, por meio de esforços cooperativos que permitam o desenvolvimento de entradas de autoridade corporativas, que auxiliarão os processos de catalogação, criação de metadados e recuperação de informação.

No Brasil, uma fonte externa interessante poderia ser a Plataforma Lattes, pois nela se registram os nomes (e suas variações) de quase a totalidade de pesquisadores brasileiros. Porém, a literatura destaca a falta de controle e de padronização nos dados (MUGNAINI *et al.*, 2012; MATIAS, 2015), visto que são os próprios autores que fazem a entrada dos dados.

Por sua vez, internacionalmente, o ORCID é promissor como fonte de dados externos, e até mesmo para ser acoplado aos sistemas dos RIs. No entanto, a literatura aponta alguns problemas, tais como o da falta de uma consolidação de políticas que precisam ser criadas pelas instituições (DOWNEY, 2019), e também a baixa adesão das universidades a essa iniciativa.

QUESTÃO 6: O repositório da sua universidade usa alguma aplicação ou ferramenta para controle da qualidade dos metadados de autoridade (campo autor)? Se sim, por favor, descreva. Nenhum dos respondentes adotava aplicação ou ferramenta para o controle de autoridade.

Nesta questão 6, todos os 5 (ou 100%) respondentes disseram que não utilizam qualquer ferramenta para o controle da qualidade dos metadados de autoridade.

As ferramentas para o controle da qualidade de metadados são ainda mais essenciais em meio digital, visto que o volume de trabalho é sempre muito grande e o controle manual dos nomes dos autores seria problemático, sendo necessário encontrar formas viáveis para a desambiguação dos nomes dos autores (SMALHEISER; TORVIK, 2009). A literatura estudada nesta pesquisa mostra que o nome pode ter um caráter referencial (MARIANI, 2014; SEABRA; ISQUERDO, 2018), pois, para os autores científicos é um meio para divulgarem as pesquisas que desenvolvem e para pleitearem bolsas de fomento (SILVEIRA, 2013), estabelecendo a ideia de Frege (1978), de que há uma relação entre referência e nome próprio dos autores acadêmicos, para não haver imprecisão na computação de sua produção. Ou então, pode-se considerar que o nome tem um sentido essencial (SEARLE, 1996), uma vez que está ligado a descrições, que têm como função o tornar uma referência intencional e unívoca, que comunica aquele conjunto de descrições.

Nessa perspectiva, o FRAD permite a descrição de um conjunto de informações sobre as pessoas, tais como país (de nascimento, residência, etc.), filiações, endereço (residencial, comercial), idioma, entre outros. E esses atributos e a abordagem conceitual da família FRBR também estão presentes no RDA, que permite codificar os dados em MARC 21 e em *Dublin Core*, por exemplo.

Sendo os bibliotecários “mediadores entre informação científica e seus leitores, atendendo as expectativas de quem a produz e de quem a utiliza” (LEITE, 2009, p. 99), e as rotinas e processos em um RI são semelhantes aos de qualquer biblioteca, é preciso que esse profissional tenha ou adquira competências que evidenciem que as tecnologias são um novo modo de sua atuação. Claro que eles não estarão sozinhos na implantação e manutenção de RIs, que devem contar com equipe multidisciplinar de especialistas (cientistas da computação, analistas de sistemas e bibliotecários), para que se possam solucionar questões de tecnologia (suporte e customização, por exemplo), assim como os problemas de gestão, avaliação de necessidades do usuário e os instrumentos de organização da informação em ambiente digital.

QUESTÃO 7: O repositório da sua universidade aceita autodepósito ou autoarquivamento pelos autores? Se sim, os metadados são revistos por bibliotecário antes da publicação no repositório?

Nesta questão 7, 4 respondentes (ou 80%) disseram que aceitam o autodepósito ou autoarquivamento, sendo que 1 (ou 20%) deles esclareceu que a quantidade desse tipo de povoamento é “insignificante” (E5, 2020). Apenas 1 respondente (ou 20%) disse que não aceita.

Sobre a revisão por bibliotecário, percebe-se que daqueles 4 respondentes (ou 80%) que aceitam o autodepósito ou autoarquivamento, apenas 1 (ou 20%) respondente disse que a maioria passa por revisão e 1 (ou 20%) respondente falou que será (no futuro). Ademais, 1 (ou 20%) respondente não esclareceu esse ponto, dizendo que “todos os trabalhos depositados são realizados através de autoarquivamento” (E3, 2020), sem acrescentar outra informação. E o último dos 4 respondentes, 1 (ou 20%) disse que a quantidade de autodepósito ou autoarquivamento é mínima, o que subentende-se que são os próprios bibliotecários que povoam o RI, na maior parte das vezes. Por fim, considera-se que o único respondente que não permite autodepósito ou autoarquivamento tem o seu RI povoado por bibliotecários.

As análises nesta questão retomam as observações proferidas na Questão 4, quando se afirma que problemas em povoamento de RIs muitas vezes estão atrelados à falta de uma política com critérios e parâmetros para o preenchimento no autodepósito ou autoarquivamento (SALO, 2008; MURUGATHAS; BALASOORIYA, 2014), e também para os bibliotecários, tendo em vista sistematizar a tarefa e não sobrecarregar o sistema ao não adotarem boas práticas para criação de metadados (SANDY; DYKAS, 2016).

Na literatura analisada, fica evidente que a conferência de metadados é essencial para o acesso, estruturação e gestão de recursos digitais em RIs (SAYÃO, 2010). E é recorrente a recomendação de políticas para regulamentar o funcionamento do RI, monitoradas por equipes institucionais capacitadas (MATIAS, 2015; COSTA; LEITE, 2017; ESPÍNDOLA; PEREIRA, 2017), seja para o povoamento do RI, seja para o treinamento para o autodepósito ou autoarquivamento (BABINI; MACHIN-MASTROMATTEO, 2015). Essas lacunas muitas vezes ocorrem devido a recursos financeiros, humanos e materiais insuficientes, assim como falta de apoio da gestão institucional para a manutenção do RI. Além da criação de políticas, baseadas em regras e instrumentos, a gestão de RIs também exige uma gestão cooperativa (ALMUZARA; ALVITE DÍEZ; BRAVO, 2012; MEY; SILVEIRA, 2009; IFLA, 2016).

QUESTÃO 8: O repositório da sua universidade é alimentado por autopovoamento? Se sim, os metadados são revistos para evitarem duplicidade de registros ou formas de entrada.

Nesta questão 8, 2 respondentes (ou 40%) disseram que realizam o autopovoamento (ou povoamento automático), sendo que um deles afirmou que todos os metadados são revisados, e quando percebe-se que ainda há duplicidade, o problema é “sanado posteriormente” (E4, 2020); e o outro respondente disse que a tarefa é feita “na medida do que é possível” (E2, 2020). Os outros 3 (ou 60%) respondentes disseram que não realizam o autopovoamento, sendo que um deles afirmou que “este projeto está nas metas” (E3) da instituição para o RI.

Aqui se retoma as afirmações de Kuramoto (2006), quando este afirma que a implementação de RIs exige a criação de mecanismos para a colheita de metadados externos, de publicações ou repositórios nacionais e internacionais. E, se por um lado o povoamento automático pode ser considerado parte inerente à implementação de um RI, por outro lado, é essencial que se faça a revisão dos metadados rastreados e coletados, uma vez que, mesmo em projetos robustos de RI, tal como da Universidade de Illinois, que possui políticas consolidadas, ainda é possível encontrar erros (STEIN; APPLGATE; ROBBINS, 2017).

Ressalta-se que os erros de povoamento no *DSpace* somente poderão ser revisados pelos administradores do RI, uma vez que, em caso de autodepósito ou autoarquivamento, o pesquisador não poderá alterar os dados de um item que arquivou, nem mesmo excluir o item para corrigir o problema (SALO, 2008). Com isso, percebe-se a relevância de políticas claras e completas para que o RI tenha sucesso, assim como a revisão dos metadados após o povoamento, seja ele automático, monitorado ou de autodepósito.

Os autores destacam que os metadados devem ser limpos a partir de funcionalidades existentes no próprio *DSpace*, tal como *MetadataValueUpdate*, ou por soluções externas, integradas ao *DSpace*, tal como o *OpenRefine* e *OpenOffice Calc*. Thompson *et al.* (2019) sugerem o uso de *scripts* no *DSpace* como alternativa automatizada para corrigirem erros no depósito.

Contudo, a literatura alerta para o fato do alto custo para a instituição ter uma equipe total e exclusivamente dedicada ao RI, razão pela qual a grande maioria delas não possui uma equipe com esse perfil (SERRANO-VICENTE; MELERO; ABADAL, 2018).

5.1.5.1 Breves considerações acerca da análise do questionário

Os projetos de implantação e manutenção de RIS exigem planejamento estratégico, incluindo o estabelecimento de missão, objetivos, metas, responsabilidades e curadoria de dados (SALO, 2008), assim como uma equipe multidisciplinar, - com analista de sistemas, profissional de comunicação/*marketing* e administrador (LEITE, 2009) -, na qual se juntam os bibliotecários, sobretudo devido à sua formação profissional. Desse modo, a formação em Biblioteconomia deve compreender competências que tornem os profissionais capazes de atuarem mais efetivamente no sucesso dos RIs (SALO, 2008), que garantam aos bibliotecários “condições de assumirem o depósito de toda a produção da universidade em que atuam” (FREITAS, 2015, p. 121).

A literatura estudada mostrou que, apesar de os RIs terem nascido dando foco no autoarquivamento (COOKE, 2007), o depósito mediado, e, algumas vezes com autopovoamento, prevalece, e, em geral, os bibliotecários são os mais requisitados para a tarefa porque, por vezes, o RI está integrado com os serviços bibliotecários já existentes na instituição. Como um exemplo de sucesso, no Brasil há o RI da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), que possui integração entre o seu RI e a rede de gestão de suas bibliotecas.

Percebe-se o que está evidenciado na literatura, a exemplo de Sandy e Dykas (2016), pois os respondentes não relataram qualquer uso uniforme de padrões e de melhores práticas, um fator essencial para a interoperabilidade. Com isso, sem essa adesão, os RIs estão, essencialmente, deixando grande parte da tarefa de recuperação para o mecanismo de pesquisa (SANDY; DYKAS, 2016).

As análises feitas a partir do referencial teórico e da revisão de literatura, também evidenciam que é, por meio de um planejamento estratégico de implantação e manutenção do RI, envolvendo toda a comunidade institucional, que se consegue sucesso nesse tipo de iniciativa. Esse planejamento engloba recursos financeiros, materiais (estrutura tecnológica) e humanos (equipe multidisciplinar), assim como a capacitação destes últimos. Dessa maneira, o RI se tornaria eficaz e eficiente em sua vocação de preservação da memória institucional e de comunicação de sua produção científica, agregando valor à comunidade acadêmica como um todo. Infelizmente, essa ainda não é uma realidade verificável. A restrição de recursos, sobretudo financeiros e humanos, é clara novamente quando se trata de desafios administrativos, pois a literatura mostra que o orçamento limitado e a falta de pessoal são os dois maiores desafios nos RIs.

Nos RIs, os metadados têm como função a descrição e identificação de documentos, tendo em vista permitirem a efetiva recuperação de informações acerca da produção acadêmica institucional. Dada à importância dos metadados no RI, para além do esquema de metadados pré-definido no *software* adotado, (por exemplo, o *Dublin Core* no *DSpace*), os gestores de repositórios devem adicionar novos campos de acordo com a especificidade da instituição, ampliando o uso de esquemas de metadados.

A experiência empírica e as avaliações de RIs na literatura têm demonstrado que a qualidade dos metadados pode ser comprometida ao depender somente dos metadados gerados pelos pesquisadores-depositantes, pois eles, em geral, não têm formação e treinamento adequados à tarefa. Portanto, é essencial fornecer algum tipo de treinamento para eles, e garantir que os formulários de envio sejam claros e fáceis de usar, assim como disponibilizar vocabulários controlados para o preenchimento dos campos. Também é preciso que os gerentes de biblioteca considerem o uso de metadados de geração automática - por exemplo, uso de *script* para autopovoamento, desde que se adotem práticas periódicas de limpeza de metadados.

Como limitação à aplicação do questionário, é preciso destacar alguns elementos. Em primeiro lugar, o tamanho da amostra não foi grande o suficiente, ficando restrito a um número limitado de variáveis, comprometendo generalizações.

Em segundo lugar, a coleta de dados ter sido realizada no período de pandemia tornou-se um dificultador, pois não foi possível um contato telefônico prévio para a confirmação e identificação dos gestores, antes do envio do convite ao questionário. Assim, não há como

garantir que os respondentes conheçam a fundo os processos que envolvem o controle de autoridades e as ferramentas do *software DSpace*, que poderiam auxiliar nesse processo.

Em terceiro e último lugar, mesmo com a aplicação do pré-teste, ainda houve certa confusão - ou desconhecimento dos respondentes - sobre o que envolve o controle de autoridade, pois alguns participantes entenderam que esse controle estava relacionado ao vocabulário (assunto e palavras-chave). Assim, os resultados não permitiram entender, de forma detalhada, todas as questões que envolvem a qualidade dos metadados de entidades pessoas. Como o questionário se baseou na pesquisa *on-line*, não foi possível explorar mais profundamente o entendimento dos respondentes acerca da manutenção de RIs.

A despeito dos elementos referentes à limitação do questionário, já apresentados, considera-se que os resultados oferecem uma exploração inicial das questões que envolvem a qualidade dos metadados e, mais especificamente, o controle de autoridade nos RIs.

Por fim, é preciso destacar que dois dos RIs pesquisados, na época do levantamento de dados da pesquisa, utilizavam a versão 4.x do *software DSpace*. Com isso, eles não teriam acesso a todos os elementos computacionais nativos do DSpace que estão listados na seção 5.2.3.

5.2 Resultados da Segunda Etapa

Esta etapa compreendeu, essencialmente, o mapeamento e descrição dos elementos computacionais de controle de autoridade para o *software DSpace* (nativos ou não), a partir de três fases: 1) Descrição da versão mais atualizada do *software DSpace*; 2) Mapeamento dos elementos computacionais (nativos ou não) de controle de autoridade para o *software DSpace*; 3) Descrição dos elementos computacionais (nativos ou não) de controle de autoridade.

5.2.1 Descrição da especificidade da versão mais atualizada do *software DSpace*

Esta fase tem o intuito de estabelecer as características atuais do *software*, evidenciando as possíveis compatibilidades com os aplicativos não nativos.

A versão estável e mais atualizada⁹ do *DSpace* é a 6.3, entendida como uma atualização das versões anteriores. Essa versão foi lançada em 2018, com o objetivo de corrigir *bugs* das versões

⁹ Informação de julho de 2020. Em 2 agosto de 2021 foi lançada a versão 7.x.

6.x. Para a atualização da versão 6.3 no RI, os desenvolvedores afirmam que não é necessária qualquer mudança no banco de dados para quem já utiliza a versão 6x.

As principais correções de erros incluem:

- a) atualização da *DSpace ORCID Integration* para usar ORCID API v2; correção da migração do banco de dados Oracle;
- b) questões gerais de *bitstream* (comando "dspace cleanup"; erro de remoção; erro de correção);
- c) duplicação de metadados;
- d) problema do plugin de autenticação;
- e) submissão de importação da PubMed;
- f) problemas fixos com o *popup* de controle de autoridade;
- g) problema dos resultados da busca que perdem o contexto Comunidade/Coleção;
- h) aprimoramento da "importação de OAI", para relatar os itens que causam problemas de indexação;
- i) melhorias nos Relatórios de Controle de Qualidade baseados em REST (dados de campo de *bitstream*; a filtragem de *bugs* para permissões de *bitstream*);
- j) capacidade de encontrar itens retirados;
- k) problemas de estatísticas;
- l) correções específicas da base de dados (migração do banco de dados Oracle; de Workflow; atualização do driver PostgreSQL JDBC, para permitir total compatibilidade com o PostgreSQL v10);
- m) correções de nível API (afetando todas as UIs);
- n) problemas de fixação de caracteres Unicode nos metadados (DONOHUE, 2018).

5.2.2 Mapeamento dos elementos computacionais (nativos e não nativos) de controle de autoridade no *DSpace*

Esta fase se refere ao levantamento dos elementos computacionais, nativos ou não do *DSpace*, para identificação daqueles que têm funções relacionadas ao controle de autoridade.

Primeiramente, do manual do *DSpace*¹⁰ foram mapeados três elementos computacionais nativos do *software*, exclusivos para o controle de autoridades:

1. *Choice Management and Authority Control*;
2. *Authority Control of Metadata Values* (dentro de *Items and Metadata*);
3. *ORCID Integration*¹¹.

Destaca-se que no *DSpace* existem configurações de sistema que também dão apoio ao controle de autoridades, tais como *Authority Control Settings* (dentro de *Configuration Reference*) e *Browse API* (dentro de *Business Logic Layer*), sem, contudo, serem exclusivas para essa tarefa, motivo pelo qual não foram descritas.

Os documentos recuperados e descritos no capítulo de Revisão de Literatura (cap. 2) foram analisados para verificar se a literatura trata sobre a existência ou não de algum elemento computacional que auxilie no controle de autoridade em repositórios, principalmente, os não nativos do *DSpace*. Pressupunha-se encontrar um número significativo de elementos computacionais, em especial, aquelas com aplicação no *DSpace*. No entanto, esse pressuposto não se concretizou, pois não foram encontrados elementos computacionais não nativos nos documentos recuperados, compatíveis com o *DSpace*, que dão apoio especificamente ao controle de autoridade para entidade pessoa.

A análise dos documentos da amostra evidenciou que existe um entendimento na literatura que o controle de autoridade, incluindo para entidades pessoa, é o requisito essencial para uma boa qualidade dos metadados. Para manter a qualidade de metadados, muitos autores recomendam a adoção de fluxos de trabalho, nos quais são incluídos níveis diferenciados de autorização e consultas a listas pré-definidas (LUBAS, 2009), revisão de metadados pelas diferentes equipes

¹⁰ Disponível em: <https://wiki.duraspace.org/display/DSDOC6x>

¹¹ VIAF e ISNI não são nativas/integradas ao *DSpace*, mas podem ser utilizadas como fonte externa para consulta das autoridades dos Repositórios Institucionais.

(STEELE; SUMP-CRETHAR, 2016), tarefas de limpeza e aprimoramento de metadados, assim como a reconciliação e desenvolvimento de listas de autoridades (NEATROUR; MYNTTI, 2019), orientado por manuais, tutoriais em vídeo, ferramentas de suporte para depósito, FAQ e instruções sobre como inserir dados bibliográficos (SERRANO-VICENTE; MELERO; ABADAL, 2018).

Sendo assim, foram mapeados os elementos computacionais que têm como objetivo melhorar a qualidade dos metadados, por meio da correção em lote, a saber: 1) *OpenRefine*; 2) *OpenOffice Calc*; 3) *Microsoft Access*; 4) *Microsoft Excel*; e 5) *Duplicate Detection Service*, descritos na subseção 5.2.3.4.

5.2.3 Descrição dos elementos computacionais (nativos e não nativos) de controle de autoridade para a última versão do *software DSpace*

Nesta fase, descrevem-se os elementos computacionais de controle de autoridade para a versão atualizada do *software DSpace*, a partir do Manual *DSpace* e dos documentos recuperados na seção 5.2.2, que fazem parte do Capítulo 2 da Revisão de Literatura. Assim, buscou-se localizar aqueles elementos que são próprios do *DSpace*, que estão embutidos nele, ou que permitem integração com o *software*, por compatibilidade, ainda que fossem soluções proprietárias.

Destaca-se que foi necessário recorrer aos manuais das versões anteriores para entender, de maneira mais detalhada, cada um dos três elementos computacionais nativos do *DSpace*, que apoiam o controle de autoridade, sendo que o conteúdo foi retirado do Manual da versão *DSpace 6.x (Current Release)*. Eles são descritos destacando-se suas características para uso nos RIs.

Como a documentação do *software* é, essencialmente, em inglês, houve a tradução de partes dessa documentação. Ressalta-se que não houve a pretensão de detalhar demasiadamente as configurações descritas na documentação técnica, criando um manual para sua implementação. A finalidade essencial dessa descrição é trazer insumos suficientes para que os elementos computacionais se tornem conhecidos, sobretudo, pelos gestores de RI.

Todos os elementos computacionais estão disponíveis na última versão disponível para *download* na página oficial do *software*.

5.2.3.1 Choice Management and Authority Control (nativo)

O *Choice Management and Authority Control* (Gerenciamento de Escolhas e Controle de Autoridade) é um elemento computacional nativo do *DSpace*, sendo um mecanismo geral para o controle de autoridade para todo e qualquer campo de metadados, incluindo qualquer campo do *Dublin Core*. A sua documentação ressalta o uso do mecanismo principalmente para o controle dos nomes de autor e de outros nomes pessoais.

Inicialmente, é preciso clarificar a terminologia, destacando algumas definições:

Autoridade: uma fonte externa de valores fixos para um determinado domínio, sendo cada valor exclusivo e identificado por uma chave. Por exemplo, o serviço de autoridade de nomes da LC, ISNI, ORCID ou VIAF.

Registro de autoridade: informações associadas a um dos valores em uma autoridade; pode incluir grafias alternativas e formas equivalentes do valor.

Chave da autoridade: um identificador persistente, correspondente a exatamente um registro na autoridade (THE DSPACE DEVELOPER TEAM, 2018, p. 27).

A Figura 6 mostra um exemplo de chave de autoridade.

Figura 6 – Exemplo de chave de autoridade



Fonte: POVEDA *et al.*, 2018, p. 12.

A seção *Metadata Management*, ou Gestão de Metadados, contempla a função ‘Gerenciamento de escolhas e controle de autoridade’ que está disponível desde a versão 1.6 do *DSpace*. Essa função “é uma estrutura configurável que lhe permite definir classes *plug-in* para controlar a

escolha de valores para campos de metadados especificados no *DSpace*” (THE DSPACE DEVELOPER TEAM, 2018, p. 26).

Isso permite que a cada campo de metadados seja atribuída uma lista de opções para gerenciar a escolha dos dados nesses campos. Dependendo do objetivo, a lista de opções exata pode ser determinada por um valor ou pode ser igual para todas as consultas. Além disso, pode ser fechada, limitada a uma lista pré-estabelecida internamente (por exemplo, tipo de documento) ou aberta, permitindo que a lista seja editada pelo usuário, incluindo uma nova opção na lista (THE DSPACE DEVELOPER TEAM, 2018).

O controle de autoridade, nessa função específica, complementa o gerenciamento de escolhas pré-definidas, fornecendo uma chave de autoridade (hoje limitada a 100 caracteres) junto ao valor escolhido. Já que, para cada opção da lista, são atribuídos valores únicos, as chaves de autoridade são também identificadores únicos (FIGURA 7), que são atribuídos no campo de metadados no registro do item no RI (THE DSPACE DEVELOPER TEAM, 2018).

Figura 7 – Autoridade controlada no *DSpace*

| Remove | Name | Value |
|--------------------------|-----------------------|--|
| <input type="checkbox"/> | dc.contributor.author | Smith, Dona |
| | | no 97030482  <input type="button" value="Lookup"/> |

Fonte: THE DSPACE DEVELOPER TEAM, 2018, p. 303.

O tipo de controle exposto na Figura 7 tem quatro vantagens:

- Forma simples e positiva de testar dois valores idênticos com as chaves de autoridade: a comparação de textos simples pode resultar em falsos positivos, por exemplo, dois autores diferentes com o nome escrito da mesma forma ou então quando o mesmo nome é escrito de maneiras diferentes.
- Ajuda para inserir valores de metadados corretos: ao submeter um registro é possível selecionar o registro a partir de uma lista com possíveis correspondências.
- Melhora da interoperabilidade: quando o repositório que utiliza *DSpace* está integrado a outros sistemas a interação entre os dois sistemas é mais eficiente.

Por exemplo, um repositório institucional *DSpace* partilhando uma autoridade de nomes com a rede social ou sistema acadêmico da universidade, permitiria construir uma lista de todos os itens *DSpace* que correspondam ao identificador de autor compartilhado entre eles, ao invés de uma correspondência de nomes, susceptível de erro.

Quando a autoridade de nomes é compartilhada com os sistemas acadêmicos, o *DSpace* pode procurar o endereço de *e-mail* de um autor para enviar uma mensagem automática sobre os seus trabalhos, que foram enviados por terceiros.

- d) As chaves de autoridade são normalmente invisíveis para o usuário, já que não fazem sentido na busca (THE DSPACE DEVELOPER TEAM, 2018).

Destaca-se que a característica de *plugin* dessa função, permite que novas autoridades sejam incluídas ou atualizadas sem impacto em outros códigos do *software*.

5.2.3.2 Authority Control of Metadata Values (nativo)

O *Authority Control of Metadata Values* (Controle de Autoridade de Valores de Metadados) é um elemento computacional nativo do *DSpace*, sendo que o valor dos metadados se refere ao conteúdo ou termos inseridos em um campo de metadados. Por exemplo, se o campo for o de uma data, a *string* “2002-11-15” seria um valor. Destaca-se que, em geral, a atribuição de valor a campos de metadados está sujeita à ambiguidade da linguagem natural, à diversidade de convenções culturais e à subjetividade das preferências pessoais. Assim, o controle de autoridade, com valores unívocos, e diretrizes de conteúdo são necessários para garantir a consistência dos pontos de acesso, dada a variabilidade dos valores que podem ser inseridos nos diversos campos de registro. Um esquema de metadados adequado deve prever um modo de controle para qualificar as informações contidas em um dado campo (THE DSPACE DEVELOPER TEAM, 2018).

Ao incluir documentos no *DSpace*, campos de metadados são preenchidos como pontos de acesso para posterior recuperação. Cada campo em *Dublin Core* tem uma designação específica: autor, título, orientador, local, editor, periódico, resumo entre outros campos. (THE DSPACE DEVELOPER TEAM, 2018).

Alguns desses campos precisam ser de preenchimento livre, tal como o resumo. Os demais, no entanto, podem ter certo nível de controle de autoridade, como, por exemplo, os campos de autor e de assunto. E, ao exercer esse controle, conseqüentemente, tem-se um cuidado com a qualidade dos metadados, que resultará numa recuperação mais precisa das informações ali contidas, pois cada registro de autoridade será único e inequívoco (THE DSPACE DEVELOPER TEAM, 2018).

Dai a importância da utilização de recursos de listas, pré-determinados, para o controle de campos específicos. A Figura 8 apresenta alguns desses recursos.

Figura 8 – Lista simples de tipos de documento

CAMPO OBRIGATÓRIO: Selecione o tipo de documento.

Tipo de documento: *

CAMPO OBRIGATÓRIO: Informe o título próprio e siglas.

Título: *

CAMPO OPCIONAL: Recomenda-se que a primeira letra do título deve ser grafada em maiúscula.

Título(s) alternativo(s):

- Selecionar
- Selecionar
- Artigo de Periódico
- Artigo de Evento
- Trabalho de Evento
- Tese
- Dissertação
- Trabalho de Conclusão de Curso (Graduação)
- Monografia (especialização)
- Memorial
- Livro
- Capítulo de Livro
- Relatório de Pesquisa
- Objeto de Aprendizagem
- Projeto de Pesquisa
- Entrevista
- Preprint
- Outro tipo de documento

Fonte: Repositório UFMG, formulário de autoarquivamento, 2019.

Observa-se, na Figura 8, um exemplo de lista fechada para o campo *dc-type*. Ao submeter o documento, o responsável seleciona entre as opções qual delas corresponde ao tipo de documento submetido.

Como já mencionado, as listas de opções ou sistema *value-pairs* podem ser aplicadas a qualquer campo do *DSpace*, no entanto, é recomendada apenas para campos com listas simples e curtas (idioma, tipos de documento, país de publicação, faculdade ou departamento, área). Para o controle de autoridade de autor e assunto, esse recurso de listas não é indicado pela complexidade e volume de dados (THE DSPACE DEVELOPER TEAM, 2018).

A opção para gerenciar grande volume de dados no *DSpace* é via arquivos XML, que ficam salvos em um diretório separado. A vantagem é que os dados podem ser trabalhados externamente ao *software*. Por exemplo, pode-se utilizar o Protégé para criar uma taxonomia,

salvá-la em *Ontology Web Language* (OWL), e, depois, transformá-la no formato de arquivo XML, adequado para ser importado para o DSpace. Atualmente, existem dois vocabulários controlados já embutidos no sistema: *The Norwegian Science Index* e *Swedish Research Subject Categories* (THE DSPACE DEVELOPER TEAM, 2018).

Até este ponto, foi visto o controle de autoridade aplicado a representações de texto em campos específicos. Contudo, o “*DSpace* também oferece suporte para adicionar chaves de autoridade e valores de confiança a um valor de texto específico introduzido num campo de metadados” (THE DSPACE DEVELOPER TEAM, 2018, p. 302). Isso significa que é possível integrar o *DSpace* com uma fonte externa de autoridade, por exemplo, o catálogo de autoridades da *Library of Congress*, a lista de autoridades VIAF ou com o ORCID.

5.2.3.3 ORCID Integration (nativo)

O ORCID (*Open Researcher and Contributor ID*) Integration (Integração com o ORCID) é um elemento computacional nativo do *DSpace*, que se liga ao banco de dados do serviço ORCID, que determina um identificador digital persistente a um pesquisador específico, ligando seu nome a uma chave de autoridade, permitindo vincular todas as suas publicações à sua identificação. Ao ser integrado ao *DSpace*, quando tal integração é estabelecida com o ORCID, o campo-chave de autoridade é impulsionado a vincular o nome do autor a seu ORCID ID (THE DSPACE DEVELOPER TEAM, 2018).

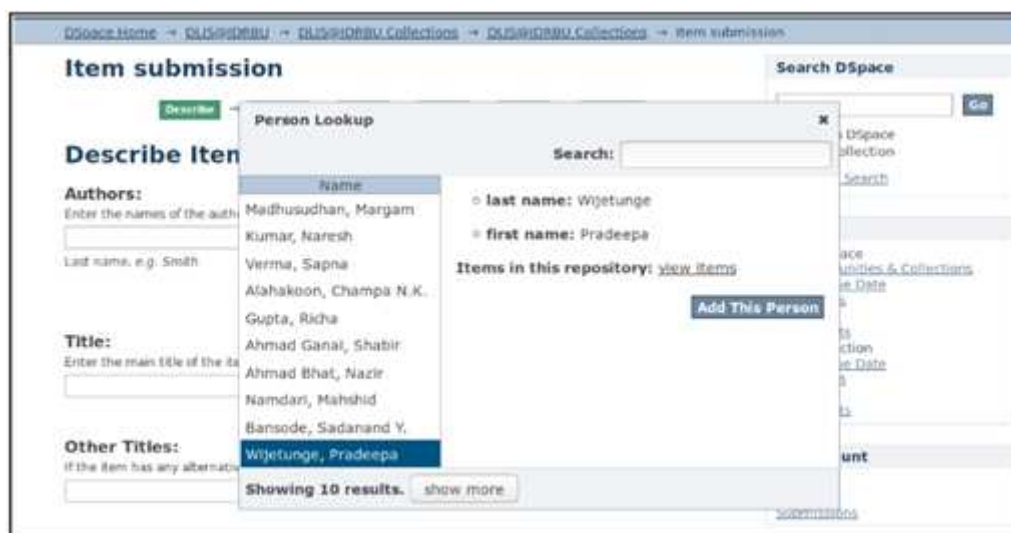
A integração do ORCID com o *DSpace* está presente desde a versão 5.x do sistema, e, a partir dessa incorporação, não mais é necessário modificar os códigos fonte para ter acesso a essa funcionalidade. Com a versão 5, os desenvolvedores podem habilitar a função de integração, que permite que as informações já existentes no sistema sejam atualizadas, possibilitando a identificação de autores de forma inequívoca, a partir do ORCID ID (THE DSPACE DEVELOPER TEAM, 2018).

A documentação destaca que a integração com o ORCID exige que se construa uma interface entre o sistema ORCID e o *DSpace*. Contudo, há iniciativas que mostram o uso do ORCID em repositórios sem ter uma integração direta do ORCID API com o próprio repositório, a exemplo

do que fez a *Duke University*, no repositório *DukeSpace*, que criou um código *DSpace* específico, que pode ser compartilhado mediante solicitação.¹²

A Figura 9 mostra um exemplo de busca por autores, usando chaves de autoridade do ORCID.

Figura 9 – Exemplo de busca no *DSpace* integrado com o ORCID



Fonte: MANDAL, 2020, p. 8.

A Figura 9 mostra a busca no *DSpace* depois da configuração do ORCID (sob o arquivo *dspace.cfg*), cuja interface permite visualizar os autores e os registros no RI, para que se escolha os autores para um item em particular (THE DSPACE DEVELOPER TEAM, 2018).

Os dados de autoridade do ORCID precisam ser atualizados via execução de *script* e não podem ser editados manualmente. Destaca-se que, com a execução de *scripts*, é possível resolver:

- Valor de metadados sem chave de autoridade, quando uma autoridade existe na base, mas não tem uma chave de autoridade relacionada.
- Registro de autoridade com uma chave de fonte externa ao *DSpace*, a chave de autoridade antiga é armazenada e uma nova chave controlada é atribuída.

Processar itens e autoridades, quando uma autoridade está vinculada a um item e o item foi excluído. É possível executar o *script* que eliminará a autoridade sem vínculo de item (THE DSPACE DEVELOPER TEAM, 2018).

¹² Exemplo disponível em: <https://dukespace.lib.duke.edu/dspace/handle/10161/17235>

A Figura 10 exemplifica a inclusão de um novo autor e como a integração com o ORCID acontece. Ao clicar sobre o botão ‘lookup’, o sistema faz uma busca por autoridades cadastradas para que o usuário selecione a autoridade correspondente.

Figura 10 – Exemplo de interface do *DSpace* integrada com ORCID

The image shows a screenshot of the DSpace interface. At the top, there is a navigation bar with buttons: Describe, Describe, Access, Upload, Review, CC License, License, Complete. Below this is the 'Describe Item' section. Under 'Authors:', there are two input fields: 'Last name, e.g. Smith' and 'First name(s) + "Jr", e.g. Donald Jr'. To the right of these fields are 'Add' and 'Lookup' buttons. Below the input fields, it says 'Enter the names of the authors of this item below.' A modal window titled 'Person lookup' is open, showing a search bar with 'Hancock' entered. Below the search bar is a list of names: Hancock, Roeland; Hancock, Yvette; Hancock, Katy; Kleiner-Hancock, Heather; Hancock, Mark; Hancock, Steven; Hancock, Karen; Hancock, Kenneth; Hancock, Travis; HANCOCK, GALEN. To the right of the list, there are details for the selected name: last name: Hancock; first name: Roeland; orcid: 0000-0001-8932-6872. Below these details is a link 'Items in this repository: view items' and an 'Add This Person' button. At the bottom of the modal, it says 'Showing 10 results.' and a 'show more' button.

Fonte: THE DSPACE DEVELOPER TEAM, 2018, p. 341.

Observa-se, na Figura 10, que, ao selecionar a autoridade, o sistema fornece outras informações do registro, como primeiro e último nome, e ORCID. Essas informações ajudam a descrever a autoridade de forma a evitar equívocos de identificação.

Destaca-se que a integração ORCID vai muito além da atribuição de identificador a um nome, pois ela possibilita uma carga de metadados importados da base original do ORCID, como nomes alternativos, endereços de *e-mail*, vínculo institucional, palavras-chave e produções. A vantagem desse recurso é que os responsáveis pela atividade, seja bibliotecário ou gestor, não precisam recorrer a uma plataforma externa para checarem os dados, pois o *DSpace* vai realizar

essa tarefa, diretamente, consultando os dados em um *metadata authority cache* (THE DSPACE DEVELOPER TEAM, 2018).

A Figura 11 destaca outros campos que podem ser importados via ORCID para que as autoridades para entidades pessoa sejam identificadas, inequivocamente. No entanto, por configuração padrão, o *DSpace* exibe apenas o primeiro e último nomes, além do código ORCID ID. Para a inclusão dos outros campos, algum desenvolvimento em JAVA precisa ser feito (THE DSPACE DEVELOPER TEAM, 2018).

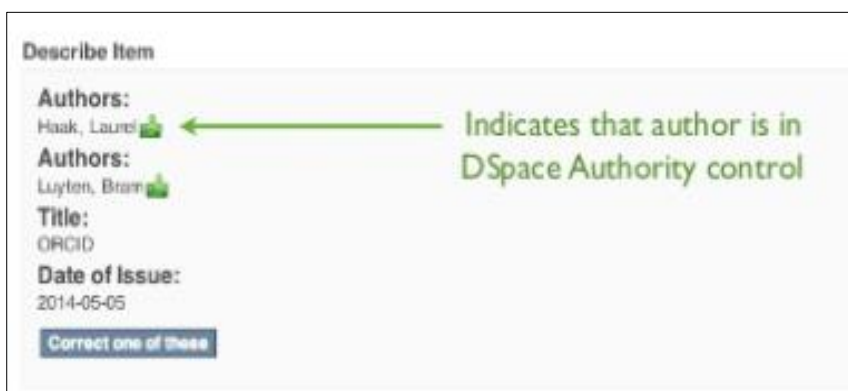
Figura 11 – Campos do ORCID para identificação das autoridades

| Schema metadata fields | | | |
|------------------------|---|---|--|
| ID | Field | Scope Note | |
| 74 | orcid.person.email | An email address for a person | |
| 76 | orcid.person.firstname | | |
| 75 | orcid.person.institution | An institution a person is associated with | |
| 77 | orcid.person.lastname | | |
| 78 | orcid.person.name_variant | Syntax: "last name, first name" | |
| 73 | orcid.person.orcidid | The orcid_id for a person | |
| 80 | orcid.record.creation-date | Date the record was created in solr | |
| 81 | orcid.record.last-modified-date | Date the record was last updated with new information | |
| 79 | orcid.record.uid | The uid of the solr record | |

Fonte: LUYTEN, 2014, *on-line*.

O *DSpace* identifica quando as autoridades são controladas internamente com uma ID ou via ORCID. A Figura 12 destaca o processo de revisão da submissão e a aprovação para a autoridade cadastrada.

Figura 12 – Autoridade controlada pelo *DSpace*



Fonte: LUYTEN, 2014, *on-line*.

A vantagem de se utilizar o ORCID como um recurso para o controle de autoridade para entidades pessoas está na possibilidade de descrever uma autoridade com base em diferentes informações, além do nome. No Quadro 9, tem-se uma comparação entre os campos *DSpace* nativos e os campos que são importados com a integração ORCID.

Quadro 9 – Campos de metadados *DSpace* versus ORCID

| DSpace | ORCID |
|--------------------|---------------------------|
| id | first_name |
| field | last_name |
| value | name_variant |
| deleted | orcid_id |
| creation_date | label_researcher_url |
| last_modified_date | label_keyword |
| authority_type | label_external_identifier |
| | label_biography |
| | label_country |

Fonte: Dados da pesquisa, 2021.

Nota-se, no Quadro 9, o quão vantajosa pode ser a integração do *DSpace* com o ORCID, se for considerado que cada um dos campos contribuirá para a identificação inequívoca das autoridades. No entanto, os dados na origem são preenchidos pelos próprios autores, que podem preencher os dados incorretamente ou deixar incompletos. Sendo assim, a integração, apesar de vantajosa, precisa contar com a colaboração dos autores para que funcione e seja eficiente.

5.2.3.4 Elementos computacionais não nativos do *DSpace*

Nesta subseção, estão apresentados os cinco elementos computacionais não nativos do *DSpace*, utilizados na correção de metadados em lote, que são opções que podem auxiliar os gestores dos RIs a melhorarem a qualidade dos metadados herdados ou já inseridos no sistema, antes da implantação de políticas que visem diminuir as inconsistências na descrição dos registros, a saber: 1) *OpenRefine*; 2) *OpenOffice Calc*; 3) *Microsoft Access*; 4) *Microsoft Excel*; e 5) *Duplicate Detection Service*, descritos a seguir.

5.2.3.4.1 *OpenRefine* (não nativo)

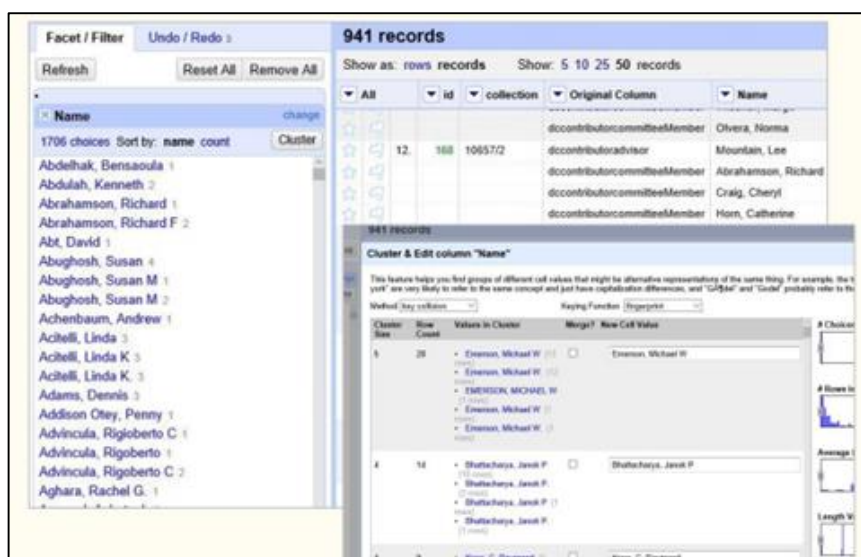
O *OpenRefine* é um elemento computacional não nativo, mas compatível com o *DSpace*, e originalmente concebido pela empresa *Metaweb Technologies, Inc* que em 2010 foi comprada pela *Google*, que renomeou a ferramenta para *Google Refine*. No entanto, desde 2012, foi

batizado de *OpenRefine*, produto gratuito, de código aberto com suporte da comunidade de usuários e disponível em mais de 15 idiomas (OPENREFINE, 2021).

Ferramenta do tipo Extração, Transformação e Carregamento (ETL) que permite: explorar, limpar e transformar, reconciliar e combinar grande volume de dados. Com uma interface intuitiva e semelhante aos editores de planilha que dividem os dados em linhas e colunas. Comporta diversos tipos de arquivo, por exemplo: TSV, CSV, +SV, *Excel* (xls y xlsx), JSON, XML, RDF como XML além de ser compatível com planilhas *Google*. (IBICT, 2016; OPENREFINE, 2021). Pode ser utilizado com os sistemas operacionais disponíveis como *Windows, Mac e Linux*.

A Figura 13 mostra um exemplo de uso do *OpenRefine* no *DSpace*.

Figura 13 – Exemplo de uso do *OpenRefine*



Fonte: LIU; DURAN; WASHINGTON, 2016, *on-line*.

Conforme explicam Liu, Duran e Washington (2016), no projeto que desenvolveram, o *OpenRefine* foi adotado para padronizar os nomes dos orientadores e dos membros dos comitês de avaliação de teses e dissertações. Segundo os autores, “todos os nomes foram consolidados em uma única coluna, usando a função de transposição” (2016, *on-line*), para, em seguida, “serem usadas as funções de faceta e agrupamento para escolher e povoar o nome preferido em todo o conjunto de dados” (2016, *on-line*).

5.2.3.4.2 *OpenOffice Calc* (não nativo)

O *OpenOffice Calc* é um elemento computacional não nativo, mas compatível com o *DSpace*, sendo uma solução de pacote com vários aplicativos, semelhante ao Pacote *Office* da *Microsoft*, mas de código aberto. Compatível com os principais sistemas operacionais *Microsoft Windows*, *Unix*, *Solaris*, *Linux* e *Mac OS X* disponíveis atualmente. A plataforma possui os seguintes componentes *Open Office: Writer; Impress; Math; Draw; Base* e *Calc*. Esse último se assemelha ao *Excel* nas suas funções básicas e foi mencionado no Artigo 17 (da revisão de literatura) para limpar dados do título, gerando, por exemplo, padronização entre maiúsculas e minúsculas em banco de dados de RIs (UNIVERSIDADE FEDERAL DE UBERLÂNDIA, [202-]; OPENOFFICE, 2009).

O *DSpace* emprega um filtro, que é uma funcionalidade denominada “*out-of-the-box*”, do *software OpenOffice.org*, para automatizar a conversão em lote (DONOHUE, 2006). Segundo o autor, adotando interfaces de programação de aplicações apropriadas, é possível que o filtro realize conversão em lote para qualquer formato *OpenOffice.org*, tanto para importar quanto para exportar dados. Por exemplo, pode-se ter, segundo Donohue (2006):

(1) Conversões de documentos textuais

Entradas: *Word*, *RTF*, *OpenOffice Writer*

Saídas: *Texto simples*, *Word*, *RTF*, *DocBook XML*, *OpenOffice Writer*, *PDF*, *HTML*

(2) Conversão de planilhas

Entradas: *CSV*, *Excel*, *OpenOffice Calc*

Saídas: *CSV*, *Excel*, *OpenOffice Calc*, *HTML*, *PDF*

(3) Conversões de Apresentação

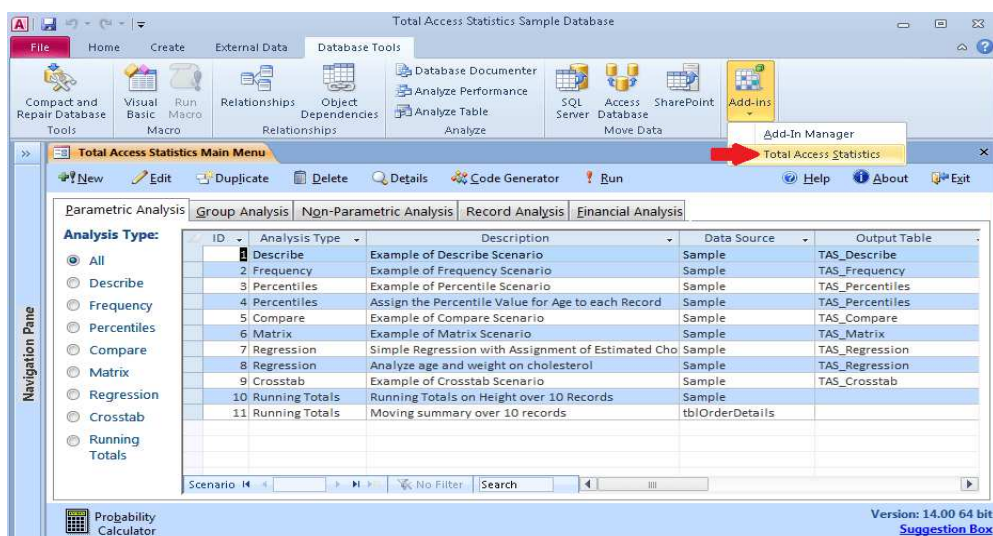
Entradas: *PowerPoint*, *OpenOffice Impress*

Saídas: *PowerPoint*, *OpenOffice Impress*, *HTML*, *PDF*

5.2.3.4.3 Microsoft Access (não nativo)

O *Microsoft Access* é um elemento computacional não nativo, mas compatível com o *DSpace*, e ele compõe o pacote *Office* da *Microsoft* é uma aplicação mais robusta que as ferramentas *OpenRefine* e *OpenOffice Calc*, anteriores, pois permite não só gerir um grande volume de dados como criar uma interface de busca para o usuário (MICROSOFT, 2021).

Figura 14 – Exemplo de banco de dados do *Access*



Fonte: ÁVILA, 2015, *on-line*.

O *Access* vai além de aplicações de edição de planilhas como o *Calc* e o *Excel*, pois é um banco de dados relacional e permite a execução de funções mais complexas, com a consolidação de colunas e células. Ele foi citado no Artigo 20 (da revisão de literatura), sendo utilizado para tratar os dados que foram exportados do *DSpace* para o *Excel* (MICROSOFT, 2021).

A Figura 15 mostra um exemplo de uso do *Microsoft Access* no *DSpace*.

Figura 15 – Exemplo de uso do *Microsoft Access*

The screenshot shows a Microsoft Access database window titled 'Remediation'. The main window displays a table with the following columns: 'id', 'collection', 'dccontributoradvisor', 'dccontributoradvisor1', and 'dccontributoradvisor2'. The data rows show various IDs (e.g., 10657/2) and names (e.g., Phillips, Scott; Lee, Rebecca E.; Arbona, Consuelo). In the foreground, a smaller window titled 'Query2' is open, showing a list of field names including 'id', 'collection', 'dccontributoradvisor', 'dccontributoradvisor1', 'dccontributoradvisor2', 'dccontributorcommitteeMember', 'dcreator', 'dcreator1', 'ddatecreated', 'ddatecreated1', 'ddateissued', and 'ddateissued1'.

Fonte: LIU; DURAN; WASHINGTON, 2016, *on-line*.

De acordo com Liu, Duran e Washington (2016), a exportação de metadados para o *DSpace* pode produzir campos repetidos que precisavam ser mesclados, e, assim, o *Microsoft Access* pode ser utilizado para fazer uma consulta, tendo em vista consolidar os valores das colunas duplicadas, tornando-a uma única coluna. Para os autores, essa consolidação é necessária para completar tarefas adicionais de limpeza (maiúscula/minúscula, por exemplo), e importar os novos metadados para o *DSpace* com consistência.

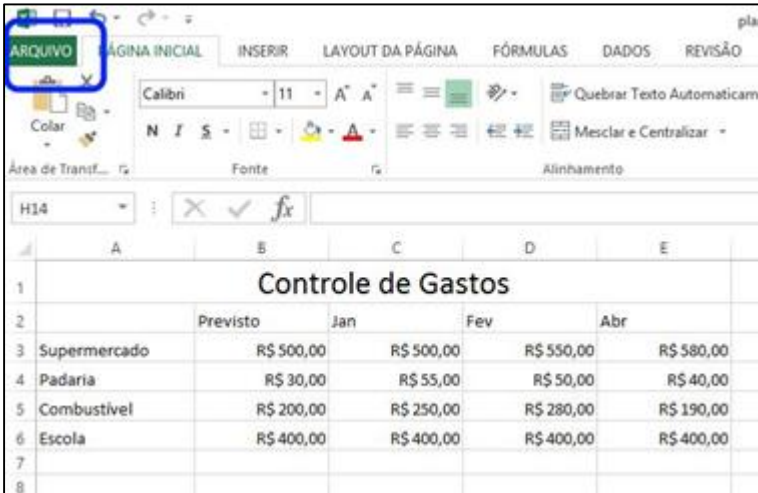
5.2.3.4.4 *Microsoft Excel* (não nativo)

O *Microsoft Excel* é um elemento computacional não nativo, mas compatível com o *DSpace*, sendo uma das aplicações do *Office* da *Microsoft*, que foi desenvolvido para criação e edição de planilhas eletrônicas, sendo organizado em linhas e colunas. Diferente do *Access*, o *Excel* pode ser utilizado como banco de dados, mas para arquivos simples (MICROSOFT, 2021).

Uma vez que a planilha do *Excel* é semelhante a uma tabela de banco de dados, pode-se usá-la tanto para exportar dados quanto para importar dados do *DSpace*. Destaca-se, entretanto, que no *Excel* não há, por padrão, relacionamento entre distintas planilhas, característica que é comum em outros aplicativos de banco de dados (*Oracle*, *SQL Server*, *Mysql*, entre outros) (MICROSOFT, 2021).

Para exportar dados do *Excel*, é preciso digitar os dados em colunas (FIGURA 16), e, depois, salvar o arquivo como CVS (*comma separated values*; separado por vírgulas).

Figura 16 – Exemplo de organização de dados no *Excel* linhas e colunas



| | Previsto | Jan | Fev | Abr |
|--------------|------------|------------|------------|------------|
| Supermercado | R\$ 500,00 | R\$ 500,00 | R\$ 550,00 | R\$ 580,00 |
| Padaria | R\$ 30,00 | R\$ 55,00 | R\$ 50,00 | R\$ 40,00 |
| Combustível | R\$ 200,00 | R\$ 250,00 | R\$ 280,00 | R\$ 190,00 |
| Escola | R\$ 400,00 | R\$ 400,00 | R\$ 400,00 | R\$ 400,00 |

Fonte: BANCO..., 2020, *on-line*.

Como pode ser observado na Figura 16, para criar um banco de dados no *Excel* é preciso que os dados estejam organizados em uma estrutura em linhas (entradas de dados) e colunas (campos armazenados). É possível, ainda, automatizar o lançamento de novos dados, caso o fluxo de informações seja grande, pois o *Excel* possui uma ferramenta específica para isso, que está na funcionalidade Formulário.

Segundo Bollini (2018, *on-line*), a “configuração do modelo de dados *DSpace*-CRIS pode ser exportada e importada usando um arquivo *Excel*”, sendo que o “*script* de importação funcionará como anexo, para que seja possível utilizá-lo para preencher o modelo inicial de dados em uma instalação *DSpace*-CRIS vazia, mas também para adicionar novos atributos ou entidades”, mas ele “não é capaz de atualizar a definição de atributos existentes”.

A autora destaca que a estrutura do arquivo *Excel* é composta de várias folhas: a) *propertiesdefinition* (definição de propriedades): tem a lista das propriedades definidas para todas as entidades, para cada uma das colunas da planilha; b) *nesteddefinition* (definição

aninhada): tem a lista de propriedades aninhadas para todas as entidades, constituída por informações estruturadas, tais como de endereço (rua, cidade, CEP, etc.) ou de afiliação (data de início/fim, função, instituição etc.); c) *tab* (aba): tem a lista de seções para cada entidade visível em páginas públicas; d) *etab*: tem a lista de seções para cada entidade, para ser usada no modo de edição; e) *box* (caixa): tem o painel usado para agregar propriedades no modo de visualização e edição dentro de uma *tab* ou *etab*; f) *tab2box*: mapeia um *box* para uma ou mais visualizações; g) *etab2box*: mapeia um *box* para editar um ou mais *etab*; h) *box2metadata*: mapeia uma propriedade para um *box*; i) *utilsdata*: tem dados que validam o conteúdo das outras folhas; j) *controlledlist* (lista controlada): tem a lista dos tipos de povoamentos permitidos, caixas de seleção e *widget*; k) *tabpolicy* (política de aba): define quais propriedades podem ser usadas, verificando a condição de acesso da *tab* com NÍVEL DE ACESSO = POLICY_ACCESS; l) *etabpolicy* (política de *etab*): igual à letra k, porém, relativo à tabulação usada no modo de edição; m) *boxpolicy* (política de caixa): igual à letra k, porém, relativo à segurança da caixa (BOLLINI, 2018).

Quando esses arquivos em *Excel* precisam evoluir, o *Access* é uma opção. No Artigo 20 da revisão de literatura, os autores mencionam o uso do *Excel* para extrair os dados do *DSpace*, que, depois, são tratados no *Access* e no *OpenRefine*.

5.2.3.4.5 *Duplicate Detection Service* (não nativo)

O *Duplicate Detection Service* (Serviço de Detecção de Duplicatas) é um elemento computacional não nativo, mas compatível com o *DSpace*, aplicativo que foi desenvolvido pela *The Library Code*, em conjunto com a Universidade de Ciências Aplicadas de Zurique (*Zürcher Hochschule für Angewandte Wissenschaften*), e estende o *DSpace* 6.2 JSPUI através de uma detecção de duplicação (MORRIS, 2019). Segundo o autor, o aplicativo provavelmente não funcionará no *DSpace* 7 e posteriores sem algum desenvolvimento adicional. No *DSpace*, a JSPUI é implementada usando *Java Servlets* (lidam com a lógica comercial) e *JavaServer Pages* (JSPs) que produzem as páginas HTML enviadas a um usuário final. Para utilizar esse aplicativo, antes será preciso instalar, manualmente, a extensão PostgreSQL “fuzzystrmatch”, seguindo as orientações em sua documentação¹³.

¹³ Disponível em: <https://github.com/the-library-code/deduplication/tree/dspace-6.2-addition>

Ao finalizar o envio, um depositante é notificado se o item tem algum similar no RI. Quando há similaridade, as possíveis duplicatas são apresentadas ao depositante, que pode decidir remover sua submissão, deixar para decidir mais tarde ou continuar com a submissão. Em geral, esse aplicativo é parte de um fluxo de trabalho, assim, se houve similaridade entre itens e, mesmo assim, o item foi publicado, o revisor receberá uma notificação sobre a existência dos itens similares, podendo decidir se aceita ou recusa uma submissão ao repositório (MORRIS, 2019).

Morris (2019) afirma que esse aplicativo pode ser adotado para apoiar os gestores na localização de conteúdos duplicados e similares. A autora destaca que, ao submeter um novo item no RI, será preciso incluir o título primeiro, e o *plugin*, que usa um algoritmo Fuzzy, buscará por títulos iguais ou similares que já estão armazenados dentro do repositório e exibi-los como uma lista. Isso evita problemas causados por erros de digitação ou até mesmo diferenças na atribuição de autoria, por exemplo.

6 CONSIDERAÇÕES FINAIS

Esta pesquisa teve como motivação tentar entender os problemas que envolviam os erros recorrentes de registro de metadados, que afetam a sua qualidade, principalmente aqueles relacionados ao controle para entidade pessoa, nos Repositórios Institucionais nacionais, que utilizam o *software DSpace*, e, mais especificamente, nos RIs mineiros. Durante o percurso da pesquisa, ficou claro que essa é uma dificuldade enfrentada em RIs em todo o mundo, que é uma herança de transição, feita sem diagnóstico e planejamento, das bibliotecas físicas para as digitais que só agravou o problema. Isso evidenciou o problema da ambiguidade que envolve a determinação de dados para entidade pessoa, que se revela na forma de registros com erros de digitação, entradas duplicadas, abreviações de nomes, entre outros, que traz prejuízos para a instituição e ao próprio autor, impedindo que os RIs alcancem com plenitude a sua vocação para a comunicação científica. A partir desse problema, este estudo teve como objetivo geral: “investigar como ocorre o controle de autoridade para entidades pessoas em repositórios institucionais que usam o *DSpace*, evidenciando os elementos computacionais nativos ou compatíveis (não nativos) com o *software*, contribuindo com as discussões acerca da desambiguação de nomes próprios”.

Em primeiro lugar, foi necessário entender o processo de desambiguação de nomes próprios, e, assim, responder ao primeiro objetivo específico: “Explorar o fenômeno da ambiguidade no âmbito da determinação de nomes próprios, tendo em vista compreender as questões para o controle de autoridade para entidades pessoas”; resposta que se encontra na seção 3.1, do Capítulo Fundamentos Teórico-Conceituais. A exploração desse objetivo demonstrou ser esse um processo complexo, que compreende variantes de nomes e erros ortográficos, mudanças de nome, pseudônimos, homônimos, abreviações do nome e informações complementares que, se não mapeadas ou controladas, impactam, direta e negativamente, o controle de autoridade. E a ausência de controle impede que se identifique um autor de forma inequívoca, e, conseqüentemente, torna inviável ou dificulta o controle da sua produção acadêmica.

Em seguida, foi investigado o controle de autoridade para entidade pessoa ao longo do tempo, para, assim, responder ao segundo objetivo específico: “Analisar e descrever as questões que envolvem o controle de autoridade, com foco no controle para entidades pessoas, trazendo elementos históricos para o entendimento da atividade nos RIs”, resultado que se encontra nas

seções 3.2 e 3.3, do Capítulo Fundamentos Teórico-Conceituais. O controle de autoridade é uma das bases da catalogação ao estabelecer pontos de acesso padronizados, normalizados e controlados para evitar a ambiguidade, e, assim, permitir a consulta pelo usuário, por meio do produto do processo de catalogação, o catálogo. A literatura mostra que, ao longo do tempo, foram várias as iniciativas para facilitar o controle de autoridade, ainda nas bibliotecas com catálogos em fichas. A cada nova tecnologia, a biblioteca foi se adaptando para oferecer à comunidade melhores serviços e produtos, seja com a automação das bibliotecas ou a criação de redes cooperativas de registros bibliográficos e catálogos *on-line*. Assim, o controle de autoridade hoje extrapola as barreiras das bibliotecas físicas e dos bibliotecários, que, mais do que nunca, exige um trabalho cooperativo e coordenado, o que envolve bibliotecários, autores (pesquisadores, alunos, docentes), usuários e as instituições, sejam as universidades, bibliotecas nacionais ou organizações independentes.

Acredita-se que as bibliotecas digitais, e entre elas os RIs, foram uma evolução natural das bibliotecas físicas, onde havia interconexões já estabelecidas e mais estáveis, para um ambiente bem mais dinâmico. Com isso, há implicações relevantes a serem consideradas, a exemplo do que ocorre com o controle de autoridade para entidade pessoa, sendo necessário respeitar questões relacionadas com descrições e fontes heterogêneas de autoridades. Considera-se que já existe tecnologia robusta o suficiente para lidar com metadados heterogêneos, sendo preciso um esforço conjunto de adaptação de todos os atores envolvidos, em especial, dos bibliotecários, e, como consequência, das bibliotecas.

Posteriormente, foi preciso entender como o controle de autoridade estava sendo efetivamente aplicado nas instituições para, assim, responder ao terceiro objetivo: “Entender como o controle de autoridade tem sido realizado em RIs, sobretudo aqueles que usam o *DSpace*, a partir de uma revisão da literatura sobre o tema”, resultados que estão no Capítulo 2. Para tanto, foi feita uma revisão da literatura para buscar informações sobre a implantação dos RIs; boas práticas, desafios e soluções para a correção dos metadados em lote e elementos computacionais para auxiliarem os gestores dos RIs no controle de autoridade e na qualidade dos metadados para entidade pessoa. Ficou claro que o controle de autoridade é um problema a ser vencido por toda a comunidade acadêmica, mas que pode ser dirimido com algumas iniciativas: usabilidade das ferramentas dos RIs, uso de menus suspensos, orientações para submissão, formulários *Web* e outros mecanismos que auxiliem o autor durante o autodepósito; equipe para a revisão de

metadados, criar um fluxo de trabalho para que os metadados possam ser revistos antes da submissão definitiva no RI; ecossistema de fluxos de trabalhos, *workflows* que atribuam funções e permissões específicas para a entrada de registros, bem como a adoção de documentações, políticas e passo a passo para o processo de catalogação, criação de listas, verificação automática de *links* e ortografia e ferramentas para autopovoamento; treinamento e capacitação de pesquisadores e profissionais; recursos de controle de autoridade para entidade pessoa com a adoção e vinculação de listas locais ao *software* do RI; e aplicativos de dados vinculados para controle de autoridade para entidade pessoa, com a adoção, por exemplo, do ORCID, como uma política acadêmica para auxiliar na identificação dos autores e facilitar a desambiguação.

A partir da revisão da literatura e da documentação do *software*, foi possível conhecer algumas das ferramentas ou alternativas computacionais nativas ou não do *DSpace*, que respondeu ao quarto objetivo específico: “Descrever informações sobre elementos computacionais nativos e compatíveis com o *software DSpace* disponíveis na sua documentação e na literatura, buscando aqueles que pudessem auxiliar no controle de autoridade para entidade pessoa”, resultados que se encontram no Capítulo 2, da Revisão de Literatura, na seção 3.4, do Capítulo de Fundamentos Teórico-Conceituais, e na seção 5.2, do Capítulo de Resultados e Discussões. Com a tradução de parte da documentação do *DSpace*, foram descritos os seus três elementos computacionais nativos: *Choice Management and Authority Control*, *Authority Control of Metadata Values* e o *ORCID Integration*, que, se utilizados desde o início do processo de implantação do RI, podem reduzir as incoerências relacionadas ao controle de autoridade para entidade pessoa, e, assim, a qualidade de seus metadados. Também foram descritos cinco elementos computacionais não nativos, que citados na revisão de literatura que podem ajudar os gestores a melhorarem a qualidade dos metadados, por exemplo, na extração de registros duplicados, padronização de maiúsculas/minúsculas, correção de erros de grafia e pontuação, entre outros. Os elementos computacionais não nativos mapeados foram os seguintes: *OpenRefine*, *OpenOffice Calc*, *Microsoft Access*, *Microsoft Excel* e *Duplicate Detection Service*, que podem auxiliar principalmente na correção de metadados em lote, herdados seja por autoarquivamento ou autodepósito.

Conhecendo os elementos computacionais, nativos ou não, partiu-se para o quinto e último objetivo específico: “Verificar de que modo e se ocorre o uso de elementos computacionais para o controle de autoridade em RIs que utilizam o *DSpace*, buscando entender o controle de

autoridade realizado para entidade pessoa”, cujos resultados se encontram na seção 5.1, do Capítulo de Resultados e Discussões. Foi aplicado um questionário para gestores de RIs que adotaram o *DSpace*, dentro do Estado de Minas Gerais, e, por meio das análises das respostas, foi possível constatar que o controle de autoridade, a adoção de padrões ou de melhores práticas ainda não é uma realidade nos RIs estaduais. E esse resultado talvez possa ser atribuído ao fato de os depósitos serem feitos pelos próprios autores, e, mesmo que a literatura recomende depósito mediado ou a revisão dos metadados por equipe especializada, essa não é uma realidade. Esse fato reforça o que é constatado na literatura, que os RIs sofrem com a escassez de recursos humanos e financeiros.

Retomando o pressuposto de pesquisa, “que a implementação desses elementos computacionais nativos e também compatíveis (não nativos) é difícil para profissionais não habilitados a resolver os possíveis problemas técnicos envolvidos, já que eles exigem um grau avançado de conhecimento em informática”, ficou confirmada em alguns estudos (SALO, 2008, 2009; MURUGATHAS; BALASOORIYA, 2014) descritos na revisão de literatura (Capítulo 2, seção 2.2), que atentam para a necessidade de uma equipe de tecnologia qualificada que possa se dedicar ao *software* e às ferramentas complementares para apoiar os gestores não apenas na instalação, personalização e suporte, mas também no amadurecimento do RI, por exemplo, oferecendo auxílio em processos de rotina, como limpeza de metadados com a criação de scripts, que exige pessoal especializado para acessar o banco de dados do *software* de gestão do RI.

E, por fim, a questão de pesquisa: “como ocorre o controle de autoridade para entidades pessoas em repositórios institucionais que usam o *DSpace*, e quais são os elementos computacionais nativos ou compatíveis (não nativos) com o *software* que podem contribuir para esse controle pelos gestores?”, foi respondida nas seções 2.2 do Capítulo 2 e anteriormente na retomada do terceiro objetivo específico, já discutido anteriormente. Especificamente sobre os elementos computacionais, esses insumos foram descritos na seção 2.2 do Capítulo 2, na seção 3.4, do Capítulo de Fundamentos Teórico-Conceituais, e na seção 5.2, do Capítulo de Resultados e Discussões, assim como já foram retomados em resposta ao quarto objetivo específico.

6. 1 Limitações da pesquisa

Uma limitação desta pesquisa refere-se ao fato de a coleta de dados ter coincidido com o início da pandemia, com a recomendação de trabalho remoto nas instituições de ensino, pois, inicialmente, estava previsto um contato telefônico com os gestores dos RIs pesquisados. Isso seria necessário para identificar com precisão o responsável pela gestão do RI, buscando saber se ele conhecia as políticas e processos que envolviam um projeto dessa natureza, inclusive sobre a necessidade ou não da indicação de um representante como o respondente da instituição, assim como motivando-o para a participação na pesquisa, abrindo um canal para dirimir possíveis dúvidas relacionadas ao questionário que seria enviado. Infelizmente, apesar das inúmeras tentativas de contato telefônico, a identificação precisa do gestor, ou de um representante mais qualificado, não foi possível. Assim, o questionário foi enviado para o *e-mail* de contato disponível no site oficial dos repositórios pesquisados, não sendo possível garantir que os respondentes conheçam a fundo os processos de controle de autoridades e se utilizavam algum elemento computacional disponível ou compatível com o *software DSpace*. Além disso, como o questionário foi aplicado *on-line*, não foi possível contatar os respondentes para esclarecer possíveis dúvidas relacionadas a ele. Apesar da impossibilidade de se saber se os respondentes eram os gestores dos RIs, pela experiência da autora sobre a dinâmica que ocorre no povoamento de repositórios mineiros, acredita-se que os respondentes eram os próprios gestores ou bibliotecários que atuam na inclusão de itens nos RIs.

Outra limitação que se pode considerar tem relação com a não identificação, nos documentos da Revisão de Literatura, de elementos computacionais não nativos do *DSpace*, mas compatíveis com ele, voltados, exclusivamente, para o controle de autoridade para entidade pessoa. Como descrito na seção 5.2.2, esperava-se encontrar na literatura um número razoável de elementos computacionais desse tipo. E essa expectativa se baseou no fato de o *DSpace* ser o *software* mais adotado pelos RIs nacionais, cerca de 83% e por cerca de 39% dos RIs no mundo, segundo o *OpenDoar* (2021). A análise da literatura levou à inferência de que o controle de autoridade, de maneira mais ampla, é um dos requisitos para uma boa qualidade dos metadados, e, assim, foram encontrados elementos computacionais não nativos do *DSpace* somente com foco na melhoria da qualidade dos metadados como um todo. E as instituições que se propuseram a resolver os problemas relacionados ao controle de autoridade criaram iniciativas locais, seja com a criação de fluxos de trabalho ou listas locais vinculadas ao sistema acadêmico. Dessa maneira, os resultados apresentados sobre os elementos computacionais

nativos e compatíveis (não nativos) com o *DSpace*, específicos para o tratamento dos problemas relativos ao controle de autoridade para entidade pessoa, se limitaram aos elementos computacionais nativos.

6.2 Estudos futuros

Como proposta de estudo futuro, sugere-se buscar responder à questão que ficou pendente nesta pesquisa, que é evidenciar os fatores que levam ao não uso dos elementos computacionais nativos, já disponíveis no *DSpace*, e até mesmo aqueles não nativos, mas que são compatíveis com o *software*, na gestão dos RIs nacionais. A revisão de literatura deixou evidente que as iniciativas mapeadas estão, prioritariamente, em literatura estrangeira, sendo necessário fazer um diagnóstico sobre a situação brasileira. Apesar da necessidade de um profissional especialista de TI para garantir o melhor desempenho possível de um sistema que usa o *DSpace*, entende-se que o sucesso de um RI depende de uma equipe multidisciplinar.

Sugere-se, também, uma pesquisa para identificar os RIs nacionais que fazem uso dos elementos computacionais identificados na seção 5.2.3, para, assim, trazer à tona as melhores práticas de gestão de RIs e facilitar a troca de informações entre os RIs nacionais. Acredita-se que o trabalho cooperativo é imprescindível para incrementar uma melhor qualidade dos metadados, e, conseqüentemente, aprimorar o controle de autoridades para entidade pessoa.

6.3 Contribuição da pesquisa

Acredita-se que os esclarecimentos sobre os elementos computacionais nativos e compatíveis (não nativos) com o *DSpace* possam auxiliar na implementação e manutenção dos RIs pelos seus gestores, e, dessa maneira, facilitar a gestão das informações sobre os autores ali publicadas. Para as instituições, o conhecimento sobre os elementos computacionais nativos, pode apoiar no controle de autoridade desde a origem do RI. Além disso, o uso dos elementos computacionais pode facilitar no processo de autodepósito pelos autores, com o uso de listas suspensas e pré-definidas. O conhecimento sobre os elementos computacionais não nativos pode poupar trabalho e tempo, sistematizando as atividades, por exemplo, da revisão dos metadados, que pode ser feita em lote, sem a necessidade de acerto individual dos registros pela equipe de bibliotecários. A implantação e a manutenção dos RIs têm como consequência benefícios para a instituição e pesquisadores. Para o primeiro, traz, sobretudo, visibilidade que

a distingue no mundo, benefício que se estende a todos que tenham seu nome ligado a ela. Para o segundo, dá visibilidade e permite fornecer indicador do impacto para a sua pesquisa, em especial, no que se refere ao mérito do trabalho e não ao nome do periódico em que sua produção foi publicada.

Outro subsídio trazido por este estudo foram os insumos para entender as questões de ambiguidade na identificação do nome do autor como um processo complexo que envolve vários fatores. Alguns exemplos são os homônimos, as mudanças de nome, as grafias múltiplas, as distintas abreviações, informações incompletas ou mesmo equivocadas sobre uma pessoa (BARBASTEFANO, 2015), renomeações, com substituição, acréscimo ou redução dos nomes (ISQUERDO, 2020), fatores que merecem uma atenção contínua dos bibliotecários na identificação do autor e no controle dessas entidades.

Os registros que referenciam as pessoas são dispositivos administrativos e políticos, utilizados pelo Estado e pelas instituições como controle das atividades do indivíduo, e dando a esse indivíduo direitos de toda natureza – política, jurídica e social. No âmbito acadêmico, o nome referencial é parte integrante de seu trabalho e garante os seus direitos de autor, dando visibilidade e reconhecimento ao conjunto de suas atividades e produção técnico-científica, e lhe permite também pleitear bolsas de fomento e captar recursos para o reconhecimento de suas pesquisas (SILVEIRA, 2013). Considera-se que, na comunicação científica, cada autor deve ser referenciado de forma unívoca, que esse referencial deve ser apreendido coletivamente, para, assim, ocorrer ambiguidade na computação de sua produção.

Acredita-se que a alternativa de criação de listas locais, desde que pensadas de forma cooperativa e interoperável, pode contribuir para o controle de autoridade de autores locais. Ou ainda, a possibilidade de criação de consórcio de instituições acadêmicas, tendo em vista contribuir com o incremento de listas de autoridade para autoria, ou a colaboração com listas já existentes, como o VIAF.

E, finalmente, entende-se que gestores, bibliotecários e equipes de TI devem se unir em prol de um controle de autoridade, em especial, para entidade pessoa, nos RIs, criando um ecossistema de fluxo de trabalho que inclua: políticas de criação e revisão de metadados, para uso de listas de vocabulários ou de autoridades, passo a passo de submissão do autodepósito, regras de catalogação, processos de verificação automática de ortografia e *links*, trilha de cursos e treinamentos para capacitar não apenas os membros da equipe do RI, como também os autores

(pesquisadores, docentes e alunos), assim como institucionalizar o uso de aplicativos de dados vinculados para entidades pessoas, a exemplo do ORCID, para auxiliarem na identificação e desambiguação dos nomes dos autores.

REFERÊNCIAS

- ALMEIDA, V. A disciplina jurídica do nome da pessoa humana à luz do direito à identidade pessoal. **Rev. Jurídica Luso-Brasileira**, ano 3, n. 3, p. 1141-1205, 2017. Disponível em: https://www.cidp.pt/revistas/rjlb/2017/3/2017_03_1141_1205.pdf. Acesso em: 5 ago. 2021.
- ALMUZARA, L. B.; ALVITE DÍEZ, M. L.; BRAVO, B. R. A Study Of Authority Control in Spanish University Repositories. **Knowl. organ.**, v. 39, n. 2, p. 95-103, 2012. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=izh&AN=73389455&lang=pt-br&site=ehost-live&authtype=ip,cookie,uid>. Acesso em: 31 jul. 2021.
- AMARAL, E. T. R. Contribuições para uma tipologia de antropônimos do português brasileiro. **Alfa: Revista de Linguística**, São Paulo, v. 55, n. 2, p. 63-82, 2001.
- AMARAL, E. T. R.; SEIDE, M. S. **Nomes próprios de pessoa: introdução à antroponímia brasileira**. São Paulo: Blucher, 2020.
- ANGJELI, A.; MACEWAN, A.; BOULET, V. **ISNI and VIAF: Transforming ways of trustfully consolidating identities**. 2014. Disponível em: <http://library.ifla.org/985/1/086-angjeli-en.pdf>. Acesso em: 30 jul. 2021.
- ASSIS, T. B. de. Análise das políticas de autoarquivamento nos repositórios institucionais brasileiros e portugueses. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 4, n. 2, p. 212–227, 2013. Disponível em: <https://doi.org/10.11606/issn.2178-2075.v4i2p212-227>. Acesso em: 14 ago. 2020.
- ASSUMPÇÃO, F. S. **Conversão de registros em XML para MARC 21: um modelo baseado em folhas de estilo XSLT**. 2013. 135 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências de Marília, 2013. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/93658/assumpcao_fs_me_mar.pdf?sequence=1&isAllowed=y. Acesso em: 5 ago. 2021.
- ASSUMPÇÃO, F. S.; SANTOS, P. L. V. A. C. A importância do controle de autoridade: uma abordagem baseada nos objetivos e nas funções dos catálogos. *In: ENCONTRO NACIONAL DE CATALOGADORES*, 2.; *ENCONTRO DE ESTUDOS E PESQUISAS EM CATALOGAÇÃO*, 3., 2012, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: e-Lis, 2012. Disponível em: <http://eprints.rclis.org/18843/>. Acesso em: 11 jul. 2020.
- ASSUMPÇÃO, F. S.; SANTOS, P. L. V. A. C.; ZAFALON, Z. R. O controle de autoridade no domínio bibliográfico: os catálogos digitais. **Biblios**, Pittsburgh, n. 68, p. 21-33, jul./set. 2017. Disponível em: <http://www.scielo.org.pe/pdf/biblios/n68/a02n68.pdf>. Acesso em: 5 ago. 2021.
- AVERKAMP, S.; LEE, J. Repurposing ProQuest metadata for batch ingesting ETDs into an institutional repository. **Code4Lib J.**, v. 7, June 2009. Disponível em: <https://journal.code4lib.org/articles/1647>. Acesso em: 5 ago. 2021.

ÁVILA, R. Como criar um banco de dados em Excel. **Blog LUZ**, 6 de maio de 2015. Disponível em: <https://blog.luz.vc/excel/como-criar-um-banco-de-dados-em-excel/>. Acesso em: 5 ago. 2021.

BABINI, D.; MACHIN-MASTROMATTEO, J. D. Latin American science is meant to be open access: Initiatives and current challenges. **Inform. Dev.**, v. 31, n. n. 5, p. 477-481, 2015. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0266666915601420>. Acesso em: 5 ago. 2021.

BANCO de Dados no Excel: Como Criar?. Excel Easy, 11 jun. 2020. Disponível em: <https://exceleasy.com.br/banco-de-dados-no-excel-como-criar/>. Acesso em: 25 jul. 2021.

BAPTISTA, A. A. *et al.* Comunicação científica: o papel da open archives initiative no contexto do acesso livre. **Enc. Bibli.** R. Eletr. Bibliotecon. Ci. Inf, Florianópolis, v. 12, n. 1, p. 1-17, 2007. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/view/377>. Acesso em: 30 jul. 2021.

BARBASTEFANO, R. G. Influência da ambiguidade de nomes na centralidade de redes de coautoria. **TransInformação**, Campinas, v. 27, n.3, p. 189-198, set./dez. 2015. Disponível em: scielo.br/j/tinf/a/bLPBq3zzFYFHZ9k9x7gnLdd/?lang=pt. Acesso em: 5 ago. 2021.

BARTON, J.; CURRIER, S.; HEY, J. M. N. Building Quality Assurance into Metadata Creation: An Analysis based on the Learning Objects and e-Prints Communities of Practice. *In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS*, 2003, Seattle. **Proceedings** [...]. Seattle: DCMI, 2003, p. p. 39-48. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/732>. Acesso em: 30 jul. 2021.

BIBLIOTECA NACIONAL (Brasil). **Catálogos**. 2021. Disponível em: <https://www.bn.gov.br/explore/catalogos>. Acesso em: 30 jul. 2021.

BOLLINI, A. **Configure the data model using an Excel file**: DSpace-CRIS, 27 Aug. 2018. Disponível em: <https://wiki.lyrasis.org/display/DSPACECRIS/configure+the+data+model+using+an+Excel+file>. Acesso em: 25 jul. 2021.

BRASIL. Lei n. 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. **Diário Oficial da União**, 20 fev. 1998. Disponível em: http://www.planalto.gov.br/ccivil_03/Leis/L9610.htm. Acesso em: 25 jul. 2021.

BRASIL. Lei n. 11.971, de 6 de Julho de 2009. Dispõe sobre as certidões expedidas pelos Ofícios do Registro de Distribuição e Distribuidores Judiciais. **Diário Oficial da União**, 7 jul. 2009. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/lei/l11971.htm. Acesso em: 25 jul. 2021.

CAMPOS, F. M. **Controle de autoridades**: novos contextos e soluções. 2003. Disponível em: <http://purl.pt/331/1/docs/comunicacao/12manha/controlodeautoridade.doc>. Acesso em: 31 jul. 2021.

CARDOSO, C. I. B.; MIRANDA, A. C. D. Modelo de controle de autoridade de nomes pessoais: uma proposta para o Repositório Institucional da Universidade Federal do Rio Grande. **ÁGORA: Arquivologia em Debate**, Florianópolis, v. 30, n. 60, p. 130-147, 2020. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/894>. Acesso em: 25 jul. 2021.

CLACK, D. H. **Authority control: principles, applications, and instructions**. ALA: Chicago; London, 1990.

COOKE, J. A visual arts perspective on open access institutional repositories. *In: COMPUTERS AND HISTORY OF ART ANNUAL CONFERENCE, 23.*, 2007, Londres. **Proceedings [...]** Londres: Goldsmiths Research, 2007.

COSTA, M. P.; LEITE, F. C. L. **Repositórios institucionais da América Latina e o acesso aberto**. Brasília: IBICT, 2017. Disponível em: <https://repositorio.unb.br/handle/10482/23202>. Acesso em: 5 ago. 2021.

COSTA, S. M. S.; KURAMOTO, H.; LEITE, F. C. L. Acesso aberto no Brasil: aspectos históricos, ações institucionais e panorama atual. *In: RODRIGUES, E.; SWAN, A.; BAPTISTA, A. A. (org.). Uma década de acesso aberto na UMinho e no mundo*. Braga: Universidade do Minho, Serviços de Documentação, 2013. p. 133-150.

CRAFT, A. R. Local Name Authorities and Linked Data: The Landscape. **Serials Rev.**, v. 45, n. 1-2, p. 61-65, 2019. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00987913.2019.1613736?journalCode=usrv> 20. Acesso em: 5 ago. 2021.

CRESWELL, J. W. **Research design: qualitative, quantitative, and mixed methods approaches**. 4. ed. Thousand Oaks: Sage, 2014.

CROW, R. **The Case for Institutional Repositories: A SPARC Position Paper**. Washington: SPARC, 2002. Disponível em: https://ils.unc.edu/courses/2014_fall/inls690_109/Readings/Crow2002-CaseforInstitutionalRepositoriesSPARCPaper.pdf. Acesso em: 5 ago. 2021.

CURRIER, S. Metadata Quality in E-Learning: Garbage in - Garbage out? **Forum CETIS**, Apr. 2004. Disponível em: <http://zope.cetis.ac.uk/content2/20040402013222/printArticle/index.html>. Acesso em: 31 jul. 2021.

DIAO, J.; HERNÁNDEZ, M. A. Transferring Cataloging Legacies into Descriptive Metadata Creation in Digital Projects: Catalogers' Perspective. **J. Libr. Metadata**, v. 14, n. 2, p. 130-145, 2014. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/19386389.2014.909670>. Acesso em: 5 ago. 2021.

DIAS, E. J. W. **OCLC: um novo conceito em cooperação bibliotecária**. Porto Alegre: FEBAB, 1977. Disponível em: <http://repositorio.febab.org.br/items/show/1067>. Acesso em: 30 jul. 2021.

DONOHUE, T. Mensagem recebida 17 de nov 2006, em Lista de Discussão sobre software, mantida por OpenOffice.org. 2006. [programador de pesquisa, no IDEALS da Universidade de Illinois].

DONOHUE, T. **Release Notes**. 2018. Disponível em: <https://wiki.lyrasis.org/display/DSDOC6x/Release+Notes#ReleaseNotes-6.3Acknowledgments>. Acesso em: 5 ago. 2021.

DOWNEY, M. Assessing Author Identifiers: Preparing for a Linked Data Approach to Name Authority Control in an Institutional Repository Context. **J. Libr. Metadata**, v. 19, n. 1-2, p. 117-136, 2019. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/19386389.2019.1590936?journalCode=wjlm> 20. Acesso em: 5 ago. 2021.

DSPACE. **Technical Specifications**: DSpace. [S. l.: s. n.], 2020. Disponível em: https://duraspace.org/wp-content/uploads/dspace-files/specsh_dspace.pdf. Acesso em: 8 fev. 2020.

DUBLIN CORE METADATA INITIATIVE – DCMI. **Using Dublin Core**. 2001. Disponível em: <https://www.dublincore.org/specifications/dublin-core/usageguide/2001-04-12/>. Acesso em: 31 jul. 2021.

DUBLIN CORE METADATA INITIATIVE – DCMI. **History**. 2020. Disponível em: <https://dublincore.org/about/history/>. Acesso em: 31 jul. 2021.

ESPÍNDOLA, P. L.; PEREIRA, A. M. Análise das políticas institucionais de catalogação em bibliotecas de ensino superior de Florianópolis (Santa Catarina, Brasil). **Biblios**, Pittsburgh v. 69, p. 73-86, 2017. Disponível em: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1562-47302017000400005. Acesso em: 5 ago. 2021.

FALK, J. D. **OCLC and RLIN**: research libraries at the scholar's fingertips. May 1989. Disponível em: <https://www.historians.org/publications-and-directories/perspectives-on-history/may-1989/oclc-and-rlin-research-libraries-at-the-scholars-fingertips>. Acesso em: 30 jul. 2021.

FERNANDES, J. D. P. B.; SOUZA, O. A contribuição do processamento técnico biblioteconômico para a acessibilidade informacional. **Revista Ciencias de la Documentación**, Santiago, v. 3, n. 4, oct./dic. 2017. Disponível em: <http://bcp.cienciasdeladocumentacion.cl/gallery/1%20v3n4%202017%20csdoc.pdf>. Acesso em: 5 ago. 2021.

FREGE, G. Sobre o sentido e a referência. In: FREGE, G. **Lógica e filosofia da linguagem**. São Paulo: Cultrix, 1978.

FREITAS, M. A. **Diretrizes para o depósito da produção científica em repositórios institucionais**. 2015. 199f. Tese (Doutorado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade de Brasília, 2015. Disponível em: https://repositorio.unb.br/bitstream/10482/19189/1/2015_MariliaAugustadeFreitas.pdf. Acesso em: 5 ago. 2021.

GALSON, G. Metadata cleanup made easy with OpenRefine. **PA Digital**, Pennsylvania, Oct. 31, 2017. Disponível em: <https://padigital.org/2017/10/31/metadata-cleanup-made-easy-with-openrefine/>. Acesso em: 5 ago. 2021.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GRINGS, L. Controle de autoridades na Biblioteca Nacional do Brasil: breve histórico e práticas atuais. **Rev. Bras. Bibl. Doc.**, São Paulo, v. 11, n. 2, p. 139-154, jul./dez. 2015. Disponível em: <https://www.bn.gov.br/sites/default/files/documentos/producao/artigo/2015/controle-autoridades-biblioteca-nacional-brasil-breve.pdf>. Acesso em: 30 jul. 2021.

GUINCHAT, C.; MENO, M. Introdução geral às ciências e técnicas da informação e documentação. Brasília: IBICT, 1994.

HERRERO PASCUAL, C. El Control de Autoridades. **Anales de Documentación**, v. 2, p. 121-136, 1999. Disponível em: <https://revistas.um.es/analesdoc/article/view/2621>. Acesso em: 30 jul. 2021.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT. **IBICT estimula adoção dos Open**. 2005. Disponível em: [https://ibict.br/sala-de-imprensa/noticias/item/Archives no Brasil 788-ibict-estimula-adoacao-dos-open-archives-no-brasil](https://ibict.br/sala-de-imprensa/noticias/item/Archives%20no%20Brasil%20788-ibict-estimula-adoacao-dos-open-archives-no-brasil). Acesso em: 31 jul. 2021.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT. **Repositórios digitais: histórico**. 2012. Disponível em: <http://sitehistorico.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais/historico>. Acesso em: 31 jul. 2021.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT. **Manual para o gerenciamento de repositórios digitais criados em DSpace**. 2016. Disponível em: https://wiki.lyrasis.org/pages/viewpage.action?pageId=112528241&preview=%2F112528241%2F112529244%2FManual_Gerenciamento_de_reposit%C3%B3rios_digitais_DSpace_final_2016.pdf. Acesso em: 30 jul. 2021.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT. **Sistema para Construção de Repositórios Institucionais Digitais (DSpace)**. 2019. Disponível em: <https://ibict.br/tecnologias-para-informacao/dspace>. Acesso em: 15 ago. 2020.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS – IFLA. Study Group on the Functional Requirements for Bibliographic Records. **Functional requirements for bibliographic records: final report**. München: K. G. Saur, 1998. Disponível em: <https://www.ifla.org/files/assets/cataloguing/frbr/frbr.pdf>. Acesso em: 5 ago. 2021.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS – IFLA. **Universal Bibliographic Control and International MARC Core Programme UNIMARC: An Introduction**. 1999. Disponível em: <https://archive.ifla.org/ubcim/p1996-1/unimarc.htm>. Acesso em: 30 jul. 2021.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS – IFLA. **Statement of international cataloguing principles**. 2009. Disponível em: https://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf. Acesso em: 30 jul. 2021.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS – IFLA. **Declaração dos Princípios Internacionais de Catalogação (PIC)**. 2016. Disponível em: https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-pt.pdf. Acesso em: 9 fev. 2020.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS – IFLA. **Authority control**. 2017. Disponível em: <https://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/9031>. Acesso em: 30 jul. 2021.

ISNI. **What is ISNI?** ©2021. Disponível em: <https://isni.org/page/what-is-isni/>. Acesso em: 30 jul. 2021.

ISQUERDO, A. N. Prefácio. *In*: AMARAL, E. T. R.; SEIDE, M. S. **Nomes próprios de pessoa**: introdução à antroponímia brasileira. São Paulo: Blucher, 2020.p. 9-25. Disponível em: https://www.blucher.com.br/livro/download_remote_file/?path=https://editor.blucher.com.br/uploads/product/presspdf/1614.pdf. Acesso em: 5 ago. 2021.

JOHNSON, R. K. Partnering with faculty to enhance scholarly communication. **Dlib Mag.**, v. 8, n. 11, Nov. 2002. Disponível em: <http://www.dlib.org/dlib/november02/johnson/11johnson.html>. Acesso em: 5 ago. 2021.

JOO, S.; HOFMAN, D.; KIM, Y. Investigation of challenges in academic institutional repositories. **Libr. Hi Tech**, v. 37, n. 3, p. 525-548, 2019. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/LHT-12-2017-0266/full/html>. Acesso em: 5 ago. 2021.

JOUDREY, D. N.; TAYLOR, A. G.; MILLER, D. P. Cataloging in Context. *In*: **Introduction to cataloging and classification**. 11. ed. Santa Barbara, California: Libraries Unlimited, 2015. p. 24-28.

JUVILER, J. What is UTF-8 Encoding? A Guide for Non-Programmers. **Blog HubSpot**, 2020. Disponível em: <https://blog.hubspot.com/website/what-is-utf-8>. Acesso em: 5 ago. 2021.

KURAMOTO, H. **Como a TI poderia auxiliar no povoamento dos RI**. 2011. Disponível em: <https://web.archive.org/web/20150804174252/http://kuramoto.blog.br/2011/10/31/como-a-ti-poderia-auxiliar-no-povoamento-dos-ri/>. Acesso em: 30 jul. 2021.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. **Ci. Inf.**, Brasília, v. 35, n. 2, p. 91-102, maio/ago. 2006. Disponível em: https://www.brapci.inf.br/_repositorio/2010/02/pdf_3dbb2fb854_0008036.pdf. Acesso em: 5 ago. 2021.

KURTZ, M. Dublin Core, DSpace, and a Brief Analysis of Three University Repositories. **Inform. Technol. Libr.**, v. 29, n. 1, p. 40-46, 2010. Disponível em: <https://search.proquest.com/docview/215828118?accountid=134127>. Acesso em: 31 jul. 2021.

LAWRENCE, S. Free on-line availability substantially increases a paper's impact. **Nature**, v. 411, n. 6836, p. 521-522, May 2001. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/11385534/>. Acesso em: 5 ago. 2021.

LEITE, F. C. L. Comunicação científica e gestão do conhecimento: enlaces conceituais para a fundamentação da gestão do conhecimento científico no contexto de universidades. **TransInformação**, Campinas, v. 19, n. 2, p. 139-151, maio/ago. 2007. Disponível em: <https://www.scielo.br/j/tinf/a/qLh7S9vyrcdZRxFdx4d6vvN/?format=pdf&lang=pt>. Acesso em: 5 ago. 2021.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília: IBICT, 2009. Disponível em: <http://livroaberto.ibict.br/handle/1/775>. Acesso em: 5 ago. 2021.

LEITE, F. C. L. *et al.* **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: IBICT, 2012. Disponível em: <https://livroaberto.ibict.br/handle/1/703>. Acesso em: 5 ago. 2021.

LEY, M. D. L. M. G. **Diretrizes para a proposição de política de povoamento de Repositório Institucional**: o contexto da Universidade Federal Fluminense (UFF). 2013. 242f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal Fluminense. Niterói: UFF, jun. 2013. Disponível em: <https://app.uff.br/riuff/handle/1/343>. Acesso em: 5 ago. 2021.

LIBRARY OF CONGRESS. **Introduction**: MARC 21 Authority. Oct. 2009. Disponível em: <https://www.loc.gov/marc/authority/adintro.html>. Acesso em: 30 jul. 2021.

LIBRARY OF CONGRESS. **About NACO**. ©2021. Disponível em: <https://www.loc.gov/aba/pcc/naco/about.html>. Acesso em: 30 jul. 2021.

LIU, X.; DURAN, A.; WASHINGTON, A. **Cleaning up legacy metadata for ETDs**: strategies, tools and a look into the future. In: UNITED STATES ELECTRONIC THESIS AND DISSERTATION ASSOCIATION CONFERENCE, 2016, Columbus. **Papers** [...]. Columbus: University of Houston, 2016. Disponível em: <https://uh-ir.tdl.org/handle/10657/1593?show=full>. Acesso em: 5 ago. 2021.

LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão de literatura. **Ci. Inf.**, Brasília, v. 31, n. 1, p. 41-52, jan./abr. 2002. Disponível em: <https://www.scielo.br/j/ci/a/WhYch5gHnWYPCPY3NPwbkcR/?lang=pt&format=pdf>. Acesso em: 5 ago. 2021.

LÓPEZ DÍEZ, M. P.; MELIÁN VINUESA, A. ACCEDA, repositorio institucional de la ULPGC: mejoras y desarrollos propios. **Ibersid**, v. 5, p. 127-133, 2011. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=73390530&lang=pt-br&site=ehost-live&authtype=ip,cookie,uid>. Acesso em: 5 ago. 2021.

LUBAS, R. Defining Best Practices in Electronic Thesis and Dissertation Metadata. **J. Libr. Metadata**, v. 9, n. 3-4, p. 252-263, 2009. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/19386380903405165>. Acesso em: 5 ago. 2021.

LUYTEN, B. **ORCID for DSpace**. 2014. Disponível em: https://pt.slideshare.net/bramluyten/orcid-identifiers-for-dspace?next_slideshow=1. Acesso em: 16 fev. 2020.

LYNCH, C. Institutional repositories: essential infrastructure for scholarship in the digital age. **ARL**, v. 226, p. 1-7, 2003. Disponível em: <https://www.cni.org/wp-content/uploads/2003/02/ar1-br-226-Lynch-IRs-2003.pdf>. Acesso em: 31 jul. 2021.

MANDAL, S. ORCID Integration with DSpace and EPrints: A Framework for ORCID Integration with DSpace and EPrints: A Framework for Research Communities Research Communities.

Library Philosophy and Practice, v. 3927, p. 1-9, 2020. Disponível em: <https://digitalcommons.unl.edu/libphilprac/3927>. Acesso em: 5 ago. 2021.

MARAIS, H. Authority control in an Academic Library Consortium Using a union catalogue maintained by a Central Office for authority control. 2004. 310 f. Thesis (Doctor of literature and philosophy in the subject Information Science) – University of South Africa, Pretoria, 2004. Disponível em: <http://uir.unisa.ac.za/handle/10500/2546>. Acesso em: 5 ago. 2021.

MARCONDES, C. H.; SAYÃO, L. F. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. **Ci. Inf.**, Brasília, v. 31, n. 3, p. 42-54, 2002. Disponível em: <http://revista.ibict.br/ciinf/article/view/947>. Acesso em: 15 ago. 2020.

MARCONDES, C. H.; SAYÃO, L. F. Software livres para repositórios institucionais: alguns subsídios para a seleção. In: SAYÃO, L. *et al.* (org.). **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009. p. 23-54. Disponível em: https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf. Acesso em: 5 ago. 2021.

MARCONI, M. A; LAKATOS, E. M. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003.

MARIANI, B. Nome próprio e constituição do sujeito. **Letras**, Santa Maria, v. 24, n. 48, p. 131-141, jan./jun. 2014. Disponível em: <https://periodicos.ufsm.br/letras/article/view/14428>. Acesso em: 5 ago. 2021.

MARTINS, D. L.; SILVA, M. F.; SIQUEIRA, J. Comparação entre sistemas para criação de acervos digitais: análise dos softwares livres DSpace, EPrints, Fedora, Greenstone e Islandora a partir de novas dimensões analíticas. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 9, n. 1, p. 52-71, mar./ago. 2018. Disponível em: <https://www.revistas.usp.br/incid/article/download/134333/140237/294617>. Acesso em: 31 jul. 2021.

MASAR, I. Bibliographic import from Scopus and WoS. [slides] *In*: OPEN REPOSITORIES, 2014, Helsinki, Finland. **Presentation** [...] Helsinki, Finland: OP, 2014. Disponível em: <https://www.doria.fi/handle/10024/97747>. Acesso em: 9 fev. 2020.

MATIAS, M. S. O. **Base referencial para o povoamento de repositórios institucionais**: coleta automatizada de metadados da Plataforma Lattes. 2015. 86f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos. São Carlos: São Carlos, 2015. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/6932>. Acesso em: 5 ago. 2021.

MATTAR, F. N. **Pesquisa de marketing**: metodologia, planejamento, execução e análise. 2. ed. São Paulo: Atlas, 1994.

MENDES, R. C. Manual Zotero básico 5.0: gerenciador de referências bibliográficas. 2019. Disponível em: http://mba.eci.ufmg.br/wp-content/uploads/Manual_Zotero.pdf. Acesso em: 5 ago. 2021.

MEY, Eliane Serrão Aleves. **Introdução a catalogação**. Brasília: Briquet de Lemos, 1995.

MEY, E. S. A.; SILVEIRA, N. C. **Catalogação no plural**. Brasília: Briquet de Lemos, 2009.

MEY, E. S. A.; SILVEIRA, N. C. Considerações teóricas aligeiradas sobre a catalogação e sua aplicação. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 1, n. 1, p. 125, 2010. Disponível em: <https://www.revistas.usp.br/incid/article/view/42309>. Acesso em: 9 fev. 2020.

MICROSOFT. **Top 10 motivos para usar o Access com Excel**. ©2021. Disponível em: <https://support.microsoft.com/pt-br/office/top-10-motivos-para-usar-o-access-com-excel-2a454445-13cc-4b39-bc2f-d27fd12ca414>. Acesso em: 30 jul. 2021.

MILLER, D. M.; HUNSAKER, A. J. Extending Name Authority Work beyond the Cataloging Department: A case study at the University of Nevada, Reno Libraries. **LRTS**, v. 62, n. 3, p. 136-148, 2018. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/6731>. Acesso em: 5 ago. 2021.

MONTE-MÓR, J. Cooperação bibliográfica nos Estados Unidos e seus reflexos no Brasil. **R. Bibliotecon. Brasília**, v. 11, n. 2, p. 257-271, jul./dez. 1983. Disponível em: <https://periodicos.unb.br/index.php/rbbsb/article/download/30271/25656/67621>. Acesso em: 30 jul. 2021.

MORRIS, C. M. **Enhance your DSpace Installation with Free Plugins from The Library Code**. 2019. Disponível em: <https://duraspace.org/enhance-your-dspace-installation-with-free-plugins-from-the-library-code/>. Acesso em: 5 ago. 2021.

MUGNAINI, R. *et al.* Normalização de nomes de autores em fontes de informação institucionais: proposta de um método automático de verificação de erros. **Em Questão**, Porto Alegre, v. 18, n. 3, p. 263-279, 2012. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/33265>. Acesso em: 14 ago. 2020.

MURUGATHAS, K.; BALASOORIYA, H. Developing an institutional repository: experiences at the library, Faculty of Medicine, University of Jaffna. **JULA**, v. 18, n. 1, p. 39-50, Jan. 2014. Disponível em: <https://jula.sjoi.info/articles/abstract/10.4038/jula.v18i1.7860/>. Acesso em: 5 ago. 2021.

NEATROUR, A.; MYNTTI, J. Western Name Authority File: a Pilot Regional Name Authority Project. **J. Libr. Metadata**, v. 19, n. 1-2, p. 19-38, 2019. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/19386389.2019.1589685?journalCode=wjlm> 20. Acesso em: 5 ago. 2021.

NEMATIANARAKI, L.; TAVASSOLI-FARAH, M. Scholarly communication through institutional repositories: proposing a practical model. **Collection and Curation**, v. 37, n.1, p. 9-17, 2018. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/CC-01-2018-002/full/html>. Acesso em: 6 abr. 2020.

NICHOLS, D. M. *et al.* Experiences in Deploying Metadata Analysis Tools for Institutional Repositories. **Cat. Classif. Q.**, v. 47, n. 3-4, p. 229-248, 2009. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01639370902737281>. Acesso em: 5 ago. 2021.

NIU, J. Evolving Landscape in Name Authority Control. **Cat. Classif. Q.**, v. 51, n. 4, p. 404-419, 2013. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01639374.2012.756843>. Acesso em: 5 ago. 2021.

ONLINE COMPUTER LIBRARY CENTER – OCLC. **About OCLC**. ©2021a Disponível em: <https://www.oclc.org/en/about.html>. Acesso em: 30 jul. 2021.

ONLINE COMPUTER LIBRARY CENTER – OCLC. **Inside WorldCat**. ©2021b. Disponível em: <https://www.oclc.org/en/worldcat/inside-worldcat.html>. Acesso em: 30 jul. 2021.

ONLINE COMPUTER LIBRARY CENTER – OCLC. **History of the OCLC Research Library Partnership**. ©2021c. Disponível em: <https://www.oclc.org/research/partnership/history.html>. Acesso em: 30 jul. 2021.

ONLINE COMPUTER LIBRARY CENTER – OCLC. **Connect authority data across cultures and languages to facilitate research**. ©2021d. Disponível em: <https://www.oclc.org/en/viaf.html>. Acesso em: 30 jul. 2021.

OPENOFFICE. **Introducing Calc**. 2009. Disponível em: <http://www.openoffice.org/documentation/manuals/userguide3/0301CG3-IntroducingCalc.pdf>. Acesso em: 30 jul. 2021.

OPENREFINE. **Documentation**. 2021 Disponível em: <https://openrefine.org/documentation.html>. Acesso em: 30 jul. 2021.

ORCID. ©2021. Disponível em: <https://orcid.org/members>. Acesso em: 30 jul. 2021.

ORTEGA, C. D. O conceito de documento em abordagem bibliográfica segundo as disciplinas constituintes do campo. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 7, n. esp, p. 41, 2016. Disponível em: <https://www.revistas.usp.br/incid/article/view/118749>. Acesso em: 15 ago. 2020.

PAPADAKIS, I. *et al.* Subject-based information retrieval within digital libraries employing LCSHs. **Dlib. Mag.**, v. 15, n. 9-10, Sept./Oct. 2009. Disponível em: <http://www.dlib.org/dlib/september09/papadakis/09papadakis.html>. Acesso em: 31 jul. 2021.

PARK, J.-R. Metadata Quality in Digital Repositories: a survey of the current state of the art. **Cat. Classif. Q.**, v. 47, n. 3-4, p. 213-228, 2009. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01639370902737240?journalCode=wccq20>. Acesso em: 25 jul. 2020.

PÉREZ-ALCAZAR, J. J. *et al.* Avaliação de Redes de Inovação usando uma ferramenta baseada em redes sociais - caso Brasileiro de Nanotecnologia. *In: CONGRESO LATINO-IBEROAMERICANO DE GESTION TECNOLÓGICA*, 14., 2011, Lima. **Anais [...]**. Lima: Pontificia Universidad Católica del Perú, 2011. Disponível em: <https://www.scielo.br/pdf/tinf/v29n1/0103-3786-tinf-29-01-00115.pdf>. Acesso em: 15 ago. 2020.

PINHEIRO, L. V. R. Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. **Reciis: Rev. Eletron. Comum. Inf. Inov. Saúde**, Rio de Janeiro, v. 8, n. 2, p. 153-165, jun. 2014. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/17107/2/8.pdf>. Acesso em: 12 fev. 2020.

POVEDA, A. P. *et al.* Sincronización en el control de autoridades repositorio/CRIS de la Universidad Carlos III de Madrid. *In: VEGA, J.A. M Ecosistemas del Acceso Abierto*. Salamanca: Ediciones Universidad de Salamanca, 2018.p. 261-272. Disponível em: <https://eusal.es/index.php/eusal/catalog/download/978-84-9012-773-5/5118/4743-1?inline=1>. Acesso em: 5 ago. 2021.

RIBEIRO-SEGUNDO, W. *et al.* Implementação de controle de autoridade em sistemas baseados em DSpace. *In: ENCONTRO NACIONAL DE ENSINO E PESQUISA EM INFORMAÇÃO*, 11., 2013, Escola Politécnica/UFBA. **Anais [...]**. Salvador: XI CINFORM, 2013. Disponível em: <http://www.cinform2013.ufba.br/modulos/submissao/Upload-164/48903.pdf>. Acesso em: 12 fev. 2020.

RIBEIRO, E. A. A perspectiva da entrevista na investigação qualitativa. **Evidência**, Araxá, n. 4, p. 129-148, 2008. Disponível em: http://www.educadores.diaadia.pr.gov.br/arquivos/File/maio2013/sociologia_artigos/tecnica_coleta_dados.pdf. Acesso em: 5 ago. 2021.

ROCHA, R. P. da. **Dublin Core**. [200-]. Disponível em: <http://www.ufrgs.br/snote/wiki/doc.php?id=190>. Acesso em: 19 jul. 2020.

ROSADO, K. M. L.; DIAS, C. C. Representação e recuperação de nomes de pessoas em catálogos de autoridades. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, v. 25, p.

1-25, 2020. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e70349>. Acesso em: 5 ago. 2021.

ROWLEY, J.; FARROW, J. **Organizing knowledge**: an introduction to managing access to information. 3rd. ed. Ashgate: Aldershot, 2000. p. 369-370.

SALO, D. Innkeeper at the Roach Motel. **Libr. Trends**, v. 57, n. 2, p. 98-123, Fall 2008. Disponível em: <https://minds.wisconsin.edu/handle/1793/22088>. Acesso em: 5 ago. 2021.

SALO, D. Name Authority Control in Institutional Repositories. **Cat. Classif. Q.**, v. 47, n. 3-4, p. 249-261, 2009. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01639370902737232>. Acesso em: 5 ago. 2021.

SANDY, H. M.; DYKAS, F. High-Quality Metadata and Repository Staffing: Perceptions of United States–Based OpenDOAR Participants. **Cat. Classif. Q.**, v. 54, n. 2, p. 101-116, 2016. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01639374.2015.1116480?journalCode=wccq> 20. Acesso em: 5 ago. 2021.

SAYÃO, L. F. Bibliotecas digitais e suas utopias. **PontodeAcesso**, Salvador, v.2, n.2, p. 2-36, ago. /set. 2008. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/2661>. Acesso em: 5 ago. 2021.

SAYÃO, L. *et al.* (org.). **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 9. Disponível em: https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf. Acesso em: 5 ago. 2021.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, v. 15, n. 30, p. 1-31, out. 2010. Disponível em: <https://www.icict.fiocruz.br/sites/www.icict.fiocruz.br/files/Uma%20outra%20face%20dos%20metadados.pdf>. Acesso em: 5 ago. 2021.

SEABRA, M. C. T. C.; ISQUERDO, A. N. A Onomástica em diferentes perspectivas: resultados de pesquisas. **Rev. Estud. Ling.**, Belo Horizonte. v. 26, n. 3, p. 993-1000, 2018. Disponível em: <http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/13700>. Acesso em: 5 ago. 2021.

SEARLE, J. R. Proper Names. *In*: MARTINICH, A. P. (Ed.). **Philosophy of Language**. 3rd. ed. Oxford: Oxford University, 1996.

SEFTON, P. Re-Discovering Repository Architecture: Adding Discovery as a Key Service. **New Review of Information Networking**, v. 14, n. 2, p. 84-101, 2009. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/13614570903359407>. Acesso em: 5 ago. 2021.

SEIDE, M. S. Prenomes cristãos: constituição, etimologia, motivação para a escolha antroponímica e conhecimento onomástico. **Rev. Estud. Ling.**, Belo Horizonte, v. 29, n. 1, p. 49-76, 2021. Disponível em:

<http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/16765/pdf>. Acesso em: 5 ago. 2021.

SERRA, M.; ROLIM, C.; BASTOS, A. P. **Universidades e desenvolvimento regional: as bases para a inovação competitiva**. Rio de Janeiro: Ideia D, 2018.

SERRANO-VICENTE, R.; MELERO, R.; ABADAL, E. Evaluation of Spanish institutional repositories based on criteria related to technology, procedures, content, marketing and personnel. **Data Technologies and Applications**, Bingley, v. 52, n. 3, p. 384-404, 2018. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/DTA-10-2017-0074/full/html>. Acesso em: 5 ago. 2021.

SEVERINO, A. J. **Metodologia do trabalho científico**. 23.ed. rev. e atual. ed. São Paulo: Cortez, 2007.

SHINTAKU, M.; MEIRELLES, R. **Manual do DSpace: administração de repositórios**. Salvador: EDUFBA, 2010. Disponível em: <https://repositorio.ufba.br/ri/bitstream/ri/769/1/Manual%20do%20Dspace%282%29.pdf>. Acesso em: 8 fev. 2020.

SHINTAKU, M.; ROBREDO, J.; BAPTISTA, D. M. Webometria dos repositórios institucionais acadêmicos. **Ci. Inf.**, Brasília, v. 40, n. 2, 2011. Disponível em: <http://revista.ibict.br/ciinf/article/view/1319>. Acesso em: 15 ago. 2020.

SILVA, A. J. **Política de catalogação para as Bibliotecas da Rede Nacional de Bibliotecas Públicas**. Lisboa: Direcção-Geral do Livro e das Bibliotecas, 2009. Disponível em: <http://rcbp.dglb.pt/pt/ServProf/Documentacao/Documents/PoliticaCatalogacaoDGLBfev2009.pdf>. Acesso em: 31 jul. 2021. Citado por ESPÍNDOLA, P. L.; PEREIRA, A. M. Análise das políticas institucionais de catalogação em bibliotecas de ensino superior de Florianópolis (Santa Catarina, Brasil). **Biblios**, Pittsburgh v. 69, p. 73-86, 2017. Disponível em: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1562-47302017000400005. Acesso em: 5 ago. 2021.

SILVA, M. C.; PAVÃO, C. G.; GABRIEL JUNIOR Ponto de acesso para pessoa: uma análise dos requisitos do FRAD e RDA. In: ENCONTRO RDA NO BRASIL, 2019, Florianópolis. **Apresentações [...]** Florianópolis: UDESC, 2019. Disponível em: <https://rdanobrasil.org/wp-content/uploads/2019/04/2019-ponto-de-acesso-para-pessoa-frad-e-rda.pdf>. Acesso em: 5 ago. 2021.

SILVA, L. C.; SERRA, L. G.; SANTAREM SEGUNDO, J. E. O ORCID como aplicação de Linked Data no Catálogo de Bibliotecas. **Informação & Tecnologia (ITEC)**, Marília/João Pessoa, v. 4, n. 2, p. 185-205, jul./dez. 2017. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/40186>. Acesso em: 5 ago. 2021.

SILVEIRA, N. C. **A trajetória da autoria na representação documental**. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, 2013. São Paulo: USP, 2013. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27151/tde-06052014-114125/publico/NairaSilveiraCorrigida.pdf>. Acesso em: 5 ago. 2021.

SMALHEISER, N. R.; TORVIK, V. I. Author name disambiguation. **ARIST**, v. 43, n.1, p. 1-43, 2009. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.2009.1440430113>. Acesso em: 6 abr. 2020.

SMITH, M. *et al.* DSpace: an Open Source dynamic digital repository. **Dlib. Mag.**, v. 9, n. 1, Jan. 2003. Disponível em: <https://dlib.org/dlib/january03/smith/01smith.html>. Acesso em: 5 ago. 2021.

SOUZA, D. L.; ORTEGA, C. D. O trabalho em rede na organização e nos serviços de informação: mapeamento e caracterização. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v.4, n.2, out. 2014. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/download/16988/13752/48095>. Acesso em: 5 ago. 2021.

SOUZA, E. G. de; BEZERRA, D. A.; COSTA, W. F. C. Descrição de recursos em uma estrutura de metadados pautada no modelo FRBR. **Em Questão**, Porto Alegre, v. 22, n. 1, p. 113-136, 2016. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/58835>. Acesso em: 5 ago. 2021.

STEELE, T.; SUMP-CRETHAR, N. Metadata for Electronic Theses and Dissertations: A Survey of Institutional Repositories. **J. Libr. Metadata**, v. 16, n. 1, p. 53-68, 2016. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/19386389.2016.1161462?journalCode=wjlm> 20. Acesso em: 5 ago. 2021.

STEIN, A.; APPLGATE, K. J.; ROBBINS, S. Achieving and Maintaining Metadata Quality: Toward a Sustainable Workflow for the IDEALS Institutional Repository. **Cat. Classif. Q.**, v. 55, n. 7-8, p. 644-666, 2017. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01639374.2017.1358786>. Acesso em: 5 ago. 2021.

TEXIER, J. *et al.* La visualización de autores en un repositorio institucional a través del enfoque Model Driven con WebRatio. **Ibersid**, v. 10, n. 1, p. 69-74, 2016. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=5566542>. Acesso em: 5 ago. 2021.

THE DSPACE DEVELOPER TEAM. **DSpace 6.x Documentation**. 2018. Disponível em: <https://wiki.duraspace.org/display/DSDOC6x>. Acesso em: 15 fev. 2020.

THOMPSON, S. *et al.* A Case Study of ETD Metadata Remediation at the University of Houston Libraries. **Libr. Resour. Tech. Ser.**, v. 63, n. 1, p. 62, 2019. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/6764/9320>. Acesso em: 5 ago. 2021.

TILLET, B. B. Considerations for authority control in the online environment. **Cat. Classif. Q.**, v. 9, n. 3, p. 1-11, 1989. Disponível em: https://www.tandfonline.com/doi/abs/10.1300/J104v09n03_01. Acesso em: 5 ago. 2021.

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG. **Repositório Institucional UFMG**. Belo Horizonte: UFMG, 2020. Disponível em: <https://repositorio.ufmg.br/>. Acesso em: 8 fev. 2020.

UNIVERSIDADE FEDERAL DE UBERLÂNDIA. Faculdade de Computação. PET Sistemas de Informação. **Apostila de LibreOffice Calc**. [202-]. Disponível em: https://aedmoodle.ufpa.br/pluginfile.php/242185/mod_resource/content/1/apostila_calc.pdf. Acesso em: 30 jul. 2021.

UNIVERSITY OF TORONTO LIBRARIES. **UTLAS**. 2018. Disponível em: <https://exhibits.library.utoronto.ca/exhibits/show/utl125/access-and-discovery/utlas>. Acesso em: 30 jul. 2021.

VEIGA, V. S. O.; MACHADO, R. R.; ALVES, A. Silva. Avaliação de usabilidade em Repositórios Institucionais: o caso da Fiocruz. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 25., 2013, Florianópolis, Santa Catarina. **Anais [...]**. São Paulo: FEBAB, 2013. Disponível em: <https://www.arca.fiocruz.br/handle/icict/8601>. Acesso em: 20 ago. 2020.

VIANA, C. L. M.; MÁRDERO ARELLANO, M. Á.; SHINTAKU, M. Repositórios institucionais em ciência e tecnologia: Uma experiência de customização do DSpace. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 3., **Anais [...]**. São Paulo: [s.n.], 2005. Disponível em: <http://eprints.rclis.org/7168/1/viana358.pdf>. Acesso em: 5 ago. 2021.

VIDOTTI, S. A. B. G.; BASTOS, F. M.; ROMANETTO, L. M. **O Repositório Institucional UNESP: compartilhando experiências e desafios**. 2017. Disponível em: <https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/19051/10Apresenta%C3%A7%C3%A3oFlavia.pdf?sequence=1>. Acesso em: 30 jul. 2021.

WALSH, M. P. Repurposing MARC metadata for an institutional repository: working with special collections and university press monographs. **Libr. Resour. Tech. Ser.**, v. 55, n. 1 p. 33-44, 2011. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5537/6814>. Acesso em: 5 ago. 2021.

WEIDNER, A.; WASHINGTON, A. M.; LIU, X. Planting Cedar: An Open Source Linked Data Vocabulary Manager at the University of Houston Libraries. **J. Libr. Metadata**, v. 19, n. 1-2, p. 53-68, 2019. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/19386389.2019.1589696?journalCode=wjlm> 20. Acesso em: 5 ago. 2021.

XIA, J. Personal name identification in the practice of digital repositories. **Program: Electronic Library And Information Systems**, v. 40, n.3, p. 256-267, 2006. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/00330330610681330/full/html>. Acesso em: 28 mar. 2020.

YASSER, C. M. An Analysis of Problems in Metadata Records. **J. Libr. Metadata**, v. 11, n. 2, p. 51-62, 2011. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/19386389.2011.570654?needAccess=true>. Acesso em: 25 jul. 2020.

ZAVALINA, O. L. **Collection-Level Subject Access in Aggregations of Digital Collections: Metadata Application and Use**. 2010. 84f. Dissertation (Docto of Philosophy in Library and Information Science) – University of Illinois at Urbana-

Champaign. Urbana e Champaign, Illinois, Urbana, 2010. Disponível em: <https://digital.library.unt.edu/ark:/67531/metadc67618/m1/1/>. Acesso em: 5 ago. 2021.

ZENG, M.; QIN, J. **Metadata**. New York, NY: Neal-Schuman, 2008. Citado por YASSER, C. M. An Analysis of Problems in Metadata Records. **J. Libr. Metadata**, v. 11, n. 2, p. 51-62, 2011. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/19386389.2011.570654?needAccess=true>. Acesso em: 25 jul. 2020.

APÊNDICE A – QUESTIONÁRIO

Controle de autoridade em Repositórios Institucionais

1. Identifique a cidade da sua Universidade.
2. Como é realizado o controle de autoridade no Repositório Institucional da sua Universidade?
3. Caso não realize o controle de autoridade, como justifica isso.
4. O repositório da sua universidade possui política para cadastro e controle de autoridade? Se sim, por favor, descreva.
5. O repositório da sua universidade utiliza fonte externa para consulta de cadastro de autoridade. Se sim, por favor, descreva. Por exemplo, o catálogo de autoridades da Biblioteca Nacional?
6. O repositório da sua universidade usa alguma aplicação ou ferramenta para controle da qualidade dos metadados de autoridade (campo autor)? Se sim, por favor, descreva. Por exemplo, algum plugin que corrija erros de digitação ou forma (maiúsculas e minúsculas).
7. O repositório da sua universidade aceita auto-depósito ou auto-arquivamento pelos autores? Se sim, os metadados são revistos por um bibliotecário antes da publicação no repositório?
8. O repositório da sua universidade é alimentado por auto-povoamento? Se sim, os metadados são revistos para evitar duplicidade de registros ou formas de entrada.

Caso tenha interesse em receber os resultados da pesquisa, por favor, compartilhe seu *e-mail*.

APÊNDICE B – TCLE

Prezado(a) Bibliotecário(a),

Você está sendo convidado para participar da pesquisa *Controle de autoridade em Repositórios Institucionais das Universidades Públicas Federais de Minas Gerais que utilizam DSpace*, que está sendo desenvolvida por Anna Carolina Leite Cota, aluna do Curso de Pós-Graduação em Gestão & Organização do Conhecimento, da Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG), sob a orientação da Profa Benildes Coura M S Maculan.

O estudo tem como objetivo mapear o controle de autoridade realizado nos repositórios institucionais das Universidades Públicas Federais de Minas Gerais que utilizam DSpace, tendo em vista demonstrar como o uso de ferramentas disponíveis neste *software* pode facilitar a gestão das informações da produção acadêmica.

Os resultados dessa pesquisa permitirão aos gestores dos repositórios conhecer as ferramentas de gestão e controle de autoridade disponíveis e que são compatíveis com o *software* DSpace, o que contribuirá na melhor gestão de indicadores de produção e impacto dos pesquisadores das universidades.

A sua participação nesta pesquisa se deve a você ser gestor/bibliotecário responsável pelo Repositório Institucional da Universidade Federal a que está vinculado. Sua participação consiste, ao aceitar colaborar com a pesquisa, em responder ao questionário eletrônico *on-line*, de forma voluntária, e concordar com a divulgação dos resultados da pesquisa em revistas e eventos na área de Ciência da Informação, assim como na dissertação de mestrado.

Caso sinta algum tipo de constrangimento em responder às questões do questionário, é importante que saiba que você pode desistir de responder à pesquisa, a qualquer momento e sem que haja qualquer prejuízo ou dano pessoal. Mas quero deixar claro que a sua participação e identidade serão mantidas em total sigilo, sendo que os dados coletados ficarão arquivados e sigilosos por cinco anos, e, após esse prazo, serão destruídos.

O pesquisador responsável, que também assina este documento, compromete-se a conduzir a pesquisa de acordo com o que preconiza a Resolução 466/12 de 12/06/2012, que trata dos

preceitos éticos e da proteção aos participantes da pesquisa. Garantimos aos participantes garantia de sigilo e confidencialidade.

Sua resposta será enviada automaticamente à pesquisadora Anna Carolina Leite Cota, que se coloca à sua disposição para quaisquer esclarecimentos que considere necessário, em qualquer etapa da pesquisa, e o pesquisador responsável pode ser contatado pelo *e-mail* carollcota@gmail.com e (31) 9 8846-7376.

O questionário terá um campo opcional, caso o participante queira deixar o contato de *e-mail*, dessa forma o pesquisador responsável se compromete a compartilhar os resultados da pesquisa tão logo seja possível.

Diante do exposto, declaro que li o documento e que poderei guardar uma cópia, conforme é meu direito, estabelecido na Resolução CNS 466/12, pois ele tem importantes informações de contato e de direitos. Atesto que esclareci todas as dúvidas relacionadas à pesquisa junto ao pesquisador responsável e que poderei entrar em contato a qualquer tempo para novos esclarecimentos, quando se fizer necessário. Estou ciente que a pesquisa em questão tem finalidade acadêmica, e dou o meu consentimento para participar da pesquisa como respondente, para a publicação dos resultados, e que tenho conhecimento de que minha participação e identidade serão mantidas em sigilo. Tenho consciência que posso abandonar minha participação na pesquisa a qualquer momento e que não receberei qualquer pagamento por esta participação.

Para participar da pesquisa é necessário que você concorde com esse Termo de Consentimento Livre e Esclarecido. Você concorda em participar desta pesquisa?

Sim

Não

Anna Carolina Leite Cota
Mestranda
PPGGOC/ECI/UFMG

Benildes Coura M. dos S. Maculan
Orientadora
PPGGOC/ECI/UFMG

Caso tenha alguma dúvida sobre a pesquisa, você pode entrar em contato o Comitê de Ética em Pesquisa (CEP) da Universidade Federal Minas Gerais:
Av. Presidente Antônio Carlos, 6627, Pampulha - Belo Horizonte - MG - CEP 31270-901
Unidade Administrativa II - 2º Andar - Sala: 2005
Telefone: (31) 3409-4592 - E-mail: coep@prpq.ufmg.br
Horário de atendimento: 09:00 às 11:00 / 14:00 às 16:00

ANEXO A – PARECER CONSUBSTANCIADO

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: CONTROLE DE AUTORIDADE EM REPOSITÓRIOS INSTITUCIONAIS DAS UNIVERSIDADES PÚBLICAS FEDERAIS DE MINAS GERAIS QUE UTILIZAM

Pesquisador: BENILDES COURA MOREIRA DOS SANTOS MACULAN

Área Temática:

Versão: 1

CAAE: 29769320.8.0000.5149

Instituição Proponente: Escola de Ciências da Informação

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.048.320

Apresentação do Projeto:

O projeto apresenta uma pesquisa exploratória que tem por objetivo investigar a gestão de repositórios institucionais que utilizam o software DSpace, nas Instituições de Ensino Superior Federais do Estado de Minas Gerais. O foco é em se identificar como o controle de autoridade tem sido feito nestes repositórios e como o uso de ferramentas disponíveis no software DSpace pode facilitar a gestão das informações relacionadas à produção acadêmica.

Dentre as 11 Universidades Federais em Minas Gerais listadas no site do MEC, através da análise de seus repositórios, identificou-se que 10 utilizam o software de interesse, DSpace. Assim, os pesquisadores pretendem entrar em contato com o gestor (bibliotecário) responsável pelo repositório e convidá-lo para participar da pesquisa através do preenchimento de um questionário sobre o processo de gestão e controle das autoridades no repositório institucional da universidade. O questionário apresenta 9 perguntas abertas a serem preenchidas pelo bibliotecário e será disponibilizado ao participante digitalmente.

Objetivo da Pesquisa:

Objetivo Primário: Mapear o controle de autoridade realizado nos repositórios institucionais das Universidades Públicas Federais de Minas Gerais que utilizam DSpace, tendo em vista demonstrar

Endereço: Av. Presidente Antônio Carlos, 6627 2º Ad SI 2005
Bairro: Unidade Administrativa II **CEP:** 31.270-901
UF: MG **Município:** BELO HORIZONTE
Telefone: (31)3409-4592 **E-mail:** coep@prpq.ufmg.br

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



Continuação do Parecer: 4.048.320

como o uso de ferramentas disponíveis neste software pode facilitar a gestão das informações da produção acadêmica.

Objetivos Secundários:

- 1) Identificar as IES Públicas de Minas Gerais que possuem repositórios institucionais e elencar aquelas que utilizam o software DSpace;
- 2) Elencar as estratégias de controle de autoridade adotadas nas IES Públicas de Minas Gerais;
- 3) Analisar a lista de autores dos repositórios institucionais que utilizam o software DSpace para determinar e descrever as principais incoerências na entrada das autoridades;
- 4) Descrever as ferramentas de controle de autoridade disponíveis na literatura que são compatíveis com o software DSpace, tendo em vista verificar aquelas que facilitam a gestão desses dados.

Avaliação dos Riscos e Benefícios:

Os riscos associados à pesquisa são mínimos: "O participante pode se sentir constrangido ao responder perguntas sobre seu processo de trabalho. Para dirimir esse risco o pesquisador garante o sigilo e o anonimato dos dados coletados, que ficarão sob a responsabilidade do pesquisador durante 5 anos e após esse prazo serão eliminados. Além disso, o participante é informado de que pode desistir de participar da pesquisa a qualquer momento, e que isso não acarretará a ele qualquer tipo de constrangimento, prejuízo ou dano."

Benefícios: Os resultados desta pesquisa serão amplamente divulgados, especialmente disponibilizados para os respondentes, e permitirão aos gestores dos repositórios conhecer as ferramentas de gestão e controle de autoridade disponíveis no DSpace, o que, se adotadas no RI, contribuirão para uma melhor gestão de indicadores de produção e impacto dos pesquisadores das universidades.

Comentários e Considerações sobre a Pesquisa:

O projeto é parte da pesquisa de mestrado sendo realizada no Programa de Pós-graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação (ECI). O financiamento é próprio e prevê gastos com ligações telefônicas e impressão dos questionários, no total de R\$ 22,00.

Endereço: Av. Presidente Antônio Carlos, 6627 2º Ad Sl 2005

Bairro: Unidade Administrativa II **CEP:** 31.270-901

UF: MG **Município:** BELO HORIZONTE

Telefone: (31)3409-4592

E-mail: coep@prpq.ufmg.br

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



Continuação do Parecer: 4.048.320

Considerações sobre os Termos de apresentação obrigatória:

Em relação aos termos entregues:

- 1) Folha de rosto: Devidamente preenchida e assinada.
- 2) Parecer consubstanciado atestando relevância científica e adequação da metodologia, aprovado ad referendum pela Câmara do Departamento de Organização de Tratamento de Informação, ECI.
- 3) Instrumentos de coleta de dados: Ok, apresentam as questões abertas a serem enviadas via questionário.
- 4) Projeto completo: Ok
- 5) TCLE: Apresentam TCLE informando claramente objetivo da pesquisa, riscos (que são mínimos) e benefícios associados, direito à recusa e desistência a qualquer momento, sem prejuízo, metodologia, armazenamento dos dados, e que não há pagamento associado à participação como carta convite, resguardando a confidencialidade dos dados, o anonimato, o direito à recusa, e desistir do projeto a qualquer momento sem qualquer prejuízo.

Recomendações:

Nenhuma recomendação a ser feita.

Conclusões ou Pendências e Lista de Inadequações:

O projeto trata adequadamente as questões éticas envolvidas na pesquisa a ser realizada. Assim, sou, S.M.J. favorável à aprovação do projeto.

Considerações Finais a critério do CEP:

Tendo em vista a legislação vigente (Resolução CNS 466/12), o CEP-UFMG recomenda aos Pesquisadores: comunicar toda e qualquer alteração do projeto e do termo de consentimento via emenda na Plataforma Brasil, informar imediatamente qualquer evento adverso ocorrido durante o desenvolvimento da pesquisa (via documental encaminhada em papel), apresentar na forma de notificação relatórios parciais do andamento do mesmo a cada 06 (seis) meses e ao término da pesquisa encaminhar a este Comitê um sumário dos resultados do projeto (relatório final).

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Endereço: Av. Presidente Antônio Carlos, 6627 2º Ad Sl 2005
Bairro: Unidade Administrativa II **CEP:** 31.270-901
UF: MG **Município:** BELO HORIZONTE
Telefone: (31)3409-4592 **E-mail:** coep@prpq.ufmg.br

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



Continuação do Parecer: 4.048.320

| Tipo Documento | Arquivo | Postagem | Autor | Situação |
|---|---|------------------------|-----------------------------|----------|
| Informações Básicas do Projeto | PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1497974.pdf | 28/02/2020 16:36:43 | | Aceito |
| Folha de Rosto | folha_de_rosto.pdf | 28/02/2020 16:36:09 | ANNA CAROLINA LEITE COTA | Aceito |
| Outros | Questionario.docx | 28/02/2020 16:35:36 | ANNA CAROLINA LEITE COTA | Aceito |
| Cronograma | CRONOGRAMA.docx | 27/02/2020 17:04:56 | ANNA CAROLINA LEITE COTA | Aceito |
| Projeto Detalhado / Brochura Investigador | Projeto.docx | 27/02/2020 17:04:43 | ANNA CAROLINA LEITE COTA | Aceito |
| TCLE / Termos de Assentimento / Justificativa de Ausência | TCLE_corrigido.docx | 27/02/2020 16:37:20 | ANNA CAROLINA LEITE COTA | Aceito |
| Orçamento | ORCAMENTO.docx | 04/02/2020 11:30:36 | ANNA CAROLINA LEITE COTA | Aceito |
| Parecer Anterior | parecer_Ana_Carolina.pdf | 04/02/2020 11:22:34 | ANNA CAROLINA LEITE COTA | Aceito |

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

BELO HORIZONTE, 26 de Maio de 2020

Assinado por:
Críssia Carem Paiva Fontainha
(Coordenador(a))

Endereço: Av. Presidente Antônio Carlos, 6627 2º Ad S/ 2005

Bairro: Unidade Administrativa II **CEP:** 31.270-901

UF: MG **Município:** BELO HORIZONTE

Telefone: (31)3409-4592

E-mail: coep@prpq.ufmg.br