

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

SARAH RÚBIA DE OLIVEIRA SANTOS

**MÉTODO AUTOMATIZADO PARA ANÁLISE DO AUTOARQUIVAMENTO NA CIÊNCIA
DA INFORMAÇÃO: ELISSCRIPT**

BELO HORIZONTE
2021

SARAH RÚBIA DE OLIVEIRA SANTOS

**MÉTODO AUTOMATIZADO PARA ANÁLISE DO AUTOARQUIVAMENTO NA CIÊNCIA
DA INFORMAÇÃO: ELISCRIPIT**

Dissertação apresentada à banca do Programa de Pós-Graduação em Gestão em Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Mestre.

Área de Concentração: Ciência da Informação

Linha de Pesquisa: Arquitetura e Organização do Conhecimento

Orientadora: Prof^a. Dr^a. Dalgiza Andrade Oliveira

BELO HORIZONTE
2021

S237m

Santos, Sarah Rúbia de Oliveira.

Método automatizado para análise do autoarquivamento na Ciência da informação [recurso eletrônico]: ELISCRIP / Sarah Rúbia de Oliveira Santos. - 2021.

1 recurso eletrônico (187 f. : il., color): pdf.

Orientadora: Dalgiza Andrade de Oliveira.
Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 147-158.

Apêndices: 159-183.

Anexos: 184-187

Exigências do sistema: Adobe Acrobat Reader.

1. Ciência da Informação – Teses. 2. Comunicação na ciência – Teses. 3. Divulgação científica – Teses. 4. Bibliometria – Teses. I. Título. II. Oliveira, Dalgiza Andrade de. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 02:004



FOLHA DE APROVAÇÃO

MÉTODO AUTOMATIZADO PARA ANÁLISE DO AUTOARQUIVAMENTO NA
CIÊNCIA DA INFORMAÇÃO: ELISCRIP T

SARAH RÚBIA DE OLIVEIRA SANTOS

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 24 de fevereiro de 2021, todos por videoconferência, pela banca constituída pelos membros:

Prof(a). Dalgiza Andrade Oliveira (Orientadora)
ECI/UFMG

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

Prof(a). Marina Nogueira Ferraz
Faculdade de Medicina/UFMG

Prof(a). Renata Kelly de Arruda
ESP-MG

Belo Horizonte, 24 de fevereiro de 2021.



ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA SARAH RÚBIA DE OLIVEIRA SANTOS

Realizou-se, no dia 24 de fevereiro de 2021, às 14:00 horas, todos por videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *MÉTODO AUTOMATIZADO PARA ANÁLISE DO AUTOARQUIVAMENTO NA CIÊNCIA DA INFORMAÇÃO: ELISSCRIPT*, apresentada por SARAH RÚBIA DE OLIVEIRA SANTOS, número de registro 2019663702, graduada no curso de BIBLIOTECONOMIA, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Dalgiza Andrade Oliveira - ECI/UFMG (Orientadora), Prof(a). Célia da Consolação Dias - ECI/UFMG, Prof(a). Marina Nogueira Ferraz - Faculdade de Medicina/UFMG, Prof(a). Renata Kelly de Arruda - ESP-MG.

A Comissão considerou a dissertação:

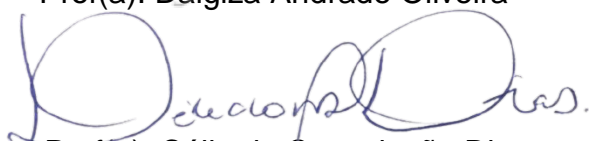
(x) Aprovada

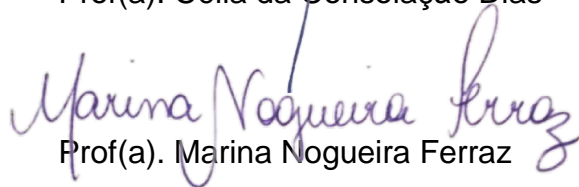
() Reprovada

A Comissão Examinadora deliberou pela recomendação da Dissertação para concorrer à indicação ao prêmio ENANCIB. Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 24 de fevereiro de 2021.


Prof(a). Dalgiza Andrade Oliveira


Prof(a). Célia da Consolação Dias


Prof(a). Marina Nogueira Ferraz


Prof(a). Renata Kelly de Arruda

AGRADECIMENTOS

Agradeço primeiramente a Deus que, em Sua infinita sabedoria, guiou o meu caminho, me fortaleceu e possibilitou mais essa conquista.

À minha família, em especial à minha mãe, Renise, que durante toda minha vida foi meu exemplo, o meu alicerce, quem sempre me incentivou e me deu todas as condições para seguir os meus sonhos. À minha namorada, quase esposa, e companheira de todos momentos, Isadora, pelo apoio incondicional, por acreditar em mim e por sempre me incentivar a ser melhor em todos os aspectos da minha vida. À minha *sister*, Raquel, que se tornou mais suportável à distância e se mostrou uma grande incentivadora em tudo que faço.

Aos meus queridos amigos, Jazz, Eddie, Kerol e Aylla, por estarem ao meu lado há tantos anos, por todo carinho, por me ouvirem quando eu preciso desabafar, pelos conselhos e por sempre acreditarem em mim. Uns amigos desses?!

Aos meus novos e não menos queridos amigos, Celsiane, Marília, Jessica, Bruna e Gabriel, vocês tornaram a mudança de Maceió para Belo Horizonte muito mais agradável, eu me diverti muito com vocês.

À minha orientadora, Dalgiza Andrade Oliveira, pelo apoio, pela confiança, por acreditar nesse projeto e por incentivar a minha independência na jornada para me tornar pesquisadora.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) por fornecer o suporte financeiro para realização desse trabalho, bem como para a minha permanência no mestrado.

À banca, Célia Dias, Marina Ferraz e Renata Arruda, pelas inestimáveis contribuições ao trabalho.

À Universidade Federal de Minas Gerais (UFMG) que, desde antes da minha aprovação no mestrado e ao longo da minha formação, ofereceu um ambiente acolhedor, agradável, motivador e repleto de oportunidades.

Ao Programa de Pós-Graduação em Gestão e Organização do Conhecimento (PPG-GOC) que acolheu a minha proposta.

Aos funcionários do PPG-GOC, em especial à Gildenara e à Sylvia, pela paciência e por me auxiliarem de forma tão atenciosa.

À comunidade Python que me ajudou a superar as dificuldades que encontrei no desenvolvimento deste trabalho.

Por último, mas não menos importante, gratidão a todos os professores que participaram da minha formação até este momento.

“Python tem biblioteca e biblioteca tem que ter Bibliotecário!”

Grupo de usuários Python de São Carlos (Grupy-Sanca)

RESUMO

O autoarquivamento é o depósito da produção científica, realizado pelo autor do trabalho, em um repositório institucional ou temático de acesso aberto. Os repositórios temáticos armazenam e fornecem acesso à literatura de uma disciplina específica e seus assuntos correlatos. O presente estudo propõe uma metodologia automatizada para analisar o autoarquivamento na Biblioteconomia e Ciência da Informação, visando avaliar as contribuições do repositório Eprints in Library and Information Science (E-LIS) para a comunicação científica aberta na área. Para tanto, foi desenvolvido um instrumento em linguagem de programação Python, o ELIScript, que permitiu coletar e analisar a produção científica do repositório. Essa produção foi caracterizada baseando-se na tipologia dos documentos depositados, suas temáticas, data de publicação, data de depósito, idioma, periódicos científicos, eventos científicos e palavras-chave. A pesquisa configurou-se como aplicada, quantitativa, descritiva e exploratória. Com a criação do ELIScript, foi possível coletar os dados de todo o acervo do E-LIS. O universo da pesquisa constituiu-se de 23.245 documentos, subdivididos entre a África (187 documentos); América, dividida em: a) América do Norte e Central (3.345) e b) América do Sul (4.125); Antártica (2); Ásia (2.169); Europa (13.253); e Oceania (164). Os resultados apontam que a produção científica do E-LIS foi depositada entre 2002-2020, e os documentos foram publicados entre 1965 e 2020. Cerca de 47% dos documentos depositados são artigos de periódicos. Os assuntos mais abordados nos documentos são “Uso da Informação e Sociologia da Informação”, “Tecnologia da Informação e Tecnologia de Biblioteca”, e “Tratamento da Informação para Serviços de Informação”. De forma mais específica, os documentos tratam sobre “Disseminação e difusão da informação”, “Repositórios (baseado ou não em acesso aberto)” e “Métodos bibliométricos”. O espanhol aparece como o idioma mais frequente das publicações, seguido do inglês, italiano, alemão e português. Os artigos publicados em periódicos e depositados no E-LIS estão indexados em bases de dados importantes da área, como a LISTA, LISA e a ISTA, e bases generalistas como Web of Science e a Scopus, alguns deles com alto fator de impacto. Verificou-se, ainda, a existência de documentos publicados ou apresentados em eventos científicos regionais, nacionais e internacionais. A partir dos dados, pode-se dizer que o E-LIS alcançou reconhecimento como uma comunidade internacional aberta de Biblioteconomia e Ciência da Informação. A criação de uma metodologia automatizada pode auxiliar no acompanhamento da produção científica depositada no E-LIS, sem que seja necessário recriar métodos para a coleta e análise de dados. Pesquisar sobre o autoarquivamento na Biblioteconomia e Ciência da Informação evidenciou a necessidade de incentivo dessa prática para preservação da memória da área, de forma que colabore como fonte de informação para trabalhos científicos e contribua com o exercício profissional de bibliotecários, arquivistas, museólogos e cientistas da informação. Ter artigos e outros documentos publicados em periódicos de acesso aberto não garante que esses trabalhos serão preservados ao longo do tempo. A melhor alternativa ainda é o depósito em repositórios digitais consoantes com o Movimento de Acesso Aberto.

Palavras-chave: ELIScript. Autoarquivamento. E-prints in Library and Information Science. Movimento de Acesso Aberto. Bibliometria.

ABSTRACT

Self-archiving is the deposit of scientific production, carried out by the author of the work, in an institutional or subject repository with open access. Subject repositories store and provide access to the literature of a specific discipline and its related subjects. This study proposes an automated methodology to analyze self-archiving in Library and Information Science, aiming to evaluate the contributions of the Eprints in Library and Information Science (E-LIS) repository for open scientific communication in the field. For such, an instrument in Python programming language, ELIScript, was developed, which allowed to collect and analyze the scientific production of the repository. This production was characterized based on the type of documents deposited, their subjects, date of publication, date of deposit, language, scientific journals, scientific events and keywords. The research was configured as applied, quantitative, descriptive and exploratory. With the creation of ELIScript it was possible to collect data from the entire E-LIS collection. The research sample consists of 23,245 documents, subdivided between Africa (187 documents); America, divided into: a) North and Central America (3,345); and b) South America (4,125); Antarctica (2); Asia (2,169); Europe (13,253); and Oceania (164). The results show that the scientific production of E-LIS was deposited between 2002 and 2020, and the documents were published between 1965 and 2020. About 47% of the documents deposited are journal articles. The subjects most covered in the documents are "Information Use and Sociology of Information", "Information Technology and Library Technology", and "Information Treatment for Information Services". More specifically, the documents deal with "Information Dissemination and Diffusion", "Repositories" and "Bibliometric Methods". Spanish appears as the most frequent language of publications, followed by English, Italian, German and Portuguese. The articles published in journals and deposited in E-LIS are indexed in important databases in the field, such as LISTA, LISA and ISTA, and general databases such as Web of Science and Scopus, some of them with a high impact factor. It was also verified the existence of documents published or presented at regional, national and international scientific events. From the data, it can be said that E-LIS has achieved recognition as an open international library and information science community. The creation of an automated methodology can assist in monitoring the scientific production deposited in E-LIS, without having to recreate methods for data collection and analysis. Researching self-archiving in Librarianship and Information Science showed the need to encourage this practice to preserve the field's memory, so that it serves as a source of information for scientific works and benefits the professional practice of librarians, archivists, museologists and information scientists. Having articles and other documents published in open access journals does not guarantee that these works will be preserved over time. The best alternative is still depositing in digital repositories according to the Open Access Movement.

Keywords: ELIScript. Self-archiving. E-Prints in the Library and Information Science. Open Access Movement. Bibliometrics.

LISTA DE FIGURAS

Figura 1 – Diagrama do processo de publicação em acesso aberto.....	20
Figura 2 – Gradação entre os sistemas de comunicação.....	35
Figura 3 – Cinco escolas de pensamento da Ciência Aberta.....	41
Figura 4 – Taxonomia da Ciência Aberta.....	44
Figura 5 – Repositório E-LIS.....	55
Figura 6 – Infraestrutura do protocolo OAI-PMH.....	56
Figure 7 – Parte do formulário de submissão de documentos do E-LIS.....	57
Figura 8 – Delineamento da pesquisa.....	64
Figura 9 – Página inicial do repositório E-LIS.....	68
Figura 10 – Quantidade de documentos distribuídos por país.....	69
Figura 11 – Busca no E-LIS por país.....	70
Figura 12 – Formatos para exportação dos metadados das publicações.....	70
Figura 13 – Ambiente computacional Jupyter Notebook.....	72
Figura 14 – Etapas para preparação do ambiente computacional.....	75
Figura 15 – Framework para tratamento e visualização de dados.....	79
Figura 16 – Página de metadados de uma publicação.....	80
Figura 17 – Página de busca com a lista de continentes e países.....	81
Figura 18 – Estrutura de uma função em Python.....	84
Figura 19 – Exemplo de uma função em Python.....	84
Figura 20 – Nuvem de palavras-chave.....	137

LISTA DE GRÁFICOS

Gráfico 1 – Comparativo entre os tipos de documentos por continente.....	104
Gráfico 2 – Quantidade de depósitos por ano.....	105
Gráfico 3 – Comparação logarítmica da quantidade de depósitos por ano.....	107
Gráfico 4 – Ano de publicação dos documentos depositados.....	108
Gráfico 5 – Comparação logarítmica do ano de publicação dos documentos depositados	109
Gráfico 6 – Idiomas das publicações.....	119
Gráfico 7 – Tipologia de documentos depositados na África.....	159
Gráfico 8 – Tipologia de documentos depositados na América do Norte e Central.....	159
Gráfico 9 – Tipologia de documentos depositados na América do Sul.....	160
Gráfico 10 – Tipologia de documentos depositados na Antártica.....	160
Gráfico 11 – Tipologia de documentos depositados na Ásia.....	161
Gráfico 12 – Tipologia de documentos depositados na Europa.....	161
Gráfico 13 – Tipologia de documentos depositados na Oceania.....	162
Gráfico 14 – Porcentagem geral das tipologias de documentos depositados no E-LIS.....	162
Gráfico 15 – Distribuição por ano de depósitos realizados na África.....	163
Gráfico 16 – Distribuição por ano de depósitos realizados na América do Norte e Central	163
Gráfico 17 – Distribuição por ano de depósitos realizados na América do Sul.....	164
Gráfico 18 – Distribuição por ano de depósitos realizados na Antártica.....	164
Gráfico 19 – Distribuição por ano de depósitos realizados na Ásia.....	165
Gráfico 20 – Distribuição por ano de depósitos realizados na Europa.....	165
Gráfico 21 – Distribuição por ano de depósitos realizados na Oceania.....	166
Gráfico 22 – Distribuição por ano de publicação na África.....	167
Gráfico 23 – Distribuição por ano de publicação na América do Norte e Central.....	167
Gráfico 24 – Distribuição por ano de publicação na América do Sul.....	168
Gráfico 25 – Distribuição por ano de publicação na Antártica.....	168
Gráfico 26 – Distribuição por ano de publicação na Ásia.....	169
Gráfico 27 – Distribuição por ano de publicação na Europa.....	169
Gráfico 28 – Distribuição por ano de publicação na Oceania.....	170
Gráfico 29 – Principais idiomas de publicação da África.....	177
Gráfico 30 – Principais idiomas de publicação da América do Norte e Central.....	177
Gráfico 31 – Principais idiomas de publicação da América do Sul.....	178
Gráfico 32 – Principais idiomas de publicação da Antártica.....	178
Gráfico 33 – Principais idiomas de publicação da Ásia.....	179
Gráfico 34 – Principais idiomas de publicação da Europa.....	179

Gráfico 35 – Principais idiomas de publicação da Oceania.....	180
Gráfico 36 – Periódicos com mais documentos depositados na África.....	181
Gráfico 37 – Periódicos com mais documentos depositados na América do Norte e Central	181
Gráfico 38 – Periódicos com mais documentos depositados na América do Sul.....	182
Gráfico 39 – Periódicos com mais documentos depositados na Antártica.....	182
Gráfico 40 – Periódicos com mais documentos depositados na Ásia.....	182
Gráfico 41 – Periódicos com mais documentos depositados na Europa.....	183
Gráfico 42 – Periódicos com mais documentos depositados na Oceania.....	183

LISTA DE QUADROS

Quadro 1 – Resultados iniciais das buscas nas bases de dados selecionadas.....	65
Quadro 2 – Resultado com a expressão de busca combinada.....	66
Quadro 3 – Especificações técnicas das aplicações computacionais.....	76
Quadro 4 – Importação de Bibliotecas e Módulos em Python.....	77
Quadro 5 – Categorias de análise.....	85
Quadro 6 – Tipologias de documentos e orientações para sua submissão.....	98
Quadro 7 – Comparativo de tipos de documentos depositados até 2007 e até 2020.....	100
Quadro 8 – Assuntos dos documentos depositados no E-LIS.....	110
Quadro 9 – Clusters temáticos do E-LIS.....	113
Quadro 10 – Assuntos específicos dos documentos divididos por continente.....	114
Quadro 11 – Distribuição de Periódicos com mais publicações depositadas.....	121
Quadro 12 – Distribuição de eventos com mais publicações depositadas.....	133
Quadro 13 – Quantidade de documentos depositados no E-LIS por assunto e continente	171

LISTA DE TABELAS

Tabela 1 – Tipologia de documento por continente.....	87
Tabela 2 – Tipologia de documentos com total e porcentagem.....	88
Tabela 3 – Depósitos por ano dividido por continente.....	90
Tabela 4 – Ano de publicação dividido por continente.....	91
Tabela 5 – Idiomas dos documentos depositados no E-LIS.....	94

LISTA DE ABREVIATURAS E SIGLAS

AMBAC	Asociación Mexicana de Bibliotecarios A. C.
APC	<i>Article processing charge</i>
ASIST	American Society for Information Science and Technology
BBB	Declarações de Budapeste, Bethesda e Berlim
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BOAI	<i>Budapest Open Access Initiative</i>
BRAPCI	Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação
CAB	Sistema de Biblioteca da Universidade de Nápoles Federico II
Capex	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CIBDA	Congreso Internacional de Bibliotecología, Documentación y Archivística
CSV	<i>Comma-Separated Value</i>
DLIST	<i>Digital Library of Information Science and Technology</i>
DOAJ	Diretório de Periódicos de Acesso Aberto
DOI	<i>Digital Object Identifier</i>
E-LIS	e-Prints in Library and Information Science
ERA	<i>Educational Research Abstracts</i>
ERIC	<i>Education Resources Information Center</i>
FAIR	<i>Findable, Accessible, Interoperable and Reusable</i>
FORPED	Fórum Discente do Programa de Pós-Graduação em Gestão e Organização do Conhecimento
FOSTER	<i>Facilitate Open Science Training for European Research</i>
HTML	<i>HyperText Markup Language</i>

IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ICSU	Conselho Internacional para a Ciência
IES	Instituição de Ensino Superior
INSDOC	Indian National Scientific Documentation Centre
ISI	Institute for Scientific Information (Thomson Scientific)
ISO	International Organization for Standardization
ISTA	Information Science and Technology Abstracts
JCR	Journal Citation Reports
LISA	Library and Information Science Abstracts
LISTA	Library, Information Science & Technology Abstracts with Full Text
LTS	<i>Long Term Support</i>
MIAR	<i>Information Matrix for the Analysis of Journals</i>
MLA	Modern Language Association
MOOC	<i>Massively Open Online Course</i>
NLM	National Library of Medicine
OAI	<i>Open Archives Initiative</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
ORCID	<i>Open Researcher and Contributor ID</i>
Pandas	<i>Python Data Analysis Library</i>
PDF	<i>Portable Document Format</i>
PEPs	<i>Index of Python Enhancement Proposals</i>
PKP	<i>Public Knowledge Project</i>
RCLIS	<i>Research in Computing, Library and Information Science</i>
RDF	<i>Resource Description Framework</i>
ROAD	<i>Directory of Open Access Scholarly Resources</i>

SCECSAL	Standing Conference of Eastern, Central and Southern Africa Library and Information Associations
SciELO	<i>Scientific Eletronic Library Online</i>
SJR	SCImago Journal Rank
TIC	Tecnologias de Informação e Comunicação
UFAL	Universidade Federal de Alagoas
UFBA	Universidade Federal da Bahia
UFMG	Universidade Federal de Minas Gerais
UFOP	Universidade Federal de Ouro Preto
UFRGS	Universidade Federal do Rio Grande do Sul
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
WoS	Web of Science

SUMÁRIO

1 INTRODUÇÃO	19
1.1 Problema.....	21
1.2 Questão de pesquisa.....	24
1.3 Objetivos.....	25
1.3.1 <i>Objetivo geral</i>	25
1.3.2 <i>Objetivos específicos</i>	25
1.4 Justificativa.....	25
2 COMUNICAÇÃO CIENTÍFICA E SUA EVOLUÇÃO	31
2.1 Canais científicos de comunicação.....	35
2.2 Produtos científicos.....	37
3 CIÊNCIA ABERTA	40
3.1 Cinco escolas de pensamento da Ciência Aberta.....	40
3.2 Ciência Aberta: mais do que um movimento.....	43
3.3 Movimento de Acesso Aberto.....	47
3.3.1 <i>Via Verde: o autoarquivamento em repositórios digitais</i>	49
3.3.2 <i>E-prints in Library and Information Science (E-LIS)</i>	54
4 ESTUDOS MÉTRICOS DA INFORMAÇÃO	61
5 PERCURSO METODOLÓGICO	64
5.1 Estratégia de busca para revisão de literatura.....	64
5.2 Caracterização da pesquisa.....	66
5.3 Ambiente de pesquisa.....	67
5.4 Universo da pesquisa.....	68
5.5 Instrumento de coleta de dados.....	69
5.6 Instrumento para análise e discussão dos dados.....	71
6 ELISCRIP: ANÁLISE AUTOMATIZADA DO AUTOARQUIVAMENTO NA CIÊNCIA DA INFORMAÇÃO	75
6.1 Preparação do ambiente computacional.....	75
6.2 Extração, tratamento e visualização de dados.....	79
6.3 Extração de dados em páginas da web.....	79
6.4 Limpeza dos dados.....	82
6.5 Análise e visualização dos dados.....	83
7 ANÁLISE E DISCUSSÃO DOS RESULTADOS	98
7.1 Distribuição por tipologia de documentos.....	98

7.2 Distribuição temporal de depósito dos documentos.....	104
7.3 Distribuição temporal de publicação dos documentos.....	107
7.4 Temática dos documentos depositados.....	110
7.5 Distribuição por idioma.....	118
7.6 Periódicos com publicações depositadas.....	120
7.7 Eventos com publicações depositadas.....	132
7.8 Palavras-chave.....	136
8 CONSIDERAÇÕES FINAIS.....	140
REFERÊNCIAS.....	147
APÊNDICE A – Tipologia dos documentos depositados no E-LIS dividido por continente.....	159
APÊNDICE B – Distribuição por ano de depósito dividido por continente.....	163
APÊNDICE C – Distribuição por ano de publicação dividido por continente.....	167
APÊNDICE D – Quantidade de documentos depositados no E-LIS dividido por assunto e continente.....	171
APÊNDICE E – Idioma das publicações por continente.....	177
APÊNDICE F – Periódicos com mais publicações depositadas.....	181
ANEXO A – Esquema de classificação JITA, versão em português – Brasil.....	184

1 INTRODUÇÃO

A comunicação científica é o conjunto de processos pelo qual os resultados de investigações científicas são tornados disponíveis. Esse conjunto perpassa todas as fases da elaboração de um trabalho científico, desde seu desenvolvimento até a disseminação da pesquisa. As formas de divulgação científica se alteraram ao longo do tempo. De acordo com Meadows (1999), os primórdios da comunicação da ciência, que tiveram impacto na comunicação científica moderna, remontam à Grécia Antiga, nos séculos V e IV a.C. Assim, a comunicação entre os cientistas evoluiu, desde as cartas utilizadas no passado até os periódicos científicos eletrônicos disponíveis na atualidade.

Os periódicos científicos têm sido considerados o meio “oficial” para a divulgação da ciência, constituindo-se como “um dos veículos que compõem o produto final, formal, consolidado da disseminação de resultados de pesquisas realizadas por estudiosos de todo o mundo” (BAPTISTA *et al.*, 2007, p. 3). No entanto, os mesmos autores apontam que, na comunidade científica em todo o mundo, tem sido discutido e questionado o modelo de negócios dos periódicos científicos em vigor, que em consequência aos altos preços das assinaturas têm limitado o acesso ao conhecimento.

Nesse contexto, um movimento crescente desde os anos 1990 buscava contornar essas cobranças e viabilizar o acesso ao conhecimento científico. Costa e Leite (2017, p. 2) apontam dois elementos que impulsionaram esse movimento. O primeiro está relacionado com o descontentamento por parte dos pesquisadores com o modelo tradicional de comunicação científica, “que impõe barreiras para o acesso daquilo que eles mesmos e seus pares produziram”. O segundo elemento corresponde ao desenvolvimento das tecnologias de informação, que possibilitaram agilizar e dinamizar os processos de comunicação científica, favorecendo o surgimento das iniciativas de acesso aberto.

O Movimento de Acesso Aberto surge, então, como resposta à limitação imposta. O Movimento de Acesso Aberto é uma iniciativa que propõe a disponibilização on-line¹ da produção científica gratuitamente, livre e sem restrições. O Acesso Aberto pode ser concretizado de duas formas, segundo Harnad *et al.* (2004): a) a partir da publicação em um periódico de acesso aberto (Via Dourada); e b) por meio do autoarquivamento realizado pelo autor (Via Verde).

Na Via Dourada, o pesquisador pode selecionar uma revista que já realize a publicação no modelo aberto – essas podem ser localizadas no Diretório de Periódicos de Acesso Aberto (DOAJ) – e publicar seu artigo nela. O periódico pode, também, ter um modelo híbrido de publicação, isto quer dizer que uma taxa de processamento de artigo (APC) é cobrada ao autor

1 Optou-se por utilizar a grafia “on-line” (e outros estrangeirismos como: web, download, script, copyright, e-mail, site, link) sem itálico ao longo do texto por ser um termo estrangeiro já incorporado à língua portuguesa, de acordo com a indicação do Vocabulário Ortográfico da Língua Portuguesa (VOLP) da Academia Brasileira de Letras e o Manual de Comunicação do Senado Federal.

para que seu artigo seja disponibilizado em acesso aberto. No periódico de acesso aberto, o artigo estará disponível nessa modalidade de acesso imediatamente.

Outra opção que o pesquisador tem é autoarquivar sua publicação, a Via Verde. Para isso, após publicar em uma revista de acesso aberto ou fechado (com pagamento de assinatura), o autor deve verificar os direitos que ele possui sobre sua publicação. O Sherpa/RoMEO é uma fonte em que essa informação pode ser consultada. Lá é possível checar as políticas da revista na qual o artigo foi publicado e se elas permitem o depósito desse trabalho em um repositório. Se for permitido, o artigo deve ser depositado em um repositório com arquitetura adequada para armazenar e preservar essa publicação. O acesso aberto pela Via Verde pode ser imediato ou tardio, dependendo da política da revista e se ela impõe um período de embargo para que o documento possa ser arquivado. Na Figura 1, o processo descrito é exemplificado:

Figura 1 – Diagrama do processo de publicação em acesso aberto



Fonte: Elaborado pela autora (2019).

O autoarquivamento é um processo consciente, realizado pelo autor, e pode ser feito em repositórios digitais institucionais e/ou temáticos. Os repositórios temáticos, foco do presente trabalho, são aqueles que armazenam e fornecem acesso à literatura de uma disciplina específica ou de uma disciplina e seus assuntos correlatos. De acordo com Adamick e Reznik-Zellen (2011), esses repositórios são sub-representados na literatura de Ciência da Informação, na medida em que, apesar de serem citados como iniciativas bem-sucedidas em comunicação científica, principalmente em relação aos repositórios institucionais, são menos abordados na literatura.

As autoras apontam que a literatura envolvendo repositórios temáticos consiste, geralmente, em análise do comportamento de pesquisadores (incluindo análises de citação), do

impacto de publicações em Acesso Aberto e de revisões de literatura, citando como exemplo os repositórios: arXiv, dos campos de Física, Ciências, Tecnologia, Engenharia e Matemática; o CiteSeerX, de Ciência da Computação; o PubMed Central, das Ciências da Vida e Biomédicas; e o E-LIS, de Biblioteconomia e Ciência da Informação, que serão melhor detalhados mais adiante².

Tais repositórios se caracterizam por fornecer um espaço seguro para armazenamento e disseminação de recursos informacionais em formato eletrônico. Além disso, alguns disponibilizam outros serviços como a atribuição de identificador digital, como o *Digital Object Identifier* (DOI) e o *Uniform Resource Identifier* (URI), que torna o documento localizável na web e citável; outros possibilitam a interação entre o depositante e os leitores, com caixas de comentários e espaços para críticas aos documentos. De tal modo, entende-se que estes espaços beneficiam tanto o pesquisador quanto as instituições ou sociedades científicas, pois promovem maior visibilidade aos resultados de pesquisas e asseguram a preservação da memória científica.

Le Coadic (2004), ao listar algumas fontes de informação na área de Ciência da Informação, aponta o E-Prints in Library and Information Science (E-LIS) como sendo um dos principais recursos abertos em Ciência da Informação e Documentação na Internet. Ele é um repositório digital de documentos técnico-científicos para as disciplinas relacionadas à Ciência da Informação. O E-LIS é o primeiro servidor eletrônico especializado em Biblioteconomia e Ciência da Informação (SANTILLÁN-ALDANA, 2009), e, atualmente, está estabelecido como o maior repositório aberto da área. Seu desenvolvimento foi estimulado pelo conceito de Acesso Aberto e funciona por meio do autoarquivamento realizado pelos autores.

1.1 Problema

Na propalada Sociedade do Conhecimento, “o sistema de produção de conhecimento está se tornando mais amplamente distribuído por meio de uma série de novos lugares e atores” (DAVID; FORAY, 2005, p. 23, *tradução nossa*³). Entende-se que a Internet e suas possibilidades, principalmente para os meios de comunicação, têm propiciado um espaço frutífero para a disseminação do conhecimento científico, uma vez que modificaram o processo, fluxo e produção de informações científicas.

O Movimento de Acesso Aberto manifesta-se, nesse cenário, como uma iniciativa que reconhece o princípio de que o conhecimento científico e acadêmico é um bem público global.

2 Na seção **3.2.1 Via Verde: o autoarquivamento em repositórios digitais**, os repositórios serão descritos.

3 “[T]he knowledge production system is becoming more widely distributed across a host of new places and actors” (DAVID; FORAY, 2005, p. 23)

Nos últimos 20 anos, o acesso aberto tem ganhado destaque e importância no âmbito das publicações científicas e de sua divulgação. Porém, este tem sido igualado apenas à publicação de artigos em periódicos de acesso aberto, negligenciando o método de autoarquivamento em repositórios digitais (HARNAD *et al.*, 2004).

Há, ainda, um entendimento equivocado sobre os documentos que são depositados em repositórios digitais, isto é, de que os documentos disponibilizados nesses repositórios não passaram pela revisão de pares ou que não possuem rigor científico. Entretanto, muitas editoras científicas, inclusive de periódicos que não estão em acesso aberto, já adicionam às suas políticas de publicação a possibilidade de que o autor deposite seu trabalho (na versão *preprint* [manuscrito pré-publicação], *postprints* [manuscrito pós-revisão de pares] ou até a versão do editor [publicada no periódico]) em repositórios digitais.

De acordo com Weitzel (2006, p. 66), ao abordar os repositórios temáticos, “a comunidade científica ainda não está suficientemente esclarecida de suas vantagens, sua função e da importância do seu engajamento para que essas iniciativas ocupem seu espaço”. Ainda assim, devido ao tempo que o E-LIS está no ar e a quantidade de documentos que foi depositada nele, entende-se que há uso e demanda por espaços como este. Dessa forma, faz-se necessário compreender do que é constituído esse repositório: quais tipos de documento são depositados, sobre quais assuntos, de que período (atualidade dos documentos), em qual idioma, para quais públicos, entre outros pontos. Em suma, que tipo de material depositam e que tipo de conteúdo está sendo disseminado.

O primeiro repositório temático foi criado em 1991. Conhecido como arXiv, o repositório que foi dedicado ao campo da física, inicialmente, conta com mais de um milhão e meio de publicações também nos campos da matemática, ciência da computação, biologia quantitativa, estatística e outros. O sucesso de repositórios temáticos como esse e a relativa ausência de trabalhos sobre repositórios temáticos (ADAMICK; REZNIK-ZELLEN, 2007; BERNIUS, 2010; BJÖRK, 2014), mais especificamente sobre o E-LIS, suscitaram o interesse em analisar as características do autoarquivamento nesse repositório e as consequências dessa atividade para a comunicação científica aberta em Ciência da Informação. Visto que mais de 23 mil publicações foram depositadas, num esforço colaborativo, no decorrer de 18 anos, e essas publicações acumulam mais de 21 milhões de downloads, supõe-se que esse material tenha alguma representação na área de Ciência da Informação.

Em consonância com o que Weitzel (2006, p. 62) aponta, é necessário “incentivar o maior número de depósitos da produção científica [em repositórios] para estimular a disseminação e uso desta produção, [...] como forma de espelhar a memória institucional ou de uma área”. Contudo, compreende-se que, para que esse incentivo seja feito, é preciso que esses espaços sejam estudados e suas contribuições, vantagens e desvantagens sejam apresentadas. O presente trabalho busca contribuir, nesse sentido, para os debates sobre o Movimento de

Acesso Aberto, por meio da discussão sobre o que é e como funciona a estratégia de autoarquivamento em repositórios digitais, bem como, averiguando quais as contribuições do E-LIS como um repositório internacional aberto de Biblioteconomia e Ciência da Informação.

Para isso, deve-se considerar abordagens para a análise do acervo disponível no repositório. Surge a oportunidade para pensar no emprego de ferramentas ou instrumentos que possam ser utilizados e reutilizados, propiciando análises periódicas do acervo, como as linguagens de programação. A aplicação de linguagens de programação, como a Python, tem se tornado cada vez mais popular. Com o avanço da Ciência Aberta, os pesquisadores estão visualizando oportunidades de compartilhar seus métodos, e os códigos provenientes deles, com toda a comunidade acadêmica (WHITE; POWELL, 2019), possibilitando a realização de uma ciência mais transparente, acessível e reprodutível.

Para exemplificar a existência de pesquisas que já utilizam a programação, realizou-se uma busca simples pelo termo “python” na base de dados Web of Science (WoS). Lá, foram localizados 73 trabalhos que citam a utilização da linguagem dentro da área de Biblioteconomia e Ciência da Informação. Buscando em um periódico específico dessa área que tem por foco os estudos métricos da informação, o *Scientometrics*, 94 resultados foram obtidos. As análises utilizando essa ferramenta são das mais diversificadas, sendo aplicadas em estudos na área de geografia, ciências biológicas e da saúde, na biblioteconomia, podendo ser utilizadas para criação de ontologias e nas próprias análises bibliométricas.

O que não se encontra com facilidade, entretanto, é o passo a passo dessa metodologia, com descrição do papel da linguagem de programação na análise. Há a possibilidade do espaço do artigo não ser suficiente ou mesmo de não ser o propósito do trabalho apresentar o código-fonte derivado do processo de programação e isto é compreensível. Contudo, para que um trabalho seja reproduzido por outros compreende-se como necessário, pelo menos, a disponibilização do código gerado e sua documentação, como fizeram, por exemplo, Ruiz-Rosero, Ramirez-Gonzalez e Viveros-Delgado (2019). Apesar de não entrar em detalhes do código no trabalho, os autores disponibilizaram o *Uniform Resource Locator* (URL) do repositório público no qual está hospedado o código completo utilizado no artigo, um manual de instruções e exemplos de comandos. Dessa forma, outros pesquisadores interessados podem aprender com esse código, avaliá-lo, testá-lo e, até mesmo, utilizá-lo em uma nova pesquisa.

Além dos pesquisadores que empregam a programação em seus processos de coleta e análise de dados, já existe uma preocupação em capacitar bibliotecários para programar. Os serviços de dados não são novos em bibliotecas acadêmicas e os bibliotecários têm lidado com dados a todo momento na *World Wide Web*, sendo um ponto de expertise desses profissionais. No entanto, o ambiente digital tem evoluído e se tornado mais amplo, novos formatos estão surgindo, novos métodos estão sendo empregados e os pesquisadores estão mudando seus hábitos em concordância com os preceitos da Ciência Aberta, o que sugere que os bibliotecários

devem se adaptar e adquirir novas habilidades para acompanhar essas mudanças, sendo a programação uma dessas habilidades. Obviamente, os bibliotecários não precisam se tornar programadores, mas entende-se que é preciso que estes compreendam conceitos básicos e acompanhem essas mudanças no processo da comunicação científica. White e Powell (2019, p. 46, *tradução nossa*⁴) fazem uma análise desse ambiente de dados e a necessidade de aquisição de novas habilidades por parte dos bibliotecários. Para eles:

a capacidade de trabalhar e fornecer acesso a dados “grandes” [big data] e abertos é importante para a biblioteconomia da ciência de dados, mas esse tipo de informação geralmente existe fora do ecossistema típico de bibliotecas. Os tecnólogos da informação que trabalham em configurações que não são de biblioteca frequentemente compartilham dados por meio de serviços da web ou interfaces de programação de aplicativos (APIs) que atendem aos consumidores de dados com habilidades básicas de programação de computadores. [...] Quando pesquisadores e estudantes de ciência de dados recorrem à biblioteca para obter assistência na aquisição, compreensão e trabalho com grandes conjuntos de dados em formatos desconhecidos, é necessário que os bibliotecários, por sua vez, obtenham as habilidades necessárias para fornecer serviço de referência e orientação técnica desejada pelos usuários. Na era digital, isso geralmente exige a criação de códigos de programa de computador.

Outras áreas do conhecimento já aplicam esses aparatos de forma mais sofisticada para realizar o mesmo trabalho que cientistas da informação estão desenvolvendo de forma mais lenta. Atualmente, existem muitas ferramentas gratuitas e disponíveis na Internet, com manuais e instruções que podem facilitar esse aprendizado, permitindo aos bibliotecários, cientistas da informação e outros profissionais da informação adquirirem novas competências.

Diante dessas informações, propõe-se a investigação dos espaços de comunicação científica – os repositórios digitais temáticos de acesso aberto –, realizando-se uma análise sobre o que está contido no E-LIS a partir da proposta de uma metodologia automatizada utilizando a linguagem de programação Python, de forma a verificar o que esse repositório oferece em termos de conteúdo e como contribui para a comunidade científica, para os profissionais da área de Ciência da Informação e para a sociedade em geral.

1.2 Questão de pesquisa

4 *The ability to work with and provide access to “big” and open data are important to data science librarianship, but these types of information often exist outside of the typical library ecosystem. Information technologists working in non-library settings frequently share data via web services or application programming interfaces (APIs) that cater to data consumers with basic computer programming abilities. [...] When data science researchers and students turn to the library for assistance in acquiring, understanding, and working with large datasets in unfamiliar formats, it is necessary that librarians, in turn, attain the skills required to provide reference service and technical guidance desired by patrons. In this digital age, that often required to writing computer program code. (WHITE; POWELL, 2019, p. 46)*

Como avaliar as contribuições do repositório E-LIS para a comunicação científica aberta em Ciência da Informação utilizando instrumentos automatizados para coleta e análise de dados?

1.3 Objetivos

Este estudo propõe atender ao objetivos, geral e específicos, apresentados a seguir.

1.3.1 Objetivo geral

Propor uma metodologia automatizada para análise de indicadores bibliométricos utilizando a linguagem de programação Python, de forma a avaliar as contribuições do E-LIS para a comunicação científica aberta em Ciência da Informação a partir dos documentos depositados no repositório.

1.3.2 Objetivos específicos

- a) Desenvolver o ELIScript, instrumento para coleta e análise de dados em linguagem de programação Python;
- b) Analisar, por meio de indicadores bibliométricos, a tipologia dos documentos depositados, assuntos, data de publicação, data de depósito, idioma, periódicos científicos com mais publicações, eventos científicos e palavras-chave;
- c) Realizar uma análise geopolítica comparativa, a partir dos documentos depositados no repositório;
- d) Identificar, caracterizar e analisar as práticas de autoarquivamento dos pesquisadores da Ciência da Informação no E-LIS.

1.4 Justificativa

A filosofia do Movimento de Acesso Aberto, que preconiza o acesso amplo e gratuito ao conhecimento científico, fornece grandes esperanças frente ao modelo de negócios científico em vigor atualmente. Na comunidade científica, principalmente nos espaços em que os pesquisadores lidam de forma mais direta com os entraves da publicação em periódicos científicos pagos e com a cobrança para acesso a publicações, nota-se um interesse em buscar maneiras de possibilitar o compartilhamento do saber produzido com recursos públicos, o que considera-se um dever dos pesquisadores e um direito da sociedade.

Conforme Khabza e Giles (2014), por meio das duas estratégias do Movimento de Acesso Aberto – Via Dourada e Via Verde –, cerca de 24% de todos os documentos acadêmicos

disponibilizados na web estão em acesso aberto. No entanto, Tennant *et al.* (2016, p. 4, *tradução nossa*⁵) afirmam que um problema central permanece:

o acesso universal ou mesmo marginal a aproximadamente 75% dos artigos não é diretamente possível a menos que o indivíduo esteja em uma posição privilegiada de trabalhar em um instituto que tenha acesso à assinatura desses artigos ou tenha dinheiro suficiente para pagar em uma base por artigo (dado que alguns periódicos fornecem esse recurso; outros não). A assinatura de todos os periódicos revisados por pares não é acessível para nenhum indivíduo, instituto de pesquisa ou universidade.

No Brasil, por exemplo, se um estudante quiser ter acesso a artigos de uma determinada área, é possível supor que ele conseguirá sem nenhum custo, uma vez que muitos periódicos brasileiros já foram criados em acesso aberto. Contudo, se o aluno necessita buscar materiais internacionais, em revistas com alto fator de impacto, talvez seja necessário que ele despenda algum dinheiro para conseguir o que deseja. Isso porque muitas editoras científicas⁶, com periódicos de renome, cobram pela sua assinatura, pelo acesso aos seus artigos, e, às vezes, até pelos custos para que se publique nelas.

Com sorte, esse estudante faz parte de uma Instituição de Ensino Superior (IES) que tem convênio com o Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e pode ter acesso a muitos artigos – mas não todos – de forma gratuita. E se esse indivíduo não é um aluno, e se ele não possui vínculo com alguma IES? Entende-se que uma pessoa não deveria ter de se justificar de nenhuma forma para ter acesso ao conhecimento científico, principalmente aquele criado por universidades públicas com recursos públicos. Importa destacar que essa problemática não afeta apenas o indivíduo, o mesmo ocorre em nível institucional. Bibliotecas que se constituem como intermediárias entre a informação especializada e os usuários não têm como arcar com as despesas para fornecer esse acesso (TENNANT *et al.*, 2016).

5 “[U]niversal or even marginal access to approximately 75% of articles is not directly possible unless one either is in a privileged position to work at an institute that has subscription access to these articles, or has enough money to pay on a per-article basis (given that journals provide this feature; some do not). Subscriptions to all peer-reviewed journals is not affordable for any single individual, research institute or university” (TENNANT *et al.*, 2016, p. 4)

6 Algumas editoras que cobram pelo acesso, e que são consideradas um oligopólio de editoras por Larivière, Haustein e Mongeon (2015), são *Reed-Elsevier*, *Wiley-Blackwell*, *Springer*, *Wolters Kluwer* e *Taylor & Francis*.

De acordo com a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) e o Conselho Internacional para a Ciência (ICSU) (UNESCO/ICSU⁷, 1999, on-line, *tradução nossa*⁸),

o acesso igualitário à ciência não é apenas um requisito social e ético para o desenvolvimento humano, mas também uma necessidade para realizar o potencial das comunidades científicas em todo o mundo e orientar o progresso científico para satisfazer as necessidades da humanidade.

Entende-se, então, que o Acesso Aberto traz vantagens para os cientistas, as bibliotecas e as instituições públicas ao disponibilizar o conhecimento científico de maneira livre e gratuita, o que despertou o interesse em desenvolver esse trabalho: uma investigação acerca de repositórios digitais em acesso aberto na Ciência da Informação. Tal propósito envolve não só o estudo do repositório E-LIS, em si, mas também o entendimento sobre comportamento dos pesquisadores frente a esse espaço por meio da coleção de materiais disponíveis nele.

Os repositórios digitais de acesso aberto oferecem um modelo de organização do conhecimento que propõe facilitar o acesso, a disseminação e a preservação do conhecimento para promover o desenvolvimento científico (COSTA; LEITE, 2017). Por isso, o autoarquivamento em repositórios digitais é uma estratégia vista de forma positiva no Movimento de Acesso Aberto, pois descentraliza a posse da produção científica, tirando-a das grandes editoras e colocando-a nas mãos daqueles que a produziram em primeiro lugar.

Poucas universidades públicas brasileiras possuem políticas de depósito das monografias, dissertações e teses em seus Repositórios Institucionais – a título de exemplo, pode-se citar a Universidade Federal de Minas Gerais (UFMG), a Universidade Federal de Alagoas (UFAL), a Universidade Federal da Bahia (UFBA), Universidade Federal de Ouro Preto (UFOP), Universidade Federal do Rio Grande do Sul (UFRGS). De acordo com Weitzel (2019), ainda é necessário estimular estudos que contribuam para a criação de políticas nacionais para depósito em repositórios, tanto institucionais como temáticos. Além dos documentos originados das atividades de cursos de pós-graduação, existem outras produções científicas eletrônicas que podem ser arquivadas, como os artigos de periódicos, artigos de conferências, capítulos de livros, pôsteres, apresentações e outros. O E-LIS constitui-se como uma alternativa para o armazenamento e divulgação desses outros tipos de publicação no âmbito da Ciência da Informação. Entender o que está disponível na coleção do E-LIS possibilitará a compreensão

7 Na sigla em inglês.

8 *“Equality in access to science is not only a social and ethical requirement for human development, but also a necessity for realizing the full potential of scientific communities world-wide and to orient scientific progress towards meeting the needs of humankind” (UNESCO/ICSU, 1999, on-line).*

sobre o que os cientistas e outros profissionais da informação consideram importante depositar para armazenar e preservar nesse repositório.

Os repositórios temáticos podem ser uma ferramenta eficaz para divulgar, disseminar e promover publicações científicas. O uso de um repositório temático não se limita apenas à instituição de origem, mas permite uma troca interinstitucional e internacional. Como apontado no próprio *site* do E-LIS, uma consulta ao seu acervo é uma experiência multilinguística e multicultural (E-LIS, 2019), que divulga artigos publicados por pesquisadores em periódicos que foram revisados por pares, possibilitando gerar maiores citações por meio de maior visibilidade. Para a comunidade de Ciência da Informação (estudantes de graduação e pós-graduação, pesquisadores, bibliotecários e outros profissionais da informação), o estudo sobre Repositórios Temáticos pode trazer uma conscientização acerca dos propósitos do Movimento de Acesso Aberto e os benefícios que ele pode acarretar para a comunicação científica.

Alguns autores afirmam que os pesquisadores em Ciência da Informação possuem boas oportunidades para pôr em prática o acesso aberto, sobretudo, no que diz respeito a menores restrições nas políticas dos periódicos da área, permitindo o autoarquivamento em repositórios digitais, bem como, incentivando a realização dessa atividade (COLEMAN, 2007; CHILIMO; ONYANCHA 2018; EMERY, 2018). A análise bibliométrica dos documentos depositados no E-LIS permitirá não só conhecer o corpo de conhecimento da Ciência da Informação produzido que está em acesso aberto nesse repositório, como também possibilitará a comparação entre as características de depósito de documentos por países e continentes com modelos diferentes de comunicação científica, entre outros estudos. Assim, entender as características dos materiais disponibilizados no E-LIS pode incentivar o autoarquivamento por mais pesquisadores, fortalecendo o acervo e aumentando essa rede colaborativa.

Conforme o exposto, entende-se como imprescindível na Ciência da Informação, uma área que lida diretamente com as Tecnologias de Informação e Comunicação (TIC) e em como elas influenciam os fluxos da ciência, discutir sobre as práticas de comunicação científica e seus desenvolvimentos nos termos do Movimento de Acesso Aberto. A discussão trazida nesse trabalho pode contribuir com aspectos científicos, metodológicos, práticos, sociais e profissionais da área da Biblioteconomia e Ciência da Informação.

As contribuições recaem, de forma mais direta, sobre o exame de um tema pouco abordado na Ciência da Informação, o autoarquivamento. Aqui, busca-se investigar e analisar um conjunto de materiais científicos abertos derivados da prática de autoarquivamento, por meio da análise bibliométrica, permitindo aos interessados conhecer o que está disponível no maior repositório aberto da área.

Para evidenciar as características da literatura científica que compõe o acervo do E-LIS, uma metodologia foi desenvolvida de maneira que essa análise seja feita automaticamente utilizando a linguagem de programação Python. A partir do desenvolvimento do ELIScript,

instrumentos de coleta e análise de dados do E-LIS, a investigação sobre os documentos depositados nesse repositório será realizada. O ELIScript permitirá que análises sobre todo o acervo contido no E-LIS sejam feitas com rapidez e com maior frequência. Além disso, o código-fonte dos instrumentos será disponibilizado em um repositório público, assim, o programa poderá ser utilizado por outros pesquisadores e empregado para outros fins, como a análise do acervo de outros repositórios. O uso de algumas ferramentas nesse estudo permitirá apresentar um fluxo de trabalho pormenorizado, descrevendo desde a preparação do ambiente computacional até a coleta e análise de dados. Assim, outros pesquisadores poderão reproduzir os passos aqui descritos de maneira mais rápida e com maior autonomia.

Do ponto de vista social, a discussão sobre o Movimento de Acesso Aberto é pertinente, pois considera que o conhecimento científico é um bem público global e tem em vista que os resultados de pesquisa financiados com impostos⁹ devem retornar de alguma forma à sociedade. Os repositórios digitais abertos têm um papel importante, nesse sentido, já que sua função é armazenar, preservar e disponibilizar as produções científicas por um longo período de tempo para que todos tenham acesso. E os pesquisadores, por meio do autoarquivamento, estão diretamente envolvidos na alimentação desses repositórios.

Quanto às contribuições práticas e profissionais, espera-se esclarecer e incentivar a prática de autoarquivamento na Ciência da Informação, de forma que mais pesquisadores desenvolvam esse hábito e ajudem a construir e fortalecer a memória da área. A metodologia, por sua vez, foi criada com o intuito de facilitar o trabalho de outros pesquisadores na Ciência da Informação, isto é, evitar o retrabalho, seja no desenvolvimento de um novo trabalho sobre o E-LIS, seja em outros repositórios.

Por fim, espera-se que o presente trabalho seja uma adição frutífera ao corpo de conhecimento em Ciência da Informação, por meio das discussões acerca das transformações do modelo de comunicação científica nos moldes do Movimento de Acesso Aberto, ressaltando a importância dos repositórios digitais temáticos e a postura dos pesquisadores frente às práticas de autoarquivamento, bem como pelo incentivo aos bibliotecários e cientistas da informação a terem um novo olhar sobre as possibilidades trazidas pela Ciência Aberta e pelo uso da linguagem de programação.

Observando-se o exposto, a presente dissertação consta dos capítulos e subcapítulos apresentados a seguir. O **primeiro capítulo** é a Introdução, na qual são apresentados o

9 Os impostos estão embutidos em todos os produtos e serviços adquiridos pela sociedade. Os recursos provenientes de “diversos setores produtivos, derivados de receitas variadas, como *royalties*, compensação financeira, licenças, autorizações, etc.” (TAVARES, 2005, p. 6) são arrecadados para os Fundos Setoriais de Ciência e Tecnologia. Esses recursos são direcionados para o Fundo Nacional de Desenvolvimento Científico e Tecnológico (FNDCT) com a finalidade dar apoio financeiro aos programas e projetos de desenvolvimento científico e tecnológico no país, e são administrados pela Financiadora de Estudos e Projetos (FINEP), ambos vinculados ao Ministério da Ciência, Tecnologia e Inovações (MCTI) (GOMES *et al.*, 2015).

problema e a questão de pesquisa; os objetivos, geral e específicos; a justificativa; e uma breve apresentação da estrutura do trabalho. O **segundo capítulo** aborda a evolução da comunicação científica, observando-se o desenvolvimento do modelo de comunicação científica a partir dos canais científicos de comunicação e dos produtos da ciência. No **terceiro capítulo**, a Ciência Aberta e os movimentos que a compõem serão abordados, tendo como destaque o Movimento de Acesso Aberto, sua criação e seus desdobramentos até a atualidade. Nele, serão discutidos, também, a estratégia da Via Verde, ou seja, de autoarquivamento realizado por autores, com foco em repositórios digitais e no E-Prints in Library and Information Science. No **quarto capítulo**, discute-se os Estudos Métricos da informação e suas aplicações. No **capítulo cinco**, está descrito o percurso metodológico adotado neste trabalho. Inicia-se com a descrição da estratégia de busca adotada na revisão de literatura, seguida pela caracterização da pesquisa, do ambiente e universo da pesquisa, bem como, dos instrumentos de coleta, análise e discussão de dados. Devido à característica particular do objetivo “desenvolver o ELIScript, instrumentos para coleta e análise de dados, em linguagem de programação Python”, optou-se por apresentar em capítulo separado, no **capítulo seis**, a criação dessa aplicação. No **capítulo sete**, serão feitas as análises e discussão dos resultados. Por fim, as **considerações finais** são apresentadas, seguidas das referências, anexos e apêndices que contribuíram para o trabalho.

2 COMUNICAÇÃO CIENTÍFICA E SUA EVOLUÇÃO

A palavra “ciência” se faz presente no cotidiano dos seres humanos. A simples observação do que é transmitido pelos sistemas de comunicação de massa – televisão, rádio, Internet – possibilita essa afirmação (CHRISTÓVÃO, 1979; TARGINO, 2000). Contudo, não há um consenso sobre o que é ciência, de fato, qual a sua amplitude e o que ela abarca. De acordo com Ziman (1979, p. 27) “[o] problema tem sido encontrar um princípio capaz de unificar a Ciência em todos os seus aspectos”.

Todavia, alguns autores tentam responder à pergunta “o que é ciência?” (CAMPBELL, 1921; ZIMAN, 1979; CHALMERS, 2013). Segundo Campbell (1921), a ciência se apresenta sob duas formas ou aspectos: a ciência prática e a ciência pura. Na primeira forma, a ciência é um corpo de conhecimento útil e prático, além de ser, também, um método para obter esse conhecimento. Em seu segundo aspecto, a ciência é o puro estudo intelectual. Ainda segundo o autor, a ciência objetiva satisfazer as necessidades da mente e não do corpo, de modo que “a ciência é um ramo de pura aprendizagem que visa a satisfação intelectual” (CAMPBELL, 1921, p. 5, *tradução nossa*¹⁰).

Ziman (1979) e Chalmers (2013), por sua vez, contestam o senso comum sobre o que é ciência. Em conformidade com Chalmers (2013), presume-se que a ciência é especial porque é baseada em fatos. Os fatos seriam, então, “reivindicações sobre o mundo que podem ser diretamente estabelecidas por um uso cuidadoso e sem preconceitos dos sentidos. A ciência deve basear-se no que podemos ver, ouvir e tocar, em vez de opiniões pessoais ou imaginações especulativas” (CHALMERS, 2013, p. 1, *tradução nossa*¹¹).

Nesse mesmo contexto, Ziman (1979, p. 22) apresenta um sumário dos diversos conceitos de ciência que não fariam “justiça à variada e complexa literatura sobre o assunto”. Alguns dos conceitos, a título de conhecimento, apontam que a ciência é “o Domínio do Meio-ambiente”, o “Estudo do Mundo Material” e o próprio “Método Experimental”. A definição mais atribuída admite que “a ciência alcança a verdade através de inferências lógicas baseada em observações empíricas” (ZIMAN, 1979, p. 20).

No entanto, essas observações baseadas em experiência também estão sujeitas à falibilidade humana. De acordo com Ziman (1979, p. 21), “[n]ossa experiência [...] indica que só nos é possível chegar a verdades parciais e incompletas; jamais conseguiremos alcançar a precisão e a perfeição que essa definição parece exigir”. Chalmers (2013), nessa mesma linha,

10 “[...] science is a branch of pure learning which aims at intellectual satisfaction” (CAMPBELL, 1921, p. 5)

11 “[...] the facts are presumed to be claims about the world that can be directly established by a careful, unprejudiced use of the senses. Science is to be based on what we can see, hear and touch rather than on personal opinions or speculative imaginings.” (CHALMERS, 2013, p. 1)

afirma que as percepções de um pesquisador são influenciadas pela sua visão de mundo e suas expectativas, de maneira que o que parece ser um fato observável para um indivíduo pode não ser para outro.

Indagações à parte, Viana (2017, p. 51) é pontual ao afirmar que

[c]ompreender que a Ciência possui estatuto empírico e se utiliza de instrumental real, sólido e concreto, que está vinculado ao fornecimento de respostas às necessidades e aos problemas humanos, significa dizer que há ciência porque a humanidade possui demandas e estas necessitam de solução, seja na descoberta de novo produto, seja na validação de uma nova teoria, realizada pelos pesquisadores/investigadores da área.

Por mais que os estudiosos, cientistas e filósofos da ciência não tenham chegado a um consenso sobre o que é ciência, Christóvão (1979) aponta que existe um aspecto fundamental sobre a ciência e que se faz presente na fala desses autores: o fato de a mesma ser comunicada.

O processo de comunicação é a essência da ciência. De acordo com Garvey (1979), a formalização do processo de comunicação científica começou há mais de 300 anos, quando a divulgação dos resultados científicos para outros pesquisadores tornou-se comum. Os pesquisadores, naquela época, costumavam se reunir em grupos, conhecidos como sociedades científicas, para discutir os novos avanços científicos e filosóficos.

Segundo Guerado (2017), as sociedades científicas, como são entendidas atualmente, existem desde, pelo menos, a Idade Média, quando a sociedade *Sodalitas Literarum Vistulana* foi fundada na Cracóvia, cidade da Polônia, em 1488, seguida por outras sociedades na Europa. Essas sociedades constituíam um mercado de conhecimento no qual havia transações de troca de informação entre seus membros. Dessa forma, o dinheiro seria usado apenas para propósitos logísticos, pois a moeda paga e recebida por seus membros para a aquisição do conhecimento seria a própria troca de conhecimento.

Até a Idade Média, a transmissão de conhecimento ocorria de pessoa para pessoa, e, em pequena escala, por meio de manuscritos enviados por correspondência. Com a introdução da prensa de tipos móveis, criação de Gutenberg, em 1455, acelerou-se a produção de documentos, mas a disseminação destes ainda não era a mais eficiente, já que os meios de transporte ainda não eram mecânicos (GILCHRIST, 2015; GUERADO, 2017). As sociedades científicas eram úteis, nesse sentido, pois se encarregavam de organizar encontros nos quais as trocas de conhecimento se tornavam possíveis, possibilitando a atualização de seus membros quanto às descobertas e inovações científicas mais recentes antes mesmo que elas fossem impressas e distribuídas (GUERADO, 2017).

Devido ao trabalho árduo de se manter atualizado quanto aos desdobramentos da ciência por meio de correspondências ou conversas informais, Meadows (1999, p. 6) afirma que a solução mais óbvia seria “fazer uma publicação impressa, com as cartas mais importantes, e distribuí-la”. A partir do século XVII surgem as revistas científicas, com a publicação do *Journal des Sçavans* em janeiro de 1665, em Paris, e o *Philosophical Transactions of the Royal Society* em Londres, no mês de março do mesmo ano (MEADOWS, 1999; GILCHRIST, 2015).

A criação das revistas científicas impressas sinalizou a formalização da comunicação da ciência. Dessa forma, existem dois tipos de comunicação científica, a comunicação formal e a informal. Segundo Meadows (1999), a comunicação informal é geralmente efêmera, tornada pública para uma audiência reduzida. Para Garvey (1979), essa comunicação ocorre por meio de um processo de troca, geralmente por apresentações em eventos ou reuniões científicas, com o intuito de transformar um trabalho em desenvolvimento em um produto para submissão a um periódico. A comunicação científica formal, por outro lado, ocorre mediante comunicação escrita, normalmente disponível por longos períodos de tempo e para um grande público (MEADOWS, 1999). Assim, os canais formal e informal servem a propósitos distintos e ambos são vitais para a divulgação da produção científica.

Com a primeira Revolução Industrial, em meados do século XVIII, a energia a vapor forneceu os meios para o aumento da produção de documentos, bem como possibilitou a distribuição mais rápida dos documentos por meios de transporte movidos a vapor. Quase imediatamente, a primeira revolução foi superada pela segunda, introduzindo o aço e as ferrovias, produtos químicos e o modelo de produção em massa. As sociedades científicas se beneficiaram dessas revoluções, não apenas porque o conhecimento podia ser distribuído de forma mais acelerada, mas também porque os cientistas podiam viajar mais rápido e participar de mais encontros científicos. Com o passar do tempo, novas fontes de energia se tornaram disponíveis: o petróleo e a eletricidade; a segunda levando à invenção da comunicação elétrica pelo telégrafo e pelo telefone. O avanço tecnológico atingiu um ritmo bastante veloz, o que levou à terceira Revolução Industrial: a revolução das comunicações, anunciando a chegada do rádio, da televisão, dos computadores, da Internet e assim por diante (GILCHRIST, 2015; GUERADO, 2017).

A evolução das tecnologias, o crescimento populacional, o aumento na escolaridade dessa população, e a própria criação de novos periódicos, acarretaram em um grande acúmulo de pesquisas, partindo de uma pequena base, há cerca de 300 anos, para a massiva quantidade de produtos bibliográficos que se tem atualmente. Os canais formais e informais se multiplicaram ao ponto de ser impossível se manter a par de tudo que estava sendo publicado ou divulgado. As TIC aparecem na década de 1960 do século passado, na forma do computador, como um acréscimo importante para o tratamento da informação, facilitando, então, a sua recuperação em meio ao caos da explosão bibliográfica. O avanço das TIC permitiu a transição dos artigos e

periódicos científicos na forma impressa para a forma eletrônica, trazendo novas possibilidades para a comunicação científica (MEADOWS, 1999). Em consonância com Guerado (2017, p. S2, *tradução nossa*¹²),

[a] introdução da Internet trouxe uma mudança crucial, à medida que a distribuição se tornou mais rápida do que a produção, e as reuniões face a face perderam terreno para a nova forma de comunicação. A Internet significou não apenas um avanço técnico, mas uma mudança de época e o surgimento de um novo paradigma na produção e distribuição da produção científica.

Segundo Baptista *et al.* (2007), as questões relacionadas ao uso das TIC na comunicação científica podem ser resumidas em três aspectos: a) o processo de comunicação científica baseado no meio eletrônico; b) os impactos que esse uso provoca na própria comunicação científica e nas comunidades científicas; e c) as reações dos pesquisadores aos modelos de negócios para a publicação científica, que não exploram as potencialidades das TIC, não permitindo o aproveitamento de seus benefícios. Barreto (1998, p. 126) afirma que

[a] comunicação eletrônica veio definitivamente libertar o texto e a informação de uma ideologia envelhecida e autoritária dos gestores da recuperação da informação, defensores de uma pretensa qualidade ameaçada, os fatais intermediários e porta-vozes que vêem (*sic*) seus poderes ameaçados cada vez mais pela facilidade da convivência direta entre os geradores e consumidores da informação.

Apesar do reconhecimento e prestígio dos periódicos no espaço científico, com o advento da Internet e das TIC, outras possibilidades de comunicação científica foram surgindo. Na década de 1980, embasando o interesse por novas formas de comunicação científica, estourava a crise dos periódicos. Mueller (2006) aponta que o que provocou essa crise foi a impossibilidade de manutenção das assinaturas das coleções de periódicos por bibliotecas universitárias, frente aos altos custos apresentados pelas editoras. A crise já afetava os países em desenvolvimento, como o Brasil, mas só se percebeu a sua repercussão quando esta alcançou, também, as universidades norte-americanas. Dessa forma, os canais eletrônicos se apresentaram como um espaço alternativo para a comunicação científica, impulsionando a criação de periódicos científicos eletrônicos e bibliotecas digitais em acesso aberto.

Com o surgimento desses espaços na Internet e as diferentes possibilidades para a comunicação científica, faz-se necessário retomar a discussão sobre os canais de comunicação

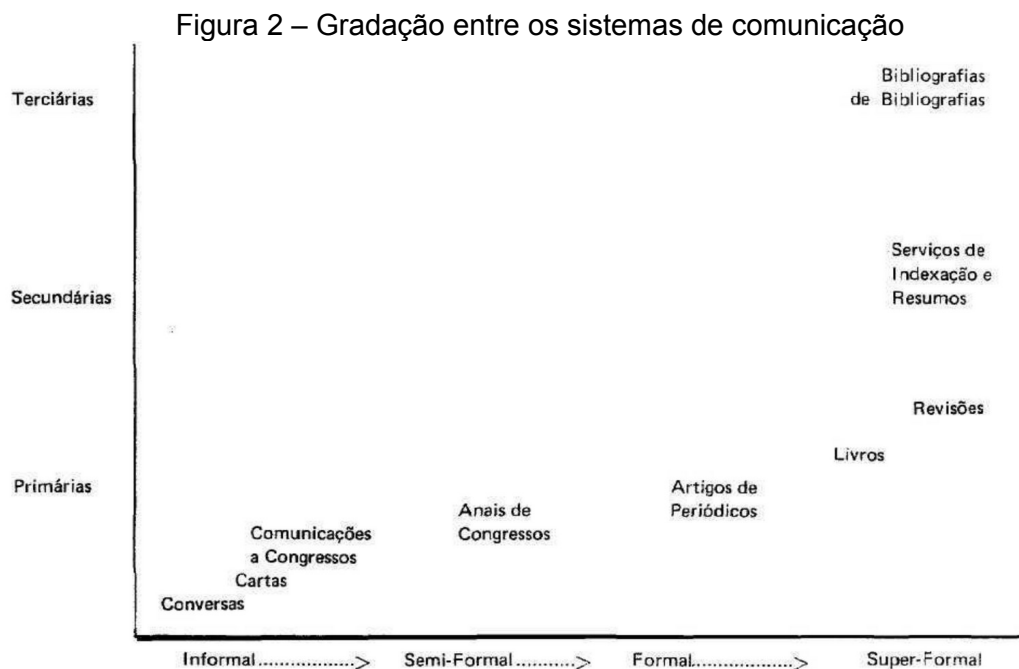
12 “The introduction of the internet brought about a pivotal change, as distribution became faster than production, and face-to-face meetings lost ground to the new form of communication. The internet signified not only a technical advance, but a change of era and the appearance of a new paradigm in the production and distribution of scientific output” (GUERADO, 2017, p. S2)

da ciência, uma vez que as formas com que os cientistas se comunicam sofreram mudanças significativas ao longo do tempo.

2.1 Canais científicos de comunicação

O processo de investigação científica, uma das principais atividades do fazer científico, perpassa diversos momentos, como a identificação do problema de pesquisa, a busca e descoberta de fontes de informação, a escrita do trabalho, a publicação dos resultados de pesquisa, entre outros. Nesse conjunto de atividades, Christóvão (1979, p. 4) afirma que “o cientista entra em contacto (*sic*) com diferentes tipos de sistemas de comunicação”. De acordo com Meadows (1999), estes podem ser subdivididos em sistemas de comunicação formal e sistemas de comunicação informal.

No entanto, essa divisão não é tão arbitrária quanto parece. Christóvão (1979) demonstra que a transição de um sistema para o outro é gradativa e não envolve só o sistema formal e o informal, mas também o semi-formal e o super-formal. Para isso, a autora elaborou um esquema (Figura 2) que relaciona as fontes de informação primárias, secundárias e terciárias com os tipos de sistema de comunicação.



Fonte: Christóvão (1979, p. 5).

Em consonância com Targino (2000, p. 18), as relações entre os sistemas de comunicação científicos “formam uma espécie de rede, na qual os cientistas e seus produtos fluem, interagindo segundo as etapas da pesquisa e a necessidade de informações que tais

etapas acarretam”. A estes quatro sistemas apresentados por Christóvão (1979), Targino (2000) acrescenta também a comunicação eletrônica. A seguir descreve-se com mais detalhe cada um destes sistemas.

Começando pelo domínio informal, discute-se a comunicação informal e semi-formal. Targino (2000, p. 19-20) afirma que esses sistemas se baseiam na “utilização de canais informais, em que a transferência da informação ocorre através de contatos interpessoais e de quaisquer recursos destituídos de formalismo, como reuniões científicas, participação em associações profissionais e colégios invisíveis”. Garvey (1979) descreve o domínio informal como aquele que abrange os procedimentos que o cientista usa para desenvolver suas informações brutas em um produto acabado, digno de ser submetido à revisão editorial de uma revista científica. De forma que o autor entende este momento como o processo de troca de informações antes da publicação. Outros autores, como Le Coadic (2004) e Meadows (1999), relacionam a informalidade com a oralidade, isto é, a comunicação oral é tida como meio para distribuição de informações em âmbito público (conferências, colóquios, seminários...) e privado (conversas, correspondências etc.). Entretanto, Targino (2000, p. 20) esclarece que “a troca informal inclui tanto recursos orais, quanto recursos escritos, como cartas, mensagens eletrônicas, entre outros”.

No que diz respeito à comunicação semi-formal, Targino (2000, p. 21), baseando-se na ideia de fluidez entre os sistemas de Christóvão (1979), descreve-a como aquela que “guarda, simultaneamente, aspectos formais e informais, e que, como a informal, possibilita discussão crítica entre os pares, o que conduz a modificações ou confirmações do teor original” do documento. São veículos que possuem, ao mesmo tempo, no entendimento de Christóvão (1979), características informais na sua forma de apresentação oral e características formais nas formas de sua divulgação, por meio de cópias ou *anais* de eventos.

O sistema de comunicação formal, por sua vez, ocorre por meio da escrita. Neste tipo de publicação, Le Coadic (2004) explica que o cientista apresenta os resultados de uma pesquisa pela primeira vez, mediante uma publicação primária. Em conformidade com Targino (2000, p. 18), alguns exemplos de publicação na comunicação formal são: “livros, periódicos, obras de referência em geral, relatórios técnicos, revisões de literatura, bibliografias de bibliografias etc.”. Como visto na Figura 2, Christóvão (1979) estabelece que neste domínio, formal/super-formal, estão inclusas as fontes primárias e secundárias. Ainda de acordo com a autora,

[f]ontes primárias seriam, por exemplo, periódicos e livros, embora, na transição do sistema formal para o super-formal, os livros pudessem ser incluídos neste último. Os livros, neste caso, seriam considerados como uma abordagem do conhecimento já aceito e absorvido pela comunidade científica. Como se pode observar, a separação não pode ser rígida. Fontes secundárias seriam os serviços de indexação e resumos, responsáveis pelos periódicos de resumos (*abstract journals*), os serviços de alerta-corrente, etc, que na escala já

mencionada, se caracterizam como veículos super-formais (CHRISTÓVÃO, 1979, p. 5-6, *grifo nosso*).

A passagem da comunicação formal para a super-formal ocorre por meio de “filtros de qualidade”. Nesse sentido, Targino (2000, p. 19) esclarece que “os livros, como conhecimento avaliado e absorvido pela comunidade científica, junto às publicações secundárias e terciárias integram o domínio superformal, o qual tem níveis variados, segundo a filtragem processada”.

Por fim, a autora volta-se a uma análise da revolução tecnológica e as mudanças decorridas desta para introduzir a comunicação eletrônica. No entendimento da mesma, a

comunicação científica eletrônica é, em sua essência, a transmissão de informações científicas através de meios eletrônicos. Pode ser vista sob duas perspectivas: (a) como um processo de mudanças estruturais induzidas tecnologicamente, ou seja, como resultante das [novas tecnologias de informação e de comunicação] NTIC; (b) como um recurso para incrementar e aperfeiçoar o contato entre cientistas (TARGINO, 2000, p. 21).

Dessa forma, os canais de comunicação eletrônica se propõem a atender um público grande, devido ao seu alcance, com a disponibilização e armazenamento de informações recentes. Pelo seu caráter ágil, a comunicação eletrônica guarda aspectos da comunicação formal e informal, possibilitando a troca de informações entre cientistas, bem como a publicação de produtos científicos que em determinado momento passaram por uma avaliação, isto é, uma filtragem.

Após a explanação sobre os tipos de sistemas de comunicação científica, no subtópico a seguir serão discutidos os produtos da ciência que são difundidos por esses canais e suas características constitutivas.

2.2 Produtos científicos

As investigações científicas realizadas por pesquisadores e as atividades governamentais, comerciais e industriais geram documentos. Esses documentos constituem a literatura branca e a literatura cinzenta. Em conformidade com Gomes, Mendonça e Souza (2000), a literatura cinzenta diz respeito aos documentos não-convencionais e semi-publicados, já a literatura branca corresponde aos materiais convencionais ou formais que foram publicados.

Nos últimos 50 anos, os computadores, bases de dados, *softwares* de publicação e a própria Internet transformaram a forma como são produzidos e disseminados os documentos. De

acordo com Lawrence (2012, p. 4, *tradução nossa*¹³), a Internet enquanto plataforma de publicação e distribuição, “tem sido particularmente eficaz na alteração da maneira como os relatórios de pesquisa e outros conteúdos produzidos por governos e outras organizações podem ser acessados e utilizados”. Antes da Internet, a literatura cinzenta era tida como um material difícil de recuperar.

Em 1997, a definição de literatura cinzenta, aceita durante a Terceira Conferência Internacional sobre Literatura Cinzenta em Luxemburgo, de acordo com Farace e Schöpfel (2010, p. 1), assinala: “aquela que é produzida em todos os níveis acadêmicos, do governo, empresas e indústria nos formatos impresso e eletrônico, mas que não é controlada por editoras comerciais”. Conforme Lawrence (2012), a literatura cinzenta era reconhecida pelo seu caráter de difícil recuperação, no entanto, as tecnologias proporcionaram sua disseminação de forma mais fácil, mudando o foco da dificuldade de recuperação da literatura cinzenta para observar sua forma, seus canais de circulação e as organizações produtoras. Na sexta edição do evento, que ocorreu na cidade de Nova Iorque no ano de 2004, adicionou-se o seguinte complemento à definição “isto é, onde a publicação não é a atividade principal do organismo produtor” (LAWRENCE, 2012, p. 2, *tradução nossa*¹⁴). Nesse mesmo sentido, Cunha e Cavalcanti (2008, p. 230) definem o verbete “literatura cinzenta” como sendo o “conjunto de documentos impressos ou mult copiados que não são nem editados, nem difundidos comercialmente. Circulam, geralmente na administração pública e privada, em congressos, reuniões e centros de pesquisa”.

A literatura branca, por seu turno, é composta por livros e publicações periódicas, são os documentos formais, publicados dentro do circuito comercial, com grande tiragem de cópias e que são difundidas de forma irrestrita. Por isso, ela é constituída por documentos convencionais disponíveis no mercado editorial, o que permite sua ampla disseminação e acesso (FERREIRA, 2017).

A linha que separa a literatura cinzenta da branca é tênue. Nas palavras de Farace e Schöpfel (2010, p. 2, *tradução nossa*¹⁵), a fronteira da literatura cinzenta

com a literatura “branca” ou “convencional” é permeável, uma vez que alguns anais de conferências são publicados por editoras comerciais como monografias ou em publicações seriadas, como os periódicos. O mesmo vale para alguns

13 “[...] *the internet has been particularly effective in changing the way research reports and other content produced by governments and other organisations (sic) are able to be accessed and used.*” (LAWRENCE, 2012, p. 4).

14 “*i.e. where publishing is not the primary activity of the producing body*” (LAWRENCE, 2012, p. 2).

15 “[...] *the borderline with “white” or “conventional” literature is permeable, since some conference proceedings are published by commercial publishers as monographs or in serial publications such as journals. The same holds true for some reports. Likewise for doctoral theses, especially in the humanities and social sciences, some are found on the commercial publishing market*” (FARACE; SCHÖPFEL, 2010, p. 2).

relatórios. Da mesma forma, para teses de doutorado, principalmente nas ciências humanas e sociais, algumas são encontradas no mercado editorial comercial.]

Como visto, os documentos da literatura branca podem ser livros ou publicações seriadas. Gomes, Mendonça e Souza (2000, p. 98) elencam alguns documentos considerados parte da literatura cinzenta, a saber: “publicações governamentais, traduções avulsas, *preprints*, dissertações, teses e literatura originada de encontros científicos, como os *anais* de congressos”. Ferreira (2017) adiciona a estes, também, os relatórios técnicos e de pesquisa. Neste estudo, os dois tipos de literatura serão considerados, dado que no repositório E-LIS não há restrição nesse sentido.

3 CIÊNCIA ABERTA

Antes de definir o que é ciência aberta, faz-se necessário entender o conceito de abertura. O que significa ser ou estar aberto? A *Open Definition* tem a função de tornar preciso o significado de “aberto” no contexto de dados ou conhecimento abertos. Ela define como material aberto aquilo que “qualquer um pode livremente acessar, usar, modificar e compartilhar para qualquer propósito (sujeito, no máximo, a requisitos que preservem a proveniência e a abertura)”. De forma mais sucinta, “dados e conteúdos abertos podem ser livremente usados, modificados e compartilhados por qualquer um para qualquer fim”. (OPEN... 2020, on-line, *tradução nossa*¹⁶).

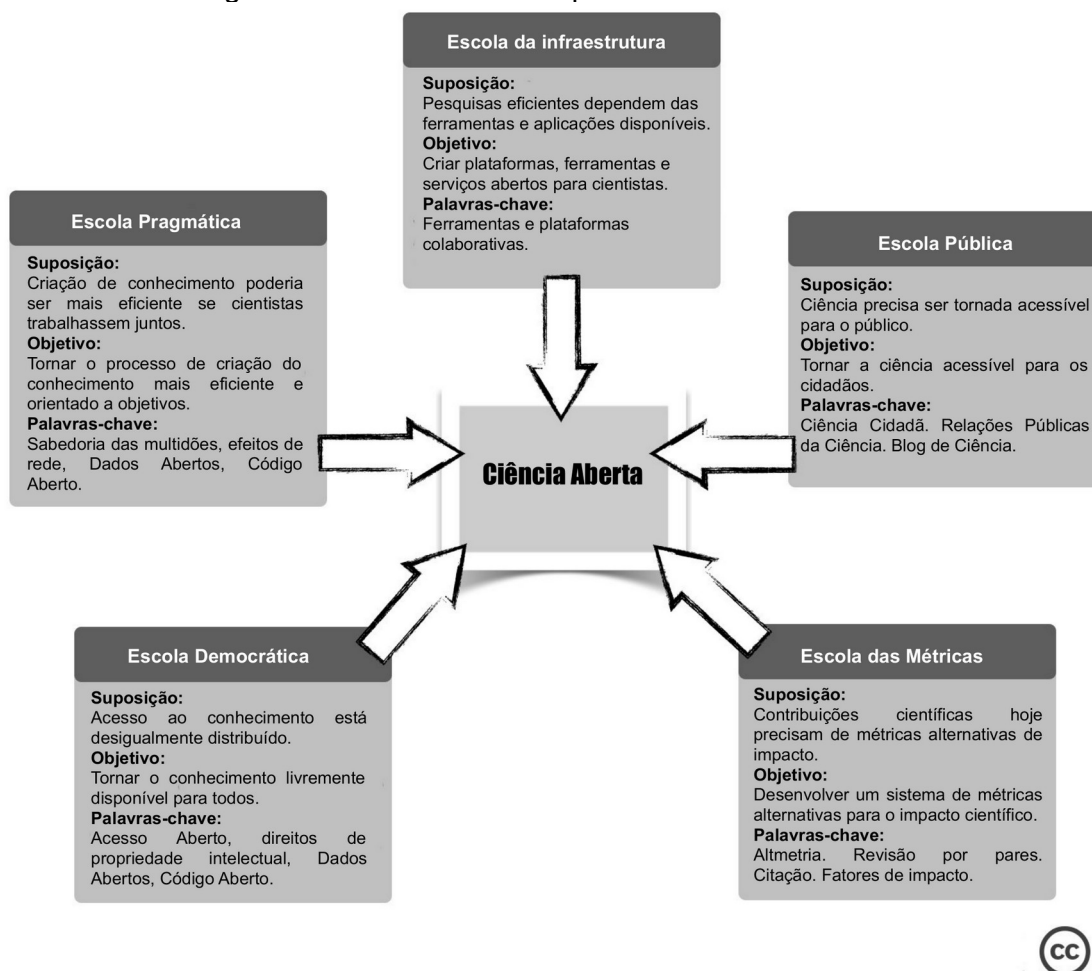
A ciência aberta, então, existe para tornar o conjunto de atividades científicas aberto para todo aquele que nele estiver interessado. O termo “Ciência Aberta”, em si, é considerado abrangente, pois não se aplica exclusivamente a uma área do conhecimento, nem tem apenas uma função dentro do espectro científico. De acordo com Fecher e Friesike (2014), esse termo é composto por uma infinidade de suposições sobre o futuro da criação e disseminação de conhecimento. No mesmo sentido, Albagli, Clinio e Raychtock (2014, p. 435) afirmam que a Ciência Aberta é “um termo guarda-chuva, que engloba diferentes significados, tipos de práticas e iniciativas, bem como envolve distintas perspectivas, pressupostos e implicações”.

3.1 Cinco escolas de pensamento da Ciência Aberta

Diante de diferentes entendimentos e dessa visão tão ampla que pode haver sobre a ciência aberta, Fecher e Friesike (2014) propõem estruturar o discurso sobre a temática em cinco escolas de pensamento, a saber: a escola pública (*the public school*), escola democrática (*the democratic school*), a escola pragmática (*the pragmatic school*), a escola da infraestrutura (*Infrastructure school*), e a escola das métricas (*the measurement school*). Cada uma delas será discutida a seguir com base nas descrições de Fecher e Friesike (2014). De forma resumida, as suposições, objetivos e palavras-chave relacionadas às cinco escolas de pensamento estão representadas na Figura 3.

16 “Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).” Put most succinctly: “Open data and content can be freely used, modified, and shared by anyone for any purpose”

Figura 3 – Cinco escolas de pensamento da Ciência Aberta



Fonte: Fecher e Friesike (2014, *tradução nossa*).

A **escola pública** concerne ao acesso que o público tem à criação do conhecimento científico. Isto quer dizer que a ciência precisa estar acessível para um público maior. Com as TIC, as ferramentas e os serviços disponíveis na web social, o cientista consegue tanto abrir e tornar seu fluxo de trabalho científico disponível e transparente, quanto disponibilizar seus resultados para não-cientistas. Dessa forma, existem duas linhas dentro dessa escola de pensamento: a) a primeira diz respeito à acessibilidade ao processo de criação de conhecimento (produção); e b) a segunda, envolve a compreensão dos resultados de pesquisa (produto). Em (a), o cidadão torna-se participante mais ativo na ciência, não tornando-se um cientista, especificamente, mas podendo emprestar sua força de trabalho para o progresso da ciência. Já em (b), nota-se a necessidade de tornar a ciência mais compreensível para quem não faz parte dessa comunidade. Essa é uma grande preocupação da divulgação científica: como tornar a ciência compreensível para a sociedade?

No que diz respeito à **escola democrática**, a preocupação recai sobre o conceito de acesso ao conhecimento. A escola democrática, diferente da escola pública, foca em promover o

acesso aos resultados da pesquisa, isto é, publicações científicas e dados de pesquisa, mas não se restringindo apenas a isso. Outros materiais incluem diferentes representações digitais, gráficas e multimídia. De forma geral, essa escola argumenta que todos os materiais envolvidos na produção de um resultado de pesquisa devem estar disponíveis abertamente. A ideia por trás dessa disponibilização corresponde ao entendimento de que o conhecimento científico é um bem público global e todos devem ter acesso a ele. Como na escola anterior, esta subdivide-se em duas linhas, a saber: a) Acesso Aberto a publicações científicas; e b) Dados Abertos.

No primeiro caso, a discussão não gira mais em torno dos cientistas, mas da sociedade e seu acesso aos resultados de pesquisa. Cribb e Sari (2010) afirmam que o conhecimento científico dobra a cada cinco anos, mas sua distribuição entre os sete bilhões de cidadãos no mundo ocorre de forma bem mais lenta. O Movimento de Acesso Aberto e as formas de acesso às publicações científicas serão discutidos com mais detalhes no tópico **3.3 Movimento de Acesso Aberto**. No segundo, o acesso aos dados abertos evita a duplicação de esforços na coleta de dados, bem como permite reutilizá-los para novos fins. A responsabilidade sobre disponibilizar esses dados recai sobre o próprio cientista, embora existam alguns impedimentos concernentes às editoras científicas e suas formas de publicação.

A **escola pragmática**, por sua vez, interessa-se pela colaboração na pesquisa. Aqueles que advogam por essa escola querem tornar a ciência, a realização das pesquisas e sua disseminação mais eficiente. Por isso, consideram a ciência como um processo passível de otimização. A ciência tem se tornado cada vez mais complexa, abarcando aspectos inter, multi e transdisciplinares, acarretando na necessidade de esforços conjuntos para a resolução de problemas. A web 2.0, nesse quesito, se caracteriza como um ambiente fértil para a prática da pesquisa conjunta e as tecnologias emergentes apontam para uma realidade cada vez mais colaborativa.

A **escola da infraestrutura** preocupa-se com a arquitetura tecnológica. Essa escola considera a Ciência Aberta como um desafio tecnológico. Todas as escolas citadas anteriormente possuem um componente tecnológico. Não é possível falar em acesso aberto, dados abertos, colaboração na ciência e abertura dos resultados de pesquisa para a sociedade sem considerar que deve existir um aparato tecnológico por trás dessas atividades. De alguma forma, as tecnologias permitiram diversas mudanças na comunicação científica. Apesar de estar contida nas outras abordagens, Fecher e Friesike (2014) consideram que existem tendências tecnológicas que possibilitam a realização da pesquisa numa escala diferente e por isso a infraestrutura merece sua própria escola. Assim, a escola da infraestrutura possui duas tendências que não são mutuamente exclusivas: a) computação distribuída, isto é, utilização do poder computacional de muitos usuários para pesquisa; e b) redes sociais e de colaboração entre pesquisadores, focando na infraestrutura que favorece a interação entre os pares. A

ciência, atualmente, é muito baseada em computação, análise de dados e colaboração (FOSTER..., 2002).

A **escola das métricas** relaciona-se com as métricas alternativas de impacto. Os pesquisadores que advogam por essa escola estão em busca de padrões alternativos que permitam mensurar o impacto das publicações científicas para além dos fatores de impacto, número de citações, entre outros. A questão é: de que forma medir o impacto de uma pesquisa no ambiente digital? Fecher e Friesike (2014, p. 40, *tradução nossa*¹⁷) colocam que

à medida que o fluxo de trabalho acadêmico migra cada vez mais para a web, usos anteriormente ocultos, como leitura, favoritos, compartilhamento, discussão e classificação, deixam rastros online e oferecem uma nova base para medir o impacto científico. O termo guarda-chuva para esse tipo de avaliação é Altmétrie.

As cinco escolas de pensamento da Ciência Aberta não são mutuamente exclusivas, elas se encontram em diversos momentos como pode ser percebido na escola da infraestrutura. No entanto, essa divisão permite observar a diversidade de tópicos que podem ser abordados dentro do termo guarda-chuva “Ciência Aberta”.

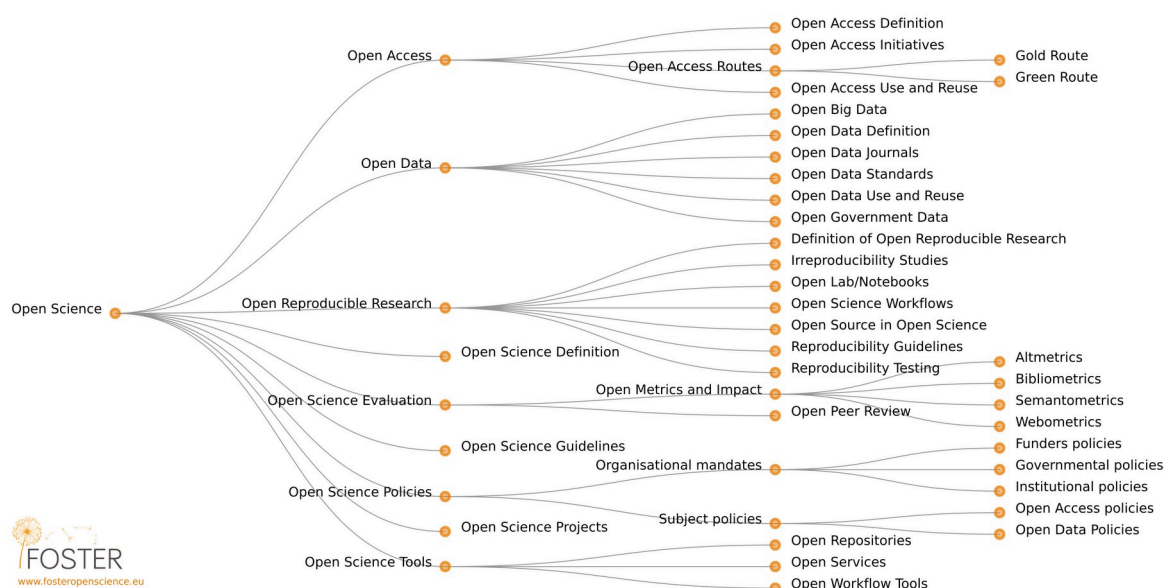
3.2 Ciência Aberta: mais do que um movimento

Para além da divisão em cinco escolas, Albagli (2019), na palestra “O que é Ciência Aberta e qual o papel das agências de fomento diante deste fenômeno?” proferida no II Encontro Capes de Ciência Aberta: direitos de propriedade intelectual e políticas institucionais¹⁸, se refere à Ciência Aberta como um “movimento de movimentos”. Para exemplificar a visão da autora, na Figura 4 estão dispostos os desdobramentos da Ciência Aberta e abordagens/temas que podem ser debatidos, estudados e praticados dentro desse movimento.

17 “As the scholarly workflow is migrates increasingly to the web, formerly hidden uses like reading, bookmarking, sharing, discussing, and rating are leaving traces online and offer a new basis by which to measure scientific impact. The umbrella term for these new impact measurements is altmetrics” (FECHER; FRIESIKE, 2014, p. 40).

18 ALBAGLI, S. O que é Ciência Aberta e qual o papel das agências de fomento diante deste fenômeno? In: ENCONTRO CAPES DE CIÊNCIA ABERTA. Tema: direitos de propriedade intelectual e políticas institucionais. dez. 2019. Disponível em: <http://capes.gov.br/conteudo/2-encontro-capes-de-ciencia-aberta/>. Acesso em: 13 dez. 2019.

Figura 4 – Taxonomia da Ciência Aberta



Fonte: Foster (2002).

Inicialmente, a Ciência Aberta ramifica-se em: Acesso Aberto, Dados Abertos, Pesquisa Reprodutível Aberta, Avaliação da Ciência Aberta, Políticas de Ciência Aberta e, por fim, Ferramentas de Ciência Aberta. Considerando cada um desses eixos como um movimento, suas particularidades serão discutidas a seguir.

O **Acesso Aberto** é o que se objetiva atingir com a disponibilização, sem restrições, dos produtos científicos. Para alcançar essa meta, existem duas iniciativas sugeridas: a Via Dourada e a Via Verde. Conforme Harnad *et al.* (2004), a primeira pode ser resumida como a publicação de artigos em periódicos de acesso aberto. A segunda consiste no depósito de artigos publicados (seja em acesso aberto, seja em periódicos com assinatura) em repositórios digitais, institucionais ou temáticos, que concordem com as diretrizes para armazenamento e preservação dessa produção por longos períodos. Dessa forma, as discussões dentro desse eixo revolvem sobre as formas de disponibilização, vantagens e estratégias para incentivar o acesso livre ao conhecimento, proporcionando seu uso e reuso.

Os **Dados Abertos** também entram na discussão sobre disponibilidade na web e, de forma mais específica, discute *Big Data*, periódicos e repositórios de dados abertos, padrões de dados abertos, como o princípio FAIR que propõe critérios de qualidade para que os dados sejam localizáveis (*Findable*), acessíveis (*Accessible*), interoperáveis (*Interoperable*) e reutilizáveis (*Reusable*), dados abertos governamentais, uso e reuso. Murray-Rust (2008) critica o fato de que algumas editoras científicas além de reterem os direitos sobre os manuscritos submetidos para publicação, também o fazem para os dados e materiais suplementares. Na

visão do autor, isso impede o reuso dos dados, tornando-se um grande impeditivo para o progresso da ciência na era digital. O autor ainda afirma que

é importante perceber que os materiais suplementares são quase sempre completamente produzidos pelos autores originais e, em muitos casos, são uma saída direta de um computador. [...] O processo editorial não agrega nada ao conteúdo e não agrega valor além de publicá-los juntamente com o texto completo (que pode ter recebido informações valiosas dos revisores e editores). É provável que muitos editores não tenham os conhecimentos técnicos necessários para avaliar a validade dos materiais suplementares por si mesmos (MURRAY-RUST, 2008, p. 53, *tradução nossa*¹⁹).

Os materiais suplementares constituem informações importantes para o artigo, mas devido à sua extensão, ou sua característica mais técnica, não cabem nesse formato. Estarem disponíveis abertamente permite a verificação dos dados, da metodologia utilizada e de como a pesquisa foi conduzida. Nas palavras de Miguel *et al.* (2014, p. 3, *tradução nossa*²⁰)

dados e materiais abertos oferecem os meios para que pesquisadores independentes reproduzam resultados publicados, testem especificações alternativas nos dados, identifiquem resultados relatados incorretamente, reutilizem ou adaptem materiais para replicação ou extensão da pesquisa anterior e entendam melhor as intervenções, medidas e contexto – todos os quais são importantes para avaliar a validade externa.

Por isso, entende-se que as editoras científicas não deveriam reter esses direitos, bem como os dados deveriam pertencer à sociedade sem nenhum custo. Esses dados, bem como dados governamentais abertos, poderiam ser depositados e disponibilizados por meio de repositórios digitais abertos, respeitando padrões estabelecidos.

Para que um estudo se torne uma **Pesquisa Reprodutível Aberta** ou uma pesquisa aberta e reprodutível, é preciso dar atenção a todas as atividades que compõem o fazer científico e observar formas de tornar esse processo transparente. A reprodutibilidade além de uma responsabilidade moral com respeito ao campo científico, também pode diminuir o fardo do

19 “It is important to realize that supporting information (SI) is almost always completely produced by the original authors and, in many cases, is a direct output from a computer. [...] The editorial process adds nothing to the content and adds no value other than publishing it alongside the full text (which may have had valuable input from the reviewers and editors). It is probable that many publishers will not have the technical expertise to evaluate the validity of the SI by themselves”. (MURRAY-RUST, 2008, p. 53).

20 “Open data and open materials provide the means for independent researchers to reproduce reported results; test alternative specifications on the data; identify misreported or fraudulent results; reuse or adapt materials (e.g., survey instruments) for replication or extension of prior research; and better understand the interventions, measures, and context—all of which are important for assessing external validity” (MIGUEL *et al.*, 2014, p. 3).

próprio pesquisador. A reprodutibilidade de uma pesquisa está relacionada não só com as tecnologias utilizadas, mas também com os hábitos adotados pelos pesquisadores para tornar esse processo mais eficiente (SANDVE *et al.*, 2013).

E como tornar uma pesquisa reprodutível? Cada área do conhecimento terá suas particularidades, no entanto, Sandve *et al.* (2013) listam 10 regras básicas para pesquisas computacionais reprodutíveis que, guardadas as proporções, podem ser aplicadas a qualquer campo. Os autores incentivam os pesquisadores, de forma resumida, a registrar cada etapa da pesquisa, observando como cada resultado foi produzido; evitar a manipulação manual de dados para que não haja adulteração de resultados; arquivar as versões exatas de todos os programas externos e scripts utilizados na pesquisa; registrar todos os resultados intermediários, quando possível, em formatos padronizados; sempre armazenar dados brutos antes de criar representações gráficas, para que outras representações possam ser feitas; conectar declarações textuais aos seus resultados para que outros pesquisadores possam chegar às mesmas conclusões; fornecer acesso público a scripts, execuções e resultados da sua pesquisa. Miguel *et al.* (2014), ao abordarem a transparência das pesquisas nas Ciências Sociais, afirmam que essas práticas ainda são incipientes na área e devem ser incentivadas para que os pesquisadores se beneficiem das mudanças que essa transparência pode trazer para o progresso científico geral.

Como será apresentado posteriormente, existem diferentes formas de mensurar a informação. Quando se fala em avaliação da ciência, considera-se a bibliometria e a cientometria como medidas adequadas. Outra métrica que tem ganhado atenção e que se torna uma ferramenta para a avaliação da ciência no movimento aberto é a Altmétrie ou as métricas alternativas. A **Avaliação da Ciência Aberta** corresponde às formas alternativas de mensuração da ciência que não se restringem apenas aos resultados de pesquisa, mas a todo o percurso que levou ao produto científico final. Conforme Fenner (2014), atualmente é possível examinar citações de artigos e outras medidas de impacto por meio de estatísticas de uso e pela web social. Nesse sentido, Albagli, Clinio e Raychtock (2014, p. 438) apontam que essas métricas alternativas pretendem medir o processo da pesquisa e seu impacto “através de comentários online, compartilhamentos, downloads, *posts* em *blogs*, *tweets*, comentários, etc.”. É importante ressaltar que as métricas alternativas não vieram para substituir a bibliometria e a cientometria, mas para complementá-las. Suas medidas de impacto são, geralmente, mais rápidas do que o acúmulo de citações e consideram não só as publicações formais, como também conjuntos de dados publicados na web, *softwares*, *blogs* de pesquisa, entre outros (PRIEM; GROTH; TARABORELLI, 2012).

Para pôr em prática tudo o que a Ciência Aberta prevê, tanto no nível individual quanto no nível institucional, é necessário que sejam desenvolvidas **Políticas de Ciência Aberta**. Essas políticas são estratégias e ações que objetivam promover os princípios e reconhecer as práticas

da Ciência Aberta. Essas políticas são, de maneira geral, criadas por instituições que realizam pesquisas, financiadores de pesquisas, governos ou editores, por meio de mandatos organizacionais e políticas especializadas (FOSTER, 2020).

Por fim, **Ferramentas de Ciência Aberta** são utilizadas para promover a abertura da ciência. Ainda que estejam em um eixo separado, devido a sua importância, essas ferramentas são utilizadas em todo o percurso para tornar o processo transparente, acessível e reproduzível. Algumas dessas ferramentas são os repositórios abertos, que têm a função de disponibilizar a pesquisa, os conjuntos de dados e tudo mais que foi utilizado para gerar os resultados da publicação, e serviços que tornam o fluxo de trabalho dos pesquisadores mais claros e fáceis de serem seguidos.

Fora as Políticas de Ciência Aberta e a avaliação por meio de métricas alternativas, todos os outros eixos serão empregados no presente trabalho. O intuito é se basear nessas diretrizes para criar um trabalho acessível, reutilizável, reproduzível e transparente, de forma que ele possa ser analisado e validado. Assim, poderá constituir um fluxo bem descrito que poderá ser reproduzido por outros pesquisadores.

3.3 Movimento de Acesso Aberto

A tradição antiga de disponibilizar os frutos da ciência e a Internet, unem-se para tornar possível a remoção das barreiras no acesso à literatura científica (CHAN *et al.*, 2002). O Acesso Aberto é definido, de acordo com Costa e Leite (2017, p. 52), como a “disponibilização das publicações científicas de modo on-line, livre e sem restrições”. Da mesma forma, Gomes, Benchimol e Barros (2018, p. 144), definem o acesso aberto como “a disponibilização gratuita de material científico (o que inclui artigos, teses, dissertações, publicações em *anais* etc.) para todos aqueles interessados em utilizá-lo”.

Ao longo dos anos, surgiram iniciativas visando contribuir para as discussões acerca do acesso livre ao conhecimento, entre elas estão a Declaração de Budapeste ou *Budapest Open Access Initiative* (BOAI) de 2002, a Declaração de Bethesda (*Bethesda statement on open access publishing*), e a Declaração de Berlim (*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*), ambas de 2003, conhecidas também como Declarações de Budapeste, Bethesda e Berlim, 3B ou BBB (COSTA; LEITE, 2017).

Dentre elas, destaca-se a BOAI por ser a primeira iniciativa a propor duas estratégias para acesso aberto: a) o autoarquivamento; e b) a publicação de periódicos em acesso aberto. O autoarquivamento (Via Verde) ocorre quando os próprios autores depositam uma cópia de suas publicações em plataformas de acesso aberto (CHAN *et al.*, 2002). A segunda estratégia (Via Dourada), de acordo com Costa e Leite (2017, p. 54), “refere-se à alteração no modelo de negócio dos periódicos já existentes e a criação de periódicos de acesso aberto, que deveriam

buscar outras fontes de financiamento que não fosse pela cobrança para o acesso aos seus conteúdos.”

Entende-se por acesso aberto à literatura científica, conforme a BOAI, a disponibilização gratuita desta literatura na Internet, permitindo a qualquer usuário ler, baixar, copiar, distribuir, imprimir e/ou recuperar. E esclarece: a única limitação quanto à reprodução e distribuição, e o único papel do copyright neste domínio é o controle por parte dos autores sobre a integridade de seu trabalho e o direito de ser propriamente reconhecido e citado (CHAN *et al.*, 2002).

A Declaração de Bethesda, por seu turno, teve o propósito de estimular a discussão sobre acesso aberto dentro da comunidade de pesquisa em Biomedicina. O objetivo principal dessa reunião foi definir passos concretos a serem tomados para promover a rápida e eficiente transição do modelo tradicional de publicação para a publicação em acesso aberto. A declaração de Bethesda propõe, assim, “o direito ao acesso livre à informação científica por meio da atribuição de licenças de uso e distribuição” (COSTA; LEITE, 2017, p. 15). Para isso, a Declaração de Bethesda afirma que uma publicação em acesso aberto deve atender a duas condições:

1) os autores e titulares dos direitos autorais devem conceder a todos os usuários direito de acesso gratuito, irrevogável, mundial e perpétuo, e uma licença para copiar, usar, distribuir, transmitir e mostrar [...], desde que reconhecida a autoria, bem como o direito de fazer um pequeno número de cópias para uso pessoal; e

2) uma versão completa de um trabalho e seus materiais suplementares [...] devem ser depositados imediatamente após publicação em, pelo menos, um repositório on-line, que seja mantido por uma instituição ou sociedade acadêmica, uma agência governamental ou outra organização bem estabelecida que procure propiciar acesso aberto, distribuição irrestrita, interoperabilidade e arquivamento a longo prazo. No caso das Ciências Biomédicas, o *PubMed Central* seria um destes repositórios (SUBER *et al.*, 2003, on-line, *tradução nossa*).²¹

Por fim, a Declaração de Berlim aponta que a missão de tornar o conhecimento disponível está pela metade se a informação não é amplamente divulgada e não está prontamente disponível para a sociedade. Esta define o Acesso Aberto como “uma fonte abrangente de conhecimento humano e patrimônio cultural que foi aprovado pela comunidade científica”

21 “1. The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use. 2. A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in a suitable standard electronic format is deposited immediately upon initial publication in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving (for the biomedical sciences, PubMed Central is such a

(BULLINGER *et al.*, 2003, on-line, *tradução nossa*²²). Em consonância com Costa e Leite (2017, p. 15), esta declaração reflete sobre “a importância da formalização de políticas de informação para o acesso aberto”. Dessa forma, a declaração sugere que a web precisa ser sustentável, interativa e transparente, por isso, o conteúdo e os *softwares* devem ser acessíveis e compatíveis.

Para que os documentos sejam acessíveis na web, surge a *Open Archives Initiative* ou em português Iniciativa de Arquivos Abertos (OAI), que desenvolve e promove os padrões de interoperabilidade para facilitar a disseminação de conteúdo eficiente, conhecidos como *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH). A OAI baseia-se no esforço de melhorar o acesso a documentos eletrônicos, como um meio de aumentar a disponibilidade dos produtos da comunicação acadêmica. (OPEN..., 2019).

Ferreira e Muniz Jr. (2005, p. 6) conseguem sintetizar a ideia das Declarações BBB unidas a OAI, ao afirmarem que o acesso aberto não se trata apenas da gratuidade no acesso às publicações científicas, mas de uma filosofia que se baseia no

auto-depósito pelos autores (garantindo-lhes o direito de dispor de seus textos como lhe convier), mecanismos de discussão e avaliação entre os pares, geração e controle de versões do texto, padronização do conjunto de metadados necessários para representação descritiva e temática dos documentos e fomento do protocolo OAI (visando garantir a interoperabilidade entre diferentes arquivos e permitindo, dentre outras funções, a recuperação e compartilhamento das informações).

Esse auto-depósito, ou a Via Verde, é visto com mais otimismo pelo Movimento de Acesso Aberto dentre as duas estratégias – isto é, a publicação em acesso aberto (Via Dourada) e o autoarquivamento em repositórios digitais (Via Verde) (SWAN, 2007; BERNIUS, 2010; COSTA; LEITE, 2017). Isto porque, de acordo com Costa e Leite (2017), essas iniciativas partem do envolvimento direto dos sistemas de informação das próprias instituições produtoras do conhecimento científico. Devido a essa posição dos autores e o próprio objetivo deste trabalho, discute-se sobre o autoarquivamento em repositórios digitais a seguir.

3.3.1 Via Verde: o autoarquivamento em repositórios digitais

Como visto anteriormente, o Movimento de Acesso Aberto possui duas formas de promover o acesso livre ao conhecimento: a Via Verde e a Via Dourada. A Via Verde, foco deste

repository)” (SUBER *et al.*, 2003, on-line).

²² “We define open access as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community” (BULLINGER *et al.*, 2003, on-line).

subtópico, se constitui por meio do depósito de um trabalho, realizado pelo próprio autor, em um repositório digital. Em conformidade com Laakso (2014), o “verde”, nesse contexto, vem da noção de que as editoras dão um “sinal verde” para que sejam disponibilizadas abertamente cópias do conteúdo de seus artigos. Logo, as cópias podem ser depositadas em repositórios digitais, que subdividem-se em repositórios institucionais ou temáticos.

Um Repositório Institucional é definido por Shreeves e Cragin (2008, p. 89, *tradução nossa*²³) como “um conjunto de serviços e tecnologias que propiciam os meios para coletar, gerenciar, acessar, disseminar e preservar materiais digitais produzidos em uma instituição”. Lynch (2003) caracteriza repositório institucional, em termos de repositórios mantidos por universidades, como um conjunto de serviços para armazenamento e disseminação de materiais digitais, que uma universidade oferece para os membros de sua comunidade. No mesmo sentido, Westell (2006, p. 213), afirma que um repositório institucional pode ser entendido como “uma coleção acessível de trabalhos acadêmicos que representa o capital intelectual de uma universidade”. Muitos desses repositórios são mantidos por faculdades e universidades, de fato, mas também podem ser administrados por agências governamentais, corporações e outras organizações.

O repositório temático, ou disciplinar, também se caracteriza por fornecer os meios para armazenar e disseminar informação digital, mas se diferencia do repositório institucional por possuir coleções de documentos pertencentes a uma disciplina (ou conjunto de disciplinas) científica específica. No entendimento de Björk (2014), repositórios temáticos também podem conter metadados de artigos, bem como dados de pesquisa. Para o autor, no entanto, os repositórios mais interessantes fornecem acesso livre ao texto completo de publicações acadêmicas e permitem que elas estejam localizáveis ou pesquisáveis na web. Bernius (2010) sugere que os repositórios temáticos fornecem as melhores condições para recuperação do conhecimento científico.

O primeiro repositório temático na Internet, o ArXiv, foi criado em 1991 por Paul Ginsparg. De acordo com Ginsparg (2004), o ArXiv foi concebido como uma interface de e-mail para criar, preservar e acessar documentos para especialistas em Física. Com os avanços das tecnologias, o repositório evoluiu e ganhou uma interface web. Atualmente, o ArXiv é hospedado pela Universidade Cornell, em Nova Iorque, nos Estados Unidos. Fornece mais de um milhão e meio de documentos eletrônicos em acesso aberto nos campos de Física, Matemática, Ciência da Computação, Biologia Quantitativa, Estatística, Engenharia Elétrica, Ciência de Sistemas e Economia. Desde então, outras iniciativas surgiram para fornecer acesso aberto a publicações científicas de áreas específicas, como o PubMed Central, o CiteSeerX, o BiorXiv, SocArXiv e o

23 “An IR is a set of services and technologies that provide the means to collect, manage, provide access to, disseminate, and preserve digital materials produced at an institution” (SHREEVES E CRAGIN, 2008, p. 89)

E-LIS. Para situar algumas dessas iniciativas, segue-se uma pequena introdução sobre essas bases de dados.

O PubMed Central é um repositório de textos completos, gratuitos, da literatura de periódicos das Ciências da Vida e Biomédicas da National Library of Medicine (NLM). O PubMed Central foi desenvolvido por volta dos anos 2000 e é gerenciado pelo Centro Nacional de Informações sobre Biotecnologia da NLM. Atualmente, contém mais de cinco milhões de registros (PUBMED..., 2019).

O CiteSeerX é uma biblioteca digital e um mecanismo de busca para literatura científica na área de Ciência da Computação. Além de ser um repositório de textos completos, o CiteSeerX busca fornecer recursos como algoritmos, dados, metadados, serviços, técnicas e *softwares* que podem ser usados para promover outras bibliotecas digitais. Criado em 1997 no NEC Research Institute, em Princeton, Nova Jersey, o serviço foi transferido para a Faculdade de Ciências da Informação e Tecnologia da Universidade Estadual da Pensilvânia em 2003. Desde então, o projeto é liderado pelo professor Lee Giles. O serviço indexa, atualmente, mais de 1,5 milhões de documentos (CITeseerX, 2019).

O BiorXiv é um serviço on-line gratuito de arquivamento e distribuição de *preprints* nas Ciências da Vida. Mais restritivo que os repositórios citados anteriormente, o BiorXiv só aceita a submissão de *preprints*. A todos eles são designados um DOI, o que permite que esses documentos sejam citados. É operado pelo Cold Spring Harbor Laboratory em Nova Iorque, uma instituição educacional e de pesquisa sem fins lucrativos. Lançado em 2013, é um serviço novo, mas popular entre os pesquisadores da área. Até setembro de 2019, mais de 57 mil documentos foram depositados nesse repositório (BIORXIV, 2019).

Por fim, o SocArXiv é um arquivo aberto das Ciências Sociais, que fornece uma plataforma de acesso aberto, sem fins lucrativos, gratuita para cientistas sociais depositarem seus trabalhos, *preprints* e documentos publicados, com a opção de vincular dados e códigos-fonte. Esse repositório está localizado na Universidade de Maryland, nos Estados Unidos e é dirigido por um comitê de cientistas sociais e membros da comunidade de bibliotecas de pesquisa. Criado em 2016, conta atualmente com cerca de 4.500 documentos disponíveis. (SOCARVIX..., 2019).

Mas como funciona o autoarquivamento? Ele consiste no depósito de um documento em um repositório digital, por isso, o autor precisa selecionar onde disponibilizará sua produção científica. Kim (2010) apresenta algumas possibilidades, como páginas pessoais na web, *sites* de grupos de pesquisa, *sites* departamentais, repositórios temáticos e repositórios institucionais. Ao fazer essa escolha, o autor não precisa escolher apenas uma das opções, podendo fazer o depósito do mesmo trabalho em mais de um local.

Além dessa escolha, existem outras variáveis envolvidas e sobre as quais o autor deve ponderar. Entre elas estão os custos envolvidos na prática do autoarquivamento; fatores

contextuais; bem como, benefícios extrínsecos e intrínsecos. Quando se fala em custos, não se trata de valor monetário, mas de preocupações com copyright, que podem envolver a necessidade de requerer permissão para a realização do depósito do documento, possibilidades de infringimento de copyright e, até mesmo, atrito entre os autores do trabalho e o desacordo por uma das partes sobre a realização do depósito. O custo também pode estar relacionado com o tempo e esforço despendidos na atividade de autoarquivamento (KIM, 2010).

Os fatores contextuais envolvem a confiança, a cultura do autoarquivamento e a influência de fatores externos. A confiança diz respeito à percepção dos autores de que seus direitos não estão garantidos se eles disponibilizarem seus trabalhos livremente na Internet, porque eles dependem das boas intenções dos leitores em lhes atribuir a autoria. Outro problema relacionado à confiança é se a manutenção dos documentos depositados será feita pelo repositório que os têm sob custódia (KIM, 2010). A cultura do autoarquivamento advém do entendimento da importância que essa prática tem para a comunicação científica. Chan (2004) indica que a inércia cultural é frequentemente citada como a razão para a lenta adoção do autoarquivamento; outro motivo comum é a falta de consciência da importância do acesso aberto. Dessa forma, é difícil criar uma cultura na qual essa prática seja comum.

O autor pode se questionar, ainda, sobre os benefícios de realizar o autoarquivamento. Esses benefícios podem ser extrínsecos e/ou intrínsecos, conforme Kim (2010). Os extrínsecos têm relação com a acessibilidade, a publicidade, a confiabilidade, a recompensa acadêmica e o reconhecimento profissional que advém do autoarquivamento. Segundo a percepção de alguns pesquisadores, os benefícios percebidos têm relação com tornar sua produção disponível de maneira segura e estável; aumentar o número de leitores e o alcance de suas publicações, gerando citações; quanto mais pesquisadores depositarem seus trabalhos em repositórios digitais, mais confiabilidade será dada aos documentos autoarquivados; tornar seu trabalho disponível abertamente pode auxiliá-lo na obtenção de títulos e/ou promoções, na forma de recompensa acadêmica; e, por fim, disponibilizar sua produção científica pode aumentar sua visibilidade no seu campo de pesquisa, gerando reconhecimento profissional (KIM, 2010).

O benefício intrínseco, por seu turno, está relacionado com o altruísmo e a reciprocidade, isto é, com o que motiva o autor a praticar o autoarquivamento. Para os pesquisadores entrevistados por Kim (2010), o autoarquivamento possibilitaria o acesso gratuito e sem restrições a documentos para países em desenvolvimento, promovendo a igualdade no acesso à informação; além disso, os entrevistados afirmaram que poder ajudar outros pesquisadores era a principal razão para realizarem o autoarquivamento. Pesquisadores que já tinham usado trabalhos autoarquivados, e se beneficiaram da prática, se sentiam motivados a depositar seus trabalhos também. No caso de pesquisadores de universidades públicas, esses se sentiam na obrigação de autoarquivar seus trabalhos, pois em seu entendimento as pesquisas feitas em universidades públicas com recurso público devem servir ao bem público. De maneira geral,

alguns professores afirmaram que o compartilhamento de informações é obrigação de todo pesquisador (KIM, 2010).

Mais do que apenas depositar o trabalho em um repositório, o autor deve estar ciente de seus direitos sobre o manuscrito e o artigo. Algumas editoras, por exemplo, propõem um acordo para a transferência dos direitos autorais do autor para a editora e, então, proíbem que um artigo publicado em sua revista seja, também, disponibilizado em acesso aberto. Em outros casos, uma editora pode colocar um período de embargo sobre suas publicações, permitindo que, após um tempo determinado, o autor disponibilize seu artigo em um serviço de sua preferência. Esses acordos têm a função de definir o que o autor pode fazer com o manuscrito no qual o artigo é baseado, uma vez que o acordo é aceito. Sobre isso, Laakso (2014) afirma que há uma tensão entre os interesses do autor em tornar sua pesquisa disponível e ampliar seu público, e as editoras que se esforçam para manter seus modelos de negócios atuais baseados em assinatura.

No entanto, na área de Biblioteconomia e Ciência da Informação esses atritos entre autor e editora não parecem ser tão fortes como em outras áreas, conforme afirmam Coleman (2007), Emery (2018) e Chilimo e Onyancha (2018). Em consonância com Coleman (2007), em seu estudo sobre o autoarquivamento e a existência de acordos de transferência de copyright em periódicos de Biblioteconomia e Ciência da Informação, 90% de 52 periódicos classificados na Thomson-Scientific (ISI)²⁴, não proíbe, de nenhuma maneira, o autoarquivamento. Em outro recorte, Emery (2018) analisa periódicos de Biblioteconomia e Ciência da Informação da editora Taylor & Francis, que permite o auto-depósito de artigos publicados em suas revistas. Nele, a autora descobriu que a taxa de depósito em Biblioteconomia e Ciência da Informação é de aproximadamente 22%. Para a autora, esse é um valor abaixo do esperado (50%), visto que a própria editora incentiva esse depósito e até realizou campanhas nesse sentido. Chilimo e Onyancha (2018), ao investigarem o grau de abertura da pesquisa em Acesso Aberto na área de Biblioteconomia e Ciência da Informação, descobriram, a partir das políticas de acesso aberto de 56 periódicos da área, que 44,1% dos artigos, em 28 dos 56 periódicos, foram publicados em revistas que autorizam o autoarquivamento da publicação em repositórios digitais.

Em conformidade com o exposto por Chilimo e Onyancha (2018), as políticas das revistas investigadas colocam a responsabilidade de garantir o acesso aberto exclusivamente nas mãos dos autores, seja por meio do pagamento de uma taxa para tornar seus artigos abertos (um modelo híbrido de Acesso Aberto) ou pelo autoarquivamento das versões posteriores à publicação de seus textos, o que significa que os artigos só estarão disponíveis abertamente se os autores tomarem a iniciativa. Coleman (2007), afirma que se houvesse um repositório temático em Biblioteconomia e Ciência da Informação disponível e, se cada autor depositasse

24 Thomson-Scientific, anteriormente conhecida como Institute for Scientific Information (ISI), oferece abordagens analíticas e dados cientométricos com base na Web of Science. Em 2016, a divisão científica da Thomson Reuters foi vendida para a Clarivate Analytics e o ISI foi reestabelecido.

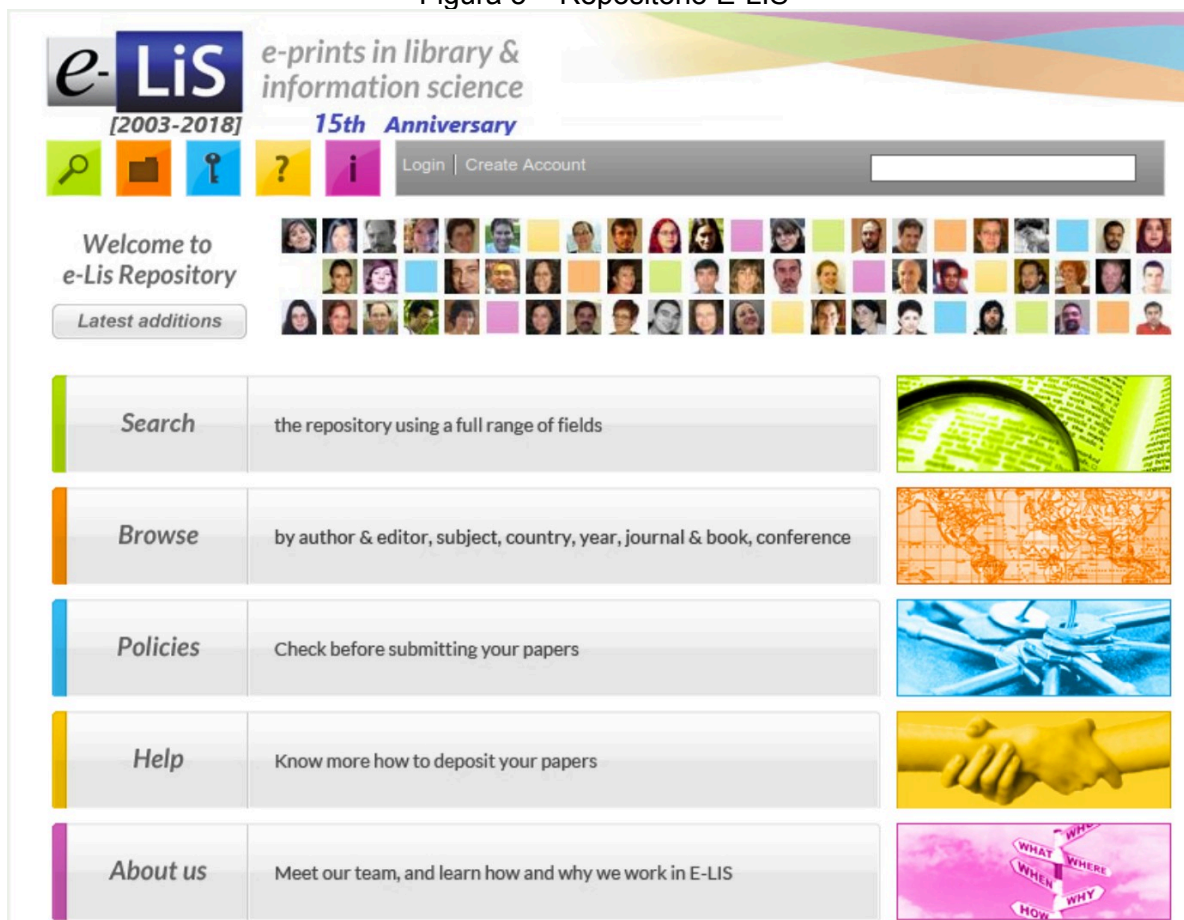
seu artigo nele, os resultados poderiam chegar próximo à marca de 50%. Para a autora, os pesquisadores deveriam se sentir encorajados a autoarquivar, uma vez que poucos periódicos proíbem. Para Emery (2018), o campo de Biblioteconomia e Ciência da Informação deve se unir para construir um espaço comum da área, para desenvolver serviços de repositórios digitais federados e de acesso aberto para a disciplina. Contudo, em que pese o exposto acima, entende-se que, felizmente, um repositório aberto de Biblioteconomia e Ciência da Informação já existe.

3.3.2 *E-prints in Library and Information Science (E-LIS)*

Em meio as iniciativas abertas que surgiram no início dos anos 2000, seguindo os postulados do Movimento de Acesso Aberto e da Iniciativa de Arquivos Abertos, o repositório E-LIS foi criado. O E-LIS (Figura 5) é um repositório digital para documentos relacionados à Biblioteconomia e à Ciência da Informação, hospedado pelo Sistema de Biblioteca (CAB) da Universidade de Nápoles Federico II, na Itália. Esse repositório foi estabelecido e é gerenciado e mantido por uma equipe voluntária de bibliotecários e cientistas da informação pertencentes a cerca de 45 países, possuindo suporte para 22 idiomas (E-LIS, 2020).

De acordo com Santillán-Aldana (2009), esse é o primeiro servidor eletrônico aberto especializado em campos relacionados à Biblioteconomia e a Ciência da Informação. Le Coadic (2004), por sua vez, cita o E-LIS como um dos principais recursos na Internet sobre Ciência da Informação e Documentação. Seu desenvolvimento foi estimulado pelo conceito de Acesso Aberto e funciona por meio de autoarquivamento realizado pelos autores. Os repositórios digitais de acesso aberto devem estar de acordo com as diretrizes propostas pela Iniciativa de Arquivos Abertos, que promovem os padrões de interoperabilidade e objetivam facilitar e tornar eficiente a disseminação dos documentos.

Figura 5 – Repositório E-LIS



Fonte: E-LIS (2019).

Para entender um pouco melhor sobre os repositórios digitais e como eles se relacionam com o acesso aberto, faz-se necessário compreender como se constitui a sua infraestrutura. Essa infraestrutura, regida pela OAI e baseada no protocolo OAI-PMH, permite a interoperabilidade entre repositórios digitais. Em consonância com Costa e Leite (2017, p. 48),

[e]sse protocolo de comunicação permite aos sistemas compartilhar metadados (dados que descrevem os recursos de informação) sem processos intermediários. Para tanto os sistemas são caracterizados como provedores de dados (sistemas que dispõem os metadados) e provedores de serviço (sistemas que coletam os metadados). A iniciativa [OAI] foi essencial para o desenvolvimento de sistemas de acesso aberto porque definiu os padrões e o protocolo que garantem a interoperabilidade entre sistemas.

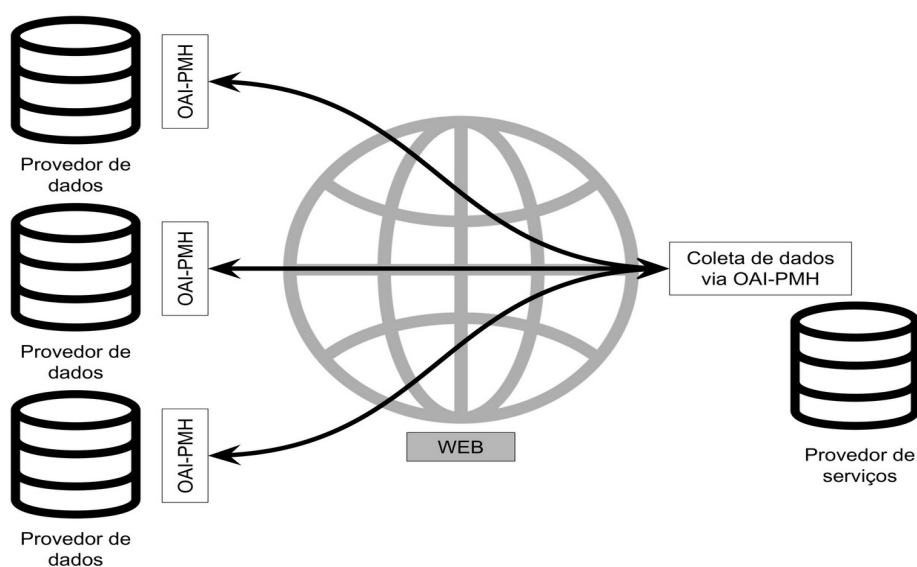
Para promover essa interoperabilidade, os repositórios têm sua infraestrutura composta de três elementos principais: 1) os provedores de dados; 2) os provedores de serviços; e 3) o protocolo OAI-PMH. Os dois provedores comunicam-se entre si por causa do protocolo (WEITZEL; LEITE; ARELLANO, 2008). É possível observar esse funcionamento na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). Nessa iniciativa, as IES exercem o papel de

provedores de dados. As teses e dissertações são depositadas nos repositórios institucionais de cada IES. Por meio do protocolo OAI-PMH, os dados dessas produções são coletados e disponibilizados por um provedor de serviço. No caso da BDTD, o provedor de serviço é o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) (BAPTISTA *et al.*, 2007). Esse exemplo e a infraestrutura do protocolo OAI-PMH são representados, de maneira mais geral, na Figura 6.

Baptista e outros autores (2007) assinalam a existência de ferramentas de acesso aberto e código aberto que implementam o OAI-PMH por padrão, algumas delas são: o EPrints, o DSpace e o Fedora. Esses *softwares* permitem a criação de repositórios digitais altamente configuráveis. O E-LIS é mantido com o Eprints, o que lhe atribui as características de infraestrutura e interoperabilidade apresentadas anteriormente.

Para além da infraestrutura, os repositórios digitais que implementam o OAI-PMH e servem como arquivos de documentos eletrônicos devem contar com: a) um mecanismo de submissão; b) um sistema de armazenamento a longo prazo; c) uma política de gestão para a submissão e preservação de documentos; e d) uma interface aberta que permita a terceiros coletar os metadados dos respectivos arquivos (BAPTISTA *et al.*, 2007). A seguir, por meio da consulta no E-LIS (2020) e informações retiradas do processo empírico de criação de conta e submissão de um documento, são descritos os elementos que devem constar em um repositório digital.

Figura 6 – Infraestrutura do protocolo OAI-PMH



Fonte: Elaborado pela autora (2020).

mecanismo de submissão do E-LIS só pode ser acessado por meio da criação de uma conta. Essa conta além de permitir o depósito de documentos no repositório, possibilita salvar buscas e

receber alertas. Para criar uma conta, será necessário informar: nome e sobrenome, endereço de e-mail, nome de usuário e uma senha. Para concluir o processo, o usuário receberá um e-mail de confirmação.

Para submeter um documento, o usuário fará *login* com o nome de usuário criado e sua senha. O processo de submissão é simples e intuitivo. Assim que o usuário entrar com sua conta, haverá um botão “*New item*”. No geral, será um processo de três passos: incluindo questões básicas sobre o tipo de documento, descrição bibliográfica e *upload* do documento.

Figure 7 – Parte do formulário de submissão de documentos do E-LIS

The screenshot shows the 'Details' step of the document submission process. The form is titled 'Edit item: Preprint #41846'. The progress bar indicates the current step is 'Details', with previous steps being 'Type' and 'Upload', and subsequent steps being 'European Union projects' and 'Deposit'. The 'Details' section includes several fields:

- Refereed:** Radio buttons for 'Yes, this version has been refereed.' and 'No, this version has not been refereed.'
- Public domain:** Radio buttons for 'Yes, it is public domain.' and 'No.'
- Authors:** A table with columns for 'Family Name', 'Given Name / Initials', and 'Person ID'. It has 4 rows and a 'More input rows' button.
- Editors:** A table with columns for 'Family Name', 'Given Name / Initials', and 'Person ID'. It has 4 rows and a 'More input rows' button.
- Title:** A large text input field.
- Subjects:** A search box for subjects, with a list of subjects below it:
 - A. Theoretical and general aspects of libraries and information.
 - B. Information use and sociology of information
 - C. Users, literacy and reading.
 - D. Libraries as physical collections.
 - E. Publishing and legal issues.
 - F. Management.
 - G. Industry, profession and education.
 - H. Information sources, supports, channels.
 - I. Information treatment for information services

Fonte: E-LIS (2020).

Na primeira página, o usuário selecionará o tipo de documento que deseja submeter (*preprint*, artigo de periódico, livro, capítulo de livro, artigo de conferência, entre outros). Na página seguinte (Figura 7) serão preenchidos os detalhes sobre o documento: o documento passou pela revisão de pares? Sim ou não; o documento está em domínio público? Sim ou não; nome dos autores; título do trabalho; assuntos, mais de um pode ser adicionado; data de publicação; resumo em inglês (*abstract*); resumo em língua vernácula; URL para dados de pesquisa relacionados à publicação; localização alternativa do documento (URL para outras plataformas nas quais o trabalho está disponível); palavras-chave; idioma do documento; país

dos autores; e a lista de referências. Na próxima página, será feito o *upload* do documento. Para garantir o acesso a longo prazo, os documentos devem ser submetidos em Formato de Documento Portátil (Portable Document Format – PDF) ou em Linguagem de Marcação de Hipertexto (HyperText Markup Language – HTML). Por fim, o usuário deve selecionar sob qual licença Creative Commons seu trabalho será disponibilizado (E-LIS, 2020).

O procedimento deve ser minucioso, pois, uma vez que o trabalho é depositado, as informações só poderão ser atualizadas pela equipe editorial do E-LIS. Antes disso, o processo pode ser salvo e recuperado em outro momento, se o usuário precisar de tempo para revisar as informações. Quando a submissão for finalizada, um editor revisará a submissão para verificar se a mesma é adequada para o repositório. Caso o documento não seja aprovado, o usuário receberá um e-mail com os motivos. Se o usuário desejar, pode também realizar o depósito por e-mail. Basta enviar o documento em um dos formatos sugerido e encaminhar uma lista com as informações bibliográficas, isto é, os metadados, para o endereço eprints@dois.it e a equipe realizará o depósito. O usuário terá a chance de comentar sobre o documento antes da publicação final no repositório (E-LIS, 2020).

Conforme o exposto, o E-LIS possui um mecanismo de submissão que dá conta de descrever documentos bibliográficos, bem como, há uma preocupação em ser um **sistema de armazenamento a longo prazo**. Sua **política de submissão** solicita que as publicações depositadas sejam da área de Biblioteconomia, Ciência da Informação e Tecnologia. Os autores devem se registrar no repositório para fazer o depósito de documentos. O repositório aceita documentos técnico-científicos publicados e não-publicados, bastando que o documento seja relevante para a área e possua a forma de um documento finalizado. Os documentos passam pela avaliação dos editores do E-LIS para que os metadados tenham sido inseridos em conformidade com o documento (E-LIS, 2020). E, por fim, os autores são responsáveis por confirmar se o documento pode, de fato, ser disponibilizado em acesso aberto, cabendo aos próprios verificarem as questões de direitos autorais.

A **política de preservação** de documentos recomenda fortemente que os autores façam a submissão dos documentos em PDF ou HTML, embora outros formatos também sejam suportados como o Microsoft Word (.doc), Microsoft Powerpoint (.ppt), Microsoft Excel (.xls) e Texto (.txt). Por “suportado”, entende-se que o E-LIS tornará o documento utilizável no futuro, usando qualquer combinação de técnicas (como migração, emulação etc.) que seja apropriada, dado o contexto de necessidade. Os formatos suportados são: PDF, HTML e txt. Nem todos os formatos proprietários podem ser suportados. No entanto, esses arquivos ainda serão preservados. É provável que, para formatos extremamente populares, mas proprietários como .doc, .xls e .ppt, o E-LIS não possa garantir completo suporte (E-LIS, 2020).

Por fim, em conformidade com o protocolo OAI, o E-LIS possui uma **interface aberta que permite a terceiros coletar os metadados** do repositório. A OAI desenvolveu um código

compartilhado para descrição de metadados (ex.: *date, author, title, journal* etc.). Assim, mesmo que os documentos estejam em diferentes formatos e localizações, se todos utilizam os mesmos descritores de metadados, eles se tornam interoperáveis. Os metadados podem, então, ser coletados e todos os documentos podem ser buscados e recuperados de uma só vez, como se estivessem em uma coleção global, acessíveis (E-LIS..., 2020). Por possuir essa característica, será possível coletar os dados do repositório e realizar as análises no presente trabalho.

Por fim, realiza-se uma breve análise quanto à presença do E-LIS na literatura de Ciência da Informação. Foram localizados trabalhos que o apresentam como a iniciativa de um repositório de acesso aberto na área de Ciência da Informação, explicam sua infraestrutura e funcionalidades, e o utilizam como fonte de informação para Revisão Sistemática de Literatura ou para estudos bibliométricos. Alguns desses estudos são apresentados a seguir.

O primeiro trabalho a mencionar o E-LIS foi publicado em 2003. Nele, Barrueco-Cruz e Subirats-Coll (2003) abordam a criação do E-LIS como parte do projeto Research in Computing, Library and Information Science (RCLIS), apresentam duas outras iniciativas de arquivos abertos para a Biblioteconomia e Ciência da Informação, o @rchiveSIC²⁵ e o Digital Library of Information Science and Technology (DLIST)²⁶, mas afirmam a necessidade de se criar um verdadeiro esforço internacional sem barreiras geográficas ou de idiomas. Os trabalhos seguintes, seguiram uma linha de campanha de divulgação para o E-LIS, para que trabalhos da área fossem depositados no repositório, além de divulgar as perspectivas do E-LIS em diferentes países (DE ROBBIO, 2003; MEDEIROS, 2004; SUBIRATS-COLL; ARENCIBIA-JORGE; DE ROBBIO, 2004; ARENCIBIA-JORGE; SANTILLÁN-ALDANA; SUBIRATS-COLL, 2005).

No Brasil, Moreno, Leite e Arellano (2006) e Weitzel (2006), ao abordarem o acesso livre a publicações e o papel dos repositórios digitais na comunicação científica, citam o E-LIS como um dos repositórios que está ganhando atenção internacionalmente. Ao reconhecer a comunidade desse repositório, Weitzel, Leite e Arellano (2008, p. 14) afirmam que a “disseminação do conteúdo dos periódicos brasileiros da área no E-LIS é fundamental para a visibilidade da produção científica nacional”.

O crescimento da coleção do E-LIS pode ser observado nos artigos subsequentes. Entre 2003 e 2007, o repositório já contava com mais de cinco mil documentos. A partir desse conteúdo, Morrison *et al.* (2007) fazem uma avaliação mais ampla das funcionalidades de busca do repositório. Em 2009, cerca de nove mil documentos estavam disponíveis (DE ROBBIO;

25 O @rchiveSIC é um arquivo francês, aberto, da Ciência da Informação e da Comunicação, criado em 2002. Possui atualmente 3.088 documentos com texto integral. Está disponível em: <http://archivesic.ccsd.cnrs.fr/>

26 O DLIST é um arquivo digital interinstitucional, disciplinar e de acesso aberto para Ciência da Informação, Biblioteconomia, Sistemas de Informação, Curadoria Digital, Informática de Museu, entre outros. Também foi criado em 2002. Está disponível em: <http://dlist.sir.arizona.edu>.

KATZMAYR, 2009). Neves e Ferreira (2014) afirmam, por seu turno, que já haviam mais de 16 mil documentos sendo disponibilizados em 2014. Atualmente, o repositório conta com mais de 21 mil documentos.

Outros trabalhos, utilizam o E-LIS como fonte de informação, seja para revisão sistemática de literatura (PUJOL; VIVÓ, 2010; MILLER, 2017), seja para estudos bibliométricos (NASEER; MAHMOOD, 2009; SANTILLÁN-ALDANA, 2009; NEVES; FERREIRA, 2014; MENÊSES, 2019).

É possível observar que muitos dos trabalhos citados foram realizados por editores do E-LIS, isto é, pessoal envolvido diretamente na manutenção do repositório. Sem tirar o mérito dessa produção, ressalta-se que o E-LIS também seja validado por pesquisadores fora dessa esfera.

4 ESTUDOS MÉTRICOS DA INFORMAÇÃO

Os Estudos Métricos da Informação são uma subárea da Ciência da Informação. De acordo com Oliveira (2018, p. 35-36), eles “representam práticas de mensuração da informação em suas diferentes modalidades, na web ou nas invenções tecnológicas, porém tendo em comum o uso de quantificações, diferenciando-se especialmente em seus objetos de estudo”. Vanti (2007), nessa mesma perspectiva, afirma que os métodos quantitativos para avaliação dos fluxos de informação e do conhecimento são variados e possuem enfoques e suportes distintos, como é o caso da bibliometria, da cientometria, da informetria, da webometria e da cibermetria.

Entretanto, as análises estatísticas sobre a literatura científica, segundo Glänzel (2003), precedem a criação do termo “Bibliometria”, e se originaram com a Lei da Produtividade (Lotka), a Lei da Dispersão Bibliográfica (Bradford) e a Lei de Zipf, agora conhecidas como as três leis clássicas da Bibliometria. Essas leis estruturaram o corpo teórico que fundamenta o *status* de ciência dos Estudos Métricos da Informação.

Para Vanti (2007, p. 45), a Lei de Lotka ou Lei do Quadrado Inverso é aquela que mensura a produtividade dos autores, por meio de “um modelo de distribuição tamanho-frequência dos diversos autores em um conjunto de documentos”. O postulado dessa lei admite que “grande quantidade de literatura científica é produzida por um pequeno número de autores e um grande número de pequenos produtores se iguala, em produção, ao reduzido número de grandes produtores” (NORONHA; MARICATO, 2008, p. 125). A Lei de Bradford, por sua vez, estuda a dispersão da publicação de artigos em periódicos. Conforme Vanti (2007), essa lei permite, a partir da medição da produtividade das revistas, “estabelecer o núcleo e as áreas de dispersão sobre um determinado assunto em um mesmo conjunto de revistas”. Por fim, a Lei de Zipf, também denominada Lei do Mínimo Esforço, “consiste em medir a frequência do aparecimento das palavras em vários textos, gerando uma lista ordenada de termos de uma determinada disciplina ou assunto” (VANTI, 2007, p. 45).

O conceito de bibliometria, supostamente, aparece na literatura pela primeira vez em Pritchard (1969). O autor buscava um termo que pudesse substituir o denominado “bibliografia estatística”, utilizado até aquele momento. A expressão era considerada insatisfatória pelo autor e outros pesquisadores, pois não era descritiva o suficiente e poderia ser confundida com a própria estatística ou com bibliografias em estatística. Isto posto, Pritchard (1969, p. 349, *tradução nossa*²⁷) define a bibliometria (*bibliometrics*) como sendo “a aplicação de métodos matemáticos e estatísticos a livros e outros meios de comunicação”.

27 “[...] the application of mathematics and statistical methods to books and other media of communication” (PRITCHARD, 1969, p. 349).

Contudo, Fonseca (1986, p. 12) afirma que há um “lamentável desconhecimento mútuo entre autores de línguas neolatinas e anglo-saxônicas”, uma vez que em francês já existia um termo precedente introduzido por Paul Otlet em 1934, o “*bibliometrie*”. De acordo com o Alvarado (2007), Otlet estava interessado na construção de uma nova disciplina científica chamada de ‘Bibliologia’, que ele define como “uma ciência geral que abranja o conjunto sistemático, ordenado, dos dados relativos à produção, conservação, circulação e utilização dos escritos e documentos de todo tipo” (OTLET, 2018, p. 11). Para o autor, a ‘bibliometria’ seria, então, uma “parte definida da bibliologia que se ocupa da medida ou quantidade aplicada aos livros (aritmética ou matemática bibliológica)” (OTLET, 2018, p. 18).

Em que pese o exposto, os estudos bibliométricos não se restringem às análises de livros, mas também levam em consideração artigos de periódicos, teses e dissertações, relatórios de pesquisas, entre outros documentos. Desse modo, eles tendem a utilizar “a publicação como medida do fluxo da Ciência” (ALVARENGA, 1996, p. 75). Conforme Souza e Oliveira (2007, p. 76), a bibliometria possibilita “a definição de uma série de indícios característicos da produção científica estudada”.

A evolução dos estudos teóricos no campo das métricas da informação e dos recursos tecnológicos ampliaram o universo dos estudos métricos, o que proporcionou, segundo Noronha e Maricato (2008, p. 123), “o surgimento de novas abordagens e das variáveis possíveis de serem analisadas”. Assim, retomando os métodos citados anteriormente, manifestam-se a cientometria, a informetria, a webometria e a cibermetria com diferentes aplicações.

De forma breve, a **cientometria** consiste no estudo dos aspectos quantitativos da ciência na condição de uma disciplina ou atividade econômica. Já a **informetria** é o estudo dos aspectos quantitativos da informação em qualquer formato (MACIAS-CHAPULA, 1998). Em vista disso, ela tem como objeto de estudo palavras, documentos, bases de dados e considera variáveis como recuperação, relevância e revocação, por exemplo. No que diz respeito à **webometria**, ela se caracteriza como “o estudo dos aspectos quantitativos da construção e uso dos recursos informacionais, estruturas e tecnologias da informação na web, utilizando abordagens bibliométricas e informétricas” (BJÖRNEBORN, 2004, p. 12, *tradução nossa*²⁸), como domínios, hiperlinks e URL. Quanto à **cibermetria**, Björneborn (2004, p. 13, *tradução nossa*²⁹) a caracteriza como “o estudo dos aspectos quantitativos da construção e uso de recursos, estruturas e tecnologias da informação em toda a Internet, utilizando abordagens bibliométricas e

28 “*The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the web, drawing on bibliometric and informetric approaches*” (BJÖRNEBORN, 2004, p. 12).

29 “*The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the whole Internet, drawing on bibliometric and informetric approaches*” (BJÖRNEBORN, 2004, p. 13).

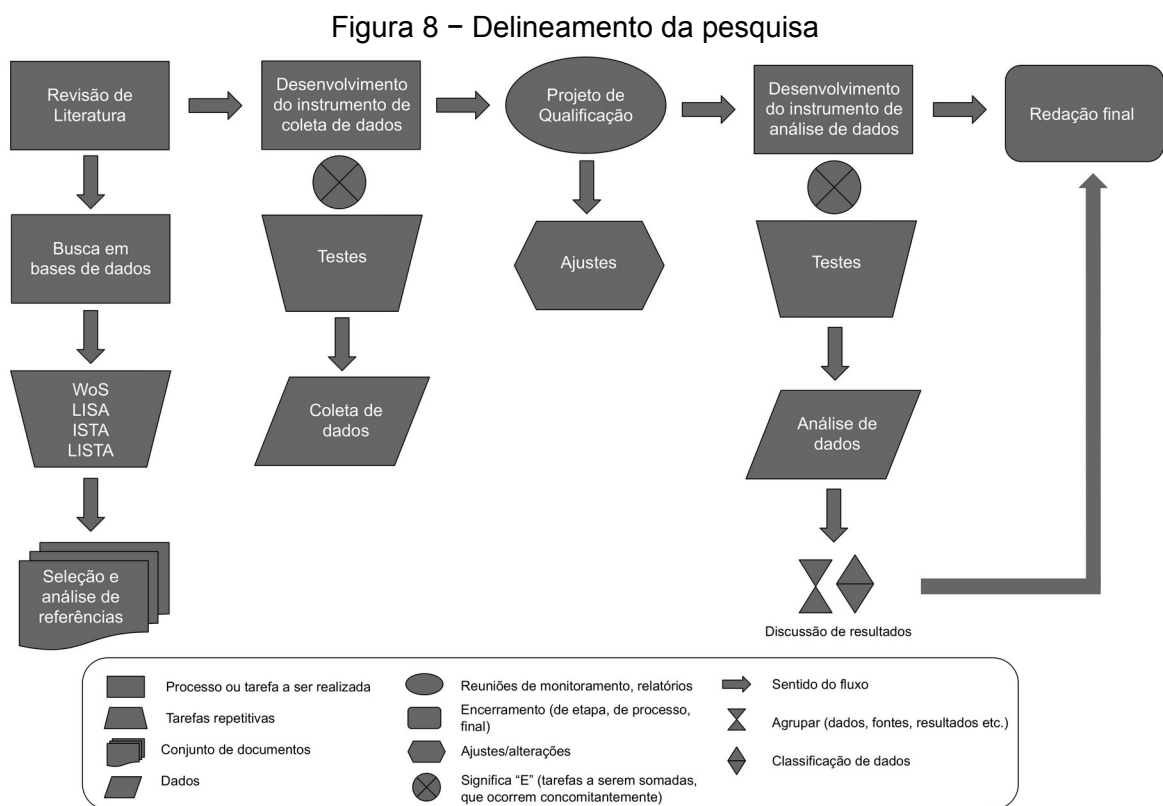
informétricas”, de forma que abrange estudos estatísticos de grupos de discussão, listas de discussão e outras comunicações mediadas por computador na Internet.

Apesar das diversas aplicações, o uso dos estudos bibliométricos não deve ser feito sem uma contextualização da área em que serão aplicados e os mesmos podem ser utilizados com uma ou mais abordagens, de forma a combiná-las. Souza e Oliveira (2007, p. 76-77) afirmam que, para além dos aspectos quantitativos, “a definição de uma série de indícios característicos da produção científica evidenciam [...] as relações interdiscursivas entre os enunciados que compõem a formação discursiva de uma determinada área do conhecimento”. O aspecto quantitativo dos Estudos Métricos da Informação “não anula seu precípuo interesse pela coletividade, mas dá segurança cognitiva para a argumentação científica desta” (OLIVEIRA, 2018, p. 46).

A realização de estudos bibliométricos pressupõe, atualmente, o uso de tecnologias de informação, pela agilidade que imprime em sua efetuação. As análises quantitativas se beneficiam das facilidades oferecidas por essas ferramentas, o que permite investigar quantidades cada vez maiores de dados, com grande rapidez e em menor tempo. Esse ambiente constitui espaço oportuno para se discutir os desenvolvimentos da Ciência Aberta, já que pode ser considerada como “o fio condutor de investigações científicas apoiadas por uma ciberinfraestrutura tecnológica e metodológica que permite o uso, reuso (*sic*) e reprodutibilidade de dados de pesquisa” (OLIVEIRA; SILVA, 2016, p. 6).

5 PERCURSO METODOLÓGICO

A metodologia de pesquisa, conforme Minayo *et al.* (2002, p. 54), “inclui as concepções teóricas de abordagem, o conjunto de técnicas que possibilitam a construção da realidade e o sopro divino do potencial criativo do investigador”. A pesquisa articula aspectos teóricos e práticos na resolução de um questionamento. Com vistas a atender os objetivos do presente trabalho, os procedimentos adotados foram representados no delineamento da pesquisa (Figura 8) e são detalhados, abaixo:



Fonte: Elaborado pela autora (2020).

Esta seção é composta pela descrição do processo busca e seleção de fontes para a Revisão de Literatura, seguida da caracterização da pesquisa, apresentação do ambiente em que esta se deu e o universo da pesquisa selecionado. Posteriormente, foi discutida a criação do instrumento de coleta e de análise de dados, o ELIScript.

5.1 Estratégia de busca para revisão de literatura

A Revisão de Literatura deste trabalho não teve como pretensão esgotar os assuntos abordados. Isso quer dizer que ela se caracterizou, pelo seu propósito, como uma Revisão de

Literatura de Base, que Noronha e Ferreira (2000) definem como uma revisão que subsidia ou serve de apoio ao desenvolvimento de um trabalho científico, neste caso, a dissertação. Aqui, optou-se pela estruturação de uma estratégia de busca e utilização de critérios específicos para a recuperação e seleção de documentos relevantes para o trabalho.

Na Revisão de Literatura realizada, todo o processo de busca foi guiado pela formulação de uma pergunta. Essa pergunta baseou-se na questão e no problema de pesquisa do presente trabalho e foi quebrada em conceitos, que permitiram selecionar os descritores utilizados. Em seguida, foram levantados os termos sinônimos e os correspondentes em inglês. Os descritores selecionados foram: “repositório digital”, “repositório temático”, “via verde”, “autoarquivamento”, “acesso aberto”, “movimento de acesso aberto”, “E-Prints in Library and Information Science”, “E-LIS” e “comunicação científica”. Optou-se pela busca dos termos em inglês nas bases internacionais, com uso de aspas em termos compostos para a recuperação da frase exata, o uso de asterisco (*) para truncagem de termos e as estratégias booleanas *AND* e *OR*, para restringir e ampliar a busca, respectivamente.

Desta forma, realizou-se a busca na WoS e nas bases de dados específicas da área de Ciência da Informação, a saber: Library, Information Science & Technology Abstracts with Full Text (LISTA), Library and Information Science Abstracts (LISA) e na Information Science and Technology Abstracts (ISTA). A estratégia de busca foi aplicada nos campos de títulos, resumos e palavras-chave, por representarem os temas principais dos artigos. Como as bases citadas não apresentaram muitos resultados sobre o E-LIS, efetuou-se a busca sobre esse tópico, também, no Google Acadêmico. Na WoS, aplicou-se o refinamento por produções em *Information Science Library Science* e a tipologia de documento selecionada foi *article*. Nas bases internacionais de Ciência da Informação, os resultados foram filtrados por “Revistas acadêmicas” ou “Revisado por especialistas”, que indicam os artigos revisados por pares.

No Quadro 1 são apresentados os resultados iniciais das buscas realizadas nas bases de dados de acordo com as expressões de busca descritas.

Quadro 1 – Resultados iniciais das buscas nas bases de dados selecionadas

EXPRESSÃO DE BUSCA	WoS	LIST A	ISTA	LISA	TOTAL
“digital repositor*” OR “subject repositor*” OR “disciplinary repositor*”	200	591	228	1.620	2.639
(“green road” OR self-archiving) AND “open access”	77	137	87	557	858
“open access movement” OR “open access initiative” OR “open archives initiative”	60	171	79	521	831
“science communication” OR “scholarly communication” OR “scientific communication”	1.116	2.359	2.208	5.232	10.915

TOTAL DE DOCUMENTOS	1.453	3.258	2.602	7.930	15.243
----------------------------	--------------	--------------	--------------	--------------	---------------

Fonte: dados da pesquisa (2019).

Devido ao alto número de documentos recuperados, 15.243 publicações, e resultados que não correspondiam às necessidades do trabalho, a busca foi realizada novamente de forma mais específica nas mesmas bases. Para isso, realizou-se uma busca utilizando a combinação das expressões de busca usadas anteriormente. No Quadro 2, são apresentados os resultados utilizando a seguinte expressão:

Quadro 2 – Resultado com a expressão de busca combinada

EXPRESSÃO DE BUSCA	WoS	LISTA	ISTA	LISA	TOTAL
<i>("digital reposito*" OR "subject reposito*" OR "disciplinary reposito*") AND (("green road" OR self-archiving) AND "open access") AND ("science communication" OR "scholarly communication" OR "scientific communication")</i>	4	4	3	152	163

Fonte: dados da pesquisa (2019).

Os resultados das buscas foram, então, exportados para um gerenciador de bibliografias. Nele, foi possível excluir as referências duplicadas (restando, ao todo, 150 documentos) e prosseguir para análise destas por meio de seus resumos. Na leitura, os documentos que não foram considerados pertinentes também foram excluídos. Ao passo que a leitura dos artigos foi realizada, ocorreu a adição de outras referências baseadas nas citações dos autores.

Com o desenvolvimento do trabalho, notou-se a necessidade de aprofundamento sobre a temática "Ciência Aberta". Ciente de que a temática "Ciência Aberta" é atual, não é discutida apenas na Biblioteconomia e Ciência da Informação, e buscando por materiais que enriqueçam a discussão, utilizou-se os artigos e referências recomendados pelo Open Science 'Massively Open Online Course' (MOOC). O Open Science MOOC³⁰ é uma plataforma de cursos on-line com foco em Ciência Aberta, que trabalha em parceria com diversas entidades que estão à frente do Movimento de Ciência Aberta como, por exemplo, a Facilitate Open Science Training for European Research (FOSTER), a Public Knowledge Project (PKP), a Creative Commons, Internet Archive, Altmetric, Open Researcher and Contributor ID (ORCID), entre outros, o que confere credibilidade aos materiais disponibilizados.

5.2 Caracterização da pesquisa

³⁰ *Open Science 'Massively Open Online Course'* (MOOC). Disponível em: <https://opensciencemooc.eu/>. Acesso em: 23 mar. 2020.

O trabalho baseia-se em uma visão de mundo pragmática, de acordo com Creswell (2014, p. 39, *tradução nossa*³¹), que “em vez de focar em métodos específicos, os pesquisadores enfatizam o problema de pesquisa e usam todas as abordagens disponíveis para entender o problema”. O desenho de pesquisa seguirá o modelo sequencial-explicativo dos métodos mistos, uma vez que será conduzida uma análise quantitativa inicialmente e, a partir dos resultados, uma análise qualitativa será feita para que os dados obtidos sejam melhor entendidos (CRESWELL, 2014).

Deste modo, o estudo é quantitativo-descritivo, de caráter exploratório. Para Lakatos e Marconi (2003), a análise quantitativa-descritiva serve ao objetivo de investigação das características de fenômenos por meio da coleta sistemática de dados. Ainda segundo as autoras, o estudo exploratório possui três finalidades: “desenvolver hipóteses, aumentar a familiaridade do pesquisador com um ambiente, fato ou fenômeno, para a realização de uma pesquisa futura mais precisa ou modificar e clarificar conceitos” (LAKATOS; MARCONI, 2003, p. 188). Assim, busca-se explorar o E-LIS, de forma a compreender o seu papel para a Ciência da Informação.

A pesquisa bibliométrica será utilizada como aporte metodológico para mensuração dos indicadores científicos. Oliveira (2018, p. 54) considera que, entre os diversos aspectos que podem ser analisados, os indicadores científicos “evidenciam os pesquisadores, as instituições, as temáticas, as áreas do conhecimento, bem como a frente de pesquisa de um campo de conhecimento, as redes de colaboração entre cientistas, grupos, instituições ou países e as redes de citação ou cocitação”.

5.3 Ambiente de pesquisa

O Ambiente de Pesquisa é o E-LIS (Figura 9). Criado em 2003, o E-LIS é um repositório digital de documentos técnico-científicos para as disciplinas relacionadas à Ciência da Informação, hospedado pelo Sistema de Biblioteca (CAB) da Universidade de Nápoles Federico II, na Itália. Estabelecido, gerenciado e mantido por uma equipe internacional de bibliotecários e cientistas da informação de cerca de 45 países e com suporte para 22 idiomas.

31 “*Instead of focusing on methods, researchers emphasize the research problem and use all approaches available to understand the problem*” (CRESWELL, 2014, p. 39)

Figura 9 – Página inicial do repositório E-LIS



Fonte: E-LIS (2019).

Atualmente, o repositório contém 23.245 documentos, entre eles artigos de periódicos, teses e dissertações, capítulos de livros, artigos de conferências, pôsteres, *preprints*, entre outros, com mais de 21 milhões de downloads efetuados.

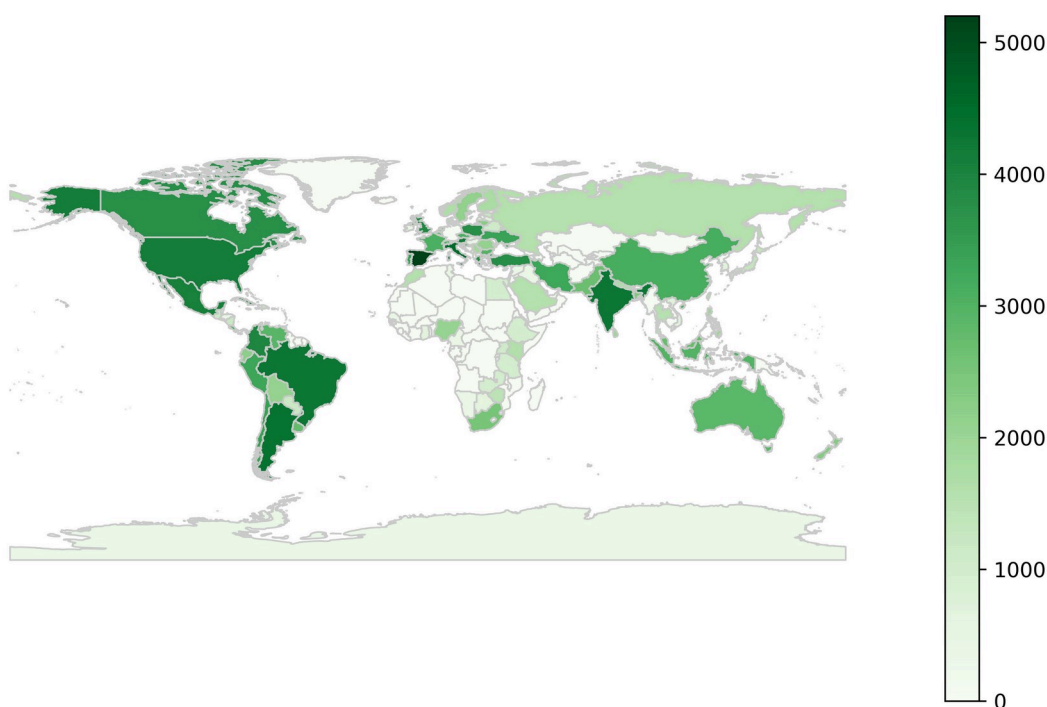
Em 2004, o IBICT declarava-se parceiro do E-LIS, tendo como objetivo comum “facilitar o arquivamento e a preservação a longo prazo das publicações científicas, provendo acesso global e fácil, estimulando a intercooperação entre pesquisadores para o maior impacto das pesquisas” (IBICT..., 2004). Miguel Ángel Márdero Arellano, servidor do IBICT e coordenador da Rede Brasileira de Serviços de Preservação Digital – CARINIANA, é editor responsável pela avaliação dos documentos em língua portuguesa submetidos no E-LIS, em companhia dos professores Rodrigo Rabello e Gildenir Carolino Santos (E-LIS, 2019).

5.4 Universo da pesquisa

O Universo da Pesquisa é composto de documentos depositados no E-LIS. No repositório, os países estão divididos por continente: a África (187 documentos); América, dividida em: a)

América do Norte e Central (3.345 documentos); e b) América do Sul (4.125 documentos); Antártica (2 documentos); Ásia (2.169 documentos); Europa (13.253 documentos); e Oceania (164 documentos). Totalizando, então, 23.245 documentos para análise (Figura 10).

Figura 10 – Quantidade de documentos distribuídos por país



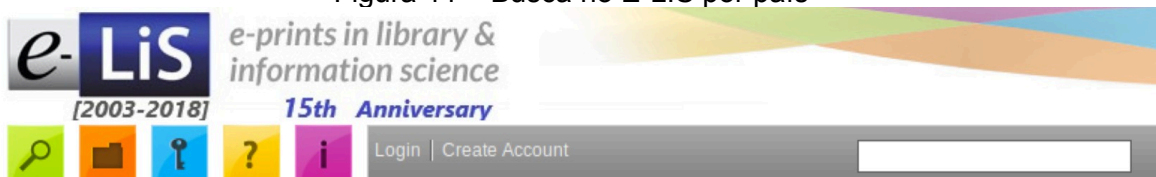
Fonte: Elaborado pela autora (2019).

No momento que o autor está realizando o depósito de um trabalho, é necessário inserir o país do autor (*Select the country/s of author/s*), entretanto, não fica especificado se é o país de origem do autor, onde ele nasceu, ou se é o país em que o autor está locado ou publicando naquela oportunidade. Por esse motivo, não é possível afirmar, de fato, que os autores são de determinado país ou se o trabalho foi publicado no país que eles assinalaram, pois depende do entendimento de cada autor no momento do preenchimento dos metadados.

5.5 Instrumento de coleta de dados

A coleta de dados será realizada diretamente no site do E-LIS. Nele, existem três tipos de busca: a simples, a avançada e a busca por itens (autor, assunto, país, ano e por conferência). Dessa forma, a busca será efetuada por itens e, logo após, por país. Na página inicial, a opção a ser selecionada é “Browse” e, em seguida, “Countries” (Figura 11).

Figura 11 – Busca no E-LIS por país



Browse Items

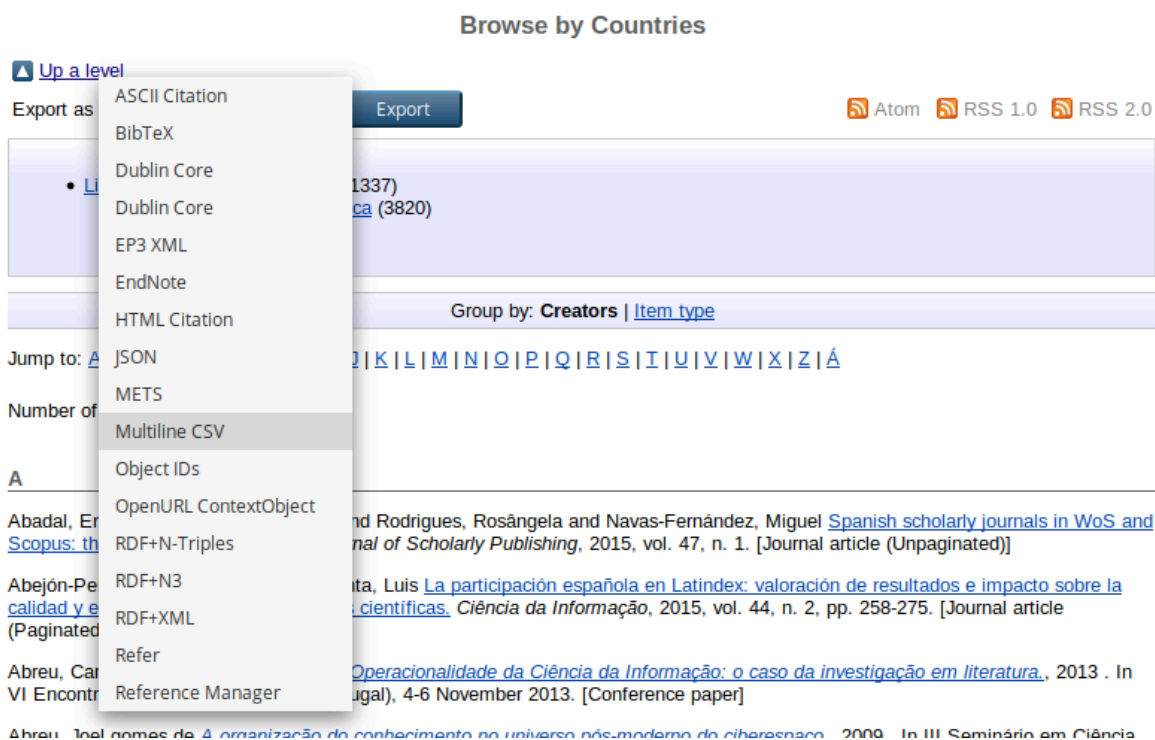
Items may be browsed by the following:

- [Author](#)
- [Subject](#)
- [Countries](#)
- [Year](#)
- [Conference](#)
- [Conference by name](#)

Fonte: E-LIS (2019).

Uma lista com os países e a quantidade de documentos disponíveis será elencada. Ao selecionar o país do qual se pretende coletar os dados, o repositório fornece a opção de exportação dos metadados das publicações recuperadas para diversos formatos, a saber: *ASCII*, *BibTeX*, *Dublin Core*, *EP3 XML*, *ENDNOTE*, *HyperText Markup Language (HTML)*, *JSON*, *METS*, planilhas *CSV*, *IDs dos objetos*, *OpenURL Context Objects*, *Resource Description Framework (RDF)*, *Refer e Reference Manager* (Figura 12).

Figura 12 – Formatos para exportação dos metadados das publicações



Fonte: E-LIS (2019).

Para fins de análise, os metadados serão exportados no formato de planilhas CSV³², o que quer dizer que os dados estão estruturados. Deste modo, a opção a ser selecionada será (*Multiline CSV*), e depois “*Export*”. O local no qual o arquivo gerado será salvo no computador deve ser determinado.

A ampla disponibilidade de dados na Internet levou os cientistas a se concentrarem mais intensamente no processo de coleta e análise de dados. Há apenas alguns anos, a capacidade de programação de computadores estava restrita a um pequeno subconjunto de pesquisadores, entretanto, o amplo acesso a informações criou uma nova geração de cientistas que desenvolver seus programas, elaboram novos *softwares* para análises e exploram novas maneiras de compartilhar e visualizar dados. De acordo com Borregaard e Hart (2016), surge um novo paradigma, no qual cientistas individuais baixam, selecionam e compartilham grandes quantidades de dados e os analisam usando pacotes de *software* reproduzíveis e scripts escritos em linguagens como Python, R e Julia.

O processo para coleta de dados no E-LIS é repetitivo e, de forma manual, tem que ser realizado para cada conjunto de dados que se tencione analisar. Aqui, como foram coletados, ao todo, 125 conjuntos de dados optou-se pela criação de um script em Python 3 para automatizar essa tarefa, isto é, o programa repetiu os procedimentos descritos nas figuras 4 e 5 automaticamente. Essa quantidade não foi definida previamente, ao realizar o download de todas as planilhas, verificou-se a quantidade de documentos baixados. Essa atividade é denominada “*Scraping*” e é utilizada para a extração de dados de páginas na web. Segundo Lowndes *et al.* (2017), ao emprestar filosofias, ferramentas e fluxos de trabalho criados principalmente para o desenvolvimento de *softwares*, pode-se melhorar drasticamente a capacidade de reprodução da ciência e isto resulta em uma ciência melhor realizada em menos tempo.

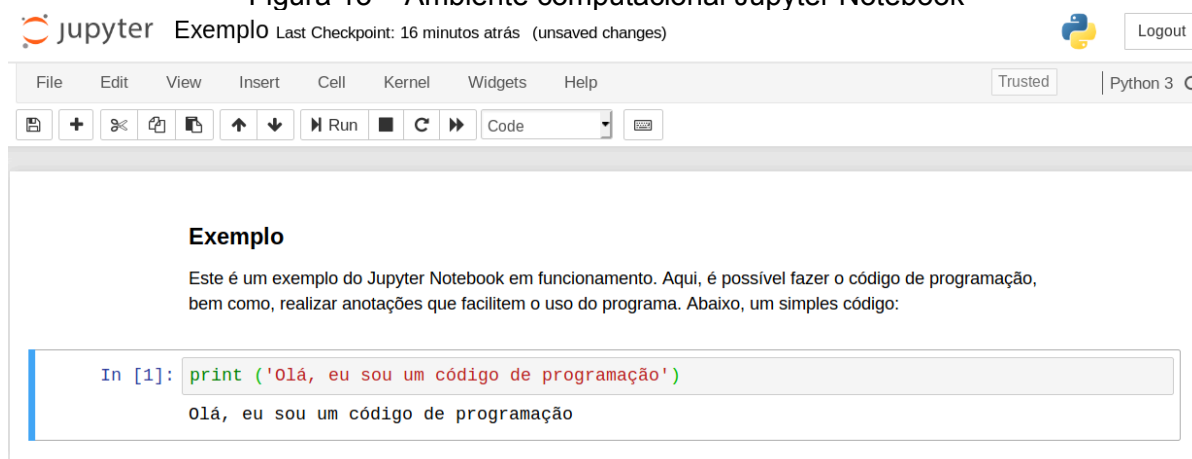
5.6 Instrumento para análise e discussão dos dados

A análise dos dados também será realizada de forma automatizada. Para isso, será empregada a linguagem de programação Python 3, que foi desenvolvida sob licença *open-source* ou de código aberto, o que a torna utilizável e distribuível. A linguagem Python foi criada por Guido van Rossum em 1991 e, desde então, tem conquistado adeptos nas mais diversas áreas do conhecimento (ROGEL-SALAZAR, 2017). Isso se dá, principalmente, porque a Python possui uma sintaxe simples, similar à língua inglesa, o que facilita seu aprendizado. Essa linguagem de programação é frequentemente utilizada na Ciência de Dados, que possui como característica a análise de grandes quantidades de informações.

32 Planilhas CSV são arquivos de texto delimitado que usam vírgula para separar valores (em inglês: *comma-separated values* – CSV).

Como ambiente computacional para a programação, será utilizado o Jupyter Notebook (Figura 13), uma ferramenta, também de código aberto, que possui uma lista ordenada de células de entrada/saída de informações, podendo conter código de programação, texto (usando *Markdown*, linguagem simples de marcação), equações matemáticas, gráficos e texto enriquecido, o que possibilitará a leitura dos dados, bem como sua representação (PYTHON... 2019). Dessa forma, será possível ler o conteúdo das planilhas CSV, extrair as informações relevantes e processá-las de forma a construir um entendimento acerca do material coletado.

Figura 13 – Ambiente computacional Jupyter Notebook



Fonte: Captura de tela do Jupyter Notebook (2019).

Mas por que utilizar essa ferramenta? A interação entre texto e código de programação transforma o trabalho **legível para humanos**. O projeto torna-se mais amigável e fácil de entender. Se bem documentado, o trabalho no Jupyter Notebook pode **descrever o fluxo de trabalho do pesquisador**, com os passos e os códigos, de forma que a análise possa ser reproduzida por outras pessoas, ou até mesmo pelo próprio pesquisador futuramente. Nele, é fácil **modificar uma análise**, sendo possível estender ou refiná-las, modificando os códigos existentes ou adicionando novos blocos de código. E, finalmente, é fácil **compartilhar as análises realizadas**, pois ele permite que o documento seja exportado em diferentes formatos, como o PDF, HTML e apresentações em *slide* (WASSER, 2018).

Seguindo os princípios da Ciência Aberta, que preconizam que os dados científicos, devem ser Localizáveis, Acessíveis, Interoperáveis e Reutilizáveis, os dados coletados e os códigos de programação criados e utilizados no presente trabalho estão disponibilizados on-line via Github. O Github é uma plataforma de hospedagem e compartilhamento de código-fonte, que permite que o trabalho seja disseminado e reproduzido por outras pessoas. Entende-se que os dados resultantes de pesquisa desenvolvida com financiamento público devem estar acessíveis, por isso este procedimento será adotado.

O ELIScript está compartilhado sob a licença **GNU General Public License v3.0** (GPL v3.0), GPL é a licença de software livre mais utilizada e tem uma forte exigência de copyleft. Isto

é, ao distribuir trabalhos derivados do ELIScript, o código-fonte do trabalho deve ser disponibilizado sob essa mesma licença.

Em termos gerais, a GPL baseia-se em 4 liberdades:

- A liberdade de executar o programa, para qualquer propósito (liberdade nº 0);
- A liberdade de estudar como o programa funciona e adaptá-lo às suas necessidades (liberdade nº 1). O acesso ao código-fonte é um pré-requisito para esta liberdade;
- A liberdade de redistribuir cópias de modo que você possa ajudar ao seu próximo (liberdade nº 2);
- A liberdade de aperfeiçoar o programa e liberar os seus aperfeiçoamentos, de modo que toda a comunidade beneficie deles (liberdade nº 3). O acesso ao código-fonte é um pré-requisito para esta liberdade.

Com a garantia dessas liberdades, a GPL permite que outras pessoas **compartilhem** o script, podendo copiá-lo e **adaptá-lo**, transformando-o e construindo em cima desse material. Sob essa licença, deve-se atribuir o crédito devido à autora do trabalho, mantendo seus direitos autorais, e indicar se alterações foram feitas no código-fonte; por fim, deverá ser compartilhado sob a mesma licença que o ELIScript foi distribuído.

O ELIScript³³ no Github está organizado em quatro pastas e dois arquivos. O esquema das pastas é apresentado abaixo:

- **db**: db, de *database*, é uma pasta com todos os arquivos coletados pelo ELIScript em CSV, isto é, todas as planilhas com os metadados das 23.245 publicações, separadas por continentes;
- **db_generated**: são as planilhas e tabelas geradas pelo ELIScript, que permitirão gerar os gráficos para análise;
- **plots**: nessa pasta estão disponíveis todos os gráficos gerados pelo ELIScript;
- **source**: os códigos-fonte do ELIScript divididos em três partes, o *ELIScript_collection* para a coleta de dados; e o *ELIScript_organization* para organização dos dados para posterior análise; e o *ELIScript_analysis* para análise de dados propriamente dita;
- **LICENSE**: nesse arquivo consta o texto da licença GPL v3.0, sob a qual o E-LIScript foi distribuído;
- **README.md**: é um arquivo de texto no formato *markdown*, com as informações sobre o ELIScript; introdução sobre o projeto, responsáveis pela criação do script, dependências a serem instaladas e agradecimentos.

33 O repositório do ELIScript está disponível publicamente no endereço: <https://github.com/sarahrubia/eliscript>. Nele, é possível ter acesso a todos os dados coletados para essa pesquisa, bem como os códigos para coleta e análise dos dados (SANTOS, 2021).

Tendo-se em vista o grau de incursão dedicado ao desenvolvimento do instrumento de coleta e análise de dados, entendeu-se por bem trabalhá-lo separadamente no capítulo seguinte.

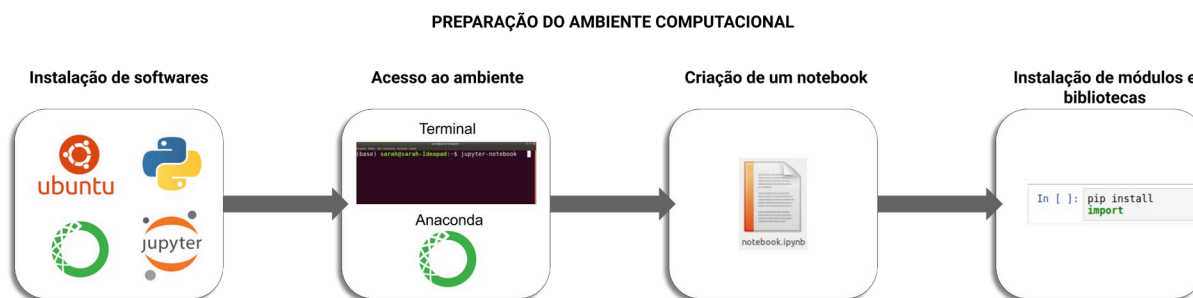
6 ELISCRIP: ANÁLISE AUTOMATIZADA DO AUTOARQUIVAMENTO NA CIÊNCIA DA INFORMAÇÃO

Neste capítulo são descritos os procedimentos para criação do ELIScript, de forma a atingir o primeiro objetivo específico desse trabalho, a saber: “a) desenvolver o ELIScript, instrumento para coleta e análise de dados em linguagem de programação Python”. Inicialmente são apresentadas as especificações técnicas e condições nas quais a aplicação foi desenvolvida. Em seguida, discute-se os processos de extração (coleta), tratamento e visualização (análise) dos dados.

6.1 Preparação do ambiente computacional

Este tópico tem por objetivo descrever as condições nas quais o presente trabalho foi realizado. Para o desenvolvimento do ELIScript, faz-se necessário preparar o ambiente computacional, o que consiste na instalação de *softwares*, acesso ao ambiente computacional, criação de um *notebook*, isto é, um arquivo de programação, e a instalação dos módulos e bibliotecas necessárias para a programação (Figura 14).

Figura 14 – Etapas para preparação do ambiente computacional



Fonte: Elaborado pela autora (2019).

Sandve *et al.* (2013) afirmam que toda pesquisa computacional deve seguir 10 regras para ser reproduzível, como descrito anteriormente. A terceira delas postula “arquive a versão exata de todos os programas usados externamente” (SANDVE *et al.*, 2013), quer dizer, para que um determinado resultado seja reproduzido, pode ser necessário utilizar programas na mesma versão utilizada originalmente, do contrário corre-se o risco de os resultados não serem iguais. Assim, as **especificações técnicas** serão apresentadas e uma breve descrição de cada ferramenta será feita.

No processo metodológico utilizou-se o Sistema Operacional Ubuntu 20.04 *Long Term Support* (LTS)³⁴, uma distribuição Linux. Ressalta-se que todo o processo de programação pode ser desenvolvido, também, nos sistemas Windows e MacOS, no entanto, as distribuições Linux são gratuitas, com comunidade ativa e suporte de longo termo. Por sua estabilidade e usabilidade, elas são amplamente utilizadas por programadores. Assim, deve-se observar particularidades de cada sistema, uma vez que alguns pontos podem diferir entre os sistemas.

A linguagem de programação utilizada foi a Python (versão 3.7.3), uma das linguagens mais populares para análise de dados devido às suas qualidades técnicas, sendo adotada em larga escala pela comunidade científica e industrial.

O Anaconda (versão 1.9.7) é um gerenciador de pacotes, um gerenciador de ambientes, uma distribuição de ciência de dados Python/R e uma coleção de mais de 7.500 pacotes de código-fonte aberto. Em seu *site*, o mesmo se define como a plataforma mais popular para Ciência de Dados, permitindo acessar e gerenciar poderosas bibliotecas, pacotes e ferramentas *open-source* de Ciências de Dados e Aprendizagem de Máquina (ANACONDA, 2020).

O Jupyter Notebook (versão 6.0.0), uma aplicação baseada na web, está contido na plataforma Anaconda. Logo, para acessá-lo é necessário antes ter instalado o Anaconda. O Jupyter Notebook cria e compartilha documentos que contenham código ativo, equações, visualizações e texto narrativo. Seus usos incluem: limpeza e transformação de dados, simulação numérica, modelagem estatística, visualização de dados, aprendizado de máquina, entre outros (JUPYTER, 2020).

Considerando as especificações de cada sistema operacional, não cabe, neste trabalho, descrever o processo de instalação de cada aplicação, no entanto, no *site* de cada uma delas encontra-se a documentação (em inglês), isto é, guias e tutoriais de como instalar o programa, suas funções e todas as descrições concernentes ao seu funcionamento. As ferramentas empregadas aqui são gratuitas e podem ser baixadas em seus respectivos *sites*.

De forma resumida, as especificações técnicas e as versões de cada aplicação são apresentadas no Quadro 3.

Quadro 3 – Especificações técnicas das aplicações computacionais

Especificações Técnicas		Versão
Sistema Operacional	Ubuntu	20.04 LTS (64-bits)
Linguagem de programação	Python	3.7.3
Plataforma de Ciência de Dados	Anaconda	1.9.7
Ambiente computacional	Jupyter Notebook	6.0.0
Navegador web	Mozilla Firefox	77.0.1 (64-bits)

34 <https://ubuntu.com/>

Fonte: dados da pesquisa (2020).

Após as instalações, procede-se para o **acesso ao ambiente de programação**, o Jupyter Notebook. Existem duas formas de entrar no programa: na primeira, o acesso pode ser feito via terminal; na segunda, o acesso é realizado por meio do Anaconda.

No terminal do Ubuntu, basta digitar Jupyter Notebook e apertar a tecla Enter, a aplicação abrirá uma aba no navegador com o ambiente. Pelo Anaconda, é necessário localizar o ícone do *software* no *menu* de aplicações e abri-lo. Na página inicial do programa, algumas ferramentas, além do Jupyter Notebook, estarão disponíveis, como o *Jupyter Lab*, o *Spyder*, *GlueViz*, *RStudio*, entre outras ferramentas de Ciência de Dados. Ao localizar o Jupyter Notebook, clique em “*Launch*” e o programa será aberto.

Com o Jupyter Notebook aberto, deve-se criar um *notebook*. O Jupyter Notebook abrirá a pasta local do computador do usuário, dessa forma, deve-se escolher a pasta na qual o *notebook* será salvo. Após essa seleção, basta clicar em “*New*” no canto superior direito e selecionar no menu suspenso a opção “Python 3”, em seguida, um *notebook* será aberto. Um *notebook* é um documento no formato *.ipynb* no qual o código de programação será escrito. Ele é composto por células de entrada que podem ser preenchidos com o código de programação, textos em linguagem de marcação *Markdown*, equações matemáticas, bem como por células de saída, em que são mostrados os resultados do código de programação, gráficos, tabelas e texto enriquecido.

Por fim, será preciso instalar os módulos e bibliotecas necessários para fazer a programação. O Anaconda já contém alguns dos pacotes, no entanto, alguns precisarão ser instalados à medida que não estejam disponíveis no gerenciador. A biblioteca, no Python, consiste em uma coleção de funções. Um programa pode utilizar recursos de uma biblioteca a fim de contribuir para atingir seu objetivo. O módulo é um arquivo Python com instruções e comandos, ele é executado quando importado. Alguns exemplos de módulos são o *NumPy*, que fornece funções que permitem a manipulação de matrizes numéricas e o *SciPy* que permite a funcionalidade normalmente esperada na computação científica, como otimização, rotinas de álgebra linear, transformação de Fourier etc. (ROGEL-SALAZAR, 2017). No Quadro 4 são descritos os módulos e bibliotecas utilizados e como são importados no Jupyter Notebook.

Quadro 4 – Importação de Bibliotecas e Módulos em Python

Extração de dados	
Bibliotecas e Módulos	Importação
Urllib.requests	<code>import urllib.request</code>
Requests	<code>import requests</code>

BeautifulSoup	from bs4 import BeautifulSoup
webdriver	from selenium import webdriver
Select	from selenium.webdriver.support.ui import Select
FirefoxProfile	from selenium.webdriver.firefox.firefox_profile import FirefoxProfile
Sleep	from time import sleep
NoSuchElementException	from selenium.common.exceptions import NoSuchElementException
OS	Import os

Organização dos dados

CHDIR	from os import chdir
Glob	from glob import glob
Python Data Analysis Library (Pandas)	import pandas as pd
OS	Import os

Análise de dados

Bibliotecas e Módulos	Importação
Python Data Analysis Library (Pandas)	import pandas as pd
Numpy	import numpy as np
Matplotlib: Visualization with Python	import matplotlib.pyplot as plt
CSV	import csv
OS	Import os
Image	from PIL import Image
Wordcloud; STOPWORDS	from wordcloud import WordCloud, STOPWORDS

Fonte: Elaborado pela autora (2020).

Para garantir a legibilidade do código, o Índice de Propostas de Aprimoramento de Python (PEPs)³⁵ será consultado. O PEP 8, especificamente, discorre sobre as convenções de estilo para código Python. Para além disso, serão observados os princípios descritos no PEP 20, que descrevem a filosofia da linguagem Python. Entre alguns dos postulados referentes ao código de programação, estão:

35 INDEX of Python Enhancement Proposals (PEPs). Disponível em: <https://www.python.org/dev/peps/>. Acesso em: 23 jun. 2020.

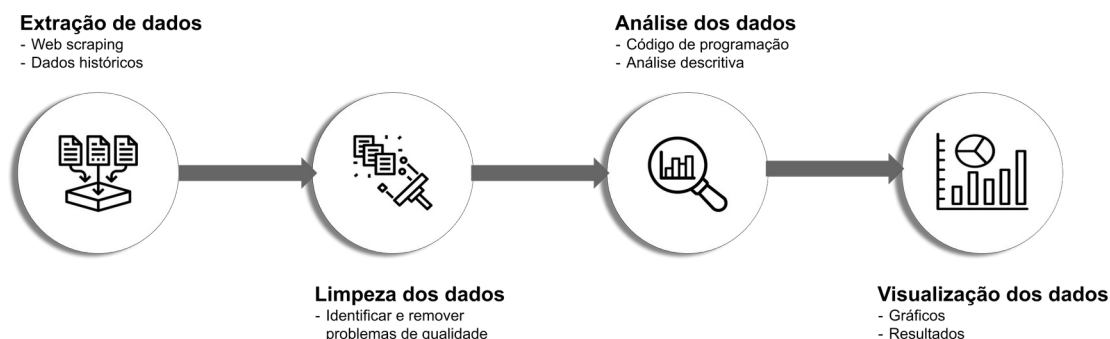
- Bonito é melhor que feio;
- Explícito é melhor que implícito;
- Simples é melhor que complexo;
- Complexo é melhor que complicado;
- Legibilidade conta.

O código será disponibilizado via Github, uma plataforma de hospedagem e compartilhamento de código-fonte, para que o trabalho possa ser reproduzido por outros pesquisadores, se necessário, ou até mesmo para ser aplicado em outras atividades.

6.2 Extração, tratamento e visualização de dados

Extrair, tratar e visualizar os dados são atividades complementares que possibilitam chegar aos resultados desse trabalho. O *framework* desses processos é apresentado na Figura 15, mostrando a estrutura sistematizada dessas etapas com uma breve descrição do que é feito em cada uma.

Figura 15 – Framework para tratamento e visualização de dados



Fonte: Elaborado pela autora (2020).

Nos subtópicos a seguir, os processos de extração de dados em páginas web, limpeza dos dados, análise dos dados e consequente visualização destes serão explicados em mais detalhes. Os códigos nesse trabalho foram divididos em três partes. O primeiro código será utilizado apenas para extração de dados. O segundo, para organização dos dados por meio de agrupamento de planilhas para facilitar as análises por continente. E o terceiro para análise dos dados, no qual será feita a limpeza dos dados e, a partir da análise, a visualização dos dados.

6.3 Extração de dados em páginas da web

No presente t3pico, ser3 descrita a extra33o dos dados do reposit3rio E-LIS. A depender do c3digo criado, documentos em texto, v3deo, 3udio, links, entre outras informa33es podem ser coletadas. Mas o que, de fato, 3 extra3do no E-LIS?

Cada documento depositado no reposit3rio gera um conjunto de metadados que 3 preenchido pelo autor no momento do autoarquivamento. Esses metadados descrevem o cont3udo do documento: t3tulo, resumo em l3ngua vern3cula e em ingl3s, sua tem3tica, onde foi publicado, qual o idioma, sua tipologia; bem como, informa33es pertinentes 3 autoria: identifica3o de autoria, vincula3o institucional dos autores, entre outras. Assim, cada publica3o tem uma p3gina com todos os seus metadados (Figura 16).

Figura 16 – P3gina de metadados de uma publica3o

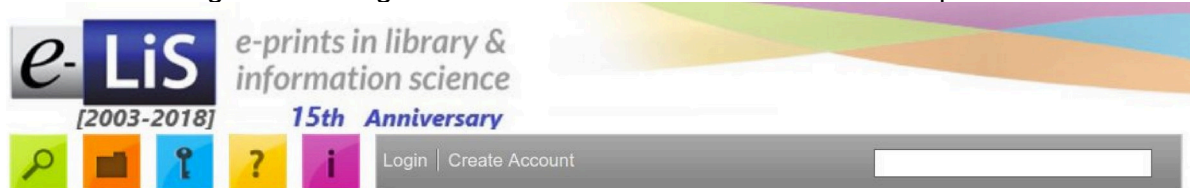
Full Metadata	
creators_name:	Santos, Sarah R3bia de Oliveira
creators_name:	Oliveira, Dalgiza Andrade
type:	journalp
datestamp:	2020-02-14 08:13:22
lastmod:	2020-02-14 08:13:22
status_changed:	2020-02-14 08:13:22
userid:	18153
metadata_visibility:	show
countries:	BR-
title:	Autoarquivamento na Ci3ncia da Informa33o: uma an3lise dos documentos depositados no reposit3rio digital e-LIS
ispublished:	pub
subjects:	BB
subjects:	HS
full_text_status:	public
keywords:	Comunica33o Cient3fica. Movimento de Acesso Aberto. Autoarquivamento. e-LIS. Bibliometria. Scientific communication. Open Access Movement. Self-archiving. Bibliometrics.
abstract:	It discusses self-archiving practices in Information Science and the contributions of the E-prints in Library and Information Science repository to open scientific communication in the field. The study is quantitative and descriptive, exploratory, and uses bibliometric analysis as a methodological support for the measurement of scientific indicators. Data collection and analysis instruments were built using Python programming language. About 34,000 documents were archived in the repository between 2002 and 2019. The most frequent document typologies are journal article (47%), conference article (18%), presentation (8%), book chapter (6%). Topics covered are 3Information Use and Sociology of Information3, 3Information Technology and Library Technology3 and 3Information Treatment for Information Services3. The results are partial and represent an overview of documents deposited in the repository.
othersabs_othabstracts:	Discute as pr3ticas de autoarquivamento na Ci3ncia da Informa33o e as contribu33es do reposit3rio E-prints in Library and Information Science para a comunica33o cient3fica aberta na 3rea. O estudo 3 quantitativo-descritivo, de car3ter explorat3rio, e utiliza33o a pesquisa bibliom3trica como aporte metodol3gico para mensura33o dos indicadores cient3ficos. Os instrumentos de coleta e an3lise de dados foram construidos utilizando a linguagem de programa33o Python. Cerca de 34 mil documentos foram arquivados no reposit3rio entre 2002 e 2019. As tipologias de documentos mais frequente s3o artigo de peri3dico (47%), artigo de confer3ncia (18%), apresenta33o (8%), cap3tulo de livro (6%). Os assuntos mais abordados s3o 3Uso da informa33o e sociologia da informa33o3, 3Tecnologia da informa33o e tecnologia de biblioteca3 e 3Tratamento da informa33o para servi3os de informa33o3. Os resultados s3o parciais e representam uma vis3o geral dos documentos depositados no reposit3rio.
date:	2020-01-31
publication:	M3tiplos Olhares em Ci3ncia da Informa33o
volume:	9
number:	2
pagerange:	1-11
linguabib:	pt
refereed:	TRUE
alldoc:	https://periodicos.ufmg.br/index.php/moci/article/view/19126/

Fonte: Captura de tela do E-LIS (2020).

Quando uma busca 3 feita no reposit3rio, uma lista com as publica33es pertinentes 3 recuperada. O E-LIS tem a funcionalidade de exportar essa lista para alguns formatos, isto 3, os metadados das publica33es nessa lista s3o estruturados em diferentes formatos para que possam ser exportados. Alguns dos formatos s3o: *ASCII*, *BibTeX*, *Dublin Core*, *EP3 XML*, *ENDNOTE*, *HTML*, *JSON*, *METS*, planilhas *CSV*, *IDs* dos objetos, *OpenURL Context Objects*, *RDF*, *Refer* e *Reference Manager*. Para a an3lise, o formato mais adequado 3 o *CSV*. Assim, um script, ou seja, um conjunto de instru33es em linguagem de programa3o, foi criado para extrair as planilhas *CSV*.

Sabendo-se que o reposit3rio possui uma p3gina com uma lista dividida por continente e pa3ses (Figura 17), no qual se encontram as planilhas com os metadados, uma fun3o foi criada dentro do script para acessar diretamente essa p3gina e recuperar a URL referente a cada pa3s.

Figura 17 – Página de busca com a lista de continentes e países



Browse by Countries

Please select a value to browse from the list below.

- [List of countries by continent](#) (22058)
 - [AFRICA](#) (154)
 - [Algeria](#) (1)
 - [Angola](#) (1)
 - [Botswana](#) (3)
 - [Cameroon](#) (2)
 - [Central African Republic](#) (1)
 - [Egypt](#) (5)
 - [Ethiopia](#) (5)
 - [Gambia](#) (1)
 - [Ghana](#) (3)
 - [Kenya](#) (15)
 - [Lesotho](#) (1)
 - [Madagascar](#) (1)
 - [Malawi](#) (1)
 - [Morocco](#) (15)
 - [Namibia](#) (2)
 - [Nigeria](#) (31)
 - [Senegal](#) (2)
 - [Seychelles](#) (1)
 - [South Africa](#) (63)
 - [Sudan](#) (2)
 - [Swaziland](#) (1)
 - [Tanzania](#) (5)
 - [Tunisia](#) (2)
 - [Uganda](#) (6)
 - [Zambia](#) (5)
 - [Zimbabwe](#) (12)

Fonte: Captura de tela do E-LIS (2020).

Após recuperar as *URL*, que ficarão salvas em uma lista, é necessário configurar o navegador web – o Mozilla Firefox será utilizado – para que ele seja acessado por meio do script e execute os comandos de exportação das planilhas. A configuração do navegador também é feita por meio de programação. Algumas preferências são definidas: a mais importante é o diretório em que o script salvará as planilhas exportadas.

Por fim, uma última função, dentro do script, utilizará as configurações definidas do navegador e a lista com as *URL* salvas, e realizará o download das planilhas. Para isso, o script:

- a) abrirá o navegador web;
- b) acessará uma *URL* na lista de *URL* recuperadas;
- c) localizará o menu suspenso com os formatos de arquivos para exportação;
- d) selecionará o formato “multiline CSV”, que corresponde às planilhas CSV;
- e) clicará no botão “*Export*”; e
- f) fará o download da planilha no diretório definido nas preferências do navegador.

O navegador será aberto apenas uma vez e será atualizado cada vez que abrir uma página. Ao final da execução, o navegador será fechado. O processo descrito foi realizado para cada URL na lista de URL salvas, que corresponde a cada planilha com os metadados das publicações de um país, até que todas tenham sido acessadas. A coleta de dados foi realizada no dia 09 de setembro de 2020.

6.4 Limpeza dos dados

A limpeza dos dados, do inglês *data cleaning*, corresponde à identificação e remoção de problemas de qualidade das planilhas. Mais do que limpar os dados, trata-se de organizá-los para que se tornem mais adequados para exploração e análise. O que pode ser considerado um problema? Como exemplo, pode-se citar a identificação, em algumas planilhas, de metadados que deveriam estar em uma coluna e estavam em outra, como os metadados correspondentes aos assuntos e aos idiomas dos documentos. Logo, por mais que os dados estivessem estruturados na forma de planilha, ainda se fez necessário padronizá-los. Deste modo, para cada categoria de análise, um pré-processamento dos dados foi realizado.

Famili *et al.* (1997, p. 4, *tradução nossa*³⁶) afirmam que “o pré-processamento de dados consiste em todas as ações executadas antes do início do processo de análise”. O pré-processamento de dados é uma tarefa que consome tempo e que, em muitos casos, é realizada de forma semiautomática. Sobre os problemas com os dados, os autores afirmam que eles sempre existem quando se lida com dados da vida real. A natureza e severidade desses problemas, no entanto, dependem de diversos aspectos: podem ser erro na entrada ou na saída dos dados; uma quantidade demasiada ou insuficiente de dados, dados corrompidos ou incompatíveis etc. Esses problemas nem sempre estão sob controle dos operadores humanos. Assim, a não realização dessa etapa pode afetar a análise dos dados e, conseqüentemente, os resultados de um trabalho.

A preparação para a análise de dados, então, consiste do entendimento da natureza dos dados e o uso de técnicas para pré-processamento destes. As técnicas de pré-processamento utilizadas nos dados do E-LIS são divididas em dois grupos: 1) transformação de dados e 2) compilação de informações. A primeira corresponde às técnicas de filtragem, ordenação e edição dos dados. A segunda, por sua vez, abrange técnicas de visualização, eliminação, seleção e amostragem de dados (FAMILI *et al.*, 1997).

As planilhas com os metadados de cada publicação que foram baixadas estarão divididas por país, como descrito anteriormente. Para possibilitar a análise comparativa entre continentes, as planilhas dos países que pertencem ao mesmo continente foram concatenadas, gerando uma

³⁶ “Data preprocessing consists of all the actions taken before the current data analysis process starts” (FAMILI *et al.*, 1997, p. 4).

planilha para cada continente. Como todos os continentes tinham, pelo menos, um país com publicações e respeitando a divisão entre América do Norte e Central e América do Sul feita no E-LIS, que representa uma divisão política dos continentes, sete planilhas foram geradas: 1) África; 2) América do Norte e Central; 3) América do Sul; 4) Antártida; 5) Ásia; 6) Europa; e 7) Oceania.

Precedendo cada análise, os processos de transformação dos dados foram empregados. Por meio da filtragem e ordenamento foi possível extrair e compilar as informações necessárias para gerar a visualização dos dados com gráficos e tabelas. Para facilitar o acompanhamento do que foi desenvolvido, as técnicas de pré-processamento dos dados serão descritas em conjunto com os processos de análise, uma vez que é difícil determinar onde termina um e onde começa o outro.

6.5 Análise e visualização dos dados

Faz-se necessário ressaltar em que se difere a análise feita com o ELIScript e o processamento de dados com um editor de planilhas como o Microsoft Excel ou o LibreOffice Calc, que também poderiam ser utilizados para análise de dados. Nesses editores seria necessário aplicar diferentes funções para cada planilha a fim de criar um gráfico para cada elemento analisado. Cada gráfico criado teria, também, que ser personalizado individualmente. Se houvesse algum problema de qualidade, como dados na coluna errada, cada erro teria que ser tratado individual e manualmente.

O ELIScript é um algoritmo. De acordo com o dicionário Oxford Languages (2020, on-line), um algoritmo é o “conjunto das regras e procedimentos lógicos perfeitamente definidos que levam à solução de um problema em um número finito de etapas”. Seguindo esse conceito, a ideia do ELIScript é realizar um conjunto de tarefas como: acessar as planilhas baixadas, realizar o pré-processamento dos dados, analisar os dados de acordo com as categorias de análise definidas, gerar a visualização e personalização dos gráficos de uma só vez. Isso é possível porque serão criadas funções que executarão conjuntos de instruções para cada uma das planilhas, sequencialmente, resultando nas análises que serão apresentadas no próximo capítulo.

Em Python, uma função é um trecho de código (um procedimento, um algoritmo), que pode ser usado várias vezes. Consiste na combinação de blocos lógicos para efetuar tarefas específicas (ROGEL-SALAZAR, 2017) e possui a seguinte sintaxe (Figura 18):

Figura 18 – Estrutura de uma função em Python

```
In [ ]: def minha_função(argumento1, argumento2, ...):
        '''Descrição da função (opcional) '''

        # Instruções a serem executadas quando a função iniciar

        return resultado da função (opcional)
```

Fonte: Captura de tela do Jupyter Notebook (2020).

Os termos **def** e **return** são palavras reservadas. O **def** vem de definição e indica que uma função será definida. O **return**, por sua vez, estabelece se um resultado será retornado ao fim da execução da função. Os argumentos são parâmetros de entrada que serão utilizados na função. Para facilitar o entendimento, o exemplo é dado na Figura 19:

Figura 19 – Exemplo de uma função em Python

```
In [1]: def adicao(valor1, valor2):
        '''Essa função calcula a soma entre dois valores'''

        soma = valor1 + valor2

        return soma

In [2]: adicao(2, 7)
Out[2]: 9
```

Fonte: Captura de tela do Jupyter Notebook (2020).

No campo de entrada *In [1]* cria-se a função *adicao*, utilizando-se a palavra reservada **def**. Essa função receberá dois argumentos, o *valor1* e o *valor2*, dentro dos parênteses. No interior dessa função está sua descrição e a instrução a ser executada: ao serem inseridos *valor1* e *valor2*, realizar a soma. O retorno, por fim, é a soma desses valores. Ao aplicar a função em *In [2]*, utiliza-se o nome definido para a função e, dentro dos parênteses, os valores a serem somados. Para um teste, os números 2 e 7 foram escolhidos. Em *Out [2]*, que é a saída do código executado em *In [2]*, está o resultado da função executada: 9.

Esse é um exemplo simples, mas que demonstra o funcionamento de uma função. O ELIScript é formado por várias funções que executarão instruções um pouco mais complexas, de forma que os dados do repositório E-LIS sejam analisados.

Algumas categorias de análise (Quadro 5) foram consideradas baseando-se nos metadados dispostos como colunas nas planilhas CSV. Essas categorias consistem na tipologia do documento, assunto do documento, data de publicação do item, data de depósito do item, idioma da publicação, periódico em que o artigo foi publicado, evento em que artigo, apresentação ou pôster foram apresentados e palavras-chave mais utilizadas.

Quadro 5 – Categorias de análise

CATEGORIAS DE ANÁLISE	Colunas CSV	DESCRIÇÃO
Tipologia do documento	<i>types</i>	Tipos de documentos depositados no repositório. Exemplo: artigos de periódicos, livros, apresentações em slides, guias e manuais, entre outros.
Data de depósito do documento	<i>datestamp</i>	Indica quando o documento foi depositado no repositório, independente de sua data de publicação.
Data de publicação do documento	<i>date</i>	Indica a atualidade dos documentos. A data em que o documento foi publicado.
Assunto do documento	<i>subjects</i>	Matéria de que trata o documento (CUNHA, 2008). Os assuntos concentram-se na área de Biblioteconomia e Ciência da Informação. A temática dos documentos é dada pela classificação JITA ³⁷ .
Idioma da publicação	<i>linguabib</i>	Podem ser depositados no E-LIS documentos em qualquer idioma. Ao todo, são suportados 38 idiomas.
Periódico científico	<i>publication</i>	“Fascículo numa série contínua sob o mesmo título, publicado a intervalos regulares, por tempo ilimitado, sendo cada fascículo numerado consecutivamente com indicação de data; publicação periódica” (CUNHA, 2008, p. 279).
Evento científico	<i>event_title conference</i>	No sentido de conferências, são reuniões de “indivíduos ou de representantes de várias entidades para discutir e/ou tomar decisões sobre assuntos de interesse comum” (CUNHA, 2008, p. 100)
Palavras-chave	<i>keywords</i>	“Palavra significativa encontrada no título de um documento, no resumo ou no texto. Essa palavra (ou grupo de palavras) caracteriza o conteúdo temático do item” (CUNHA, 2008, p. 274)

Fonte: Elaborado pela autora (2020).

Tendo em vista essas categorias, a seguir, serão apresentadas as funções criadas no código do ELIScript para organização, análise e visualização dos dados. Apesar de mencionar a criação de gráficos, estes serão apresentados apenas no capítulo 7, durante a análise e discussão dos resultados.

a) Tipologia do documento

³⁷ O esquema de classificação JITA foi desenvolvido a partir de uma fusão de Plano de Classificação do *News Agent Topic* e o esquema de classificação RIS (extinto) da revisão da Ciência da Informação originalmente concebido por Dagobert Soergel (CRUZ *et. al.*, 2018). É utilizado para representar documentos da Biblioteconomia e Ciência da Informação.

Para a análise de tipologia de documentos, seis funções foram criadas no ELIScript. A primeira, *open_csv_file(names)*, como o nome sugere, abre os arquivos CSV nos quais estão armazenados os metadados das publicações depositadas no E-LIS. Para isso, ela utiliza o argumento “*names*”, que é uma lista com o nome dos arquivos e a sua extensão, por exemplo: “*áfrica.csv*”. O resultado dessa função é uma lista *open_file*, isto é, a lista de arquivos abertos para processamento e análise, que será utilizada em outros momentos do código.

A segunda função, *count_doc_type_values(open_file, typology_dict, continents)*, recebe três argumentos: *open_file*, *typology_dict* e *continents*. Observe-se que o resultado da primeira função, *open_file*, serve como argumento para a segunda função. Essa função quantifica os documentos por tipologia. Como? Ela utiliza os documentos abertos em *open_file*, localiza a coluna que identifica o tipo do documento, *types*, quantifica os documentos pela sua tipologia e os adiciona em uma lista. O argumento *typology_dict* é um dicionário composto por chaves e valores e é apresentado da seguinte forma: dicionário = {“chave” : “valor”}. No E-LIS e nas planilhas, os tipos de documento estão identificados pelo nome em inglês e com letra minúscula, exemplo: “*bibliography*”. No entanto, para a visualização dos dados, é preferível que a nomenclatura esteja traduzida como “Bibliografia”, assim, toda vez que o código localizar “*bibliography*” na coluna de tipologias, essa chave será substituída pelo valor “Bibliografia”. O argumento *continents*, por sua vez, é uma lista com o nome dos continentes e será usado para atribuir o nome de cada continente ao conjunto de dados respectivo, dessa forma, será possível identificar que certos dados correspondem a um continente específico. Por fim, essa função retornará uma tabela, *doc_type_table*, com a contagem de tipos de documento por continente (Tabela 1).

Tabela 1 – Tipologia de documento por continente

Tipologia	África	Antártica	Ásia	Europa	América do Norte	Oceania	América do Sul
Anais de evento	0	0	7	55	21	0	19
Apresentação	26	0	88	1102	380	36	194
Artigo de conferência	50	0	556	2582	532	38	673
Artigo de periódico	43	0	1008	6256	1624	44	1886
Bibliografia	1	0	5	21	41	0	30
Capítulo de livro	6	1	99	891	105	1	68
Conjunto de dados	0	0	0	8	1	0	1
Guia / manual	2	0	31	131	29	2	48
Livro	9	0	31	201	48	4	98
Material instrucional de biblioteca	1	0	0	12	2	0	4
Matéria de jornal/revista	1	0	74	66	49	6	124
Outro	7	0	25	167	53	4	199
Preprint	18	1	114	452	202	7	278
Projeto / Plano de negócios	1	0	3	23	14	1	16
Pôster de conferência	2	0	28	316	55	0	129
Relatório	4	0	44	294	65	15	89
Relatório técnico	0	0	7	23	7	0	14
Relatório técnico departamental	0	0	0	1	2	0	1
Resenha Crítica	0	0	5	252	9	1	32
Tese/Dissertação	11	0	37	356	106	5	216
Tutorial	5	0	7	44	0	0	6

Fonte: Dados da pesquisa (2020).

A partir dos dados da tabela gerada, é possível calcular a soma total por tipos de documento e também a porcentagem. Para isso, criou-se a função `get_rows_sum_percentage(doc_type_table, round_value)`, que soma os valores de cada linha da tabela. Ela tem como argumentos `doc_type_table`, a Tabela 1 (resultado da função anterior), e `round_value`, que será um valor numérico inteiro para definir o número de casas decimais para a porcentagem. Dessa forma, a tabela `doc_type_table` é utilizada como o argumento para gerar somas e porcentagens, e o valor 1 define as casas decimais das porcentagens (Tabela 2).

Tabela 2 – Tipologia de documentos com total e porcentagem

Tipologia	África	Antártica	Ásia	Europa	América do Norte	Oceania	América do Sul	Total	%
Anais de evento	0	0	7	55	21	0	19	102	0,4
Apresentação	26	0	88	1102	380	36	194	1826	7,9
Artigo de conferência	50	0	556	2582	532	38	673	4431	19,1
Artigo de periódico	43	0	1008	6256	1624	44	1886	10861	46,7
Bibliografia	1	0	5	21	41	0	30	98	0
Capítulo de livro	6	1	99	891	105	1	68	1171	5
Conjunto de dados	0	0	0	8	1	0	1	10	0
Guia / manual	2	0	31	131	29	2	48	243	1
Livro	9	0	31	201	48	4	98	391	1,7
Material instrucional de biblioteca	1	0	0	12	2	0	4	19	0,1
Matéria de jornal/revista	1	0	74	66	49	6	124	320	1,4
Outro	7	0	25	167	53	4	199	455	2
Preprint	18	1	114	452	202	7	278	1072	4,6
Projeto / Plano de negócios	1	0	3	23	14	1	16	58	0,2
Pôster de conferência	2	0	28	316	55	0	129	530	2,3
Relatório	4	0	44	294	65	15	89	511	2,2
Relatório técnico	0	0	7	23	7	0	14	51	0,2
Relatório técnico departamental	0	0	0	1	2	0	1	4	0
Revisão de literatura	0	0	5	252	9	1	32	299	1,3
Tese/Dissertação	11	0	37	356	106	5	216	731	3,1
Tutorial	5	0	7	44	0	0	6	62	0,3
							TOTAL	23245	100

Fonte: dados da pesquisa (2020).

A primeira e a segunda funções poderiam ser criadas como uma única função que, ao final, resultaria em uma tabela com a contagem dos tipos de documentos, mais o total por tipos de documento e, também, a porcentagem. Contudo, optou-se por fazê-las separadamente por dois motivos: 1) melhor visualização das etapas; 2) uma função separada de soma e porcentagem pode ser utilizada para quantificar outras categorias de análise. As duas funções foram utilizadas para organizar os dados. Elas extraem os dados importantes de cada uma das planilhas e as reúnem para análise, apresentando-as na forma de uma única tabela.

A partir dessa organização, ou dessa limpeza, pode-se prosseguir para a visualização dos dados em outros formatos. Para isso, mais três funções foram criadas: `plot_doc_type(doc_type_table, continents)`; `plot_type_percent(doc_type_table, column)`; `doc_type_stacked(doc_type_table, colors)`. As três funções recebem como argumento a tabela `doc_type_table`. Na função `plot_doc_type(doc_type_table, continents)`, o outro argumento é `continents`. Como apontado anteriormente, esse argumento é uma lista com o nome dos continentes. Essa função analisa cada coluna da tabela individualmente e gera um gráfico para cada uma, resultando em um gráfico com as tipologias de documento para cada continente (Apêndice A).

Já a função `plot_type_percent(doc_type_table, column)` recebe como argumento “`column`”. Considerando que trata-se de um gráfico com dados de porcentagem, a coluna que será adicionada como argumento é a de porcentagem. Assim, o código observará essa coluna quando desenhar o gráfico. Diferente do código anterior que gerou um gráfico por continente, aqui será gerado apenas um gráfico com valores gerais de porcentagem para cada documento, apresentando uma visão global das tipologias de documento encontradas no E-LIS.

Por fim, a função *doc_type_stacked(doc_type_table, colors)* considera o argumento *colors* para o último gráfico sobre tipos de documento. Esse argumento é uma lista com o nome de cores, que pode aparecer em três formatos: 1) como o nome da cor em inglês: 'red', 'green'; 2) como códigos hexadecimais: #FF0000 (vermelho); #008000 (verde); ou como códigos RGB (**R**ed – Vermelho, **G**reen – Verde, **B**lue – Azul): (255,0,0) representa vermelho e (0,128,0), verde. Pela uniformidade, optou-se por utilizar os códigos hexadecimais. Assim, essa função cria um gráfico empilhado considerando a porcentagem de cada tipo de documento por continente e permite que se defina quais cores estarão presentes no gráfico.

b) Data de depósito do documento

Para a análise do período em que houve autoarquivamento por parte dos autores ou editores no repositório E-LIS, considerou-se a coluna *datestamp*. Os dados dessa coluna aparecem no formato AAAA-MM-DD HH:mm:ss, isto é, ano-mês-dia hora:minutos:segundos. Optou-se por considerar apenas os anos, pois uma análise com mês, dia e horário tornaria os dados muito dispersos.

A função *count_datestamp_values(open_file, continents)* tem como argumentos os arquivos abertos, *open_file*, e a lista *continents* com o nome dos continentes. Essa função é similar à função que conta valores dos tipos de documento, no entanto, considera a coluna *datestamp* e, dela, extrai apenas os dados de ano (AAAA). O resultado dessa função é uma tabela, *datestamp_counts*, com a contagem de depósitos por ano, dividido por continente (Tabela 3).

A partir da tabela obtida, foram criadas três funções para gerar gráficos. A primeira, *datestamp_plot_by_continent(datestamp_counts, continents)*, recebe como argumento a tabela gerada na função anterior e a lista com o nome dos continentes. Essa função considera as colunas, ou seja, os dados por continente, e gera um gráfico para cada um deles com os dados de depósito por ano. O argumento com o nome dos continentes permite identificar qual gráfico pertence à qual continente e serve como legenda para cada gráfico (Apêndice B).

A função *datestamp_ww_plot(datestamp_counts, column_name, list_colors)*, por sua vez, tem como argumentos a tabela com a contagem dos anos de autoarquivamento; uma lista com o nome das colunas que serão consideradas para o gráfico, nesse caso, *column_name* é igual a *continents*; e uma lista de cores com códigos hexadecimais, *list_colors*. Essa função traça um gráfico de linhas, com uma linha para cada continente, comparando os anos em que os documentos foram depositados.

Tabela 3 – Depósitos por ano dividido por continente

Ano	África	Antártica	Ásia	Europa	América do Norte	Oceania	América do Sul
2002	0.0	0.0	0.0	1.0	0.0	0.0	0.0
2003	0.0	0.0	5.0	233.0	44.0	0.0	0.0
2004	2.0	0.0	104.0	858.0	399.0	5.0	115.0
2005	14.0	0.0	114.0	664.0	231.0	10.0	103.0
2006	24.0	0.0	288.0	938.0	340.0	22.0	148.0
2007	13.0	0.0	271.0	1366.0	344.0	8.0	270.0
2008	12.0	0.0	154.0	1074.0	344.0	22.0	244.0
2009	13.0	0.0	49.0	997.0	230.0	17.0	129.0
2010	16.0	0.0	98.0	856.0	134.0	25.0	264.0
2011	20.0	0.0	81.0	833.0	139.0	29.0	233.0
2012	9.0	0.0	87.0	939.0	130.0	6.0	533.0
2013	23.0	0.0	174.0	757.0	239.0	3.0	360.0
2014	10.0	0.0	99.0	788.0	107.0	4.0	271.0
2015	8.0	2.0	174.0	724.0	226.0	0.0	296.0
2016	2.0	0.0	98.0	516.0	84.0	4.0	218.0
2017	7.0	0.0	55.0	358.0	80.0	1.0	222.0
2018	6.0	0.0	117.0	480.0	166.0	2.0	278.0
2019	2.0	0.0	141.0	606.0	65.0	5.0	299.0
2020	6.0	0.0	60.0	265.0	43.0	1.0	142.0

Fonte: dados da pesquisa (2020).

Tendo-se em conta que a Europa tem muito mais depósitos que os outros continentes, considerou-se necessário criar um gráfico com valores logarítmicos para representar e melhorar a visualização das diferenças entre os outros continentes. Dessa forma, a função `datestamp_logaritimic_plot(datestamp_counts, list_colors)` foi criada. Essa função opera da mesma forma que a anterior, mudando apenas os valores inteiros para valores logarítmicos.

c) Data de publicação do documento

A função `count_date_values(open_file, continents)` foi criada no ELIScript para organizar os dados referentes ao ano de publicação. Essa função recebe os argumentos `open_file` e `continents`, seleciona os dados da coluna “`date`”, extrai apenas os 4 primeiros dígitos correspondentes ao ano e cria uma tabela com a contagem de valores por ano dividido por continentes, `df_date_counts` (Tabela 4).

Tabela 4 – Ano de publicação dividido por continente

Ano	África	Antártica	Ásia	Europa	América do Norte	Oceania	América do Sul
1965	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1966	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1969	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1974	0.0	0.0	1.0	1.0	0.0	0.0	1.0
1975	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1976	0.0	0.0	0.0	1.0	0.0	0.0	1.0
1977	0.0	0.0	0.0	0.0	1.0	0.0	0.0
1978	0.0	0.0	0.0	4.0	0.0	0.0	1.0
1979	0.0	0.0	0.0	0.0	1.0	0.0	0.0
1980	0.0	0.0	1.0	2.0	1.0	0.0	0.0
1981	0.0	0.0	2.0	1.0	2.0	0.0	1.0
1982	0.0	0.0	4.0	1.0	0.0	0.0	2.0
1983	0.0	0.0	7.0	3.0	0.0	0.0	1.0
1984	0.0	0.0	4.0	4.0	0.0	0.0	1.0
1985	0.0	0.0	5.0	9.0	1.0	0.0	1.0
1986	0.0	0.0	9.0	11.0	1.0	0.0	1.0
1987	0.0	0.0	9.0	17.0	0.0	0.0	2.0
1988	0.0	0.0	9.0	7.0	3.0	0.0	1.0
1989	0.0	0.0	7.0	6.0	2.0	0.0	2.0
1990	0.0	0.0	4.0	21.0	6.0	0.0	1.0
1991	2.0	0.0	7.0	15.0	4.0	0.0	5.0
1992	0.0	0.0	8.0	26.0	9.0	0.0	3.0
1993	0.0	0.0	4.0	12.0	17.0	0.0	4.0
1994	0.0	0.0	15.0	41.0	20.0	0.0	3.0
1995	1.0	0.0	25.0	64.0	32.0	0.0	5.0
1996	1.0	0.0	21.0	104.0	28.0	0.0	11.0
1997	0.0	0.0	19.0	134.0	30.0	0.0	14.0
1998	0.0	0.0	29.0	139.0	39.0	1.0	20.0
1999	0.0	0.0	26.0	260.0	48.0	0.0	26.0
2000	0.0	0.0	38.0	289.0	68.0	2.0	47.0
2001	4.0	0.0	52.0	364.0	124.0	3.0	45.0
2002	4.0	0.0	70.0	546.0	124.0	5.0	68.0
2003	4.0	0.0	67.0	688.0	176.0	3.0	105.0
2004	8.0	0.0	95.0	723.0	189.0	11.0	127.0
2005	18.0	0.0	130.0	802.0	219.0	13.0	203.0
2006	25.0	0.0	183.0	932.0	338.0	11.0	226.0
2007	13.0	0.0	154.0	935.0	258.0	13.0	240.0
2008	9.0	0.0	129.0	897.0	229.0	22.0	184.0
2009	21.0	0.0	101.0	852.0	209.0	17.0	261.0
2010	11.0	0.0	102.0	695.0	161.0	29.0	252.0
2011	5.0	0.0	80.0	788.0	128.0	10.0	304.0
2012	22.0	0.0	113.0	648.0	116.0	6.0	345.0
2013	6.0	1.0	97.0	659.0	177.0	2.0	234.0
2014	10.0	0.0	92.0	539.0	132.0	3.0	211.0
2015	4.0	1.0	105.0	437.0	148.0	1.0	225.0
2016	6.0	0.0	83.0	393.0	79.0	4.0	241.0
2017	5.0	0.0	82.0	377.0	106.0	2.0	223.0

Ano	África	Antártica	Ásia	Europa	América do Norte	Oceania	América do Sul
2018	2.0	0.0	93.0	352.0	74.0	4.0	216.0
2019	6.0	0.0	70.0	331.0	39.0	1.0	221.0
2020	0.0	0.0	17.0	117.0	5.0	1.0	39.0

Fonte: dados da pesquisa (2020).

Ao realizar essa contagem, observou-se quatro valores que não correspondiam ao padrão de ano (AAAA): 0003, 0015, 0201 e 0203, depositados na América do Norte e Central, América do Sul e Europa. Ao consultar os documentos e não apenas seus metadados, verificou-se que foram publicados em 2015 (0015) e em 2013 (0203), no entanto, não foi possível identificar a data nos outros dois documentos. Ressalta-se, aqui, a importância do cuidado no preenchimento dos metadados na submissão em qualquer repositório, para precisão e integridade da descrição do item, pois o que for inserido nos campos será utilizado para localizar o registro no futuro. Para elaboração dos gráficos, as datas que não seguiam o padrão foram desconsideradas.

Com a tabela *df_date_counts*, prossegue-se para a criação de alguns gráficos. A função *date_plot_by_continent(df_date_counts, continents)* recebe, além da tabela com a contagem de valores por ano, a lista com o nome dos continentes, que servirá como legenda para os gráficos gerados. Assim, um gráfico com anos de publicação de cada continente será gerado (Apêndice C).

Na função *date_ww_plot(df_date_counts, column_name, list_colors)*, por sua vez, apenas um gráfico é criado. Essa função recebe como argumento a tabela com a contagem de valores dos anos, o nome da coluna que será utilizada para criar o gráfico, que corresponde ao nome dos continentes, e uma lista com os códigos hexadecimais de cores. Obtém-se, então, um gráfico com a comparação da quantidade de publicações por ano entre os continentes. Uma versão logarítmica dessa função, *date_logaritimic_plot(df_date_counts, list_colors)*, também foi criada.

d) Assunto do documento

O assunto dos documentos, nas planilhas com os metadados dos documentos, é localizado na coluna *subjects*. Ao extrair os dados dessa coluna, observou-se que os assuntos dos documentos são representados pelas notações do esquema de classificação JITA³⁸, apresentado em anexo (Anexo A), que aparecem no formato de dois caracteres, uma letra

³⁸ A nome da classificação JITA é um acrônimo derivado da letra inicial do nome de cada autor do esquema: José Manuel Barrueco-Cruz, Imma Subirats-Coll, Thomas Krichel e Antonella De Robbio (MORRISON *et al.*, 2007).

maiúscula e um ponto (X.) ou duas letras maiúsculas (XX). Os assuntos não são exclusivos, isto é, um mesmo documento pode ser classificado em mais de um assunto.

Esse esquema é utilizado para indexação e busca na área de Biblioteconomia e Ciência da Informação. O esquema de classificação JITA é simples, alfabético e possui apenas dois níveis hierárquicos. Aqui, utilizou-se a tradução e adaptação para o português do esquema elaborada por Viana (2006), com os devidos acréscimos nos casos em que houve atualização do esquema de classificação e as notações não constavam na tradução (por exemplo, as notações FJ. *Knowledge management*, HT. *Web 2.0*, *Social networks*, IL. *Semantic Web* e LT. *Mobile Devices*). Para facilitar o entendimento e evitar a necessidade de consulta ao esquema de classificação, criou-se um dicionário de assuntos, que permite representar as notações pelo nome dos assuntos por extenso.

Dessa forma, a função *get_subjects(open_file, subjects_dict, continents)* foi criada. Ela recebe como argumentos as planilhas abertas em *open_file*, o dicionário com as chaves-valores {notação : assunto}, *subjects_dict*, e a lista com o nome dos continentes, *continents*. A função, então, seleciona a coluna *subjects*, extrai os dados dela, substitui as notações do esquema de classificação JITA pelo nome do assunto por extenso, conta os valores para cada assunto e cria uma tabela com a contagem desses valores por continente, denominada *subjects_counts*.

Utilizando a tabela *subjects_counts* e a lista *continents* como argumento, criou-se a função *plot_subjects_by_continent(subjects_counts, continents)* para traçar gráficos de assuntos para cada continente. Por questões de visualização dos dados, foram selecionados os 15 assuntos mais recorrentes nos documentos depositados em cada continente. Se assim não fosse, os gráficos ficariam ilegíveis. Vale lembrar que é possível aumentar ou diminuir esse valor diretamente no código. A lista completa dos assuntos divididos por continente pode ser localizada no Apêndice D.

e) Idioma da publicação

Da mesma forma que ocorreu na análise dos assuntos dos documentos, a coluna correspondente aos dados sobre o idioma das publicações possui códigos atribuídos a cada um deles. Identificou-se 38 idiomas no total, representados por códigos da norma técnica International Organization for Standardization (ISO) 639-1, que atribui identificadores para os nomes de línguas diversas. A ISO 639-1 fornece um código de duas letras, designado para representar as principais línguas do mundo. Em geral, mas não necessariamente, o código é uma abreviação do nome da língua em inglês (ISO 639-1, 2002). Para facilitar a visualização das informações no gráfico, criou-se um dicionário, *lang_dict*, com as chaves-valores {código ISO 639-1 : idioma}.

A função *get_languages(open_file, lang_dict)*, então, recebe como argumentos os arquivos abertos, *open_file*, e o dicionário *lang_dict*, que permitirá a decodificação dos idiomas. Com os arquivos abertos, seleciona-se a coluna *linguabib*; os dados são, então, extraídos; os códigos no formato ISO 639-1 são convertidos para o nome da língua por extenso; os valores de cada língua são contados; e uma tabela é gerada (Tabela 5).

Tabela 5 – Idiomas dos documentos depositados no E-LIS

Idioma	África	Antártica	Ásia	Europa	América do Norte	Oceania	América do Sul
Espanhol	12	1	7	4579	1850	5	2852
Inglês	160	0	1820	3185	1496	156	228
Italiano	0	0	3	1391	22	0	0
Alemão	1	1	1	1370	0	1	0
Polonês	0	0	0	467	2	0	0
Grego	0	0	0	446	0	0	0
Turco	0	0	2	407	0	0	0
Catalão	0	0	0	361	0	0	3
Português	0	0	0	242	3	1	1058
Sérvio	0	0	0	223	0	0	0
Ucraniano	0	0	0	213	0	0	0
Francês	11	0	1	175	15	1	5
Tcheco	0	0	0	135	0	0	0
Búlgaro	0	0	0	106	1	0	0
Croata	0	0	0	103	0	0	0
Russo	0	0	1	56	1	0	0
Lituano	0	0	0	34	0	0	0
Esloveno	0	0	0	27	0	0	1
Basco	0	0	0	24	0	0	0
Romeno	1	0	0	21	1	0	0
Galego	0	0	0	15	0	0	1
Bósnio	1	0	0	7	0	0	0
Chinês	0	0	110	6	0	0	0
Macedônio	0	0	0	2	0	0	0
Húngaro	0	0	1	2	0	0	0
Holandês	0	0	0	1	0	0	1
Servo-croata	0	0	0	1	0	0	0
Norueguês	0	0	0	1	0	0	0
Indonésio	0	0	95	1	0	0	0
Marata	0	0	2	0	0	0	0
Árabe	4	0	1	0	0	0	0
Persa	0	0	62	0	0	0	0

Idioma	África	Antártica	Ásia	Europa	América do Norte	Oceania	América do Sul
Canarês	0	0	33	0	0	0	0
Malaiala	0	0	26	0	0	0	0
Malaio	0	0	8	0	0	0	0
Nepalês	0	0	3	0	0	0	0
Dinamarquês	0	0	0	0	0	0	1

Fonte: dados da pesquisa (2020).

Para criação de gráficos, duas funções foram criadas. A primeira, *plot_ww_language_circle(languages_sum_percentage, column, list_color)*, cria um gráfico circular com as porcentagens dos idiomas mais utilizados nas publicações disponibilizadas no E-LIS. Como argumentos, uma tabela com a soma total de cada idioma, *languages_sum_percentage*, o nome da coluna dessa tabela que será considerada para elaboração do gráfico, *'total'*, e uma lista de códigos hexadecimais para colorir o gráfico são fornecidos. Nessa função, optou-se pela exibição dos 10 idiomas principais. O cálculo da porcentagem é feito automaticamente pela função, considerando apenas os 10 idiomas selecionados, ou seja, é uma porcentagem relativa a esses 10 e não ao geral.

Também foi possível estabelecer os idiomas mais comuns por continente, utilizando a segunda função: *plot_language_by_continent(languages_df)*. Apesar da existência da tabela 5, em que observa-se os idiomas para todos os continentes, é difícil verificar quais os idiomas mais presentes nas publicações por continente, porque só é possível ordenar uma das colunas pelo maior valor. Assim, utilizando-se a tabela *languages_df*, selecionou-se uma coluna por vez e os dados foram ordenados em ordem decrescente para extração dos 10 principais idiomas. Em alguns casos, é possível que haja menos idiomas, pois não há grande diversidade de línguas nos documentos publicados naquele continente, seja por homogeneidade no idioma, seja pela quantidade de publicações. A partir dessas informações, um gráfico para cada continente foi gerado (Apêndice E).

f) Periódico científico

Para identificação dos periódicos com mais publicações depositadas no E-LIS, a função *count_journals_values(open_file, continents)* foi criada. Seguindo a mesma fórmula das anteriores, essa função recebe os arquivos abertos, coleta os dados da coluna “*publication*” ou periódicos, para documentos que foram publicados nesse formato, conta os valores para cada continente e os adiciona a uma tabela denominada *journals_df*.

Com a tabela, prossegue-se para a criação dos gráficos. A função *plot_journals_by_continent(journals_df, continents)* dá conta disso, no entanto, os periódicos possuem nomes grandes, muitas vezes com mais de cinco palavras, o que não favorece a criação de gráficos. Apesar disso, os gráficos foram criados e estão disponíveis no Apêndice F. Para apresentação dos resultados, optou-se, então, pela apresentação na forma de tabelas. As tabelas foram elaboradas a partir da função *journals_table(journals_df, continents)*. Essa função seleciona a coluna de cada continente na tabela *journals_df*, ordena os valores e os adiciona a uma nova tabela, assim, é possível verificar quais periódicos têm mais publicações depositadas pelos seus autores. Uma tabela foi criada para cada continente, uma vez que os principais periódicos de um continente podem não ser os mesmos de outro.

g) Evento científico

Existem quatro tipos de documentos, considerando as tipologias que são depositadas no E-LIS, que podem ser publicadas em eventos: anais de evento, artigo de conferência, pôster de conferência e apresentações. Ao tentar localizar a coluna correspondente aos eventos nas planilhas, duas opções foram encontradas: “*conference*” e “*event_title*”. Poderia-se supor que a primeira opção é a correta, mas como isso não está explícito e a segunda opção existe, considerou-se por bem analisar as duas.

As funções *count_events_values(open_file, continents)* e *count_conf_values (open_file, continents)*, executam as mesmas tarefas, com a diferença que uma atua na coluna *conference* e a outra na *event_title*. Assim, as funções, a partir dos documentos abertos, buscam a coluna especificada, contam os valores por evento e os divide por continentes. Ao final, esses dados são adicionados a uma tabela, *conference_df* e *events_df*, respectivamente. Os resultados mostram que a nomenclatura *event_title* foi pouco utilizada, 11 títulos de eventos foram localizados. Isso pode indicar a renomeação do campo de metadados de *event_title* para *conference* ao longo do tempo. Quando um usuário submete um dos tipos de documentos listados, a opção de título do evento aparece como “*Conference*”.

A função *plot_conf_by_continent(conference_df, continents)* foi criada, mas como ocorre com os periódicos, os nomes dos eventos são extensos e não ficam bem representados em gráficos. Optou-se, assim, pela criação da função *conference_table(conference_df, continents)*, para elaboração de tabelas com a contagem dos valores por evento, obtendo-se uma tabela para cada continente.

h) Palavras-chave

Por fim, a partir das palavras-chave inseridas pelos autores no momento de submissão de um documento, elaborou-se uma nuvem de palavras. Nuvem de palavras, *wordcloud*, é uma técnica para mostrar palavras mais frequentes dentro de um texto.

A função *get_keywords(open_file, continents)*, então, busca nas planilhas abertas a coluna *keywords*, extrai todas as palavras-chave e as adiciona a uma lista. Para cada continente uma lista é gerada, assim, obtêm-se uma lista de listas, isto é, uma lista geral contendo as sublistas de palavras-chave de cada continente. Para uma análise geral das palavras-chave, fez-se necessário juntar o conteúdo de todas as sublistas em uma única lista.

Com a lista criada, utiliza-se o módulo *wordcloud* para gerar uma nuvem de palavras. Esse módulo tem um algoritmo simples, possibilita maior personalização da imagem criada e preenche todos os espaços disponíveis, tornando a nuvem de palavras mais agradável visualmente (MUELLER, 2020). Com a existência desse módulo, não foi necessário desenvolver um programa para criar uma nuvem de palavras do início, apenas utilizar as funcionalidades dele e aplicar os dados da pesquisa.

Deste modo, por meio das funções descritas para cada categoria de análise e as tabelas geradas, foram criados os gráficos e outras representações visuais para subsidiar a discussão dos resultados no capítulo 7.

7 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Neste capítulo, busca-se atender aos objetivos específicos b) analisar, por meio de indicadores bibliométricos, a tipologia dos documentos depositados, assuntos, data de publicação, data de depósito, idioma, periódicos científicos com mais publicações, eventos científicos e palavras-chave; c) realizar uma análise geopolítica comparativa, a partir dos documentos depositados no repositório; e d) identificar, caracterizar e analisar as práticas de autoarquivamento dos pesquisadores da Ciência da Informação no E-LIS.

A partir dos dados gerados pelo ELIScript, são apresentados os resultados da pesquisa considerando-se as categorias de análise apresentadas.

7.1 Distribuição por tipologia de documentos

Ao iniciar o processo de submissão de um item no E-LIS, o primeiro passo é selecionar o tipo de documento que será submetido. No repositório, há uma orientação para cada item, informando as condições para que o documento se enquadre naquela tipologia. No Quadro 6, essas orientações são apresentadas.

Quadro 6 – Tipologias de documentos e orientações para sua submissão

Tipologia dos documentos	Orientação para submissão
Preprint	Selecione se o seu título não foi publicado.
Tese/Dissertação	Selecione para qualquer tipo de monografia, como teses, projetos de teses e dissertações.
Livro	Selecione para livros ou outras monografias.
Capítulo de Livro	Selecione para uma parte de um livro ou outras monografias.
Bibliografia	Selecione para monografias que se enquadrem claramente na categoria de bibliografia.
Guia/Manual	Selecione para livros ou outras monografias que se enquadrem claramente na categoria de guia/manual.
Tutorial	Selecione para todos os tutoriais.
Material Instrucional de Biblioteca	Selecione para documentos que ensinam sobre questões bibliotecárias.
Anais de Evento	Selecione apenas se estiver depositando anais inteiros da conferência.
Pôster de Conferência	Selecione apenas se estiver depositando um único pôster da conferência.
Apresentação	Selecione apenas se você está depositando uma única

Tipologia dos documentos	Orientação para submissão
	apresentação de conferência sem um artigo de conferência.
Projeto/plano de negócios	Selecione apenas se você estiver depositando itens como projeto/plano de negócios.
Relatório	Selecione apenas se você estiver depositando itens como relatório.
Relatório técnico departamental	Selecione apenas se estiver depositando itens como relatórios técnicos departamentais.
Relatório técnico	Selecione apenas se estiver depositando itens como relatórios técnicos.
Artigo de periódico não paginado	Selecione esta opção para artigos de periódico on-line não paginado (por exemplo, um periódico em html).
Artigo de periódico paginado	Selecione esta opção para artigos de um periódico paginado (impresso ou on-line em pdf).
Resenha crítica	Selecione apenas se você está depositando uma resenha crítica de outro documento.
Matéria de jornal/revista	Selecione isto para todos os artigos de jornais/revistas
Em coleção	Selecione apenas se você estiver depositando um grupo de documentos que foram coletados ou pertencem à mesma série.
Conjunto de dados	Selecione apenas se estiver depositando uma coleção ou agrupamento logicamente significativo de dados semelhantes ou relacionados, geralmente reunidos como uma questão de registro ou para pesquisa.
Outro	Selecione para algo dentro do escopo do repositório, mas não coberto pelas outras categorias.

Fonte: E-LIS (2020, *tradução nossa*).

Algumas das descrições convergem e não explicitam a diferença entre cada tipo de documento. Por exemplo, existem três tipos de relatórios: Relatório, Relatório Técnico Departamental e Relatório Técnico. A descrição, em inglês, dos dois últimos é igual, a não ser pelo plural “*technical reports*” e “*technical report*”. Supõe-se que, no primeiro, faltou o termo “*department*”, que restringiria esses relatórios técnicos a relatórios produzidos em um contexto departamental. No caso do Relatório, não há especificação do tipo de relatório, dando margem para que qualquer tipo de relatório seja depositado. Esse tipo de omissão e a falta de orientações mais detalhadas podem dificultar o processo de autoarquivamento para o autor.

No caso do Artigo de Periódico Paginado e o Artigo de Periódico Não Paginado, a única diferença, como os nomes explicitam, é que um foi publicado em um periódico com paginação e o outro não. Para diferenciar, as orientações do repositório indicam que o Artigo de Periódico Não Paginado é aquele publicado em páginas HTML, isto é, em páginas on-line nas quais não se adiciona a numeração. Para fins de análise, e conforme sugestão da banca do II Fórum Discente

do Programa de Pós-Graduação em Gestão e Organização do Conhecimento (FORPED/PPG-GOC), ambos serão considerados como uma única categoria “Artigo de Periódico”.

O último trabalho a avaliar a coleção completa do E-LIS foi realizado por Santillán-Aldana (2009), com dados de 2007, quando o repositório tinha cerca de 7.000 documentos. Em seu trabalho, Santillán-Aldana (2009) contabiliza a quantidade pelos tipos e apresenta suas respectivas porcentagens. De forma comparativa, apresenta-se os valores do artigo de Santillán-Aldana e os valores referentes a todos os documentos depositados desde 2003 até 2020 obtido com o ELIScript (Quadro 7).

Quadro 7 – Comparativo de tipos de documentos depositados até 2007 e até 2020

Tipo de documento	Ano (2007)*	% (2007)	Ano (2020)**	% (2020)
Artigo de periódico	3323	48	10861	47
Artigo de conferência	1662	24	4431	19
Apresentação	699	10	1826	8
<i>Preprint</i>	233	3	1072	5
Tese/Dissertação	192	3	731	3
Capítulo de livro	161	2	1171	5
Relatório técnico departamental	111	2	4	0
Matéria de jornal/revista	89	1	320	1
Pôster de conferência	84	1	530	2
Resenha Crítica	80	1	299	1
Livro	56	1	391	2
Guia/Manual	50	1	243	1
Relatório técnico	42	1	51	0
Bibliografia	28	0	98	0
<i>Anais</i> de evento	25	0	102	0
Projeto / Plano de negócios	21	0,4	58	0
Material instrucional de biblioteca	20	0,4	19	0
Tutoriais	17	0,4	62	0
Ementa	1	0	0	0
Relatório	0	0,4	511	2
Conjunto de dados (<i>Dataset</i>)	0	0,4	10	0
Outro	0	0,4	455	2
TOTAL:	6894	100	23245	100

*Valores do artigo de Santillán-Aldana (2009);

** Valores obtidos com o ELIScript.

Fonte: dados da pesquisa (2020).

Em relação a 2007, observa-se um aumento percentual de 137% na quantidade de documentos depositados até 2020. Ao todo foram coletados dados de 23.245 documentos, no entanto, o E-LIS mostrava nos dados estatísticos em sua página inicial, no momento da coleta, que existiam 22.313 documentos depositados. Na página de consulta por países, por outro lado,

o número corresponde a 22.135 documentos. É possível que essa dissonância esteja relacionada com o tempo de atualização das informações no *site*. Durante todo o período da pesquisa, os dados entre a página inicial e a página de consulta por países, dos quais os dados para a pesquisa foram extraídos, não sincronizaram.

Santos e Oliveira (2019), no estágio inicial de criação do ELIScript, haviam identificado metadados de 34.534 documentos. No entanto, verificou-se em testes posteriores que esse total está incorreto. Isso ocorre porque algumas funções do código de análise não podem ser executadas mais de uma vez seguidamente. Ao executá-lo repetidamente, sem realizar a reinicialização, a função criada para somar valores não substitui o primeiro valor, ela soma o primeiro valor ao último valor obtido, gerando erros, ou a adição de valores que não deveriam ser somados. Dessa forma, cabe salientar que o código do ELIScript deve ser executado de uma só vez para gerar valores corretos.

Dentre os dados coletados até 2020, 47% dos documentos correspondem a artigos de periódico, valor próximo ao observado por Santillán-Aldana (2009) em 2007, 48%. Identificar que quase 50% dos documentos depositados são artigos de periódicos foi esperado, uma vez que os artigos se configuram como o meio “oficial” de comunicação da ciência (BAPTISTA *et al.*, 2007). Mas faz-se importante ressaltar que diversos tipos de documentos são aceitos para submissão e podem contribuir com atividades de outros autores, pesquisadores, estudantes e profissionais.

Em seguida, aparecem os Artigos de Conferência e as Apresentações, itens que acompanham o ciclo de comunicação científica, fazendo parte do estágio inicial das pesquisas. Santillán-Aldana (2009) aponta o esforço dos editores do E-LIS como sendo responsável pela grande quantidade de artigos de conferências depositados até 2007. Os editores do repositório costumavam solicitar em eventos de Biblioteconomia e Ciência da Informação que os apresentadores depositassem seus trabalhos no E-LIS. Apesar do aumento na quantidade de documentos depositados em comparação a 2007, a porcentagem para essas duas tipologias diminuiu, o que indica o crescimento no autoarquivamento de outras tipologias de documento.

A porcentagem de *Preprints* depositados, por exemplo, aumentou 2% em comparação aos dados de 2007. Os *preprints* são documentos distribuídos antes da publicação e até mesmo antes da finalização do processo de revisão por pares (SØNDERGAARD; ANDERSEN; HJØRLAND, 2003). Para Goldschmidt-Clermont (2002, on-line, *tradução nossa*), eles se caracterizam como “cópias quase impressas de manuscritos submetidos para publicação em periódicos. Eles são documentos temporários cuja função é preencher a lacuna de tempo criada pelos atrasos na publicação de artigos”.

A urgência na disponibilização de *preprints* pode se dar pela necessidade de estabelecer prioridade sobre um trabalho ou ideia, mas também para dar acesso rápido e fácil a novas

descobertas. Uma situação que pode demonstrar essa urgência foi o surgimento do COVID-19³⁹, pois, suscitou a necessidade acelerada de disseminação de pesquisas em busca de compartilhar informações e entender as características desse novo vírus, quais sintomas desencadeia, como age no corpo, quais as implicações da infecção e como curá-la. Muitas pesquisas foram e estão sendo distribuídas por servidores de *preprints*, sem revisão por pares, para agilizar a comunicação e auxiliar no entendimento dessa doença e de tantos outros tópicos.

A Biblioteconomia e a Ciência da Informação não possuem as mesmas características aceleradas das Ciências da Saúde ou das Ciências Exatas, mas a distribuição de um *preprint* pode ser feita com intuito de receber comentários e pareceres antes da submissão do manuscrito para um periódico. Dessa forma, os *preprints* devem ser entendidos como um meio e não como um fim em si mesmo. Ele serve, sim, para dar acesso a novas ideias, estabelecer prioridade, receber *feedback*, mas não deve ser considerado como um artigo finalizado. O *preprint* é um material que deve passar pela revisão por pares para ser validado cientificamente e, assim que publicado, pode e deve ser destruído, referenciando-se apenas a publicação, uma vez que o *preprint* pode conter informações incompletas e/ou incorretas (GOLDSCHMIDT-CLERMONT, 2002).

Nota-se que praticamente todas as tipologias de documento estão contempladas na amostra geral, observando-se as devidas proporções. No entanto, a categoria “Ementa” aparece na análise de Santillán-Aldana (2009), mas não existe como categoria de depósito em 2020. Se observados os itens permitidos para submissão no repositório, vê-se que a categoria “Ementa”, *Syllabus* em inglês, não está mais presente. Durante o perscrutamento dos dados, nenhum documento foi identificado com essa tipologia, o que impossibilitou verificar que tipo de ementa estava sendo depositada. Além disso, os relatórios técnico-departamentais tinham o total de 111 depósitos na análise do autor, mas em 2020, apenas quatro foram localizados. Não é possível determinar qual a causa disso, se foi remanejamento dos documentos para outra categoria (visto que existem mais duas categorias de relatórios: relatórios técnicos e relatórios), ou se os documentos foram excluídos (pelos usuários ou pelos editores).

Uma tipologia que levantou dúvidas na análise foi “*Review*”. Inicialmente, traduziu-se para “Revisão de Literatura” (SANTOS; OLIVEIRA, 2019). Após uma inspeção mais aprofundada tendo como base as instruções de submissão do próprio repositório e alguns dos documentos depositados sob essa tipologia, foi observado que tratam-se, na verdade, de Resenhas Críticas.

Pelo fato de o E-LIS ter uma lista bem compreensiva de tipos de documento, uma categoria que se destacou foi “Outro”, que passou de zero documentos em 2007 para 455 em 2020. Analisando os itens depositados, verificou-se a existência de arquivos de imagem, vídeo, áudio, arquivos compactados, muitos deles caracterizados como “não publicado”. Nesses casos,

39 COVID-19 é a doença causada por um novo coronavírus denominado SARS-CoV-2. A Organização Mundial de Saúde (OMS) soube deste novo vírus em 31 de dezembro de 2019, após um relatório de um grupo de casos de "pneumonia viral" em Wuhan, na República Popular da China (WHO, 2020).

faz sentido que estejam na categoria “Outro”, no entanto, foram identificados também itens como artigos de conferência, resenhas críticas, guias/manuais, matérias de jornais/revistas, tipologias já existentes no repositório, o que revela um descuido no processo de autoarquivamento.

Outro ponto intrigante é trazido pela tipologia Conjuntos de Dados. Apenas 10 itens foram localizados nessa categoria. Considerando-se o contexto da crescente quantidade de dados de pesquisa sendo gerada, parece que o E-LIS está dando um passo adiante na busca por possibilitar o acesso aos dados de pesquisa oriundos dos itens publicados no repositório. Apesar da categoria Conjunto de Dados existir no E-LIS, localizou-se uma comunidade no Zenodo com o nome *eLIS_data*⁴⁰. O Zenodo é um repositório de acesso aberto de uso geral desenvolvido sob o programa europeu OpenAIRE e operado pela Organização Europeia para a Pesquisa Nuclear, também conhecida como CERN, acrônimo para *Conseil Européen pour la Recherche Nucléaire*. Ele permite que os pesquisadores depositem conjuntos de dados, *software* de pesquisa, relatórios e outros artefatos digitais relacionados à pesquisa (ZENODO, 2020).

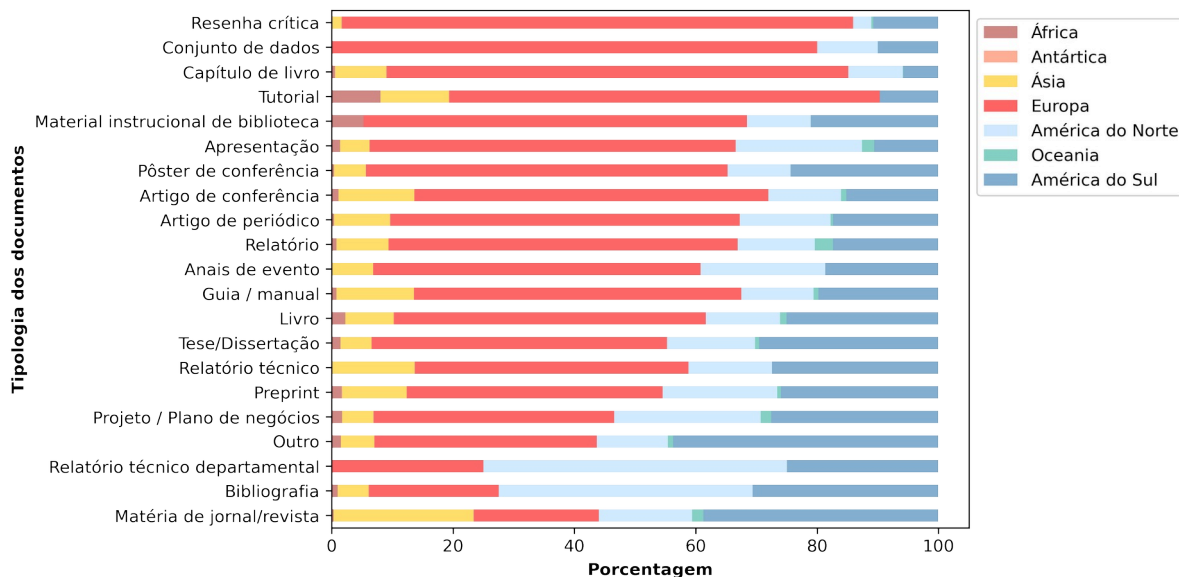
Trabalhando na perspectiva de dados ligados entre si (*Linked Data*), a comunidade do E-LIS no Zenodo possibilita a vinculação de dados publicados nesse repositório com publicações depositadas no E-LIS por meio do protocolo OAI-PMH. E qual a vantagem de publicar o conjunto de dados no Zenodo? Além de ser um repositório maior, incluindo diversas áreas do conhecimento e focado na disponibilização de dados de pesquisa, o Zenodo atribui um DOI para cada conjunto de dados publicado, funcionalidade não implementada no E-LIS.

A comunidade do E-LIS no Zenodo, *eLIS_data*, foi criada em 2017 e lá encontra-se apenas um conjunto de dados categorizado como material suplementar de um artigo de conferência, esse conjunto foi publicado em 2019. Não foi localizada nenhuma menção a essa comunidade do Zenodo no próprio E-LIS, nem nenhum tipo de incentivo para que autores publiquem seus conjuntos de dados nessa comunidade. Entretanto, no artigo mais recente dos editores do E-LIS, De Robbio, Subirats e Peset (2020) afirmam que foi incluído um novo campo no E-LIS, como adição ao conjunto de metadados de artigos de periódico, denominado Link para Dados de Pesquisa (*Link to Research Data*) e esse link deve ser utilizado para vincular o artigo com o conjunto de dados disponível no Zenodo, para aqueles documentos que possuem um conjunto de dados como ferramenta útil para a validação da pesquisa na Biblioteconomia e Ciência da Informação.

No gráfico 1, por fim, faz-se uma comparação entre os tipos de documentos depositados por autores de cada continente. Considera-se que cada barra corresponde a 100% de uma tipologia de documento e cada cor corresponde a um continente.

40 E-LIS data community. Zenodo. 2020. Disponível em: https://zenodo.org/communities/elis_data/. Acesso em: 27 out. 2020.

Gráfico 1 – Comparativo entre os tipos de documentos por continente



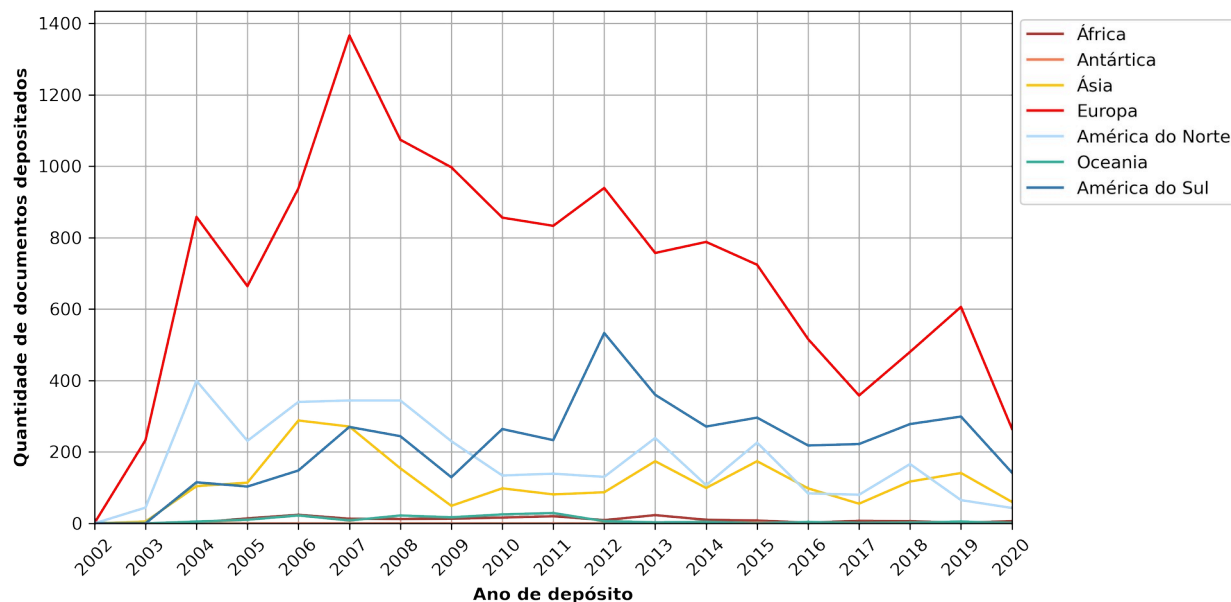
Fonte: dados da pesquisa (2020).

Com esta representação visual, nota-se a predominância da Europa no depósito de praticamente todos os tipos de documento, com exceção dos Relatórios Técnico Departamentais e Bibliografias, sendo ultrapassado pela América do Norte e Central; e Matéria de Jornal/Revista e Outro, que foi mais depositado por pesquisadores da América do Sul. Apenas a Europa (vermelho), a América do Norte e Central (azul-claro), e a América do Sul (azul-escuro) realizaram depósito de todas as tipologias de documentos. A Ásia (amarelo) também aparece em destaque, embora não tenha sido localizado nenhum depósito de Conjunto de Dados, nem de Relatório Técnico Departamental. A Antártica só possui dois documentos depositados, um Capítulo de Livro e um *Preprint*, tornando-se imperceptível no gráfico.

7.2 Distribuição temporal de depósito dos documentos

Como visto anteriormente na Revisão de Literatura (subtópico **3.3.2 E-prints in Library and Information Science**), a evolução da coleção do E-LIS pode ser observada nos artigos de Morrison *et al.* (2007), De Robbio e Katzmayr (2009), e Neves e Ferreira (2014). No Gráfico 2, é possível acompanhar o crescimento apontado pelos autores, bem como a comparação da quantidade de documentos depositados por continente, com os dados coletados pelo ELIScript.

Gráfico 2 – Quantidade de depósitos por ano



Fonte: dados da pesquisa (2020).

O período registrado em que houve depósito de documentos no repositório compreende os anos de 2002 até 2020. Entre 2003 e 2007, o repositório já contava com mais de cinco mil documentos. Em 2009, cerca de nove mil documentos estavam disponíveis (DE ROBBIO; KATZMAYR, 2009). Neves e Ferreira (2014) afirmam, por seu turno que já ultrapassavam mais de 16 mil documentos sendo disponibilizados em 2014. Esses dados foram retirados de alguns artigos e confirmados a partir de análise quantitativa.

Até 2004, os depósitos realizados no repositório estavam em uma curva crescente, experimentando o primeiro declínio em 2005. Em 2007 ocorre o maior pico de autoarquivamento, com mais de 3,5 mil documentos sendo depositados no mesmo ano. A partir desse período, no entanto, há um declínio na quantidade de depósitos, com picos em 2012, 2014 e 2019 (SANTOS; OLIVEIRA, 2019).

Conforme Morrison e outros autores (2007), a Equipe Editorial do E-LIS era responsável por promover o repositório, trabalhando em contato com organizadores de eventos e conferências de Biblioteconomia e Ciência da Informação em mais de 40 países. Os autores mencionam a inclusão de trabalhos apresentados nos eventos: International Workshop on Webometrics, Informetrics and Scientometrics e no COLLNET Meeting, na França; Annual Meeting of the American Society for Information Science and Technology (ASIST), nos Estados Unidos; e International Bielefeld Conference, realizado na Alemanha, como uma intervenção bem-sucedida, pois elevou a quantidade de depósitos no E-LIS até 2006. Essa iniciativa de divulgação tinha como objetivo aumentar a consciência dos autores sobre a prática de

autoarquivamento e incentivá-los a realizar submissões de outros documentos no futuro. Não se localiza na literatura mais menções sobre esses esforços nos anos seguintes.

O cessar dessas atividades de divulgação pode ter influenciado na diminuição da quantidade de depósitos, já que os editores do E-LIS não realizam mais chamadas para que os pesquisadores e profissionais da Biblioteconomia e Ciência da Informação façam depósitos no repositório. Conforme Singeh, Abrizah e Karim (2013), entende-se que somente a conscientização dos pesquisadores acerca da importância do Acesso Aberto e dos repositórios digitais pode motivá-los a contribuir com materiais para o repositório. O trabalho de divulgação deve ser contínuo e começar desde a graduação, nos cursos de Biblioteconomia, Arquivologia e Museologia, passando pelos Programas de Pós-Graduação da Ciência da Informação, sendo também incentivado por bibliotecários e outros profissionais da informação em eventos da área.

De forma geral, houve uma grande queda na quantidade de depósitos realizados. Embora o autoarquivamento continue ocorrendo, este acontece numa frequência menor. Conforme Kim (2010), o depósito de documentos pode ser feito em uma variedade de plataformas, desde repositórios temáticos ou institucionais até em sites de redes sociais. Com o surgimento das redes sociais acadêmicas, como o ResearchGate, o AcademiaEdu, e os gerenciadores de bibliografia que também permitem a criação de um perfil para leitura e disponibilização de documentos, como o Zotero e o Mendeley, muitos autores perceberam a oportunidade de compartilhar seus trabalhos com seus pares e seguidores de outra forma que não nos repositórios digitais. De acordo com Couto e Ferreira (2019, p. 7),

pesquisadores criam perfis para divulgar o seu trabalho e acompanhar o trabalho dos colegas em *sites* de redes sociais acadêmica. Geralmente, pode-se seguir outros pesquisadores ou tags por assunto, colecionando seguidores e formando uma rede de contatos interessados. Presume-se que os seguidores de um cientista, nesse tipo de *site*, sejam outros cientistas da mesma área, interessados em seu trabalho e dispostos a ler seus textos.

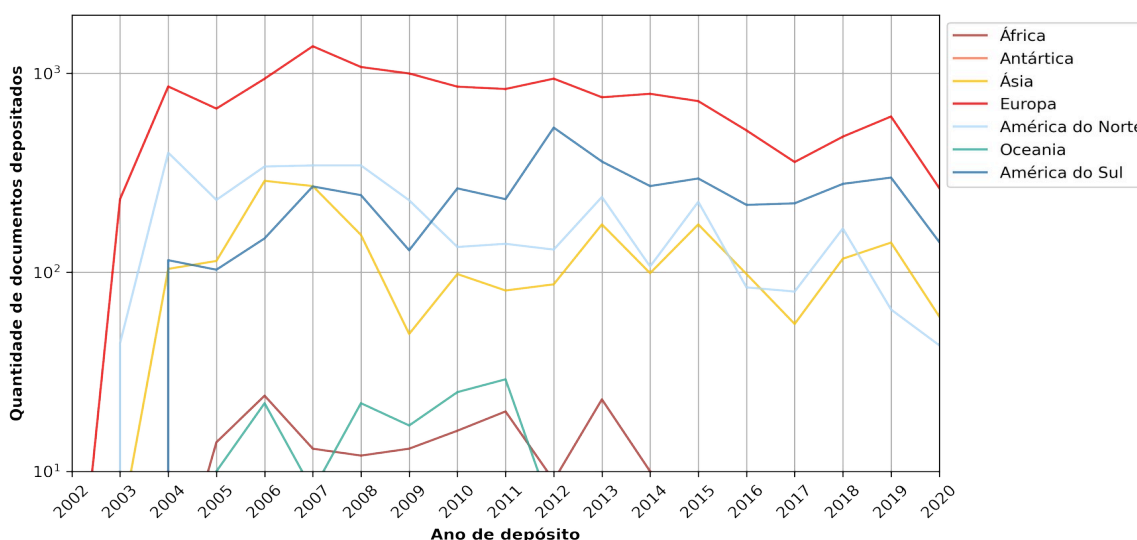
Desse modo, as redes sociais acadêmicas têm funcionado como uma espécie de portfólio para os pesquisadores, fornecendo, inclusive, indicadores de interesse como quantidade de leitores e de downloads dos documentos. Esse tipo de retorno e a possibilidade de interação nesse formato de rede social podem influenciar os pesquisadores na escolha de onde depositar seus trabalhos, podendo diminuir a quantidade de depósitos em repositórios temáticos ou institucionais que não possuam esses atributos, embora o depósito em um não impossibilite a disponibilização no outro.

Como apontado anteriormente, o aumento de depósitos no início do repositório não foi orgânico, por esse motivo uma queda na quantidade de autoarquivamentos deveria ser esperada. Ademais, não se pode afirmar, com certeza, qual o motivo da diminuição de depósitos no E-LIS: se foi a falta de divulgação, o desconhecimento do repositório por parte dos

pesquisadores e profissionais da informação, a escolha por realizar o depósito de documentos em outras plataformas, ou conjunto desses três fatores.

Vale ressaltar, todavia que a quantidade de depósitos flutuou em torno dos milhares na Europa; na América do Sul, América do Norte e Central e Ásia, em torno das centenas; e na África, Antártica e Oceania, os valores foram baixos, em torno de dezenas e, por vezes, até abaixo disso, como demonstrado no Gráfico 3.

Gráfico 3 – Comparação logarítmica da quantidade de depósitos por ano



Fonte: dados da pesquisa (2020).

Os depósitos na Ásia e América do Norte e Central começaram a ser realizados em 2003. Em 2004, trabalhos de outros continentes aparecem, como da América do Sul, África e Oceania. Os dois trabalhos depositados na Antártica ocorreram em 2015 e não houve alteração até então. A Europa foi o único continente em que trabalhos foram autoarquivados em 2002. Como a data de lançamento do E-LIS foi em 2003 e o repositório foi criado na Itália, supõe-se que esses primeiros depósitos, antes do lançamento oficial, constituíam testes para averiguar a viabilidade do repositório. De Robbio, Subirats e Peset (2020) estimam que, no geral, cerca de 100 documentos são depositados por mês no E-LIS.

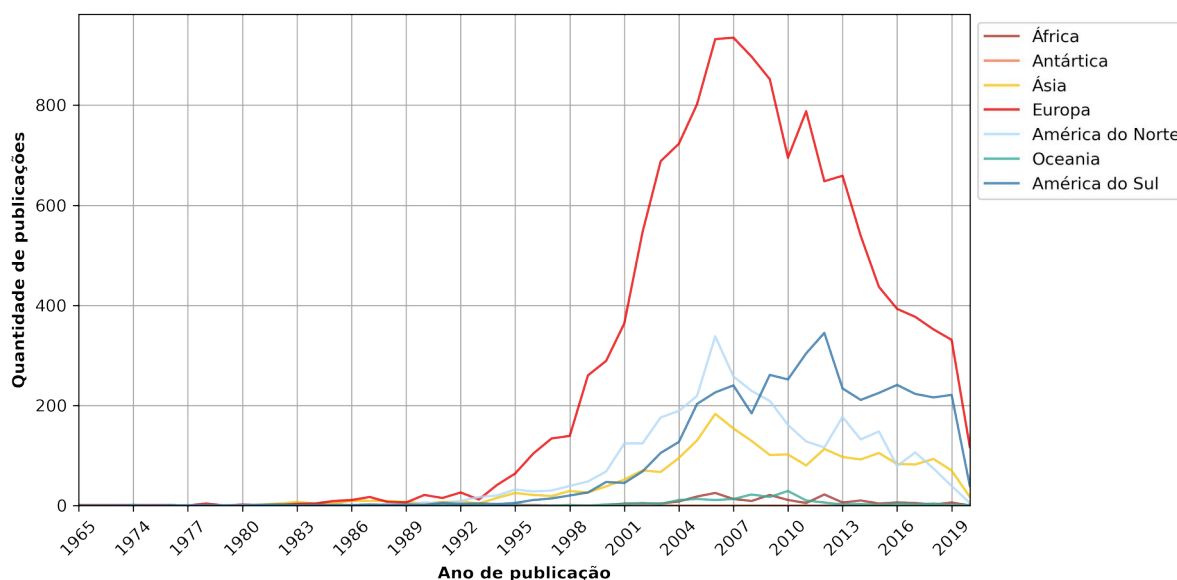
7.3 Distribuição temporal de publicação dos documentos

Nenhum dos trabalhos que analisaram a produção científica disponível no E-LIS, seja o acervo inteiro ou a produção específica de um país como Brasil, Portugal ou Cuba, abordou a atualidade dos documentos (ARENCEBIA-JORGE; SANTILLÁN-ALDANA; SUBIRATS-COLL, 2005; MORRISON *et al.*, 2007; WEITZEL; LEITE; ARELLANO, 2008; SANTILLAN-ALDANA, 2009). Fazer o depósito de documentos é importante para a preservação, para além disso, do

ponto de vista do usuário, considera-se que ter acesso a documentos atuais também pode incentivá-los a buscar fontes de informação no repositório.

A distribuição temporal de publicação dos documentos corresponde à atualidade dos documentos, isto é, ao ano de publicação. No preenchimento dos metadados referentes à data de publicação, é possível que o autor adicione dia, mês e ano. No entanto, observou-se que muitos autores optaram por adicionar apenas o ano, ou mês e ano. Para a análise, apenas o ano foi considerado, por ser o dado que constava em todas as submissões. No Gráfico 4, têm-se a comparação da quantidade de publicações por ano entre os continentes.

Gráfico 4 – Ano de publicação dos documentos depositados



Fonte: dados da pesquisa (2020).

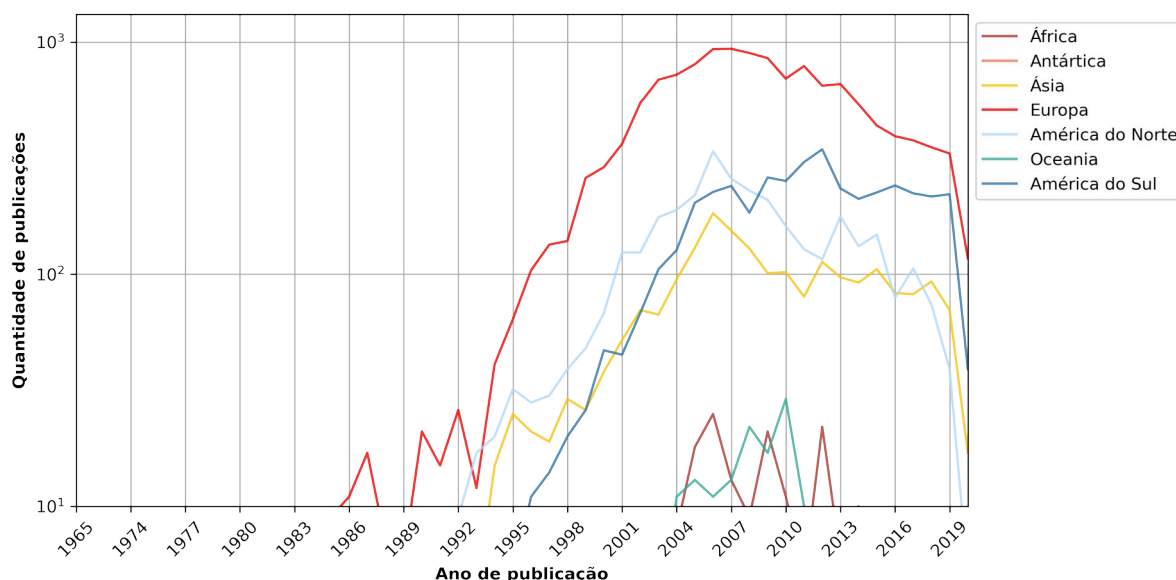
O período de publicação dos documentos compreende os anos entre 1965 e 2020, evidenciando que o repositório continua a ser utilizado por pesquisadores para realização do depósito, e que parte desses trabalhos são atuais. Apesar de haver um documento datando de 1965, não parece ser comum o depósito de documentos antigos no E-LIS. Observa-se o autoarquivamento esporádico de documentos publicados entre 1965 e 1983, sem representação muito evidente no gráfico.

Percebe-se, em comparação, a relação entre o ano de publicação e o ano de depósito dos documentos. Laakso (2014) aponta que o momento em que autoarquivamento pode ser realizado depende das políticas editoriais do periódico, no caso de artigos ou dados publicados em revistas científicas. O conteúdo pode ser depositado no mesmo momento em que é publicado no periódico ou algum tempo após a publicação. Conforme os dados apresentados, 2007 foi ano com mais depósitos e também é o ano com mais publicações, o que pode indicar a realização do autoarquivamento dos trabalhos no mesmo ano em que eles foram publicados.

A partir de 1993, há um aumento na distribuição dos documentos publicados na Europa, América do Norte e Central, América do Sul e Ásia. Nesses continentes, a maior parte dos documentos foram publicados entre 2007, considerando o pico, e 2019. Os documentos na África e na Oceania, por sua vez, foram publicados a partir de 1995 e 1998, respectivamente. Evidencia-se, ainda, um crescimento no depósito de trabalhos publicados a partir de 2003 até 2015 nesses dois continentes. Na Antártica, um trabalho foi publicado em 2013 e o outro em 2015.

Como a Europa possui muito mais documentos publicados por ano depositados no E-LIS, há dificuldade em visualizar as diferenças entre a data de publicação dos documentos entre os outros continentes no Gráfico 4. Similar à comparação da data de depósito dos documentos, foi criada uma visualização logarítmica para evidenciar essas diferenças (Gráfico 5).

Gráfico 5 – Comparação logarítmica do ano de publicação dos documentos depositados



Fonte: dados da pesquisa (2020).

Em mais detalhe, observa-se que a Ásia, América do Sul e América do Norte e Central, possuem um comportamento semelhante, assim como a Oceania e a África se aproximam, no que concerne à data de publicação dos trabalhos. Na Europa, localiza-se mais de 500 trabalhos publicados em cada ano a partir de 2002 até 2019; a maior quantidade, 935 documentos, foi publicada em 2007. Até o momento da coleta de dados, realizada em setembro de 2020, já haviam 117 documentos depositados nesse continente. Na América do Sul, 2012 é o ano com mais publicações, 345 ao todo. Da mesma forma, o ano que se destaca na África é 2012, com 22 trabalhos. Na Ásia e na América do Norte e Central, a maior quantidade de documentos foi

publicada em 2006, 183 e 338 trabalhos, respectivamente. Por fim, a Oceania teve 29 documentos publicados em 2010, ano com mais publicações.

Conforme apresentado, os documentos depositados no E-LIS foram publicados num período extenso, entre 1965 e 2020, com destaque a partir dos anos 2000, em que todos os continentes têm trabalhos publicados. A depender do tipo de pesquisa a ser realizada e a temática abordada, o E-LIS pode se caracterizar como uma fonte de informação útil, pois possui trabalhos relativamente atuais. Relativamente porque, conforme Vergueiro (1995) salienta, diferentes áreas de conhecimento e, até mesmo, diferentes campos dentro da mesma área de conhecimento, sofrem desatualizações em seus conteúdos com maior ou menor frequência.

7.4 Temática dos documentos depositados

Por meio da classificação JITA, esquema utilizado para definir o assunto dos documentos no repositório, pôde-se averiguar quais as temáticas referentes aos documentos depositados no E-LIS.

No Quadro 8, estão as notações da classificação JITA, os assuntos e a respectiva quantidade de documentos que os abordam. Inicialmente, optou-se por apresentar esses dados com base nas classes principais de assunto, isto é, no primeiro nível hierárquico, assim como, uma visão geral do repositório, não havendo distinção entre os continentes.

Quadro 8 – Assuntos dos documentos depositados no E-LIS

ASSUNTO	QUANTIDADE
B. Uso da informação e sociologia da informação	1764
L. Tecnologia da informação e tecnologia de biblioteca	1594
I. Tratamento da informação para serviços de informação	1164
H. Fontes de informação, suportes e canais	1161
A. Aspectos teóricos e gerais de bibliotecas e informação	1219
C. Usuários, nível de instrução e leitura	1124
G. Indústria, profissão e formação	728
F. Gerenciamento	700
E. Publicação e questões legais	532
J. Serviços técnicos em bibliotecas, arquivos e museus	521
D. Bibliotecas como coleções físicas	476
K. Técnicas de Armazenamento	57

Fonte: dados da pesquisa (2020).

Osorio (2014), ao analisar a cobertura de assuntos dos documentos do E-LIS, verificou que o assunto mais comum no repositório era “Uso da Informação e Sociologia da Informação”. Conforme identificado, esse continua a ser o tema mais abordado no repositório. Além dele, “Tecnologia da Informação e Tecnologia de Biblioteca” e “Tratamento da Informação para Serviços de Informação” são os assuntos mais abordados nos documentos arquivados.

As três temáticas estão de acordo com o que Freire (2006) coloca como os olhares que refletem a Ciência da Informação, a saber: i. Recuperação da Informação, incluindo a representação da informação, as tecnologias de processamento da informação e os serviços da informação; ii. Comunicação da Informação, que envolve as TIC, especialmente as digitais; e iii. Estudos da Cognição, que implicam nos estudos de usuários e capacitação informacional. A partir dessas temáticas, é possível observar o que Saracevic (1996) caracteriza como razão da existência e da evolução da Ciência da Informação: a interdisciplinaridade, sua ligação inexorável com a tecnologia da informação e sua preocupação com o usuário. O tratamento da informação, como “núcleo duro” da Biblioteconomia, se faz muito presente nos documentos depositados.

A seguir, uma breve descrição do que cada classe de assunto aborda.

- **Uso da informação e sociologia da informação:** compreende os estudos de uso, disseminação e impacto da informação na sociedade. Está relacionado com as políticas de informação, economia da informação e sociedade da informação, considerando as necessidades de informações dos usuários;
- **Tecnologia da informação e tecnologia de biblioteca:** corresponde aos assuntos de teor tecnológico e suas relações com a biblioteca e a recuperação da informação. Aborda desde os meios de comunicação até as formas de registro da informação digital, englobando assuntos como a Internet, a rede de computadores, recuperação automática de texto, sistemas de busca, automação de bibliotecas e outros;
- **Tratamento da informação para serviços de informação:** nessa classe, estão os assuntos sobre tratamento e organização do conhecimento. O Tratamento da Informação “pode ser entendido, de modo geral, como um conjunto de ações, vinculadas à Organização do Conhecimento e da Informação, voltadas à análise, representação, descrição e síntese da informação, em ambientes sistêmicos” (MARTÍNEZ-ÁVILA; GRACIOSO, 2020, p. 51);
- **Fontes de informação, suportes e canais:** como o título sugere, essa classe engloba os documentos ou fontes de informação, em diferentes suportes e formatos, incluindo também os canais (impressos e digitais) de transmissão da informação como as bases de dados, os *sites* e os repositórios;

- **Aspectos teóricos e gerais de bibliotecas e informação:** aqui são classificados os trabalhos que abordam a Biblioteconomia e Ciência da Informação como um campo, suas teorias e especificidades, e suas relações com outras áreas do conhecimento;
- **Usuários, nível de instrução e leitura:** aborda os estudos de usuários, atividades de treinamento e competência informacional. Os estudos de usuários se baseiam no entendimento sobre as necessidades destes, suas preferências, opiniões e avaliações acerca dos serviços de informação que são ou poderão vir a ser oferecidos nas unidades de informação (FIGUEIREDO, 1990);
- **Indústria, profissão e formação:** relativo aos aspectos profissionais da Biblioteconomia e Ciência da Informação, incluindo as indústrias de informação, de *software*, computadores e telecomunicações. Aborda, ainda, elementos educacionais e curriculares para a formação de profissionais dessa área;
- **Gerenciamento:** envolve questões de marketing, relações públicas, financiamento e gestão de pessoas vinculadas aos sistemas e unidades de informação;
- **Publicação e questões legais:** sobre publicação impressa e eletrônica, discute a propriedade intelectual, venda de livros e censura de publicações;
- **Serviços técnicos em bibliotecas, arquivos e museus:** aborda os serviços realizados nas bibliotecas, como atividades de circulação de livros (empréstimos, devoluções, reservas, comutação), gestão de periódicos, preservação de documentos, digitalização entre outros;
- **Bibliotecas como coleções físicas:** menciona os tipos de bibliotecas físicas, a saber: bibliotecas mundiais, nacionais, públicas, acadêmicas, escolares, governamentais, privadas, especializadas, científicas, técnicas, médicas ou da saúde, além de instituições como os arquivos e museus;
- **Técnicas de Armazenamento:** instituições de guarda de documentos, como bibliotecas, arquivos museus e elementos de planejamento desse espaço, considerando aspectos arquitetônicos dos ambientes, questões de locomoção, segurança e preservação do acervo.

Geralmente, um documento tem mais de um assunto atribuído. Nesse sentido, Osorio (2014) identificou *clusters* temáticos na coleção do E-LIS, isto é, temáticas que se relacionam com outras mais frequentemente. De acordo com o autor, existem cinco *clusters* que representam as temáticas abordadas pelos documentos do repositório. Essas relações são apresentadas no Quadro 9.

Quadro 9 – Clusters temáticos do E-LIS

CLUSTER	NOTAÇÃO JITA	CLASSES
1	I, L, C	Tratamento da informação para serviços de informação; Tecnologia da informação e tecnologia de biblioteca; Usuários, nível de instrução e leitura
2	F, D, H, B	Gerenciamento; Bibliotecas como coleções físicas; Fontes de informação, suportes e canais; Uso da informação e sociologia da informação
3	K, E	Técnicas de Armazenamento; Publicação e questões legais
4	J, A	Serviços técnicos em bibliotecas, arquivos e museus; Aspectos teóricos e gerais de bibliotecas e informação
5	G	Indústria, profissão e formação

Fonte: Osorio (2014, p. 7).

Observa-se que os *clusters* apresentam um nível significativo de compreensão lógica. Por exemplo, no *cluster* 1 é possível verificar as relações entre o tratamento da informação, as tecnologias da informação e os usuários. O profissional da informação faz uso das técnicas de organização e tratamento da informação para representar os itens documentais e disponibilizá-los em uma base de dados digital, de forma que os usuários possam recuperá-los futuramente. No *cluster* 2, verifica-se as relações entre a gestão de unidades de informações e das fontes de informação. Pode-se inferir que alguns dos documentos abordam os processos e políticas de seleção de materiais para compor o acervo de uma biblioteca. Já no *cluster* 3, estão as questões de armazenamento de documentos e os aspectos legais que devem ser considerados. Cita-se, por exemplo, o caso das práticas de autoarquivamento. O autor deve estar ciente de seus direitos de propriedade sobre o item antes de fazer seu depósito em um repositório digital. No *cluster* 4, elementos de tratamento da informação e os aspectos teóricos da informação estão presentes. E isolado no *cluster* 5 está a classe “Indústria, profissão e formação”, que não parece se conectar com os assuntos apresentados anteriormente.

A seguir, levando em consideração o segundo nível hierárquico, assuntos que estão subordinados aos apresentados acima, identifica-se as temáticas específicas mais abordadas por continente (Quadro 10).

Quadro 10 – Assuntos específicos dos documentos divididos por continente

CONTINENTE	ASSUNTO	QUANT.
África	Repositórios (baseado ou não em OAI)	24
	Disseminação e difusão da informação	21
	Métodos bibliométricos	18
	Propriedade intelectual: os direitos de autor, propriedade, direitos autorais e copyleft	17
	Biblioteconomia e Ciência da Informação (CI) como um campo	11
América do Norte e Central	Disseminação e difusão da informação	252
	Informação na sociedade	250
	Uso e impacto da informação	199
	Formação	189
	Repositórios (baseado ou não em OAI)	180
América do Sul	Formação	273
	Informação na sociedade	237
	Uso e impacto da informação	227
	Disseminação e difusão da informação	227
	Biblioteconomia e Ciência da Informação (CI) como um campo	214
Antártica	Teoria da informação e teoria de bibliotecas	1
	Hiperlinks de referência	1
	Biblioteconomia e Ciência da Informação (CI) como um campo	1
	Bibliotecas de Instituições de Ensino Superior	1
Ásia	Métodos bibliométricos	253
	Internet, inclusive WWW	124
	Catálogo, controle bibliográfico	106
	Disseminação e difusão da informação	103
	Repositórios (baseado ou não em OAI)	90
Europa	Bibliotecas de Instituições de Ensino Superior	882
	Repositórios (baseado ou não em OAI)	834
	Disseminação e difusão da informação	740

CONTINENTE	ASSUNTO	QUANT.
	Métodos bibliométricos	696
	Formação	695
	Digitalização	36
	Nenhum destes, mas na seção Tecnologia da informação e tecnologia de bibliotecas	22
Oceania	Internet, inclusive WWW	16
	Catálogo, controle bibliográfico	11
	Bibliotecas de Instituições de Ensino Superior	9

Fonte: dados da pesquisa (2020).

No geral, os assuntos mais abordados são "Disseminação e Difusão da Informação", "Repositórios (baseado ou não em OAI)", "Métodos Bibliométricos" e "Bibliotecas de Instituições de Ensino Superior". Chang, Huang e Lin (2015) analisaram a evolução dos temas de pesquisa na área de Biblioteconomia e Ciência da Informação, entre 1995 e 2014, com base em análises de palavras-chave, acoplamento bibliográfico e co-citação. Dividindo o período da pesquisa em quatro fases (1995-1999, 2000-2004, 2005-2009, 2010-2014), os pesquisadores identificaram que os Estudos Bibliométricos consistiam no tema mais abordado nas quatro fases e, apenas entre 2000-2004, ficou em segundo lugar, atrás de Aplicação da Tecnologia da Internet, mas ainda foi um assunto largamente estudado.

A "Disseminação e Difusão da Informação" é um dos assuntos mais recorrentes e aparece como um dos cinco assuntos mais trabalhados na África, América do Norte e Central, América do Sul, Ásia e Europa. Os "Métodos Bibliométricos" e os "Repositórios (baseado ou não em OAI)" são dois dos temas mais estudados na África, Ásia e Europa. "Disseminação e Difusão da Informação" e "Métodos Bibliométricos" estão dentro da classe geral "Uso da Informação e Sociologia da Informação", que aparece em primeiro lugar no assunto mais abordado de forma geral no repositório. Já os "Repositórios (baseado ou não em OAI)" pertencem à classe "Fontes de Informação, Suportes e Canais", que consta em quarto lugar.

O assunto específico "Bibliotecas de Instituições de Ensino Superior" aparece três vezes: na Oceania, Antártica e Europa. Inclusive, é o tema mais abordado na Europa. O assunto está dentro da classe geral "Bibliotecas como coleções físicas" do esquema JITA, que aborda funcionalidades orientadas ao usuário, de direção e gerenciamento, incluindo questões socioeconômicas e legais (VIANA, 2006).

Esses são os assuntos principais considerando tanto a ocorrência de cada tema dentre aqueles mais abordados por cada continente, quanto mensurando a quantidade de documentos

que tratam de cada um desses assuntos. Tendo-se em vista a quantidade de documentos, pode-se localizar mais documentos que tratam sobre "Disseminação e difusão da informação" (1.343 documentos), "Repositórios (baseado ou não em OAI)" (1.128) e "Métodos bibliométricos" (967), que são três dos cinco temas principais abordados nos documentos depositados na Europa.

A comparação de temas e tendências de pesquisa é um trabalho que deve ser realizado com cautela. A razão está na forma como cada pesquisa foi conduzida, bem como, nos instrumentos utilizados para categorizar as temáticas das bases de dados estudadas. Dora e Kumar (2017) e Abadal e Guallar (2020) identificaram trabalhos que classificavam os documentos com a classificação JITA, como é o caso do E-LIS, outros com o Tesouro de Biblioteconomía y Documentación criado pelo Conselho Superior de Investigações Científicas (CSIC), algumas investigações usaram a análise de co-palavras, análise de palavras do título e métodos de análise de co-citação, assim como métodos bibliométricos simples como análise de palavras-chave para representar os tópicos de pesquisa. Dessa forma, as comparações dependem dos métodos e/ou instrumentos utilizados por pesquisador.

Sabendo-se que Kawalec (2013) utilizou a classificação JITA para identificar as tendências da pesquisa em Biblioteconomia e Ciência da Informação na Espanha, com base na produção científica publicada entre 2000 e 2010 derivada dos Programas de Pós-graduação em Ciência da Informação de 12 universidades públicas da Espanha, uma comparação pode ser feita. Verificou-se na pesquisa da autora que os assuntos mais abordados na produção científica espanhola foram "Fontes de informação, suportes e canais", "Uso da Informação e Sociologia da Informação", "Indústria, profissão e formação" e "Tratamento da informação para serviços de informação". Como a Espanha é o país com mais documentos depositados, identificou-se uma aproximação entre o examinado pela autora e os subtópicos mais populares da Europa no E-LIS.

De forma mais específica, Kawalec (2013) também constatou os subtópicos que mais aparecem dentro das temáticas principais. Os subtópicos identificados foram: informação na sociedade; políticas de informação; métodos bibliométricos; periódicos eletrônicos; recursos eletrônicos; materiais de arquivo; e repositórios baseados ou não em OAI. Similar ao apresentado por Kawalec (2013) e os dados do ELIScript, Olmeda-Gomez, Ovalle-Perandones e Perianes-Rodríguez (2017) e Dora e Kumar (2019) identificaram os métodos bibliométricos como uma forte tendência de pesquisa na Espanha. Esse tópico aparece como um dos mais populares também na África, Ásia e Europa.

Outros dois assuntos que se destacam são: "Disseminação e difusão da informação" e "Repositórios (baseado ou não em OAI)". Os dois temas estão intimamente relacionados. A disseminação da informação, nas palavras de Cunha e Cavalcanti (2008, p. 130) "é a difusão de informações ou documentos, a partir de um ponto central de armazenamento". Esse ponto de armazenamento pode ser uma unidade ou um sistema de informação, como uma biblioteca ou um arquivo, como também pode ser um repositório digital. Ao estudar artigos indexados na base

de dados LISA entre 1990 e 2010, a partir de uma análise de co-ocorrência de palavras para identificar tendências na pesquisa de Biblioteconomia e Ciência da Informação, Mittal (2011) encontrou pesquisas sobre tópicos como bibliometria e estudos de usuários, direitos autorais, bibliotecas digitais, repositórios institucionais, recursos eletrônicos, acesso aberto e Web 2.0. Identificando-se, assim, correspondência com alguns dos tópicos mais populares no E-LIS.

Referente à “Indústria, profissão e formação”, estão os subtópicos relacionados à educação, formação e aspectos curriculares (KAWALEC, 2013). Abadal e Guallar (2020) apontam que pesquisas mais diretamente relacionadas à profissão de bibliotecário e à documentação parecem ter sido menos significativas nos últimos anos, embora continuem aparecendo em estudos. Nesse sentido, Onyancha (2018) entende que a Ciência da Informação costuma receber mais atenção em pesquisa do que a Biblioteconomia, há mais ênfase na compreensão das propriedades da informação e em como gerenciá-la (ou seja, Ciência da Informação) em oposição a organizar e fornecer acesso a coleções de materiais (isto é, Biblioteconomia). Entretanto, nos documentos do E-LIS, o tópico “Formação” aparece entre os cinco subtópicos em destaque da América do Norte e Central, América do Sul e Europa, conforme dados do ELIScript.

No caso da Europa, o destaque sobre formação profissional pode ter sido provocado por mudanças no ensino superior induzidas pelo Processo de Bolonha e pelo Espaço Europeu do Ensino Superior (KAWALEC, 2013). O Processo de Bolonha busca conferir maior coerência aos sistemas de Ensino Superior em toda a Europa, estabelecendo o Espaço Europeu do Ensino Superior para facilitar a mobilidade dos estudantes e do corpo técnico, fazendo com que o ensino superior seja mais inclusivo e acessível, tornando-se mais atrativo e competitivo a nível mundial. O Programa Esquema de Ação da Região Europeia para a Mobilidade de Estudantes Universitários (ERASMUS), com o qual o PPGGOC/UFMG tem vínculo, é uma das iniciativas que desencadeou a cooperação mais intensa entre as instituições europeias de ensino superior. Por causa dessas transformações no ensino superior europeu (COMISSÃO..., 2021), entende-se que os currículos e a formação de bibliotecários e outros profissionais da informação precisaram ser discutidos.

Os tópicos sobre Educação e Formação também têm sido discutidos de forma abrangente na América do Norte e Central, e na América do Sul. Abdullahi, Karisiddappa e Asundi (2014) apontam que a educação na Biblioteconomia e Ciência da Informação aparece como um assunto que tem ganhado destaque na América Latina. Isso ocorre porque a inovação no currículo é um elemento importante na formação de profissionais da informação, que lidam constantemente com as transformações nos processos de mediação da informação e do conhecimento.

As escolas de Biblioteconomia e Ciência da Informação na América Latina possuem uma grande tradição no que se refere à transmissão de conhecimento e desenvolvimento de competências para realizar essa atividade, no entanto, deve-se ter em conta não apenas a

tradição, mas a necessidade corrente de atualização, adaptabilidade e articulação para internacionalização da produção científica, dos programas de graduação e pós-graduação, bem como dos corpos docente e discente dos países latino-americanos (ARELLANO; PEÑA, 2019). É essa busca por atualização que pode culminar em mais trabalhos científicos sobre a formação dos profissionais da informação.

Como apontado anteriormente, a comparação entre temas e tendências na Biblioteconomia e Ciência da Informação, com base na literatura, deve ser feita com cautela. Foi possível localizar algumas correspondências entre as temáticas dos documentos depositados no E-LIS e aquelas de universidades públicas espanholas, com documentos indexados pela WoS e pela base de dados LISA. O acervo do E-LIS será capaz de refletir a produção científica da área enquanto houver esforços para alimentar o repositório. Como uma iniciativa de voluntários, mantida por voluntários, cabe aos pesquisadores e profissionais da informação compreenderem a importância de se ter um acervo representativo e atualizado da área de forma aberta e gratuita na Internet.

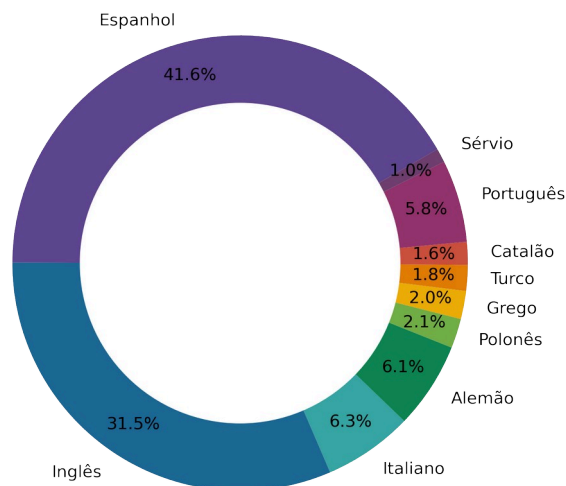
7.5 Distribuição por idioma

De Robbio, Subirats e Peset (2020) apontam que, dentre os mais de 22.000 documentos em acesso aberto disponíveis no E-LIS, podem ser encontrados documentos em 27 idiomas diferentes. De acordo com uma consulta no E-LIS, os autores podem escolher entre 38 idiomas para caracterizar sua produção.

Com o ELIScript foram identificados documentos em 37 idiomas, a saber, em ordem alfabética: alemão, árabe, basco, bósnio, búlgaro, canarês, catalão, chinês, croata, dinamarquês, esloveno, espanhol, francês, galego, grego, holandês, húngaro, indonésio, inglês, italiano, lituano, macedônio, malaiala, malaio, marata, nepalês, norueguês, persa, polonês, português, romeno, russo, sérvio, servo-croata, tcheco, turco e ucraniano.

No Gráfico 6, são apresentadas as porcentagens dos dez principais idiomas encontrados nas publicações disponíveis no E-LIS.

Gráfico 6 – Idiomas das publicações



Fonte: dados da pesquisa (2020).

O espanhol ocupa o primeiro lugar com 41,6% das publicações. Stockemer e Wigginton (2019) identificaram o inglês como a 'língua comum da ciência' ou a língua mais usada na publicação científica por pesquisadores que não têm a língua inglesa como língua materna. Entre as variáveis apontadas pelos pesquisadores para usar o idioma estão: idade do pesquisador; reputação percebida do inglês como o idioma da ciência; pressão institucional percebida para publicar em inglês; local de trabalho; área de estudo; e a região geográfica em que o pesquisador está baseado. Nesse sentido, identificar mais trabalhos em espanhol foi inesperado.

Morrison *et al.* (2007) caracterizam o inglês como o idioma principal do repositório. A interface do E-LIS está em inglês, além disso, cada documento submetido em outro idioma que não seja o inglês deve ser acompanhado de um resumo em inglês. Todavia, observou-se que a Espanha e os países falantes da língua espanhola da América do Sul e da América do Norte e Central são os responsáveis pela maior quantidade de documentos do repositório.

A Biblioteconomia e Ciência da Informação, na condição de ciências sociais, investigam fenômenos locais e regionais na sociedade relativos à informação, embora não se limitem a eles. Albarillo (2014) afirma que a linguagem é parte integrante desses fenômenos sociais, e sendo o Uso da Informação uma das principais temáticas abordadas nos documentos depositados no E-LIS, tema com característica evidentemente social, é compreensível que grande parte dos documentos que compõem o acervo do E-LIS estejam no idioma dos países com mais publicações depositadas, como é o caso da Espanha, países da América do Sul e América do Norte e Central.

Rey-Rocha *et al.* (1999) apontam que o fato de artigos estarem em espanhol, ou em qualquer outro idioma que não seja o inglês, não diminui a qualidade da publicação, embora a comunidade científica e seus métodos de avaliação subestimem essa qualidade porque os documentos estão no “idioma errado”, isto é, não estão no idioma definido como padrão. Os pesquisadores não-anglófonos devem confiar em avaliações científicas desenvolvidas localmente (como por exemplo o Qualis Periódicos⁴¹ no Brasil) que representam os interesses acadêmicos e idiomas regionais, em vez de serem avaliados por meio de métricas baseadas apenas em indicadores internacionais, que tendem a favorecer publicações em inglês (ALBARILLO, 2014).

Tendo em vista toda a discussão sobre o “idioma oficial da ciência”, o inglês aparece em segundo lugar, com 31,5%. Os dois idiomas compreendem, juntos, 73% dos documentos depositados no repositório. Conforme estudo de Morisson *et al.* (2007), esses dois idiomas já eram destaque no repositório desde 2007. Seguidos do italiano com 6,3%, alemão, 6,1% e português com 5,8%. Entre os países que falam português, localizou-se apenas documentos do Brasil e de Portugal, sendo a maior parte proveniente do Brasil. Evidencia-se que o repositório é conhecido pelos pesquisadores e profissionais da informação brasileiros, mas o autoarquivamento ou o depósito de documentos ainda é uma prática que precisa ser incentivada, principalmente considerando que os periódicos de Ciência da Informação do país estão em acesso aberto e permitem o autoarquivamento em repositórios digitais. Outros idiomas localizados foram polonês (2,1%), grego (2%), turco (1,8%), catalão (1,6%), outro idioma falado na Espanha, e sérvio (1%). Identifica-se, assim, que o E-LIS funciona como uma plataforma que promove a visibilidade de publicações em diferentes idiomas.

Um gráfico simples para averiguar os idiomas mais presentes nas publicações de cada continente foi gerado (Apêndice E). Na América do Norte e Central, Europa e América do Sul, os três continentes com mais depósitos, o idioma mais utilizado é o espanhol. Já na África, Ásia e Oceania, é o inglês. Na Antártica, por fim, encontra-se um documento em alemão e outro em espanhol.

7.6 Periódicos com publicações depositadas

Alguns gráficos foram gerados para apresentação dos resultados referentes aos periódicos científicos com mais publicações depositadas (Apêndice F), entretanto, optou-se pela apresentação na forma de tabela, uma vez que os títulos dos periódicos são muito extensos e não ficam bem representados em gráficos.

⁴¹ Qualis Periódicos é um sistema usado para classificar a produção científica dos programas de pós-graduação do Brasil no que se refere aos artigos publicados em periódicos científicos. O Qualis afere a qualidade dos artigos e de outros tipos de produção, a partir da análise de qualidade dos periódicos científicos (CAPES, 2021).

No Quadro 11, são apresentados os periódicos com mais publicações depositadas, divididos por continente.

Quadro 11 – Distribuição de Periódicos com mais publicações depositadas

CONTINENTE	PERIÓDICO	QUANT.
África	Continental Journal of Information Technology	6
	Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics	5
	Journal of Balkan Libraries Union	4
	- Academic Library Development - South African Journal of Library and Information Science - Revista Iberoamericana de Ciencia, Tecnología y Sociedad - IAALD Quarterly Bulletin - RESI : Revue de l'Ecole des Sciences de l'Information	2
América do Norte e Central	ACIMED	472
	Bibliotecas. Anales de Investigación	105
	e-Ciencias de la Información	58
	Anales de Documentación	44
	Liber : Revista de Bibliotecología	30
América do Sul	Revista Interamericana de Bibliotecología	278
	Información, Cultura y Sociedad	234
	RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação	112
	Códices	109
	Serie Bibliotecología y Gestión de Información	106
Ásia	Malaysian Journal of Library & Information Science	44
	Annals of Library and Information Studies	35
	Webology	32
	Pakistan Journal of Information Management and Libraries	27
	International Journal of Pure and Applied Mathematics	20
Europa	El profesional de la información	610
	Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare	463
	Comunicar	300
	Anales de Documentación	291

	AIDAInformazioni : rivista di Scienze dell'informazione	264
	Journal of Creative Industries and Cultural Studies – JOCIS	4
	D-Lib Magazine	4
Oceania	Gateways	4
	Anales de Documentación	3
	Library Management	2

Fonte: dados da pesquisa (2020).

• Periódicos com mais publicações na África

Dentre os periódicos com mais publicações no continente africano, identificou-se o Continental Journal of Information Technology, ISSN 2141-4033, periódico publicado na Nigéria bianualmente, com seis documentos. Além dele, tem-se o Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics, ISSN 1137-5019, publicado na Espanha, com cinco documentos. Os sites dos dois periódicos estavam indisponíveis no momento da consulta, em novembro de 2020. Numa breve busca, localizou-se alguns trabalhos publicados na primeira revista em 2017 (v. 10, n. 1), no entanto, esses artigos estavam disponíveis na base de dados Zenodo e não em uma página específica do periódico, demonstrando a preocupação dos autores em autoarquivar o trabalho em um repositório digital e promover o acesso às suas publicações.

O terceiro periódico da lista é o Journal of Balkan Libraries Union, ISSN 2148-9823. Esse é um periódico revisado por pares publicado anualmente. Conforme a política de submissão de manuscritos, recebe artigos em inglês e turco. O periódico está em acesso aberto e permite a prática de autoarquivamento⁴². Esse periódico é indexado nas bases LISA, Index Copernicus, Ulrich's, DOAJ, Google Scholar, E-LIS, CiteSeerX, Semantic Scholar e J-Gate Portal.

Os periódicos restantes da África, tiveram dois documentos depositados cada: South African Journal of Library and Information Science⁴³, ISSN 2304-8263; Revista Iberoamericana de Ciencia, Tecnología y Sociedad⁴⁴, ISSN 1850-0013; The International Association of Agricultural Information Specialists (IAALD) Quarterly Bulletin; e RESI: Revue de l'Ecole des Sciences de l'Information⁴⁵, ISSN 1113-4844.

⁴² <https://dergipark.org.tr/tr/pub/jblu>

⁴³ <https://sajlis.journals.ac.za/pub/index>

⁴⁴ <http://www.revistacts.net/acerca-de-cts/#institucional>

⁴⁵ <http://www.esi.ac.ma/?rub=65>

Ao consultar as políticas de submissão de manuscrito dos quatro periódicos, verificou-se que todos publicam artigos em acesso aberto e permitem o autoarquivamento em repositórios digitais. O IAALD Quaterly Bulletin foi publicado entre 1956-2007, os outros três continuam publicando artigos, com edições disponibilizadas em 2020. Observou-se que, apesar de poucos trabalhos disponibilizados na África, as revistas consultadas permitem e incentivam o autoarquivamento. No caso dos periódicos desativados, notou-se a importância de que trabalhos publicados, mesmo que em acesso aberto, sejam depositados em repositórios temáticos ou institucionais, a fim de preservar a produção e memória da Biblioteconomia e Ciência da Informação.

Por fim, foi identificado que o Academic Library Development trata-se, na verdade, de um livro, e os dois trabalhos publicados nele são capítulos. Ambos os capítulos foram digitalizados e disponibilizados no repositório. Como não foi encontrado o ISSN do periódico, e a pesquisa pelo título da revista no Portal de Periódicos CAPES, no Sherpa/RoMEO e no DOAJ não retornaram resultado, fez-se necessário consultar os registros diretamente e também acessar o E-LIS para verificação do procedimento de preenchimento dos metadados para capítulos de livros.

Após exame, foi identificado que não houve problema por parte dos autores no preenchimento dos metadados dos documentos, mas falta de coesão na nomenclatura dos campos dos metadados do repositório. O mesmo campo "*publication*" corresponde ao nome do periódico, no caso de artigos, e nome do livro, no caso de capítulos de livros. Em uma análise que considera esse campo, os dois tipos podem se misturar e interferir na análise. Alguns cuidados poderiam ser tomados, nesse caso, como filtrar nas planilhas apenas os documentos dentro da tipologia de artigos de periódicos, o que retornaria apenas os periódicos no campo "*publication*". Contudo, para isso ser implementado, seria necessário um conhecimento prévio de que um mesmo campo pode corresponder a dois valores diferentes, o que não foi o caso.

- **Periódicos com mais publicações na América do Norte e Central**

Os artigos depositados por pesquisadores da América do Norte e Central, no E-LIS, foram publicados principalmente nas revistas ACIMED, ISSN 1561-2880; Bibliotecas. Anales de Investigación, ISSN 1683-8947; e-Ciencias de la Información, ISSN 1659-4142; Anales de Documentación, ISSN 1575-2437; Liber: Revista de Bibliotecología, o ISSN dessa revista não foi localizado.

O ACIMED foi fundado em 1993 e, em 2012, teve o nome mudado para Revista Cubana de Información em Ciencias de la Salud⁴⁶, ISSN 2307-2113. É publicada quadrimestralmente e indexada pela DOAJ, Latindex, Scopus, SciELO Citation Index (WoS) e Redalyc. O escopo da

46 <http://www.acimed.sld.cu/index.php/acimed>

revista é voltado para as relações entre a Biblioteconomia e a Ciência da Informação com a área da Saúde. Tem o índice H (h-index) 9, que corresponde à quantificação da produtividade e o impacto de cientistas baseando-se nos seus artigos mais citados; e pontuação de 0.17 no SCImago Journal Rank (SJR), que mede a influência de um determinado periódico pela média do número de citações recebidas nos últimos três anos que antecedem o ano analisado e que são ponderadas dependendo da área do conhecimento e do prestígio do periódico (SCIMAGO, 2020). O periódico *Bibliotecas. Anales de Investigación*⁴⁷, por sua vez, é a revista mais antiga da América Latina a tratar sobre Bibliologia (ciência da história e composição dos livros) e Informação, publicada desde 1963. Tem periodicidade quadrimestral e aceita manuscritos em espanhol. As duas revistas são editadas em Cuba.

A *e-Ciencias de la Información*⁴⁸, publicada pela Universidad de Costa Rica, aceita manuscritos em espanhol, inglês, francês e português. A revista está em acesso aberto e permite o autoarquivamento de publicações. De acordo com consulta feita no Sherpa/RoMEO (2020), pode-se fazer o depósito da versão publicada no periódico e, também, da versão aceita, em repositórios digitais e em sites não-comerciais, mas não é permitida a distribuição da versão submetida do trabalho, que seria aquela sem as contribuições dos pareceristas do artigo. A única condição para o depósito é que se deve indicar que o artigo foi publicado na *e-Ciencias de la Información*.

O periódico *Anales de Documentación*⁴⁹, publicado pela Universidad de Murcia, na Espanha, tem periodicidade semestral e aceita artigos em espanhol, português e inglês. Esse caso mostra pesquisadores de um continente publicando em periódicos de outro continente, ou seja, pesquisadores da América do Norte e Central publicando em um periódico da Espanha, na Europa. Ele está indexado nas bases de dados internacionais WoS, SCOPUS, LISA, LISTA, Latindex-Catalog, Latindex-Directory, DOAJ, ERIH PLUS, na base de dados espanhola ISOC, nos repositórios internacionais E-LIS e Redalyc, e nos repositórios espanhóis da Recyt; Temaria; e-revist@s. A revista teve, ainda, reconhecimento pelo selo Fecyt 2014 na avaliação de critérios objetivos de qualidade em revistas científicas, sendo renovado em 2019. Também em 2019, obteve SJR de 0.13. Conforme o Sherpa/RoMEO (2020), pode-se realizar o depósito dos artigos em todas as versões (submetida, aceita e publicada), desde que as versões submetidas e aceitas direcionem para a versão publicada. A *Anales de Documentación* aparece como um dos principais periódicos com documentos depositados no E-LIS não apenas na América do Norte e Central, como também na Europa e na Oceania. Foi o único periódico a se repetir no Quadro 11.

47 <https://dialnet.unirioja.es/servlet/revista?codigo=24661>

48 <https://revistas.ucr.ac.cr/index.php/eciencias>

49 <https://revistas.um.es/analesdoc>

Por fim, a revista *Liber: Revista de Bibliotecología*⁵⁰ foi publicada entre 1999 e 2003 e teve suas atividades encerradas. Era uma revista técnico-profissional, gerida pela Asociación Mexicana de Bibliotecarios A. C. (AMBAC). Ao todo, 30 trabalhos publicados na revista foram depositados no E-LIS. Apesar da quantidade tímida, reforça-se a ideia de que o repositório é um espaço tanto para pesquisadores quanto profissionais da informação.

- **Periódicos com mais publicações na América do Sul**

Na América do Sul, as revistas com mais artigos depositados no E-LIS foram: *Revista Interamericana de Bibliotecología*, ISSN 2538-9866; *Información, Cultura y Sociedad*, ISSN 1851-1740; *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, ISSN 1678-765X; *Códices*, ISSN 1794-9815; e *Serie Bibliotecología y Gestión de Información*, ISSN 0719-0832.

A *Revista Interamericana de Bibliotecología*⁵¹ é a primeira da lista. É publicada na Colômbia, na Universidad de Antioquia, tem frequência quadrimestral e não realiza nenhum tipo de cobrança para publicação de artigos. É uma revista de acesso aberto dirigida a bibliotecários, arquivistas e especialistas em Ciência da Informação com foco em discussões no âmbito latino-americano. A revista é indexada pelas bases Scopus, LISTA, ISTA, LISA, Library Literature and Information Science, vLex, DOAJ, DIALNET. Seu índice H é 8, e obteve um score SJR de 0.16 em 2019.

Em seguida aparece o periódico argentino *Información, Cultura y Sociedad*⁵². Tem frequência semestral e aceita manuscritos em espanhol e português. Está indexada pela Redalyc, SciELO, DOAJ, Latindex, REDIB, Dialnet, EBSCOhost, GALE, EuroPub, LatinRev e ERIH PLUS. Baseando-se em suas políticas editoriais, está de acordo com as diretrizes de Acesso Aberto e incentiva os autores a depositarem em repositórios digitais, além dos artigos, os dados que respaldam os resultados da investigação.

O primeiro periódico brasileiro aparece em terceiro lugar. A *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*⁵³ está vinculada ao Sistema de Bibliotecas da UNICAMP (SBU) e, desde 2019, passou a adotar o sistema de publicação contínua, isto é, não há periodicidade determinada, os artigos são publicados conforme são aprovados. O periódico é indexado pela Academic OneFile da Gale, BRAPCI, EBSCO, Edubase, Infobila, Informe

50 <https://www.latindex.org/latindex/ficha?folio=14439>

51 <https://revistas.udea.edu.co/index.php/RIB/index>

52 <http://revistascientificas.filo.uba.ar/index.php/ICS/index>

53 <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci>

Acadêmico, Google Acadêmico e Scopus da Elsevier. Além disso, pode ser encontrada em diretórios como o Diadorim, DOAJ, Dialnet, Latindex, Sherpa/RoMEO e UlrichsWeb.

Em 2018, o Brasil foi reconhecido pelo Relatório da Science-Metrix⁵⁴, *Analytical Support for Bibliometrics Indicators: open access availability of scientific publications*, como o país com a maior quantidade de publicações científicas abertas no mundo: cerca de 74% da produção nacional está disponível gratuitamente na Internet. Nesse relatório identificou-se, ainda, que 42% das publicações estavam em acesso aberto verde, 41% das publicações em acesso aberto dourado, e 11% indeterminado (SCIENCE-METRIX, 2018). Devido a essa quantidade tão expressiva de documentos brasileiros em acesso aberto, buscou-se localizar outras publicações brasileiras no E-LIS. O segundo periódico brasileiro aparece em 7º lugar. O Ciência da Informação⁵⁵, ISSN 1518-8353, que possui 85 documentos depositados. Os próximos periódicos a aparecerem são o Perspectivas em Ciência da Informação, ISSN 1981-5344, em 13º; o DataGramZero ISSN 1517-3801, que foi descontinuado, está em 23º; e o TransInformação ISSN 2318-0889 em 24º. Tendo-se em vista a representatividade do Brasil no cenário mundial de publicações em acesso aberto e sabendo-se que todos os periódicos brasileiros de Biblioteconomia e Ciência da Informação estão em acesso aberto dourado, entende-se que a quantidade de documentos depositados em acesso aberto verde poderia ser maior. No entanto, vê-se que existe uma preocupação por parte de pesquisadores brasileiros em realizar o autoarquivamento, sendo o E-LIS um dos destinos para suas publicações.

Logo após aparece a revista Serie Bibliotecología y Gestión de Información⁵⁶, do Chile, que publica textos em castelhano, português e inglês, sendo disponibilizados no idioma em que foram submetidos. Tem como foco as produções sobre o livro e a leitura. É Indexado pelas bases Academic Search Complete, Dialnet, DOAJ, E-LIS, Latindex, Information Matrix for the Analysis of Journals (MIAR), Directory of Open Access Scholarly Resources (ROAD), Portal de Revistas Chilenas e Ulrich's.

Observa-se que o Anales de Documentación e o Bibliotecología y Gestión de Información lista o E-LIS como uma das bases que os indexam, no entanto, no próprio repositório não foram localizadas informações sobre a indexação de periódicos, apenas a indicação de que suporta o protocolo OAI com base na URL <http://eprints.rclis.org/cgi/oai2> (E-LIS, 2020). Essa informação, embora não confirmada, é interessante porque indica a possibilidade de artigos de periódicos serem indexados automaticamente, uma vez que sejam publicados nas revistas, a partir do protocolo OAI-PMH. Isso aumentaria consideravelmente a quantidade de documentos

54 Empresa norte-americana dedicada a avaliar atividades ligadas à ciência e tecnologia.

55 <https://revista.ibict.br/ciinf>

56 <https://seriebibliotecologia.utem.cl/>

arquivados no repositório, resultando em uma rica coleção de documentos de Biblioteconomia e Ciência da Informação.

Por fim, foram identificados nomes de periódicos como “Códices” e “Biblios”, que são muito comuns para periódicos da área. Nesse caso, entende-se que os periódicos precisam ser referenciados junto aos seus ISSN. Na busca por essas revistas, foram localizados outros periódicos com os mesmos nomes tanto na Biblioteconomia e Ciência da Informação, como em outras áreas. Apesar de ter um campo disponível para inserção de dados do ISSN no momento de submissão do documento no E-LIS, nem sempre este é preenchido, como foi o caso desses dois periódicos. Dos 94 localizados, disponíveis na Biblios, apenas um teve o campo ISSN preenchido.

- **Periódicos com mais publicações na Ásia**

Na Ásia, os periódicos com mais publicações depositadas são: Malaysian Journal of Library & Information Science, ISSN 1394-6234; Annals of Library and Information Studies, ISSN 0975-2404; Webology, ISSN 1735-188X; Pakistan Journal of Information Management and Libraries, ISSN 2409-7462; International Journal of Pure and Applied Mathematics, ISSN 1314-3395.

O primeiro periódico na lista é o Malaysian Journal of Library & Information Science⁵⁷. Essa publicação é feita quadrimestralmente pela University of Malaya. Entre 1996 e 2008 esse periódico era publicado tanto na forma impressa, quanto eletrônica. Em 2009, optou por manter apenas a versão eletrônica. Tem por objetivo servir como espaço para comunicação entre profissionais de Biblioteconomia e Ciência da Informação na região da Ásia-Pacífico. O periódico é indexado pelas bases Scopus, Social Sciences Citation Index, Library, LISTA, LISA, Library Literature and Information Science. Seu fator de impacto JCR é 1,0, ranqueando em 57º entre os 89 periódicos analisados em Biblioteconomia e Ciência da Informação. Seu SJR é de 0.441, em 2018 o periódico foi ranqueado em 69º dos 227 periódicos de Biblioteconomia e Ciência da Informação.

Em segundo está o Annals of Library and Information Studies⁵⁸, periódico mais antigo da Índia sobre Biblioteconomia e Ciência da Informação. Publicado quadrimestralmente, ele é indexado nas bases: Emerging Sources Citation Index, Scopus, IBZ Online, INSPEC, LISTA, ISTA, LISA e DOAJ. Em 1954, o Indian National Scientific Documentation Centre (INSDOC) lançou a primeira edição do Annals of Library Science e seu primeiro editor foi Dr. Shiyali Ramamrita Ranganathan, o renomeado pesquisador da área de Biblioteconomia. Mais tarde, em

57 <https://mjlis.um.edu.my/>

58 <http://op.niscair.res.in/index.php/ALIS>

2001, o periódico teve seu nome alterado para *Annals of Library and Information Studies*, para acompanhar as mudanças no campo. O periódico não cobra taxas para submissão de artigos e disponibiliza suas publicações em acesso aberto, apesar disso, possui um sistema de assinatura anual para compra da versão impressa da revista.

O *Webology*⁵⁹ aparece em terceiro lugar. É um periódico bianual dedicado ao campo da *World Wide Web*, envolvendo discussões sobre praticamente tudo que envolve a web e Ciência da Informação, servindo como fórum de discussão e experimentação. Está indexado nas bases: Scopus, Aerospace Database, Index Islamicus, LISTA, Communication Abstracts, LISA e DOAJ. Em 2018, teve SJR de 0.18, figurando no terceiro quartil de revistas de Biblioteconomia e Ciência da Informação avaliadas pela SCImago.

Em seguida aparece o *Pakistan Journal of Information Management and Libraries*⁶⁰, indexado pela Scopus, LISA, LISTA, ProQuest, Ulrich's, DOAJ, E-LIS, Literature in the Humanities and Social Sciences (IBR), International Bibliography of Periodical Literature in the Humanities and Social Sciences (IBZ) e Cabell's Directory of Publishing Opportunities. Tem índice H 6 e SJR de 0.18. Mais uma vez, o E-LIS aparece como um dos indexadores de um periódico.

O quinto periódico, o *International Journal of Pure and Applied Mathematics*⁶¹, parou de receber submissões de novos artigos em 2018. Até 2016, tinha índice H 23 e recebeu, em 2019, SJR de 0.14, pois esse indicador tem relação com os 3 últimos anos a partir ano de análise, quando o periódico ainda estava em atividade. Como foi observado, na Ásia, um dos temas com mais publicações são os estudos bibliométricos, dessa forma, faz sentido que um periódico sobre matemática e ciência da computação esteja entre os principais do continente. Além dele, o *Webology* também coloca ênfase nos estudos métricos da informação.

Nota-se nas orientações para autores das revistas asiáticas a predominância da publicação em inglês. Não foi identificado em nenhuma das revistas com mais publicações depositadas, a opção de depositar trabalhos no idioma nativo, o que demonstra a intenção de internacionalização da ciência produzida no continente.

- **Periódicos com mais publicações na Europa**

No continente europeu, os periódicos com mais publicações depositadas são: *El profesional de la información*, ISSN 1699-2407; *Mitteilungen der Vereinigung Österreichischer*

59 <https://www.webology.org/>

60 <http://journals.pu.edu.pk/journals/index.php/pjiml/index>

61 <https://ijpam.eu/>

Bibliothekarinnen & Bibliothekare (VÖB-Mitteilungen), ISSN 1022-2588; Comunicar, ISSN 1988-3293; e o AIDAInformazioni: rivista di Scienze dell'informazione, ISSN 1594-2201.

Dentre as revistas com mais publicações depositadas no E-LIS, a *El profesional de la Información*⁶², atualmente apenas *Profesional de la Información* sem o artigo “EI”, é a segunda melhor posicionada no SJR, com 0.48. Além disso, possui fator de impacto JCR de 1.580. Está indexada nas grandes bases de dados da Biblioteconomia e Ciência da Informação, como LISTA, LISA, ISTA, Library & Information Science Source (EBSCO), na Library Science Database (ProQuest), como também na Scopus, na Social Sciences Citation Index (Social SCI) da WoS, na Journal Citation Reports (JCR), Academic Search Premier, FRANCIS, Fuente Academica Plus, Applied Science & Technology Source, Business Source Premier, Computer & Applied Sciences e INSPEC. Suas publicações são feitas em fluxo contínuo, à medida que os artigos estejam prontos para serem publicados.

O *Profesional de la Información* é o único periódico, dentre os listados no Quadro 11, que cobra assinatura e taxa para acesso a artigos. O periódico foi publicado em acesso aberto entre 1992 e 2016. Apesar da mudança no modelo de negócios, alguns artigos continuam a ser publicados abertamente. Ele também é o único que adiciona um período de embargo para disponibilização da versão publicada em repositórios digitais, 6 meses. Entretanto, a revista permite o depósito da versão aceita para publicação em qualquer repositório, site institucional ou pessoal do autor. A versão submetida sem o parecer dos revisores da revista, por sua vez, não pode ser disponibilizada. Apesar do período de embargo, esse ainda é o periódico com mais publicações depositadas no E-LIS, de forma geral. O que demonstra o interesse dos pesquisadores de tornar acessível trabalhos que foram publicados em revistas que colocam seus trabalhos atrás de *paywall*, cobrando assinatura. Reconhece-se, assim, a importância da prática de autoarquivamento para a democratização do acesso ao conhecimento.

A *VÖB-Mitteilungen*⁶³ é uma revista alemã indexada nas bases Scopus, Library, LISTA e DOAJ. Tem índice H 3 e SJR de 0.2. É um periódico de acesso aberto diamante que, de acordo com a SCImago (2020), segue os princípios do "Acesso Aberto Justo" (*Fair Open Access*), não cobrando APC, nem outras taxas pela publicação de artigos. Como visto na Introdução, é possível que no Acesso Aberto Dourado haja uma cobrança de APC para viabilizar a publicação de um artigo, o que não é comum no Brasil, por exemplo. O Acesso Aberto Diamante seria, nesse caso, uma forma mais “pura” de acesso aberto, sem nenhum tipo de cobrança. Não houve nenhuma outra menção ao acesso aberto diamante nos periódicos do Quadro 11, embora todos os outros periódicos em acesso aberto não cobrem taxas para publicação de artigos e se identifiquem como pertencentes ao modelo dourado.

62 <http://www.elprofesionaldelainformacion.com/index.html>

63 <https://www.univie.ac.at/voeb/publikationen/voeb-mitteilungen/>

A Comunicar: Revista Científica de Comunicación y Educación⁶⁴, é a melhor colocada aqui, com índice H 32, SJR 1.09 e fator de impacto de JCR 3.375. É um periódico indexado pelas bases Scopus, Social Sciences Citation Index, Academic Search Premier, Fuente Academica Plus, IBZ Online, Education Abstracts, Communication & Mass Media Index, EBSCO Education Source, Educational research abstracts (ERA), Education Resources Information Center (ERIC), Linguistics & Language Behavior Abstracts, Modern Language Association (MLA) Database, PAIS International, Psycodoc, Psycinfo, Sociological abstracts, DOAJ, e DIALNET. É uma revista de acesso aberto que permite o autoarquivamento dos trabalhos em repositórios digitais, com exceção da versão submetida para avaliação. As condições para depósito do trabalho é que o autor faça a vinculação do documento depositado com a versão publicada e seu DOI, bem como o link para a página do periódico.

Finalmente, a revista AIDA Informazioni: rivista di Scienze dell'informazione⁶⁵, publicada na Itália, foi classificada pela Agência Nacional Italiana para a Avaliação de Sistemas Universitários e de Pesquisa (ANVUR), como um periódico “Classe A”. Essa classificação é utilizada para identificar periódicos que se comparam a periódicos analisados a partir da bibliometria como parte do quartil superior (Q1), em áreas que a análise de citações não é apropriada, como as ciências sociais, direito, arquitetura, história, filosofia, antiguidades, filologia, estudos literários e pedagogia (ANVUR, 2021). Conforme apontado na análise dos idiomas do ELIScript, cada país também possui seus indicadores de qualidade para avaliar sua produção científica.

- **Periódicos com mais publicações na Oceania**

Na Oceania, os periódicos com mais publicações depositadas são: Journal of Creative Industries and Cultural Studies – JOCIS, ISSN 2184-0466; D-Lib Magazine, ISSN 1082-9873; Gateways, ISSN 1836-3393; Library Management, ISSN 0143-5124.

A Journal of Creative Industries and Cultural Studies – JOCIS⁶⁶ é uma revista científica comercial gerida pela Media XXI e indexada pela Dialnet. Está classificada na área de Humanidades e Ciências Sociais em geral e aborda os campos da arte e audiovisuais, cinema, rádio e televisão. Apenas o último número da revista está disponível para acesso sem custos, todos os outros só permitem o acesso mediante a pagamento ou assinatura do periódico. Além disso, o número do periódico que pode ser acessado tem informações de que o trabalho não pode ser reproduzido, armazenado em sistemas de recuperação e transmitido de nenhuma

64 <https://www.revistacomunicar.com/index.php>

65 <http://www.aidainformazioni.it/?lang=en>

66 <https://www.mediaxxi.com/journal-of-creative-industries-and-cultural-studies/>

maneira, o que vai contra as diretrizes do acesso aberto. Esse foi o único periódico localizado com essas condições.

A D-Lib Magazine⁶⁷ foi um periódico focado em bibliotecas digitais. A publicação de novos números foi suspensa em 2017. Atualmente, o site continua no ar apenas para manter os artigos publicados nos números anteriores. Consta no site, no entanto, um pedido de sugestões para arquivamento de longo prazo para essas publicações. Vê-se, mais uma vez, a importância do autoarquivamento e a necessidade dos autores pensarem em formas de preservar a sua produção científica, seja em um repositório digital como o E-LIS, comprometido com a preservação a longo prazo, ou um site pessoal, no qual o pesquisador fará a manutenção do seu conteúdo.

O periódico Gateways: International Journal of Community Research and Engagement⁶⁸ é indexado pelo DOAJ, OpenAIRE, Google Acadêmico, Ulrich'sWeb, Sherpa/RoMEO, Web of Science e COPE. A revista publica estudos de caso sobre iniciativas de engajamento de comunidades; análises de ambientes políticos; e reflexões teóricas que contribuem para o engajamento acadêmico com essas comunidades. Todas as publicações devem ser feitas em inglês, podendo ter variações britânicas, australianas ou americanas na ortografia. O periódico está em acesso aberto e incentiva a prática de autoarquivamento em sites institucionais ou pessoais e em repositórios digitais.

Por fim, o Library Management⁶⁹ é um periódico gerenciado pela Emerald Group Publishing, uma editora comercial. No caso de publicações feitas na Library Management, o autor publica em acesso aberto dourado mediante pagamento da APC, a taxa de processamento de artigos. Caso o autor precise publicar em acesso aberto, mas não receba financiamento para arcar com essa taxa, o periódico sugere que o autor disponibilize a versão aceita do seu trabalho, assim que for aceita, em um repositório digital, seguindo a via acesso aberto verde. A Library Management é indexada pelas bases The Publication Forum (Finlândia); BFI (Dinamarca); CORE Rankings; Scopus; Academic Search Alumni Edition; Academic Search Complete; Academic Search Premier; Current Abstracts; Education Full Text; Emerging Sources Citation Index; Information Management & Technology Abstracts; ISTA; INSPEC; LISA; LISTA; Library Literature and Information Science; Library Literature and Information Science Full Text; e The Informed Librarian.

A partir dos dados apresentados, foi possível verificar que praticamente todos os periódicos com mais documentos depositados estão em acesso aberto e permitem o depósito de seus artigos em repositórios digitais, como é o caso do E-LIS. Com exceção do periódico

67 <http://www.dlib.org/>

68 <https://epress.lib.uts.edu.au/journals/index.php/ijcre/index>

69 <https://www.emerald.com/insight/publication/issn/0143-5124>

Profesional de la Información, que cobra assinatura e taxas para acesso de artigos avulsos, mesmo que alguns trabalhos ainda se encontrem em acesso aberto no seu site; do Library Management que cobra taxas para a publicação de artigos, embora os disponibilize em acesso aberto; e o JOCIIS, que se caracteriza como um periódico comercial. O Profesional de la Información foi o único periódico a colocar um período de embargo antes que o autor possa realizar o autoarquivamento da versão publicada em um repositório.

Alguns dos periódicos apresentados da Biblioteconomia e Ciência da Informação parecem possuir enfoque em pesquisas regionais e/ou nacionais, publicando em seus idiomas nativos e, poucas vezes, adicionando a opção para publicação em inglês. A Ásia foi o único continente em que os periódicos só aceitavam a submissão de trabalhos em inglês. Para além disso, observou-se que muitos dos periódicos são conceituados na área de Biblioteconomia e Ciência da Informação e têm reconhecimento de bases como a Scopus, WoS, LISTA, LISA e ISTA, bem como possuem pontuação no ranking da SCImago e índices de impacto como o índice H e o JCR, conferindo confiabilidade aos documentos depositados no repositório.

Como Coleman (2007) aponta, realmente não foi observado grandes entraves para a realização de depósitos na Ciência da Informação, inclusive no caso da Profesional de la Información, que mesmo o autor não podendo depositar o artigo publicado imediatamente em um repositório digital, ele pode submeter a versão aceita para publicação. Couto e Ferreira (2019) afirmam que alguns pesquisadores costumam infringir as leis de direitos autorais ao disponibilizar seus artigos em repositórios digitais ou sites de redes sociais, muitas vezes de forma desproposita. A partir dos dados apresentados, isso não parece ter acontecido no E-LIS, principalmente pela grande quantidade de periódicos que permitem o autoarquivamento.

7.7 Eventos com publicações depositadas

Existem quatro tipos de documentos que podem ser publicados em eventos, considerando as tipologias que são depositadas no E-LIS, são elas: *anais* de evento, artigo de conferência, pôster de conferência e apresentações.

Inicialmente, foi encontrado um problema na análise dos eventos. No preenchimento dos metadados, os depositantes colocavam o número do evento antes do nome do evento, por exemplo a sétima edição do Encontro Brasileiro de Bibliometria e Cientometria, os autores inseriam o título como “VII Encontro Brasileiro de Bibliometria e Cientometria”. Se cada evento tem um número antes, em uma quantificação automática todos eles ficarão separados por edição. Uma possibilidade para a análise seria remover os dados que representam a edição do evento antes do título do evento, no entanto, esses números ou a quantidade de caracteres não são padronizados e alguns autores não colocaram o número da edição, o que inviabilizou essa abordagem. Tendo-se essas questões em mente, procedeu-se para a análise buscando verificar

qual evento e, possivelmente, qual edição desse evento obteve mais documentos depositados no E-LIS.

No Quadro 12, optou-se por apresentar os dados como foram preenchidos nos metadados no momento da submissão do documento.

Quadro 12 – Distribuição de eventos com mais publicações depositadas

CONTINENTE	EVENTO	QUANT.
África	XXth Standing Conference of Eastern, Central and Southern Africa Library and Information Associations (SCECSAL)	11
	Tswane University of Technology Web 2.0 Library Workshop	3
	International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting	3
	6 th E-ICOLC (International Coalition of Library Consortia in Europe)	2
América do Norte e Central	69 th Annual Meeting of the American Society for Information Science and Technology (ASIST)	48
	III Conferencia Internacional de "Acceso abierto, preservación digital y datos científicos"	37
	68 th Annual Meeting of the American Society for Information Science and Technology (ASIST)	35
	XXXIV Jornadas Mexicanas de Biblioteconomía	32
	British Columbia Library Association Conference 2006 : Sharing a Vision: The Power of Collaboration	18
América do Sul	2o. Congreso Internacional de Bibliotecología, Documentación y Archivística (CIBDA)	26
	II SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO da Universidade Federal de Londrina (UEL)	13
	Seminario Dilemas de la Biblioteca Actual: La creatividad frente a la crisis	26
	IV Jornada Temas Actuales en Bibliotecología	20
	III Jornada Temas Actuales en Bibliotecología	12
Ásia	International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting	31
	Trends and Strategic Issues for Libraries in Global Information Society	5
	BOSLA One day Seminar on Resource Management	5
	Asia-Pacific Conference on Library and Information Education and Practice	5
	IAALD World Congress	4

CONTINENTE	EVENTO	QUANT.
Europa	XIV Jornadas Nacionales de Información y Documentación en Ciencias de la SaludVI	46
	VI Encontro Ibérico EDICIC	36
	10ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών	35
	6ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών	35
	12ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών	34
Oceania	The 8 th ACM/IEEE-CS Joint Conference on Digital Libraries	4
	Library Perspectives	2
	68 th Annual Meeting of the American Society for Information Science and Technology (ASIST)	2
	ALIA National Library and Information Technicians Conference	1
	Workshop on the Open Archives Initiative (OAI) and Peer Review journals in Europe (OAI1)	1

Fonte: dados da pesquisa (2020).

De início, o que se pode observar é a falta de padronização dos metadados. Para a numeração da edição do evento, algumas vezes os dados são preenchidos com números ordinais (inclusive em inglês), outras vezes cardinais, noutras com números romanos. Tem eventos com todas as letras em maiúsculo, outros com o nome em minúsculo. Eventos com nome por extenso ou abreviado. Logo, entende-se que os eventos listados no Quadro 12 não representam de forma concreta os eventos com mais publicações.

Na América do Norte e Central, América do Sul e Europa, nota-se que, dentre os eventos com mais publicações, existem eventos iguais, mas em edições diferentes, exemplo: *69th Annual Meeting of the American Society for Information Science and Technology (ASIST)* e o *68th Annual Meeting of the ASIST*. A dúvida é se a quantificação deveria ser feita por evento ou por edição do evento. Esse tipo de problema poderia ser evitado se houvesse um campo para inserir apenas a edição do evento e se houvessem instruções claras para o preenchimento desses metadados. Assim, seria possível verificar tanto o total de documentos depositados por evento, como por edição individualmente.

No formulário de submissão de artigos de conferência e *anais* de conferência existem os campos *volume* e *number*, como se o artigo tivesse sido publicado em um periódico. As instruções informam “insira o número do volume da revista em que o seu depósito apareceu”⁷⁰ e

⁷⁰ “Enter the volume number of the journal in which your deposit appeared” (E-LIS, 2020).

“insira o número do volume da revista em que o seu depósito apareceu”⁷¹, respectivamente. No formulário para pôsteres de conferência e apresentações não há nenhuma opção referente à edição do evento, nem de volume e número, como nas anteriores. A instrução para inserir o título da conferência apenas informa “por favor, insira o nome completo da conferência”⁷² (E-LIS, 2020, *tradução nossa*). Na falta de um campo específico para a edição do evento e de instruções adequadas para o preenchimento dos metadados, a decisão de como preencher as informações sobre o evento fica a cargo de cada autor, o que resulta na falta de padronização.

Tendo-se em vista as ponderações feitas, observa-se que na **África**, o evento com mais depósitos foi a vigésima edição da Standing Conference of Eastern, Central and Southern Africa Library and Information Associations (SCECSAL). O evento foi realizado em Nairóbi no Quênia em junho de 2012, tendo como tema a “informação para o desenvolvimento sustentável em um ambiente digital” (IFLA, 2012).

Na **América do Norte e Central**, o evento com mais depósitos e que aparece duas vezes na lista, em suas edições 68 e 69, é a Annual Meeting of the American Society for Information Science and Technology (ASIST). O encontro anual da ASIST ocorreu em 2005 e 2006, em Charlotte na Carolina do Norte e em Austin no Texas, respectivamente. Em 2005, o tema foi “Provocando sinergia: reunindo pesquisa e prática”; já em 2006, o tema girava em torno das “realidades da informação: moldando o futuro digital para todos”, em que se questionava quais as implicações das tendências digitais no futuro da informação (ASIST, 2005, 2006).

O segundo Congresso Internacional de Bibliotecología, Documentación y Archivística (CIBDA) aparece como o evento com mais documentos depositados na **América do Sul**. A conferência foi realizada em La Paz na Bolívia, em 2003, e teve como tema a “Informação como fator de desenvolvimento: um desafio para a Sociedade da Informação” (RAMIREZ, 2003).

Na **Ásia**, o International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting foram realizados em maio de 2006 em Villers-lès-Nancy na França. O evento tinha como foco a visualização de dados em estudos informétricos, cientométricos e webométricos. Identificou-se que o evento ainda é realizado e o último ocorreu em novembro de 2019 em Dalian na China (COLLNET, 2019).

Na **Europa**, a XIV Jornadas Nacionales de Información y Documentación en Ciencias de la Salud, também conhecida como Bibliosalud, foi realizada em 2011 em Cádiz na Espanha. Em uma busca simples pelo evento, foi identificada uma postagem no Facebook que indicava o E-LIS como o repositório no qual seriam depositadas as comunicações orais e os pôsteres do evento. O link divulgado não estava mais disponível, embora os trabalhos ainda possam ser

71 “Enter the issue number of the journal in which your deposit appeared” (E-LIS, 2020).

72 “Please enter the full name of the conference” (E-LIS, 2020).

recuperados no repositório. O tema do evento foi “25 anos projetando o futuro”, era a comemoração de 25 anos da conferência, que estava sendo realizada desde 1986 (PEÑA; TINOCO; LORENZO, 2011). Um evento que se destacou foi o Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών (Conferência Grega de Bibliotecas Acadêmicas) que aparece em suas edições 6, 10 e 12, entre os cinco eventos com mais documentos depositados na Europa. Os eventos ocorreram em 1997, 2001 e 2003, respectivamente. Devido à limitações de idioma, não foi possível localizar o tema dos eventos.

Já na **Oceania**, o evento com mais documentos depositados foi a oitava ACM/IEEE-CS *Joint Conference on Digital Libraries*, que ocorreu em Pitsburgo na Pensilvânia, em 2008. O evento transcorreu nos Estados Unidos, no entanto, quantidade considerável de pesquisadores da Oceania participaram dele. Apesar de não ter um tema definido, o evento buscava discutir as relações entre a computação e as bibliotecas, e como as TIC impulsionaram a representação e o tratamento da informação. Cerca de 117 artigos completos e 61 resumos expandidos foram recebidos nessa edição (JCDL, 2008).

No geral, observa-se que os documentos dos eventos foram depositados entre 2003 e 2012. O evento com mais documentos depositados, a saber: o encontro anual da ASIST, coincide com o período das iniciativas de incentivo ao depósito no E-LIS para pesquisadores da Biblioteconomia e Ciência da Informação. Na lista, foram identificados eventos regionais, nacionais e internacionais. Como a Antártica só tem um capítulo de livro e um preprint, o continente não tem documentos publicados ou apresentados em eventos e por isso não está presente no quadro.

7.8 Palavras-chave

Por fim, a última categoria de análise apresentada é “palavras-chave”. No momento de submissão de um documento, solicita-se que os autores insiram os termos que melhor descrevem o trabalho. A instrução para inserção de dados esclarece: “digite aqui qualquer palavra-chave que você acha que ajudará os pesquisadores a encontrar o seu depósito. Tenha o cuidado de inserir palavras-chave específicas relevantes para o seu depósito e não seja muito genérico”⁷³ (E-LIS, 2020, *tradução nossa*).

A partir das palavras-chave inseridas pelos autores criou-se uma nuvem de palavras (Figura 20). Nuvem de palavras, ou *wordcloud*, é uma técnica utilizada para evidenciar as palavras mais frequentes dentro de um texto. Neste caso, as palavras-chaves dos 23.245 documentos foram reunidas em uma lista para gerar essa representação.

73 “Enter here any keywords you think will aid searchers in finding your deposit. Please be careful to enter specific keywords relevant to your deposit, and don't be too general” (E-LIS, 2020).

Figura 20 – Nuvem de palavras-chave



Fonte: dados da pesquisa (2020).

Apesar de existirem mais trabalhos em língua espanhola no repositório, ao analisar as palavras-chave, observa-se a incidência de mais termos em inglês. Esse pode ser um reflexo da exigência de preenchimento do resumo e palavras-chave em inglês em todas as submissões.

Optou-se por apresentar apenas uma nuvem de palavras geral, considerando todos os documentos depositados no E-LIS. Os temas centrais são “biblioteca”, termo no singular e no plural, e “informação”. Como um repositório de Biblioteconomia e Ciência da Informação, os termos são representativos e ressaltam os elementos principais sobre os quais os pesquisadores e profissionais da informação se dedicam.

A Biblioteconomia tem por objetivo a formação, informação e recreação através de todos os tipos de documento e se utiliza de instrumentos para a organização e administração dos diferentes tipos de bibliotecas (nacionais, públicas, infantis, escolares, universitárias e especializadas). Ela responde pelos problemas suscitados pelos acervos, desde sua formação, passando pelo tratamento de todos os documentos que compõem o acervo, até a sua disponibilização. Responde, ainda, pela própria biblioteca como serviço organizado, dando conta

da administração da biblioteca e atividades administrativas, bem como pelos leitores, os usuários (FONSECA, 2006; CUNHA; CAVALCANTI, 2008).

A Ciência da Informação, por sua vez, tem por objetivo estudar a gênese e comunicação da informação e tem forte relação interdisciplinar (FONSECA, 2006). Borko (1968), em seu trabalho seminal *Information Science: What is it?*, afirma que a Ciência da Informação investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo, e os meios de processamento para garantir acessibilidade e usabilidade ideais. Assim, preocupa-se com a origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. Nesse sentido, entende-se que a Ciência da Informação se utiliza das tecnologias para explorar as questões que concernem o conhecimento humano na sociedade, desde sua origem até a sua disseminação. As duas áreas compartilham teorias e práticas no que concerne à organização e recuperação de informação de documentos em diferentes formatos.

Alguns dos objetos ou focos de estudo da Biblioteconomia e Ciência da Informação podem ser observados nas outras palavras em destaque: “acesso aberto”, “dados”, “pesquisa”, “universidade”, “web”, “digital”, “bibliotecas digitais”, “serviços de usuário”. Muitos desses termos estão diretamente relacionados, como por exemplo: acesso aberto e dados; web, digital e bibliotecas digitais; as bibliotecas digitais e os serviços de usuário; ou as bibliotecas digitais, a pesquisa e a universidade.

O acesso aberto foi bastante discutido no presente trabalho. Nota-se que os pesquisadores da Ciência da Informação estão envolvidos nas investigações sobre as implicações do acesso aberto para a ciência e que esses trabalhos são disponibilizados de forma livre em repositórios de fácil acesso. Essa palavra-chave está relacionada com os repositórios digitais, que é um dos principais assuntos discutidos na África, na América do Norte e Central, na Ásia e na Europa. A “web”, por sua vez, perpassa a discussão de acesso aberto, de bibliotecas digitais, de serviços da informação, de uso da informação, de dados, de universidade e pesquisa, entre outros. Por causa da Internet, as TIC se desenvolveram e forneceram novas maneiras de criar, armazenar, tratar, representar, organizar e disponibilizar a informação.

Todos esses assuntos devem levar em consideração os usuários. Já em 2007, Santillan-Aldana (2009) identificou um número considerável de documentos relacionados aos estudos de usuários no E-LIS. Conforme identificado na análise dos assuntos, muitos documentos abordam o uso da informação em relação aos serviços realizados por bibliotecas e o acesso a fontes de informação em diferentes suportes. Não se pode pensar informação, seu tratamento e sua disseminação, sem pensar no usuário que fará uso desta.

As palavras-chave demonstram a relação com as temáticas mais discutidas nos trabalhos depositados no E-LIS, como também evidenciam assuntos indissociáveis da Biblioteconomia e da Ciência da Informação. Para Saracevic (1996), alguns dos enfoques da pesquisa e prática

profissional da Ciência da Informação são os registros do conhecimento, necessidades e usos da informação, a comunicação humana; temas perpassados pelas tecnologias da informação.

8 CONSIDERAÇÕES FINAIS

A comunicação científica de resultados de pesquisa pode se dar por vias formais e informais, no meio impresso ou digital. Essa comunicação é um processo fundamental para disseminação de pesquisas e desenvolvimento da sociedade, uma vez que permite gerar conhecimentos capazes de transformar a realidade ao seu redor. Nesse entendimento, a ciência e os resultados de pesquisa devem ser considerados bens públicos globais, isto é, um bem ao qual todas as pessoas deveriam ter acesso sem nenhuma restrição. O Movimento de Acesso Aberto busca viabilizar esse acesso a partir da disseminação de informações científicas de forma aberta, livre e gratuita.

Para atender ao objetivo do Movimento de Acesso Aberto existem duas estratégias, uma de publicação em periódicos de acesso aberto – a via dourada, e outra de autoarquivamento de publicações em repositórios digitais – a via verde. O autoarquivamento, foco deste trabalho, é uma prática realizada, geralmente, pelo próprio autor com vistas a preservar, disponibilizar para acesso e aumentar a visibilidade de sua produção científica. Percebendo-se a importância de um movimento como esse para a preservação da memória de uma área, surgem questionamentos acerca da prática de autoarquivamento na Biblioteconomia e Ciência da Informação.

Dentro dos padrões recomendados para arquivos abertos em meio digital, o E-LIS foi criado para facilitar o acesso a documentos da Biblioteconomia e Ciência da Informação. Esse repositório é reconhecido, atualmente, como a maior base de dados de documentos em acesso aberto com texto completo da área. Tendo isso em vista, o presente trabalho buscou caracterizar o acervo que compõe o repositório, de forma a avaliar as contribuições deste para a comunicação científica aberta em Ciência da Informação.

Para isso, foi desenvolvida um método automatizado para análise dos indicadores bibliométricos utilizando a linguagem de programação Python: o ELIScript. Esse instrumento é utilizado para a coleta e análise de dados dos registros disponíveis no E-LIS. A ideia principal foi produzir uma ferramenta que permitisse realizar a coleta das informações bibliográficas dos documentos de forma rápida, fácil e estruturada, permitindo, assim, a análise periódica do acervo do repositório.

O ELIScript foi criado para possibilitar as análises nessa dissertação e servir como uma ferramenta para análise periódica do acervo do E-LIS. Com ele, pôde-se coletar os dados pertinentes a todo o acervo do repositório, fazer a limpeza e organização desses dados, e analisá-los de acordo com as categorias de análise propostas. Embora o instrumento seja capaz de realizar a coleta de dados do acervo inteiro em poucas horas e as análises em alguns segundos, gerando todos os gráficos e tabelas apresentados aqui, o seu desenvolvimento envolveu um longo período de testes e de falhas até chegar ao produto final. Para além de sua

criação, o ELIScript foi aplicado e sua viabilidade confirmada, o que permitiu levantar dados significativos sobre o acervo, bem como acerca do cenário das práticas de autoarquivamento na área de Biblioteconomia e Ciência da Informação.

Com a recuperação dos registros utilizando o ELIScript, propôs-se analisar, por meio de indicadores bibliométricos, a tipologia dos documentos depositados, assuntos, data de publicação, data de depósito, idioma, periódicos científicos com mais publicações, eventos científicos e palavras-chave. E, por meio dessas categorias de análise, realizar uma análise geopolítica comparativa considerando todos os continentes. Em meio a essas análises, foi possível identificar, caracterizar e analisar as práticas de autoarquivamento dos pesquisadores da Ciência da Informação no E-LIS.

Por meio dos dados levantados de 23.245 publicações, foi identificado que os artigos de periódico (47%), artigos submetidos para eventos científicos (19%) e apresentações (8%) são os tipos de documento mais frequentemente depositados. Pesquisadores em todos os continentes realizaram o autoarquivamento, embora alguns mais do que outros. As práticas de autoarquivamento por pesquisadores da Biblioteconomia e Ciência da Informação compreendem o período entre 2002 e 2020, isto é, desde a criação do repositório até o momento da coleta de dados. Em todos os anos houve autoarquivamento.

Entretanto, a África, a Antártica e a Oceania registraram depósitos de forma esporádica. Na África, o autoarquivamento foi realizado entre 2004 e 2014. Na Oceania, entre 2005 e 2012. E, na Antártica, apenas em 2015. Por outro lado, na América do Norte e Central, América do Sul, Ásia e Europa, por sua vez, os depósitos continuaram a ser realizados até 2020. Os pesquisadores da América do Norte e Central, e da Ásia, começaram a realizar seus depósitos de documentos em 2003. Já na América do Sul, esse processo se iniciou em 2004. A Europa foi o único continente com depósitos em 2002, supõe-se que os primeiros documentos depositados foram um teste para averiguar a viabilidade do repositório, criado na Itália.

Para verificar a atualidade dos documentos depositados no E-LIS, procedeu-se para a análise da distribuição temporal das publicações. Os documentos depositados foram publicados entre 1965 e 2020. Observou-se que mais trabalhos podem ser encontrados entre 1983 e 2020. No período entre 1965 e 1983, há um número pouco expressivo de trabalhos publicados. A partir de 1983, no entanto, percebe-se um crescimento. A maior parte dos trabalhos foi publicada por volta de 2007, ano em que mais depósitos foram realizados, evidenciando uma relação entre o ano de depósito e o ano de publicação. Isso quer dizer que os autores costumam depositar seus trabalhos no mesmo ano em que eles são publicados.

Foi possível observar, também, os assuntos mais abordados nos documentos depositados. Os documentos do E-LIS são classificados de acordo com o esquema de classificação JITA, um instrumento desenvolvido especificamente para classificar documentos pertinentes à Biblioteconomia e Ciência da Informação. Esse esquema subdivide-se em dois

níveis hierárquicos. Entre os assuntos identificados no primeiro nível hierárquico estão o “Uso da Informação e Sociologia da Informação” (1.764 documentos), a “Tecnologia da Informação e Tecnologia de Biblioteca” (1.594 documentos), e “Tratamento da Informação para Serviços de Informação” (1.164 documentos). Percebe-se, pelos assuntos principais, que esses assuntos refletem alguns olhares da Biblioteconomia e Ciência da Informação, área que se ocupa da recuperação da informação e sua comunicação, fazendo uso de instrumentos que permitam o tratamento e representação da informação para posterior disseminação. Vale ressaltar que os assuntos não são exclusivos, um mesmo documento pode ter mais de um assunto atribuído no E-LIS.

Se for considerado o segundo nível hierárquico, ou seja, as temáticas específicas atribuídas aos documentos, os assuntos mais abordados são: “Disseminação e difusão da informação” (1.343 documentos) em primeiro lugar, seguido de “Repositórios (baseado ou não em OAI)” (1.128 documentos) e, por fim, “Métodos bibliométricos” (967 documentos). Observa-se, a partir desses dados, a preocupação basilar da área com formas de armazenar e preservar a informação para permitir sua disseminação para os usuários futuramente. Com o crescimento da informação digital, aumenta também o interesse pelos repositórios digitais, suas funcionalidades e aplicabilidade para manutenção da informação digital. Esses dados permitiram confirmar que os assuntos mais abordados nos documentos depositados no E-LIS são reflexo evidente da área, achando-se correspondências entre os assuntos do acervo com aqueles de documentos publicados em revistas científicas renomadas da Biblioteconomia e Ciência na Informação, que representam perspectivas e tendências da área.

O corpo editorial do E-LIS apresenta o repositório como uma ferramenta capaz de fornecer uma experiência multicultural e multilíngue. Os autores podem caracterizar sua produção escolhendo entre 38 idiomas ao submeter um documento ao E-LIS. Na pesquisa foram identificados documentos em 37 idiomas. Os dez idiomas principais foram elencados e observou-se a prevalência do espanhol como idioma mais presente nas publicações. A Europa é responsável pela maior quantidade de depósitos no repositório em língua espanhola, visto que o maior número de depósitos deriva da Espanha. Mesmo nos periódicos de maior fator de impacto identificados na pesquisa, geralmente publicados em inglês, o espanhol é utilizado como idioma principal de publicação. Em seguida, estão o inglês, o italiano, o alemão e o português.

A pesquisa possibilitou identificar, ainda, os periódicos com mais publicações depositadas no E-LIS. Os cinco periódicos com mais publicações são o *El Profesional de La Información* (610 artigos), o *ACIMED* (472 artigos), o *VÖB-Mitteilungen* (463 artigos), o *Comunicar* (300 artigos) e o *Anales de Documentación* (291). Desses, apenas o *ACIMED*, agora denominado *Revista Cubana de Información en Ciencias de la Salud*, não é editada em um país europeu. Uma das características observada é que grande parte dos periódicos com publicações depositadas no E-LIS estão indexados em bases de dados importantes da Biblioteconomia e Ciência da

Informação, como a LISTA, a LISA e a ISTA. Além delas, os periódicos estão indexados por bases generalistas como a Web of Science e a Scopus, indicando a relevância dos documentos depositados.

Com relação aos eventos com mais publicações depositadas, foram identificados problemas na análise devido, principalmente, à falta de padronização para inserção dos títulos e o número da edição dos eventos. No momento da submissão de um documento apresentado ou publicado em evento, não existe um campo específico para adicionar a edição do evento, levando os autores a inserirem a edição do evento junto ao título deste. Apesar disso, foi possível verificar a existência de eventos regionais, nacionais e internacionais. O evento com mais publicações depositadas foi o 69th Annual Meeting of the American Society for Information Science and Technology (ASIST), na América do Norte e Central.

Por fim, também foram analisadas as palavras-chave mais recorrentes no formato de nuvem de palavras. Nessa análise, foi possível identificar agrupamentos de assuntos que estão relacionados com as temáticas mais presentes nos documentos do repositório. Apesar de existirem mais trabalhos em língua espanhola no repositório, ao analisar as palavras-chave, observou-se a incidência de mais termos em inglês, como *library*, *open access*, *university*, *research* e outros. Os termos com maior representação foram “biblioteca”, “informação”, “acesso aberto”, “usuário”, “digital”, “serviço” e “dados” (em inglês). Por mais gerais que sejam esses termos, eles representam o propósito do repositório que é reunir documentos pertinentes para as áreas de Biblioteconomia e Ciência da Informação. Eles evidenciam o cerne da área, que é a preocupação com as formas de trabalhar com os fluxos da informação, independente do meio.

Concluiu-se, a partir desse estudo, que a criação de uma metodologia automatizada pode auxiliar no acompanhamento da produção científica depositada no E-LIS, sem que seja necessário recriar métodos para a coleta e análise de dados. Esses processos podem ser realizados de forma otimizada com o ELIScript quando houver necessidade de avaliar o acervo. Pode-se, assim, manter dados atualizados periodicamente sobre a quantidade de depósitos realizados, permitindo saber se houve aumento ou queda na quantidade de depósitos, se os documentos são atuais, bem como verificar os tipos de documentos que estão sendo depositados e as temáticas deles.

A partir do ELIScript foi possível identificar como o E-LIS contribui para a comunicação científica aberta em Biblioteconomia e Ciência da Informação. Essa contribuição se dá por meio da disponibilização de documentos produzidos por pesquisadores e profissionais da área de forma livre para a toda a sociedade. Seu acervo e, especificamente, seus artigos científicos passaram pela revisão por pares e foram publicados em periódicos indexados por bases de dados conceituadas da Biblioteconomia e Ciência da informação. Essas bases não podem ser acessadas sem assinatura, seja em nível pessoal ou institucional, mas no E-LIS é possível ter acesso ao texto completo desses artigos de forma gratuita.

Numa sociedade que passou a ser conhecida como Sociedade da Informação e do Conhecimento, falar em acesso à informação, disseminação do conhecimento e tecnologias digitais é reconhecer o papel da Biblioteconomia e Ciência da Informação como mediadoras do conhecimento humano. As pesquisas nessa área buscam dar conta do intenso fluxo de informações que circulam de forma impressa e digital, de forma a possibilitar o acesso democratizado, universal, global e total da informação por meio das unidades e sistemas de informação. Prezar por espaços como os repositórios digitais de acesso aberto é garantir que a população tenha acesso ao que é produzido de melhor pelas universidades públicas.

Dessa forma, o E-LIS serve não apenas como um repositório para preservar a memória da área, mas também como fonte de informação especializada e conceituada para trabalhos científicos e para o exercício dos profissionais da informação, que permitirá o desenvolvimento de pesquisas e a atuação profissional com vistas à promover o acesso ao conhecimento. Como uma base de dados da área, é interessante poder localizar documentos publicados em diferentes periódicos, indexados por diferentes bases de dados, apresentados nos mais diversos eventos científicos em um único lugar. A experiência é, de fato, multicultural. Por meio da análise dos dados de pesquisa, entende-se que o E-LIS não só pode contribuir para a área de Biblioteconomia e Ciência da Informação, como já contribui. O E-LIS alcançou reconhecimento como uma comunidade internacional aberta de Biblioteconomia e Ciência da Informação.

Na introdução foi perguntado “Por que o cidadão precisa se justificar para ter acesso ao conhecimento?”. E essa é uma pergunta que persiste, visto que o autoarquivamento não é tão realizado quanto deveria, considerando as possibilidades identificadas para sua concretização. Pergunta-se, ainda: há quem seja contra ter acesso gratuito ao conhecimento científico? Iniciativas como a do SciHub – site que fornece acesso a milhões de publicações de forma gratuita, mas ilegal –, indicam que não. O Movimento de Acesso Aberto e a Iniciativa de Arquivos Abertos fornecem as estratégias e ferramentas necessárias para viabilizar a promoção do conhecimento livre, então, deve-se questionar o que falta para concretização desse ideal. O autoarquivamento, por exemplo, depende dos pesquisadores e de sua conscientização acerca da importância dessa prática. O quanto nós estamos dispostos a nos esforçar para fazer parte dessa mudança? São questões que precisam ser levantadas. A inércia cultural e a falta de consciência sobre a importância dessa prática costumam ser citadas como a razão para a lenta adoção do autoarquivamento, mas isso precisa ser mudado, o hábito de fornecer acesso ao que é produzido, principalmente pelas universidades públicas, precisa ser internalizado e, também, difundido.

Na Ciência da Informação, foi observado que o incentivo é necessário para que a prática de autoarquivamento se torne comum. O incentivo não deve vir apenas dos editores do E-LIS, mas de professores da graduação de Biblioteconomia, Arquivologia e Museologia, dos Programas de Pós-Graduação em Ciência da Informação e áreas correlatas, de profissionais da

informação que devem estar cientes da importância de armazenar e preservar o conhecimento produzido e, mais do que isso, de viabilizar o acesso a esse acervo. Em 2007, ano em que foram depositados mais de 3,5 mil documentos, notou-se um grande trabalho de divulgação para que o repositório fosse utilizado, o que surtiu efeito. Atualmente, mais de 100 documentos são depositados a cada mês, vê-se que esse número poderia ser ainda maior se comparado aos dados de 2007. Os periódicos de Biblioteconomia e Ciência da Informação, que seguem as recomendações do Movimento de Acesso Aberto, já fornecem a possibilidade de publicação em acesso aberto e, além disso, incentivam a prática de autoarquivamento em repositórios digitais.

Apesar da identificação de trabalhos recentes sendo depositados no repositório, verificou-se uma queda na quantidade de depósitos nos últimos anos. Ter artigos e outros documentos publicados em periódicos de acesso aberto não garante que esses trabalhos serão preservados ao longo do tempo. A melhor alternativa ainda é o depósito das publicações e outros documentos em repositórios digitais. Considerando-se o exposto, entende-se que os objetivos propostos foram atingidos. A seguir, algumas limitações da pesquisa e sugestão de trabalhos futuros serão apresentadas.

8.1 Limitações da Pesquisa

Apesar dos objetivos alcançados, devido ao conhecimento limitado da autora sobre programação, o ELIScript não pôde ser desenvolvido com análises mais aprofundadas e relacionais, como, por exemplo: assuntos mais discutidos em determinado ano (relação entre assunto e ano de publicação).

Uma análise dos autores e instituições com mais publicações depositadas também não foi realizada, pois exigiria uma padronização de todos os campos de autores e instituições das mais de 23 mil publicações. O ideal seria aplicar um controle de autoridade para padronizar os sobrenomes dos autores, por exemplo, mas esse é um trabalho que demanda muito tempo e não seria viável para o tempo de desenvolvimento dessa dissertação.

Outro problema localizado é que as análises dependem dos metadados preenchidos, geralmente, pelos próprios autores. Em alguns momentos foram localizados metadados insuficientes ou imprecisos, o que interferiu na análise. No mesmo sentido, observou-se que alguns formulários e instruções para submissão de documentos no E-LIS eram ambíguos, desatualizados ou incompletos, como certos campos de metadados que deixaram de existir, baseando-se em análises de outros autores, como Santillan-Aldana (2009).

8.1 Sugestão de trabalhos futuros

O ELIScript não é um trabalho que se encerra com essa dissertação. Devido às limitações encontradas ainda há muito o que se implementar na ferramenta. A criação de novas categorias de análise, como identificação de autoria, co-autoria e instituições; e a análise inter-relacionada das categorias já analisadas no presente trabalho são sugestões de trabalho futuro.

Sugere-se, também, a criação de um *dashboard*, uma espécie de painel interativo, em que o usuário pode inserir os dados coletados do E-LIS e visualizar as informações agrupadas pelas categorias de análise, sem necessariamente ter que mexer com o código de programação. Nesse painel, o usuário poderia optar pela visualização dos dados de forma geral (análise global do repositório) ou com informações segmentadas por categorias e continentes.

O E-LIS teve um novo campo incorporado aos formulários de submissão de documentos, *Link to Research Data*, para vincular os dados que subsidiaram a pesquisa às publicações depositadas no repositório (DE ROBBIO, SUBIRATS E PESET, 2020). Será interessante investigar como essa função é utilizada e como é feita a gestão de dados de pesquisa por pesquisadores e profissionais da área de Biblioteconomia e Ciência da Informação.

REFERÊNCIAS

- ABADAL, E.; GUALLAR, J. Research on Library and Information Science in Spain: diagnosis 2020. **Profesional de la información**, [s. l.], v. 29, n. 4, 2020. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/83162>. Acesso em: 14 jan. 2021.
- ABDULLAHI, I.; KARISIDDAPPA, C. R.; ASUNDI, A. Y. Philosophical and Pedagogical Foundations of LIS Curriculum Design. In: ABDULLAHI, I.; KARISIDDAPPA, C. R.; ASUNDI, A. Y. (ed.). **LIS Education in Developing Countries: The Road Ahead**. Boston: DeGruyter, 2014. p. 131-145.
- ADAMICK, J.; REZNIK-ZELLEN, R. Representation and recognition of subject repositories. **D-Lib Magazine**, [s.l.], v. 17, n. 9-10, p. 1-9, 2011. DOI 10.1045/september2010-adamick. Disponível em: <http://www.dlib.org/dlib/september10/adamick/09adamick.html>. Acesso em: 27 maio 2019.
- ALBAGLI, S. O que é Ciência Aberta e qual o papel das agências de fomento diante deste fenômeno? In: ENCONTRO CAPES DE CIÊNCIA ABERTA, 2., 2019, Brasília. Disponível em: <http://capes.gov.br/conteudo/2-encontro-capes-de-ciencia-aberta/>. Acesso em: 13 dez. 2019.
- ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, 2014. DOI <https://doi.org/10.18617/liinc.v10i2.749>. Disponível em: <http://revista.ibict.br/liinc/article/view/3593>. Acesso em: 02 jul. 2020.
- ALBARILLO, F. Language in Social Science Databases: English Versus Non-English Articles in JSTOR and Scopus. **Behavioral & Social Sciences Librarian**, [s. l.], v. 33, n. 2, p. 77–90, 2014. Disponível em: <https://doi.org/10.1080/01639269.2014.904693>. Acesso em: 10 jan. 2021.
- ALVARADO, R. U. **A lei de Lotka e a produtividade dos autores**. 2007. 240f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.
- ALVARENGA, L. **A institucionalização da pesquisa educacional no Brasil: estudo bibliométrico dos artigos publicados na Revista Brasileira de Estudos Pedagógicos 1944-74**. 1996. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade Federal de Minas Gerais, Belo Horizonte, 1996. Disponível em: <http://hdl.handle.net/1843/FAEC-85SHVE>. Acesso em: 16 out. 2019.
- ANACONDA. **Anaconda: the world's most popular data science platform**. 2020. Disponível em: <https://www.anaconda.com/>. Acesso em: 03 jul. 2020.
- ANVUR. **Rating of Scientific Journals: Classification of “Scientific” and “Class A” journals**. 2021. Disponível em: <https://www.anvur.it/en/activities/rating-of-scientific-journals/>. Acesso em: 18 jan. 2021.
- ARELLANO, F. F. M.; PEÑA, J. M. P. Core Curriculum in LIS Latin American Schools. In: IFLA WLIC: Libraries: dialogue for change, 2019. Athens. **Proceedings** [...]. Athens: IFLA, 2019. p. 1-10. Disponível em: <http://library.ifla.org/id/eprint/2848>. Acesso em: 16 jan. 2021.
- ARENCIBIA-JORGE, R. SANTILLÁN-ALDANA, J.; SUBIRATS-COLL, I. Iniciativas de acceso abierto en Ciencias de la Información y Documentación: evolución y perspectivas de E-LIS. **Revista Española de Documentación Científica**, [s.l.], v. 28, n. 2, p. 221-232, 2005. Disponível em: <http://hdl.handle.net/10760/6633>. Acesso em: 24 maio 2019.

ASIST. Inside ASIS&T: Annual meeting 2006 just around the corner. **Bulletin of the American Society for Information Science and Technology**, [s. l.], jun/jul. 2006. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/bult.2006.1720330102>. Acesso em: 28 nov. 2020.

ASIST. Inside ASIS&T: ASIS&T 2005 Annual Meeting Second Plenary Speaker Added to Lineup. **Bulletin of the American Society for Information Science and Technology**, [s. l.], oct/nov. 2006. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/bult.2005.1720320102>. Acesso em: 28 nov. 2020.

BAPTISTA, A. A. *et al.* Comunicação científica: o papel da Open Archives Initiative no contexto do acesso livre. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, n. esp., p. 1-17, 2007. DOI 10.5007/1518-2924.2007v12nesp1p1. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12nesp1p1>. Acesso em 8 maio 2019.

BARRETO, A. D. A. Mudança estrutural no fluxo do conhecimento: a comunicação eletrônica. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 122–127, out. 1998. DOI <https://doi.org/10.1590/S0100-19651998000200003>. Disponível em: <http://revista.ibict.br/ciinf/article/view/792/822>. Acesso em: 17 maio 2019.

BARRUECO-CRUZ, J. M.; SUBIRATS-COLL, I. RCLIS: towards a digital library for Information Science. **Biblioteconomia i Documentació**, [s. l.], p. 1-11, 2003. Disponível em: <http://bid.ub.edu/11barru2.htm>. Acesso em: 24 maio 2019.

BERNIUS, S. The impact of open access on the management of scientific knowledge. **Online Information Review**, [s. l.], v. 34, n. 4, p. 583-603, 2010. DOI <https://doi.org/10.1108/14684521011072990>. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/14684521011072990>. Acesso em: 21 maio 2019.

BIORXIV. **About bioRxiv**. 2019. Disponível em: <https://www.biorxiv.org/about-biorxiv>. Acesso em: 17 out. 2019.

BJÖRK, B. Open Access Subject Repositories: An Overview. **Journal of the American Society for Information Science**, [s. l.], v. 65, n. 4, p. 698-706, 2014. DOI <https://doi.org/10.1002/asi.23021>. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23021>. Acesso em: 27 maio 2019.

BJÖRNEBORN, L. **Small-world structures across an academie web space**: a library and information science approach. 2004. 339f. Dissertation (PhD in Library and Information Science) – Department of Informations Studies, Royal School of Library and Information Science, Copenhagen, 2004. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.2353&rep=rep1&type=pdf>. Acesso em: 16 out. 2019.

BORKO, H. Information science: what is it?. **American documentation**, [s. l.], v. 19, n. 1, p. 3-5, 1968. Disponível em: <https://doi.org/10.1002/asi.5090190103>. Acesso em: 13 dez. 2020.

BORREGAARD, M. K.; HART, E. M. Towards a more reproducible ecology. **Ecography**, [s. l.], v. 39, n. 4, p. 349-353, 2016. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/ecog.02493>. Acesso em: 17 out. 2019.

BULLINGER, H. *et al.* **Berlin Declaration on open access to knowledge in the sciences and humanities**. 2003, on-line. Disponível em: <https://openaccess.mpg.de/Berlin-Declaration>. Acesso em: 21 maio 2019.

CAMPBELL, N. R. **What is Science?** London: Methuen, 1921. 186 p.

- CAPEES. **Qualis periódicos**. 2021. Disponível em: <https://sucupira.capes.gov.br/sucupira/public/index.xhtml>. Acesso em: 10 jan. 2021.
- CHALMERS, A. F. **What is this thing called science?: an assessment of the nature and status of science and its methods**. 4th ed. Brisbane: University of Queensland Press, 2013.
- CHAN, L. Supporting and enhancing scholarship in the digital age: The role of open-access institutional repositories. **Canadian Journal of Communication**, [s. l.], v. 29, n. 3, p. 277–300. Disponível em: <https://cjc-online.ca/index.php/journal/article/view/1455>. Acesso em: 24 out. 2020.
- CHAN, L. *et al.* **Budapest open access initiative**. 2002, on-line. Disponível em: <https://www.budapestopenaccessinitiative.org/translations/portuguese-translation>. Acesso em: 18 maio 2019.
- CHANG, Y. W.; HUANG, M. H.; LIN, C. W. Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. **Scientometrics**, [s. l.], v. 105, 2015, p. 2071-2087. Disponível em: <https://doi.org/10.1007/s11192-015-1762-8>. Acesso em: 19 nov. 2020.
- CHILIMO, W. L.; ONYANCHA, O. B. How open is open access research in Library and Information Science? **South African Journal of Library and Information Science**, [s. l.], v. 84, n. 1, p. 11-19, 2018. DOI <https://doi.org/10.7553/84-1-1710>. Disponível em: <http://sajlis.journals.ac.za/pub/article/view/1710/1488>. Acesso em: 27 maio 2019.
- CHRISTÓVÃO, H. T. Da comunicação informal à comunicação formal: identificação da frente de pesquisa através de filtros de qualidade. **Ciência da Informação**, Brasília, v. 8, n. 1, p. 3-36, 1979. Disponível em: <http://www.brapci.ufpr.br/documento.php?dd0=000005538&dd1=854c6>. Acesso em: 08 out. 2019.
- CITeseerX. **About CiteSeerX**. 2019. Disponível em: <http://citeseerx.ist.psu.edu/>. Acesso em: 17 out. 2019.
- COLEMAN, A. Self-archiving and the copyright transfer agreements of ISI-ranked Library and Information Science journals. **Journal of the American Society for Information Science and Technology**, [s. l.], v. 58, n. 2, p. 286-296, 2007. DOI <https://doi.org/10.1002/asi.20494>. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20494>. Acesso em: 27 maio 2019.
- COLLNET. **Global Interdisciplinary Research Network for the Study of all Aspects of Collaboration in Science and in Technology**. 2019. Disponível em: <http://www.collnet.de/>. Acesso em: 28 nov. 2020.
- COMISSÃO Europeia. **O Processo de Bolonha e o Espaço Europeu do Ensino Superior**. 2021. Disponível em: https://ec.europa.eu/education/policies/higher-education/bologna-process-and-european-higher-education-area_pt. Acesso em: 14 jan. 2021.
- COSTA, M. P.; LEITE, F. C. L. **Repositórios institucionais da América Latina e o acesso aberto à informação científica**. Brasília: IBICT, 2017. Disponível em: <http://hdl.handle.net/10760/31109>. Acesso em: 27 maio 2019.
- CRESWELL, J. W. **Research Design: qualitative, quantitative, and mixed methods approaches**. 4. ed. Los Angeles: SAGE, 2014.
- CRIBB, J.; SARI, T. **Open Science: sharing knowledge in the global century**. Collingwood: CSIRO Publishing, 2010. 232 p.

CRUZ, J. M. B.; *et. al.* Tabela de Classificação JITA. **Boletim Técnico do PPEC**, Campinas, SP, v. 3, n. 1, p. 15 p., 2018. Disponível em: <https://econtents.bc.unicamp.br/boletins/index.php/ppec/article/view/9136>. Acesso em: 8 jan. 2021.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008. 451p.

DAVID, P. A.; FORAY, D. Economic Fundamentals of the Knowledge Society. **Policy Futures in Education**, [s. l.], v. 1, n. 1, p. 20-49, 2005. DOI <https://doi.org/10.2304/pfie.2003.1.1.7>. Disponível em: <https://journals.sagepub.com/doi/10.2304/pfie.2003.1.1.7>. Acesso em: 27 maio 2019.

DE ROBBIO, A. E-LIS: un open archive per library and information science. **AIB Notizie**, [s. l.], v. 15, n. 2, p. 1-2, 2003. Disponível em: <http://hdl.handle.net/10760/5366>. Acesso em: 24 maio 2019.

DE ROBBIO, A.; KATZMAYR, M. The management of an international open access repository: the case of E-LIS. **GMS Medizin-Bibliothek-Information**, [s. l.], v. 9, n. 1, p. 1-8, 2009. Disponível em: <http://hdl.handle.net/10760/13220>. Acesso em: 24 maio 2019.

DE ROBBIO, A.; SUBIRATS, I.; PESET, F. E-LIS between old and new forms of Grey Literature encompasses new forms of relationship between librarians in the different country. **Grey Journal (TGJ)**, [s. l.], v. 16, n. 2, 2020. Disponível em: <http://search.ebscohost.com.ez27.periodicos.capes.gov.br/login.aspx?direct=true&db=lih&AN=143256931&lang=pt-br&site=ehost-live&authtype=ip,cookie,uid>. Acesso em: 10 nov. 2020.

E-LIS. **E-Prints in Library & Information Science**. 2019. Disponível em: <http://eprints.rclis.org/>. Acesso em: 8 maio 2019.

E-LIS. **Policies**. 2020. Disponível em: <http://eprints.rclis.org/policies.html>. Acesso em: 8 maio 2019.

EMERY, J. How green is our valley?: five-year study of selected LIS journals from Taylor & Francis for green deposit of articles. **Insights: the UKSG journal**, [s. l.], v. 31, p. 1-9, 2018. DOI <https://doi.org/10.1629/uksg.406>. Disponível em: <https://insights.uksg.org/articles/10.1629/uksg.406/>. Acesso em: 27 maio 2019.

FAMILI, A. *et al.* Data Preprocessing and Intelligent Data Analysis. **Intelligent Data Analysis**, [s. l.], v. 1, n. 1, jan. 1997, p. 3-23. DOI 10.3233/IDA-1997-1102. Disponível em: <https://content.iospress.com/articles/intelligent-data-analysis/ida1-1-02>. Acesso em: 03 jul. 2020.

FARACE, D.; SCHÖPFEL, J. (org.). **Grey literature in library and information studies**. Berlin: Walter de Gruyter, 2010. Disponível em: <https://doi.org/10.1515/9783598441493>. Acesso em: 03 jul. 2020.

FENNER, M. Altmetrics and Other Novel Measures for Scientific Impact. *In*: FECHER, B.; FRIESIKE, S. **Opening science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing**. Springer: Cham, 2014. p. 179-189. DOI https://doi.org/10.1007/978-3-319-00026-8_2. Disponível em: <https://link.springer.com/content/pdf/10.1007%2F978-3-319-00026-8.pdf>. Acesso em: 02 jul. 2020.

FECHER, B. FRIESIKE, S. Open science: one term, five schools of thought. *In*: FECHER, B. FRIESIKE, S. **Opening science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing**. Springer: Cham, 2014. p. 17-47. DOI https://doi.org/10.1007/978-3-319-00026-8_2. Disponível em:

<https://link.springer.com/content/pdf/10.1007%2F978-3-319-00026-8.pdf>. Acesso em: 02 jul. 2020.

FERREIRA, S. A. **Produção científica sobre Biblioteca Pública nos programas de pós-graduação em Ciência da Informação no Brasil**. 2017. 164 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte. 2017. Disponível em: <http://hdl.handle.net/1843/BUOS-AMWGHP>. Acesso em: 08 out. 2019.

FERREIRA, S. M. S. P.; MUNIZ JR., J. S. O movimento do livre acesso e a democratização de conteúdos científicos: um projeto de editoração eletrônica de revistas de Ciências da Comunicação. *In*: CONGRESSO INTERNACIONAL LATINO-AMERICANO DE PESQUISA DA COMUNICAÇÃO, 3., São Paulo. 2005. **Anais** [...]. São Paulo: [s. n.], 2005. Disponível em: <https://www.researchgate.net/publication/268424277>. Acesso em: 22 maio 2018.

FIGUEIREDO, N. M. **Metodologias para promoção do uso da informação**: técnicas aplicadas particularmente em bibliotecas universitárias e especializadas. São Paulo: Nobel: Associação Paulista de Bibliotecários, 1990.

FONSECA, E. N. **Introdução à biblioteconomia**. Brasília: Briquet de Lemos Livros, 2007.

FONSECA, E. N. (org.) **Bibliometria**: teoria e prática. São Paulo: Cultrix, 1986.

FOSTER, I. The Grid: A New Infrastructure for 21st Century Science. **Physics Today**, [s. l.], v. 55, n. 2, 2002. Disponível em: <https://doi.org/10.1063/1.1461327>. Acesso em: 23 jun. 2020.

FOSTER Facilitate Open Science Training For European Research. **Open Science Policies**: What is it?. [S. l.: s. n.], 2020. Disponível em: <https://book.fosteropenscience.eu/en/02OpenScienceBasics/09OpenSciencePolicies.html>. Acesso em: 23 mar. 2020.

FREIRE, G. H. *et al.* Ciência da informação: temática, histórias e fundamentos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, p. 6-19, 2006. DOI <https://doi.org/10.1590/S1413-99362006000100002>. Disponível em: http://www.scielo.br/scielo.php?pid=S1413-99362006000100002&script=sci_abstract&lng=pt. Acesso em: 17 out. 2019.

GARVEY, W. D. **Communication**: the essence of science: facilitating information exchange among librarians, scientists, engineers and students. Oxford: Pergamon, 1979. 332 p.

GILCHRIST, A. Reflections on knowledge, communication and knowledge organization. **Knowledge Organization**, [s. l.], v. 42, n. 6, p. 456-469, 2015. DOI <https://doi.org/10.5771/0943-7444-2015-6-456>. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2015-6-456/reflections-on-knowledge-communication-and-knowledge-organization-jahrgang-42-2015-heft-6>. Acesso em: 27 maio 2019.

GINSPARG, P. Scholarly information architecture, 1989-2015. **Data Science Journal**, [s. l.], v. 3, p. 29-37, 2004. DOI <https://doi.org/10.2481/dsj.3.29>. Disponível em: https://www.jstage.jst.go.jp/article/dsj/3/0/3_0_29/article-char/ja/. Acesso em: 27 maio 2019.

GLÄNZEL, W. **Bibliometrics as a research field**: a course on theory and application of bibliometric indicators. Statistics. Course Handouts, 2003. Disponível em: https://www.cin.ufpe.br/~ajhol/futuro/referencias/01%23_Bibliometrics_Module_KUL_BIBLIOMETRICS%20AS%20A%20RESEARCH%20FIELD.pdf. Acesso em: 08 out. 2019.

GOLDSCHMIDT-CLERMONT, L. Communication patterns in high-energy physics. **High Energy Physics Libraries Webzine**, n. 6, 2002. Disponível em: <https://cds.cern.ch/record/546422/files/sis-2002-163.html>. Acesso em: 21 set. 2020.

GOMES, D. L.; BENCHIMOL, A. C.; BARROS, T. H. B. O uso de ferramentas de busca e acesso a artigos científicos pelos pesquisadores brasileiros. **Informação & Sociedade: Estudos**, João Pessoa, v. 28, n. 1, 2018. DOI <https://doi.org/10.22478/ufpb.1809-4783.2018v28n1.38113>. Disponível em: <http://www.periodicos.ufpb.br/index.php/ies/article/view/38113>. Acesso em: 22 maio 2018.

GOMES, S. L. R.; MENDONÇA, M. A. R.; SOUZA, C. M. Literatura cinzenta. *In*: CENDÓN, B. V.; KREMER, J. M.; CAMPELLO, B. S. (org.) **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: Ed. UFMG, 2000. Cap. 6, p. 96-103.

GOMES, V. C. *et al.* Os fundos setoriais e a redefinição do modelo de promoção de ciência, tecnologia e inovação no Brasil: uma análise à luz do CT-Agro. **Revista de Administração**, São Paulo, v. 50, n. 3, p. 353-368, set. 2015. DOI <https://doi.org/10.5700/rausp1205>. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0080-21072015000300353&lng=en&nrm=iso. Acesso em: 24 jun. 2020.

GUERADO, E. Scientific societies and the third industrial revolution – The future role of the OTC. **Injury**, [s. l.], v. 48, p. S1-S4, 2017. DOI [https://doi.org/10.1016/S0020-1383\(17\)30788-X](https://doi.org/10.1016/S0020-1383(17)30788-X). Disponível em: [https://www.injuryjournal.com/article/S0020-1383\(17\)30788-X/abstract](https://www.injuryjournal.com/article/S0020-1383(17)30788-X/abstract). Acesso em: 19 ago. 2019.

HARNAD, S. *et al.* The access/impact problem and the green and gold roads to open access. **Serials Review**, [s. l.], v. 30, n. 4, p. 310-314, 2004. DOI <https://doi.org/10.1080/00987913.2004.10764930>. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00987913.2004.10764930>. Acesso em: 27 maio 2019.

IBICT é parceiro do E-LIS. 2004. Disponível em: <http://sitehistorico.ibict.br/Sala-de-Imprensa/noticias/2004/ibict-e-parceiro-do-E-LIS>. Acesso em: 27 maio 2019.

IFLA. **SCECSAL XX Conference**. 2012. Disponível em: <https://www.ifla.org/DE/events/scecsal-xx-conference>. Acesso em: 08 nov. 2020.

ISO 639:1. **Codes for the representation of names of languages — Part 1: Alpha-2 code**. 2002. Disponível em: <https://www.iso.org/standard/22109.html>. Acesso em: 13 set. 2020.

JCDL. **Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries**. Association for Computing Machinery, New York, NY, USA. Disponível em: <https://dl.acm.org/doi/proceedings/10.1145/1378889?tocHeading=heading14>. Acesso em: 28 nov. 2020.

JUPYTER. **Project Jupyter**. 2020. Disponível em: <https://jupyter.org/>. Acesso em: 03 jul. 2020.

KAWALEC, A. Research trends in library and information science based on Spanish scientific publication 2000 to 2010. **Malaysian journal of library & information science**, [s. l.], v. 18, n. 2, p. 1-13, 2017. Disponível em: <http://mjs.um.edu.my/index.php/MJLIS/article/view/1864>. Acesso em: 14 jan. 2021.

KHABSA, M.; GILES, C. L. The number of scholarly documents on the public web. **PloS one**, [s. l.], v. 9, n. 5, p. 1-6, 2014. DOI <https://doi.org/10.1371/journal.pone.0093949>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4015892/>. Acesso em: 19 ago. 2019.

KIM, J. Faculty Self-Archiving: Motivations and Barriers. **Journal of the American Society for Information Science and Technology**, [s. l.], v. 6, n. 9, p. 1909-1922, 2010. Disponível em: <https://doi.org/10.1002/asi.21336>. Acesso em: 24 out. 2020.

LAAKSO, M. Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed. **Scientometrics**, [s. l.], v. 99, n. 2, p. 475-494, 2014. DOI <https://doi.org/10.1007/s11192-013-1205-3>. Disponível em: <https://link.springer.com/article/10.1007/s11192-013-1205-3>. Acesso em: 27 maio 2019.

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

LARIVIÈRE, V.; HAUSTEIN, S.; MONGEON, P. The oligopoly of academic publishers in the digital era. **PloS one**, [s. l.], v. 10, n. 6, p. 1-15, 2015. DOI <https://doi.org/10.1371/journal.pone.0127502>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127502>. Acesso em: 05 out. 2019.

LAWRENCE, A. Electronic documents in a print world: grey literature and the internet. **Media International Australia**, [s. l.], n. 143, p. 122-131, 2012. DOI <https://doi.org/10.1177/1329878X1214300114>. Disponível em: <https://core.ac.uk/download/pdf/30678667.pdf>. Acesso em: 08 out. 2019.

LE COADIC, Y. F. **A ciência da informação**. 2. ed. rev. e atual. Brasília: Briquet de Lemos, 2004. 124p.

LOWNDES, J. S. S. *et al.* Our path to better science in less time using open data science tools. **Nature ecology & evolution**, [s. l.], v. 1, n. 6, p. 1-7, 2017. DOI <https://doi.org/10.1038/s41559-017-0160>. Disponível em: <https://www.nature.com/articles/s41559-017-0160/>. Acesso em: 17 out. 2019.

LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. **Portal: Libraries and the Academy**, [s. l.], v. 3, n. 2, p. 327-336, 2007. DOI <https://doi.org/10.1353/pla.2003.0039>. Disponível em: <https://muse.jhu.edu/article/42865/summary>. Acesso em: 27 maio 2019.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, 1998. DOI <https://doi.org/10.1590/S0100-19651998000200005>. Disponível em: <http://www.scielo.br/pdf/%0D/ci/v27n2/macias.pdf>. Acesso em: 16 out. 2019.

MARTÍNEZ-ÁVILA, D.; GRACIOSO, L. S. Tratamento temático da informação a partir dos trabalhos publicados nos anais do capítulo brasileiro de ISKO: pontos de partida, identidade nacional e agentes epistêmicos. *In*: SIMÕES, M. G.; LIMA, G. Á. (coord.). **Do tratamento à organização da informação: reflexões sobre concepções, perspectivas e tendências**. Coimbra: Imprensa da Universidade de Coimbra, 2020. p. 49-82.

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

MEDEIROS, N. A repository of our own: the E-LIS e-prints archive. **OCLC Systems & Services: International digital library perspectives**, [s. l.], v. 20, n. 2, p. 58-60, jun. 2004. DOI <https://doi.org/10.1108/10650750410539040>. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/10650750410539040>. Acesso em: 21 abr. 2019.

MENÊSES, R. V. **A literatura sobre Ciência Aberta na Ciência da Informação: um estudo na LISTA e E-LIS**. Dissertação (Mestrado em Ciência da Informação) — Universidade de Brasília, Faculdade de Estudos Sociais Aplicados, Departamento de Ciência da Informação e

Documentação, 2019. Disponível em: <https://repositorio.unb.br/handle/10482/37293>. Acesso em: 09 out. 2020.

MIGUEL, E. *et al.* Promoting transparency in social science research. **Science**, [s. l.], v. 343, n. 6166, p. 30-31, 2014 DOI <https://doi.org/10.1126/science.1245317>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4103621/>. Acesso em: 20 mar. 2020.

MILLER, R. E. Literature Suggests Information Professionals Have Adopted New Roles. **Evidence Based Library and Information Practice**, [s. l.], v. 12, n. 1, p. 137-139, 2017. DOI <https://doi.org/10.18438/B8B05R>. Disponível em: <https://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/28739>. Acesso em: 24 maio 2019.

MINAYO, M. C. S. *et al.* (Org.). **Pesquisa social: teoria, método e criatividade**. 21. ed. Petrópolis: Vozes, 2002.

MITTAL, R. Library and information science research trends in India. **Annals of Library and Information Studies**, [s. l.], v. 58, 2011. p. 319-325. Disponível em: <http://nopr.niscair.res.in/handle/123456789/13481>. Acesso em: 16 jan. 2021.

MORENO, F. P.; LEITE, F. C. L.; ARELLANO, M. Á. M. Acesso livre a publicações e repositórios digitais em ciência da informação no Brasil. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, p. 82-94, jan/abr. 2006. DOI <https://doi.org/10.1590/S1413-99362006000100007>. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/447/258>. Acesso em: 27 maio 2019.

MORRISON, B. H. *et al.* E-LIS: The Open Archive for Library and information Science. **The Charleston Advisor**, [s. l.], v. 9, n. 1, p. 56-61, 2007. Disponível em: <http://eprints.rclis.org/10158/>. Acesso em: 27 maio 2019.

MUELLER, A. **Wordcloud for Python documentation**. 2020. Disponível em: https://amueller.github.io/word_cloud/index.html. Acesso em: 18 set. 2020.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 27-38, 2006. DOI <https://doi.org/10.1590/S0100-19652006000200004>. Disponível em: <http://revista.ibict.br/ciinf/article/view/1138/1293>. Acesso em: 27 maio 2019.

MURRAY-RUST, P. Open data in science. **Serials Review**, [s. l.], v. 34, n. 1, p. 52-64, 2008. Disponível em: <https://doi.org/10.1080/00987913.2008.10765152>. Acesso em: 20 mar. 2020.

NASEER, M. M.; MAHMOOD, K. LIS research in Pakistan: An analysis of Pakistan Library and Information Science journal 1998-2007. **Library Philosophy and Practice**, [s. l.], p. 1-10, 2009. Disponível em: <https://search-proquest.ez27.periodicos.capes.gov.br/docview/224102701?accountid=134127>. Acesso em: 24 maio 2019.

NEVES, B.; FERREIRA, C. Caracterização da produção científica portuguesa em Ciência da Informação disponibilizada em acesso aberto no E-LIS. **Cadernos BAD**, [s. l.], v. 1, n. 2, p. 95-98, 2014. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/download/1184/1191>. Acesso em: 27 maio 2019.

NORONHA, D. P.; FERREIRA, S. M. S. P. Revisões de Literatura. *In*: CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. M. **Fontes de informação para pesquisadores e profissionais**. Editora UFMG, 2000. Cap. 14, p. 191-198.

NORONHA, D. P.; MARICATO, J. M. Estudos métricos da informação: primeiras aproximações. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 13, n. 1, p. 116-128, 2008. DOI <https://doi.org/10.5007/1518-2924.2008v13nesp1p116>. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1137>. Acesso em: 16 out. 2019.

OLIVEIRA, E. F. T. **Estudos Métricos da Informação no Brasil**. São Paulo: Cultura Acadêmica, 2018. Disponível em: <http://www.marilia.unesp.br/Home/Publicacoes/estudos-metricos-da-informacao-no-brasil---e-book.pdf>. Acesso em: 9 abr. 2018.

OLMEDA-GÓMEZ, C.; OVALLE-PERANDONES, M. A.; PERIANES-RODRÍGUEZ, A. Co-word analysis and thematic landscapes in Spanish information science literature, 1985–2014. **Scientometrics**, [s. l.], v. 113, n. 1, p. 195-217, 2017. Disponível em: <https://link.springer.com/article/10.1007/s11192-017-2486-8>. Acesso em: 14 jan. 2021.

ONYANCHA, O. B. Forty-five years of LIS research evolution, 1971–2015: An informetrics study of the author-supplied keywords. **Publishing research quarterly**, [s. l.], v. 34, n. 3, p. 456-470, 2018. Disponível em: <https://link.springer.com/article/10.1007/s12109-018-9590-3>. Acesso em: 14 jan. 2021.

OPEN Archives Initiative. 2019. on-line. Disponível em: <https://www.openarchives.org/>. Acesso em: 23 maio 2019.

OPEN Definition. 2020. on-line. Disponível em: <https://opendefinition.org/>. Acesso em: 19 mar. 2020.

OSORIO, N. L. An Analysis of Subject Coverage and Worldwide Involvement of E-LIS: the International Repository for Library and Information Science. **Library Philosophy and Practice**, [s. l.], n. 1067, 2014. Disponível em: <https://digitalcommons.unl.edu/libphilprac/1067/>. Acesso em: 13 jan. 2021.

OTLET, P. **Tratado de documentação**: o livro sobre o livro, teoria e prática. Brasília: Briquet de Lemos, 2018. 698p.

PEÑA, E. P.; TINOCO, C. M.; LOREZON, J. M. E. Crónica de las XIV Jornadas Nacionales de Información y Documentación en Ciencias de la Salud. **Revista Española de Documentación Científica**, [s. l.], v. 34, n. 3, p. 483–486, 2011. Disponível em: <http://redc.revistas.csic.es/index.php/redc/article/view/711>. Acesso em: 28 nov. 2020.

PRIEM, J.; GROTH, P.; TARABORELLI, D. The Altmetrics Collection. **PLoS ONE**, [s. l.], v. 7, n. 11, e48753. Disponível em: <https://doi.org/10.1371/journal.pone.0048753>. Acesso em: 23 mar. 2020.

PRITCHARD, A. Statistical bibliography or bibliometrics. **Journal of documentation**, [s. l.], v. 25, n. 4, p. 348-349, 1969.

PUBMED Central. **PMC overview**. 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/about/intro/>. Acesso em: 17 out. 2019.

PUJOL, E. F.; VIVÓ, L. A. Las tesis doctorales en españa (1997-2008): Análisis, estadísticas y repositorios cooperativos. **Revista Española de Documentación Científica**, [s. l.], v. 33, n. 1, p. 63-89, 2010. DOI <https://doi.org/10.3989/redc.2010.1.711>. Disponível em: <https://search-proquest.ez27.periodicos.capes.gov.br/docview/212061030?accountid=134127>. Acesso em: 24 maio 2019.

PYTHON Introduction. 2019. Disponível em: https://www.w3schools.com/python/python_intro.asp. Acesso em: 09 ago. 2019.

RAMIREZ, J. Evolución de las tecnologías de la información. *In*: CONGRESO INTERNACIONAL DE BIBLIOTECOLOGÍA, DOCUMENTACIÓN Y ARCHIVÍSTICA, 2., La Paz, 2003. **Anales** [...]. La Paz: [s. n.], 2003. Disponível em: <http://eprints.rclis.org/4657/>. Acesso em: 28 nov. 2020.

REY-ROCHA, J. *et al.* English versus Spanish in science evaluation. **Nature**, [s. l.], v. 397, n. 14, 1999. DOI <https://doi.org/10.1038/16132>. Disponível em: <https://www.nature.com/articles/16132>. Acesso em: 10 jan. 2021.

ROGEL-SALAZAR, J. **Data Science and Analytics with Python**. DOI <https://doi.org/10.1201/9781315151670>. Boca Raton: CRC Press, 2017. 376 p.

RUIZ-ROSETO, J.; RAMIREZ-GONZALEZ, G.; VIVEROS-DELGADO, J. Software survey: ScientoPy, a scientometric tool for topics trend analysis in scientific publications. **Scientometrics**, [s. l.], 121, 1165–1188, 2019. DOI 10.1007/s11192-019-03213-w. Disponível em: <https://doi.org.ez27.periodicos.capes.gov.br/10.1007/s11192-019-03213-w>. Acesso em: 07 mar. 2020.

SANDVE, G. K. *et al.* Ten simple rules for reproducible computational research. **PLoS computational biology**, [s. l.], v. 9, n. 10, 2013. DOI <https://doi.org/10.1371/journal.pcbi.1003285>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3812051/>. Acesso em: 20 mar. 2020.

SANTILLÁN-ALDANA, J. The open access movement and the library world seen from the experience of the E-LIS project. **OCLC Systems & Services: International digital library perspectives**, [s. l.], v. 25, n. 2, p. 135-147, maio 2009. DOI <https://doi.org/10.1108/10650750910961938>. Disponível em: <https://www.emeraldinsight.com/doi/full/10.1108/10650750910961938>. Acesso em: 24 maio 2019.

SANTOS, S. R. O.; OLIVEIRA, D. A. Autoarquivamento na Ciência da Informação: Uma análise dos documentos depositados no repositório digital e-LIS. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v. 9, n. 2, 2019. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/19126>. Acesso em: 03 maio. 2021.

SANTOS, S. R. O. **ELIScript**. Github, 2021. DOI 10.5281/zenodo.4743129. Disponível em: <https://github.com/sarahrubia/eliscript>. Acesso em: 11 maio. 2021.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, 1996. Disponível em: http://www.brapci.inf.br/_repositorio/2010/08/pdf_fd9fd572cc_0011621.pdf. Acesso em: 13 out. 2019.

SCIENCE-METRIX. **Analytical Support for Bibliometrics Indicators**: open access availability of scientific publications. 2018. Disponível em: https://www.science-metrix.com/sites/default/files/science-metrix/publications/science-metrix_open_access_availability_scientific_publications_report.pdf. Acesso em: 16 jan. 2021.

SCIMAGO. **SJR — SCImago Journal & Country Rank [Portal]**. 2020. Disponível em: <http://www.scimagojr.com>. Acesso em: 24 nov. 2020.

SHREEVES, S. L.; CRAGIN, M. H. Introduction: Institutional repositories: Current state and future. **Library Trends**, [s. l.], v. 57, n. 2, p. 89-97, 2008. DOI <https://doi.org/10.1353/lib.0.0037>. Disponível em: <https://muse.jhu.edu/article/262023/summary>. Acesso em: 27 maio 2019.

SILVA, L. C.; SEGUNDO, J. E. S.; SILVA, M. F. Princípios de FAIR e melhores práticas do *linked data* na publicação de dados de pesquisa. *In*: ENCONTRO NACIONAL DE PESQUISA EM

CIÊNCIA DA INFORMAÇÃO, 19, 2018, Marília. **Anais [...]**. Marília: ANCIB, 2018. Disponível em: <http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/view/1598>. Acesso em: 17 out. 2019.

SINGEH, F. W.; ABRIZAH, A.; KARIM, N. H. A. What inhibits authors to self-archive in open access repositories? A Malaysian case. **Information Development**, [s. l.], v. 29, n. 1, p. 24-35, 2013. Disponível em: <https://doi.org/10.1177/0266666912450450>. Acesso em: 12 jan. 2020.

SOCARVIX. **About**. 2019. Disponível em: <http://socarxiv.org>. Acesso em: 17 out. 2019.

SØNDERGAARD, T. F.; ANDERSEN, J.; HJØRLAND, B. Documents and the communication of scientific and scholarly information: revising and updating the UNISIST model. **Journal of Documentation**, [s. l.], v. 59, n. 3, p. 278-320, 2003. DOI 10.1108/00220410310472509. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/00220410310472509/full/html>. Acesso em: 18 jan. 2021.

SOUZA, E. D.; OLIVEIRA, D. A. A análise documentária no grupo Temma: dos indícios às evidências da formação de unidades discursivas. **Ciência da Informação**, Brasília, v. 36, n. 2, 2007. DOI <https://doi.org/10.1590/S0100-19652007000200008>. Disponível em: <http://revista.ibict.br/ciinf/article/view/1178>. Acesso em: 16 out. 2019.

STOCKEMER, D.; WIGGINTON, M. J. Publishing in English or another language: An inclusive study of scholar's language publication preferences in the natural, social and interdisciplinary sciences. **Scientometrics**, [s. l.], v. 118, n. 2, p. 645-652, 2019. Disponível em: <https://link.springer.com/article/10.1007/s11192-018-2987-0>. Acesso em: 10 nov. 2020.

SUBER, P. *et al.* **Bethesda statement on open access publishing**. 2003, on-line. Disponível em: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:4725199>. Acesso em: 21 maio 2019.

SUBIRATS-COLL, I.; ARENCIBIA-JORGE, R.; DE ROBBIO, A. E-prints for Library and Information Science (E-LIS): la tecnología al servicio de la bibliotecología y las ciencias de la información. **Centro Nacional de Información de Ciencias Médicas**, [s. l.], v. 12, n. 6, p. 1-10, 2004. Disponível em: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352004000600007. Acesso em: 27 maio 2019.

SWAN, A. Open access and the progress of science. **American Scientist**, [s. l.], v. 95, n. 3, 2007. DOI <https://doi.org/10.1511/2007.65.373>. Disponível em: https://eprints.soton.ac.uk/263860/1/American_Scientist_article.pdf. Acesso em: 27 maio 2019.

TARGINO, M. G. Comunicação Científica: uma revisão de seus elementos básicos. **Informação & Sociedade: Estudos**, João Pessoa, v. 10, n. 2, p. 1-27, 2000. Disponível em: <http://www.periodicos.ufpb.br/ojs2/index.php/ies/article/view/326>. Acesso em: 27 maio 2019.

TAVARES, W. M. L. **Os fundos setoriais de ciência e tecnologia e seu impacto sobre o setor**. Brasília: Câmara dos Deputados, 2005. Disponível em: https://royaltiesdopetroleo.ucam-campos.br/wp-content/uploads/2017/03/os_fundos_setoriais.pdf. Acesso em: 24 jun. 2020.

TENNANT, J. P. *et al.* The academic, economic and societal impacts of Open Access: an evidence-based review. **F1000Research**, [s. l.], v. 5, p. 1-55, 2016. DOI <https://doi.org/10.12688/f1000research.8460.1>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4837983/>. Acesso em: 19 ago. 2019.

UNESCO/ICSU. **Draft declaration on science and the use of scientific knowledge**, 1999. on-line. Disponível em: <https://www.nature.com/wcs/m00s.html>. Acesso em: 12 ago. 2019.

VANTI, N. **Links hipertextuais na comunicação científica: análise webométrica dos sítios acadêmicos latino-americanos em Ciências Sociais**. 2007. 292 f. Tese (Doutorado em

Comunicação e Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007. Disponível em: <https://www.lume.ufrgs.br/handle/10183/11175>. Acesso em: 16 out. 2019.

VERGUEIRO, W. **Seleção de materiais de informação**: princípios e técnicas. Brasília: Briquet de Lemos, 1995.

VIANA, A. F. **A produção científica acerca da temática LGBT**: um estudo propedêutico nas teses e dissertações na UFMG. 2017. 145 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte. Disponível em: <http://hdl.handle.net/1843/BUOS-AV2FZ5>. Acesso em: 2 set. 2018.

VIANA, C. L. M. **JITA**: versão em português. 2006. Disponível em: <http://eprints.rclis.org/8163/>. Acesso em: 04 set. 2019.

WASSER, L. **Earth analytics python course**: earthlab/earth-analytics-python-course: Version 1.0. Zenodo, 2018. DOI 10.5281/zenodo.2209415. Disponível em: <https://www.earthdatascience.org/courses/earth-analytics-python/>. Acesso em: 17 out. 2019.

WEITZEL, S. R. O mapeamento dos repositórios institucionais brasileiros: perfil e desafios. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 24, n. 54, p. 105-123, 2019. DOI <https://doi.org/10.5007/1518-2924.2019v24n54p105>. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2019v24n54p105>. Acesso em: 05 out. 2019.

WEITZEL, S. R. O papel dos repositórios institucionais e temáticos na estrutura da produção científica. **Em Questão**, Porto Alegre, v. 12, n. 1, p. 51-71, 2006. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/19>. Acesso em: 24 maio 2019.

WEITZEL, S. R.; LEITE, F. C. L.; ARELLANO, M. Á. M. E-LIS: um repositório digital para a biblioteconomia e ciência da informação no Brasil. *In*: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 15., 2008, São Paulo. **Anais [...]**. São Paulo: CRUESP, 2008. p. 1-16. Disponível em: <http://eprints.rclis.org/12537/>. Acesso em: 21 abr. 2019

WESTELL, M. Institutional repositories: Proposed indicators of success. **Library Hi Tech**, [s. l.], v. 24, n. 2, p. 211–226, 2006. DOI <https://doi.org/10.1108/07378830610669583>. Disponível em: <https://www.emeraldinsight.com/doi/abs/10.1108/07378830610669583>. Acesso em: 27 maio 2019.

WHITE, P.; POWELL, S. Code-Literacy for GIS Librarians: A Discussion of Languages, Use Cases, and Competencies. **Journal of Map and Geography Libraries**, [s. l.], v. 15, n. 1, p. 45–67, 2019. Disponível em: <https://doi.org/10.1080/15420353.2019.1660754>. Acesso em: 07 mar. 2019.

WHO – World Health Organization. **Coronavirus disease (COVID-19)**. 2020. Disponível em: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19#:~:text=symptoms>. Acesso em: 08 jan. 2021.

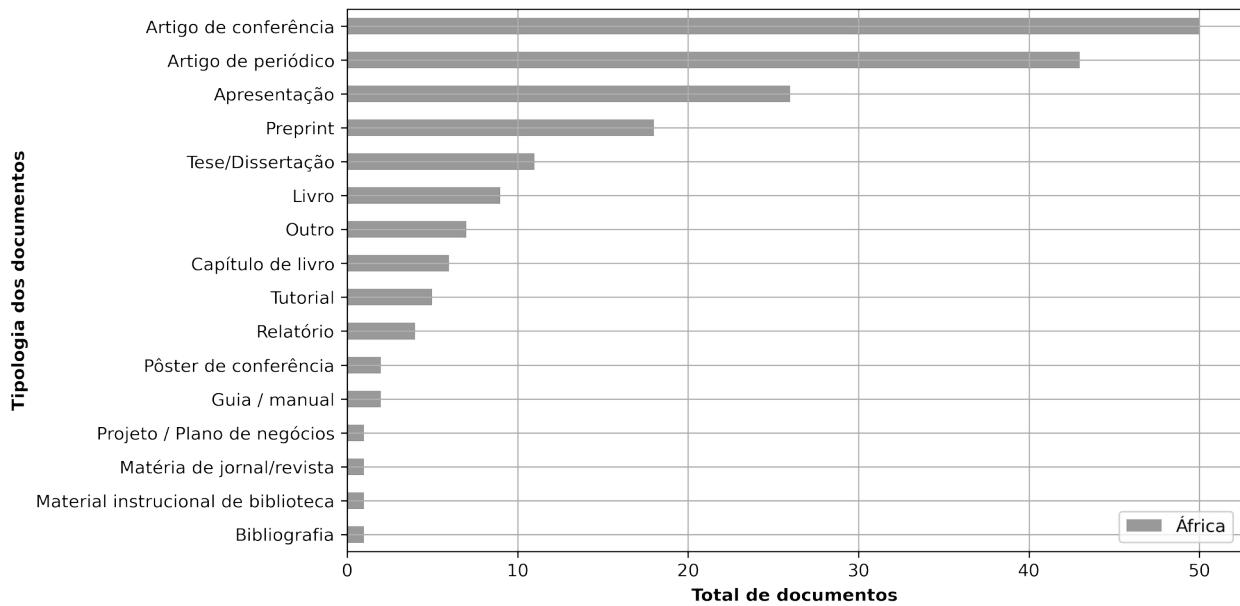
ZENODO. **Research Shared**. 2020. Disponível em: <https://zenodo.org/>. Acesso em: 21 set. 2020.

ZIMAN, J. **Conhecimento público**. Belo Horizonte: Itatiaia; São Paulo: Ed. da Universidade de São Paulo, 1979. 164 p.

APÊNDICE A – Tipologia dos documentos depositados no E-LIS dividido por continente

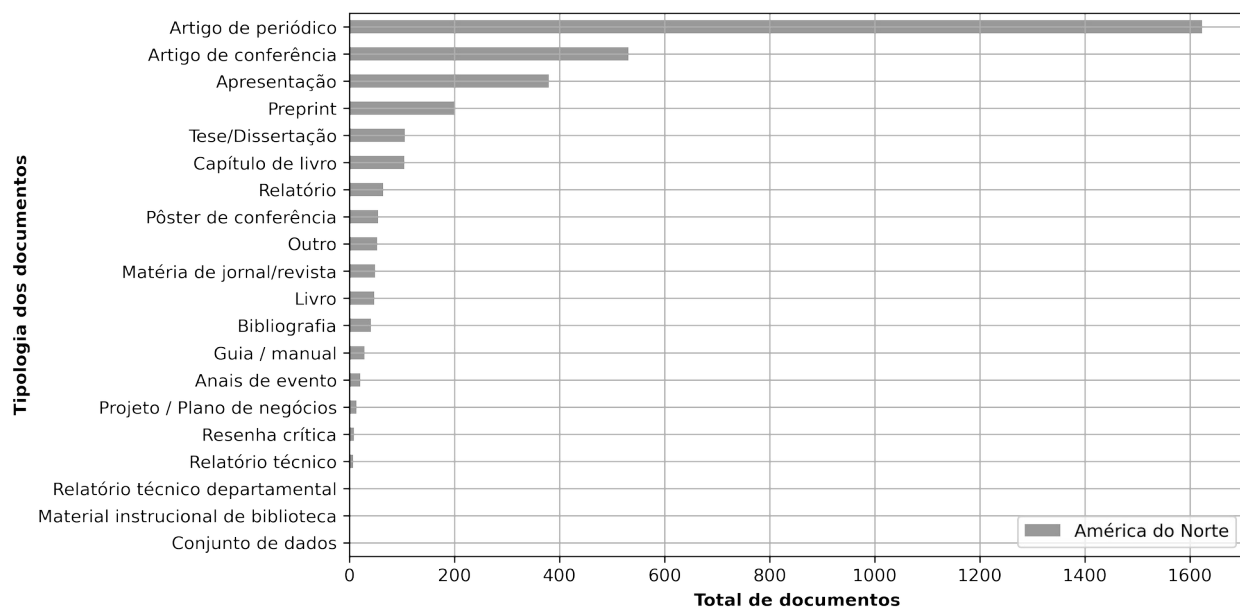
Por meio do ELIScript foi possível gerar gráficos com as tipologias de documentos mais depositados em cada continente. A seguir esses gráficos são apresentados apenas a título de conhecimento, pois os dados foram discutidos anteriormente.

Gráfico 7 – Tipologia de documentos depositados na África



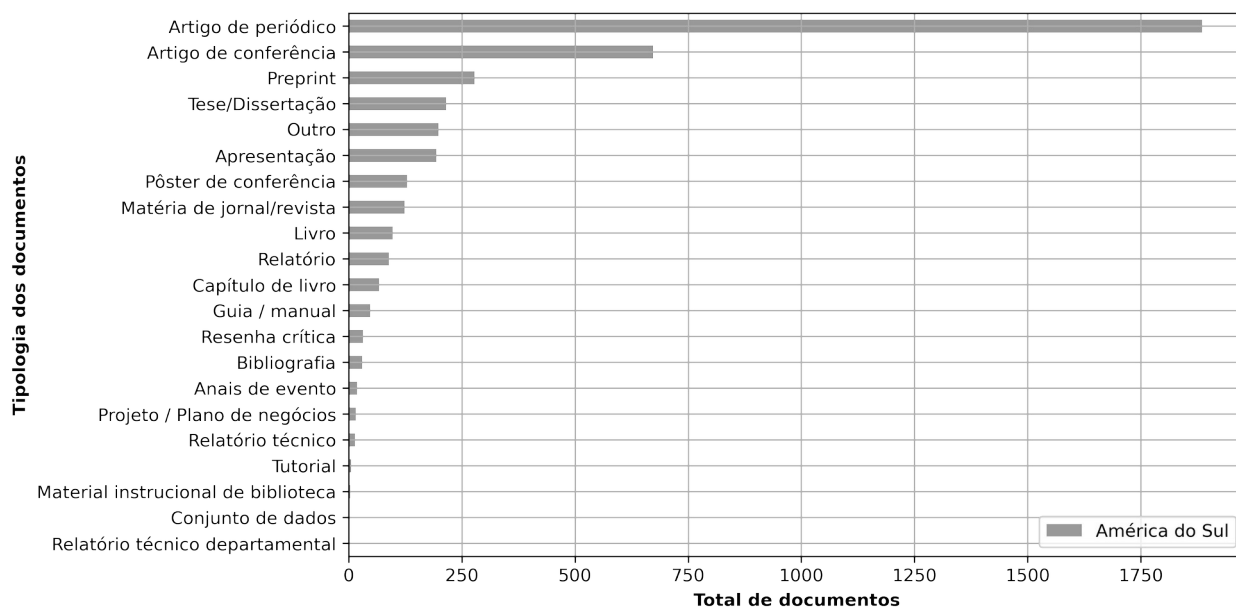
Fonte: dados da pesquisa (2020).

Gráfico 8 – Tipologia de documentos depositados na América do Norte e Central



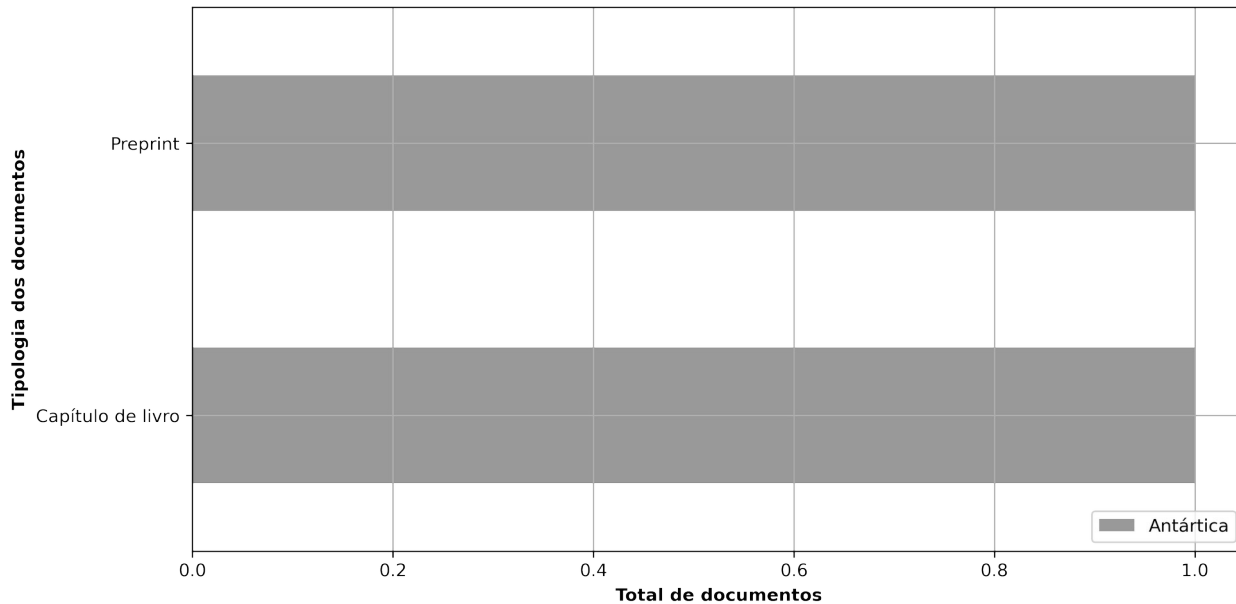
Fonte: dados da pesquisa (2020)

Gráfico 9 – Tipologia de documentos depositados na América do Sul



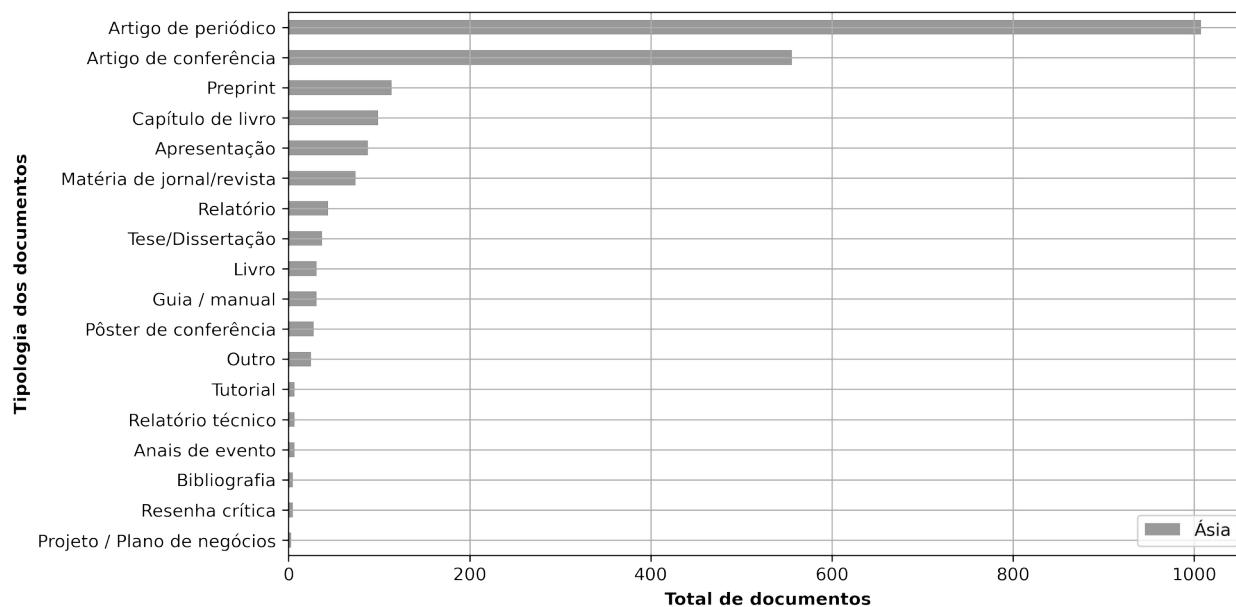
Fonte: dados da pesquisa (2020).

Gráfico 10 – Tipologia de documentos depositados na Antártica



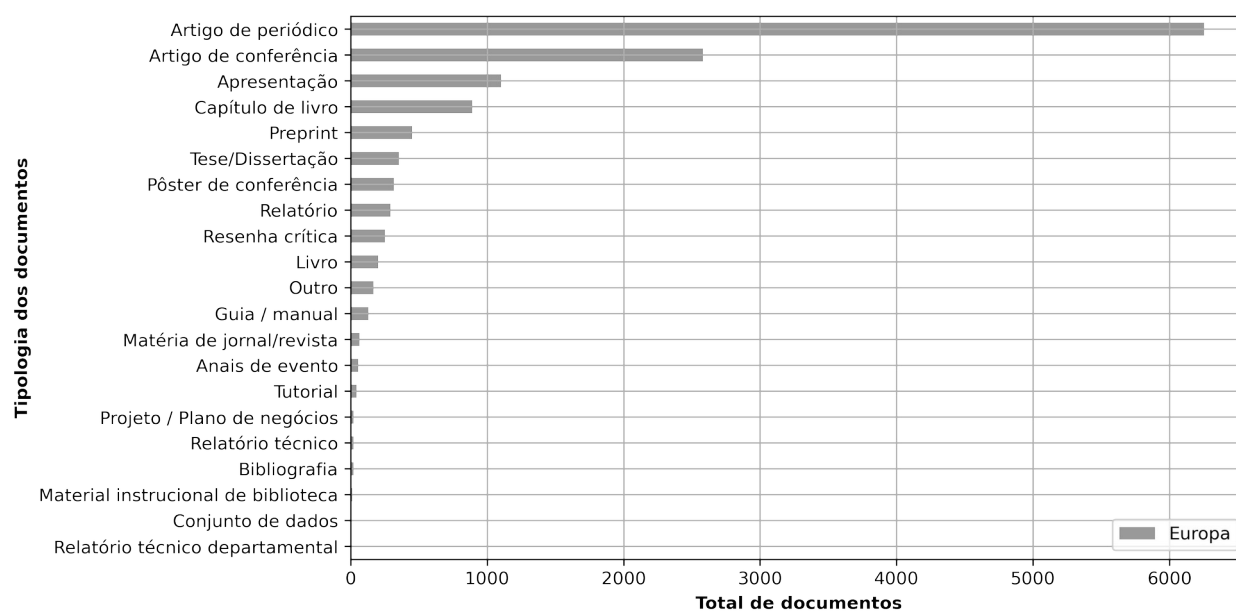
Fonte: dados da pesquisa (2020).

Gráfico 11 – Tipologia de documentos depositados na Ásia



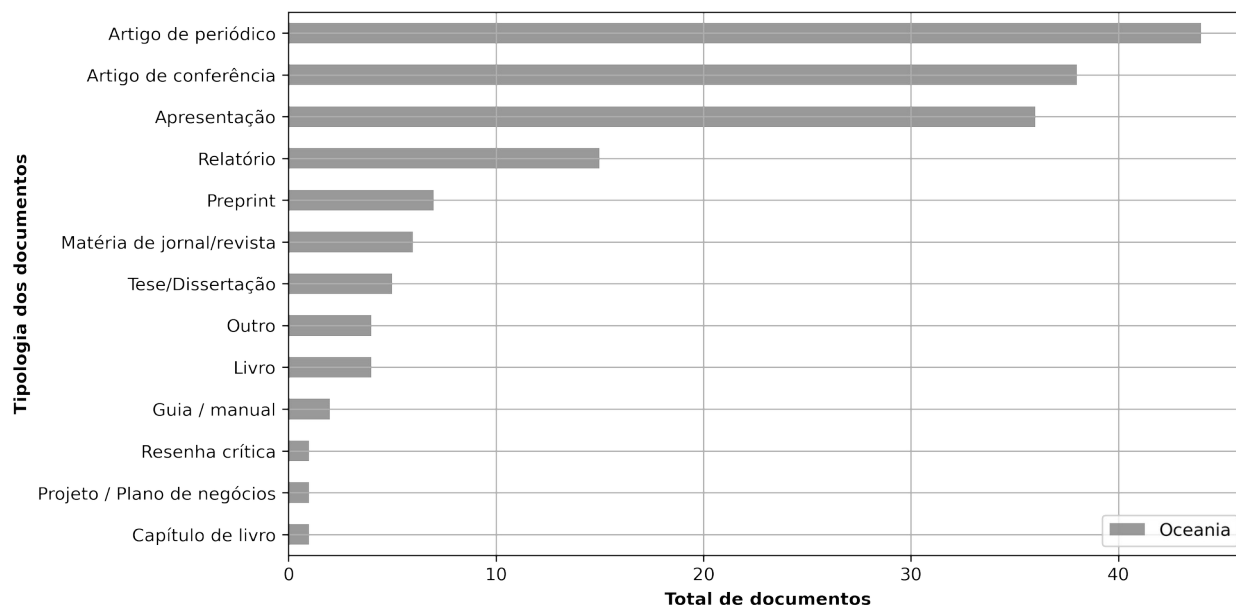
Fonte: dados da pesquisa (2020).

Gráfico 12 – Tipologia de documentos depositados na Europa



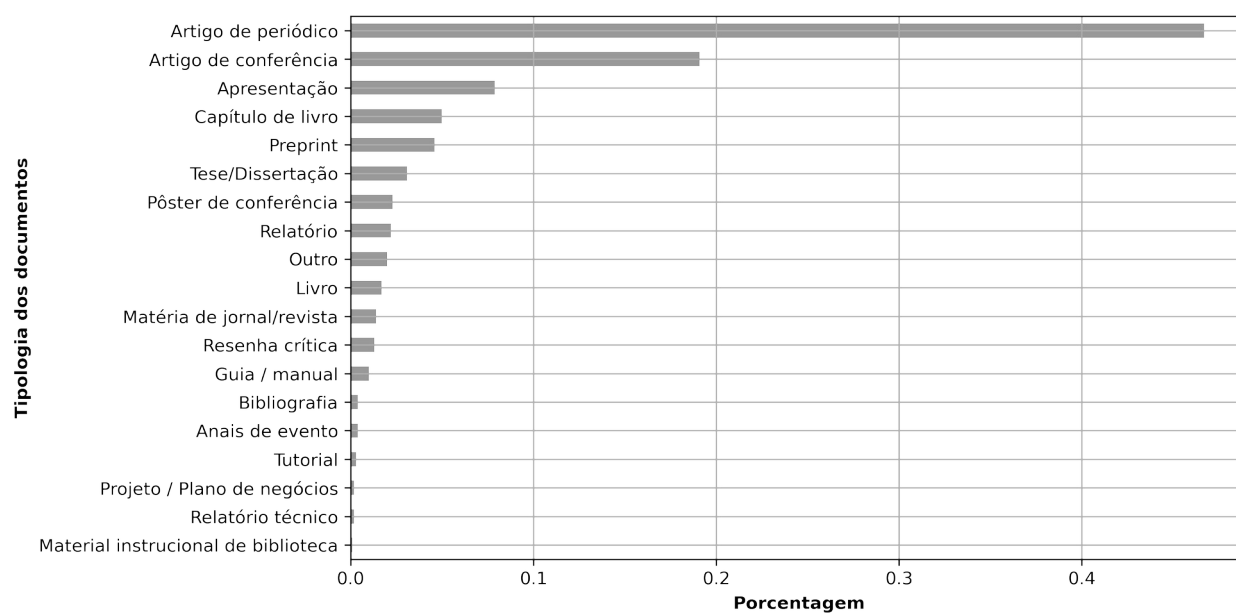
Fonte: dados da pesquisa (2020).

Gráfico 13 – Tipologia de documentos depositados na Oceania



Fonte: dados da pesquisa (2020).

Gráfico 14 – Porcentagem geral das tipologias de documentos depositados no E-LIS

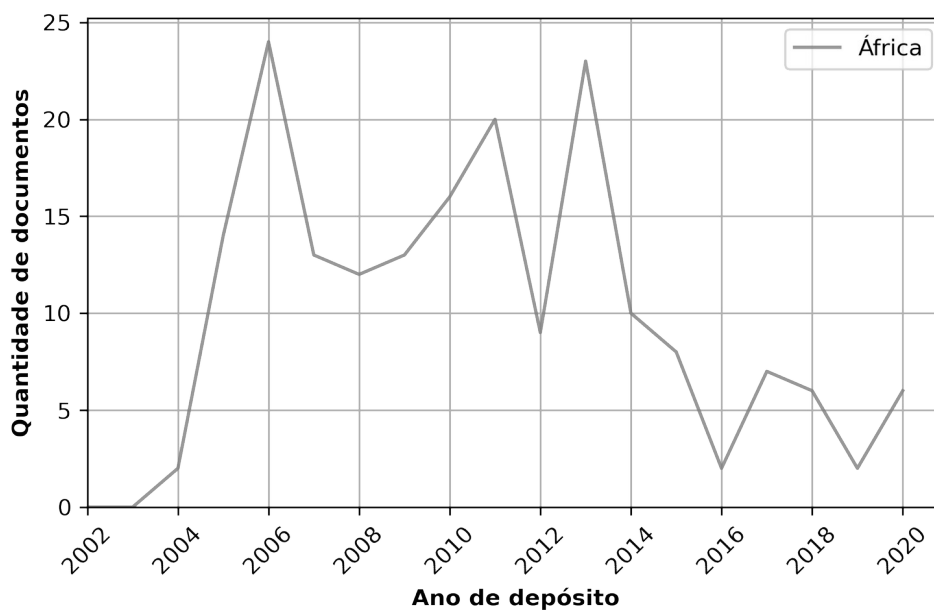


Fonte: dados da pesquisa (2020).

APÊNDICE B – Distribuição por ano de depósito dividido por continente

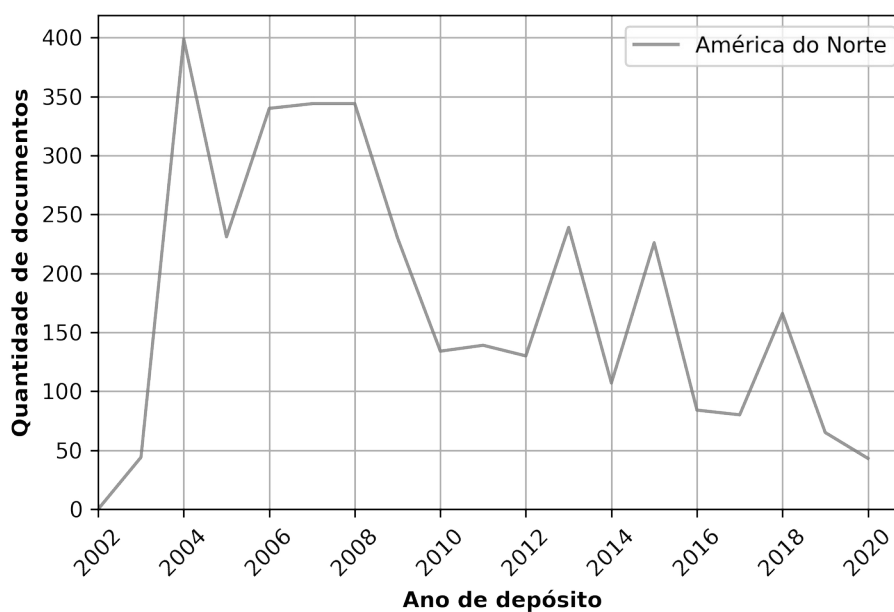
São apresentados os gráficos gerados pelo ELIScript referentes ao ano de depósito por continente.

Gráfico 15 – Distribuição por ano de depósitos realizados na África



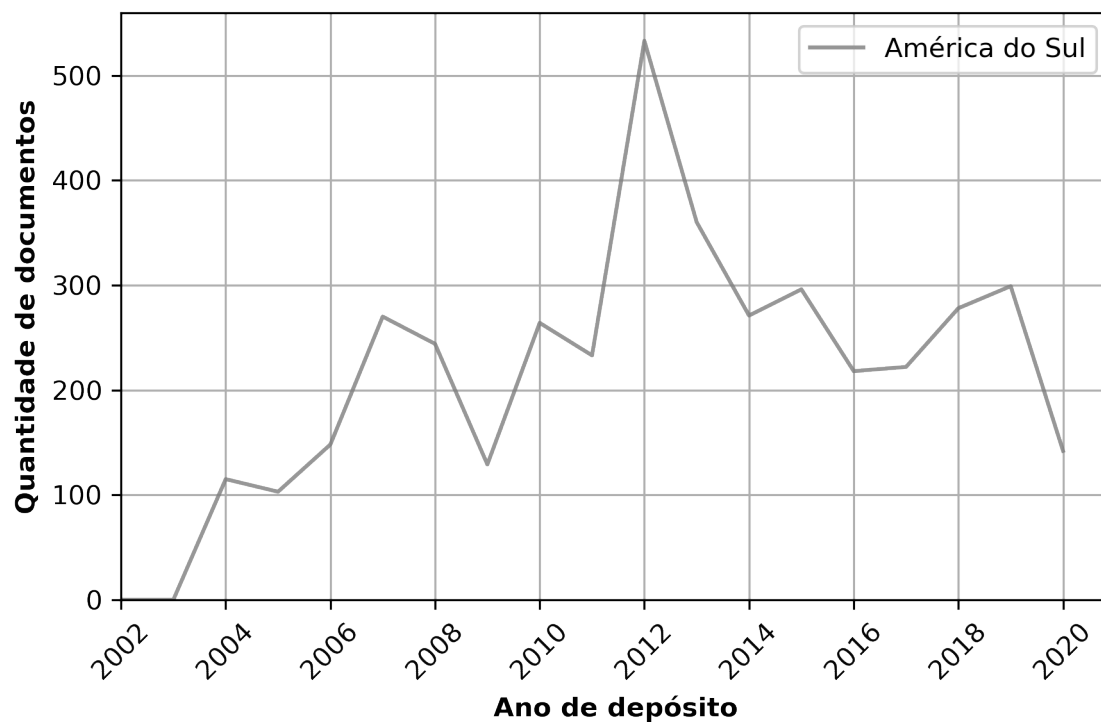
Fonte: dados da pesquisa (2020).

Gráfico 16 – Distribuição por ano de depósitos realizados na América do Norte e Central



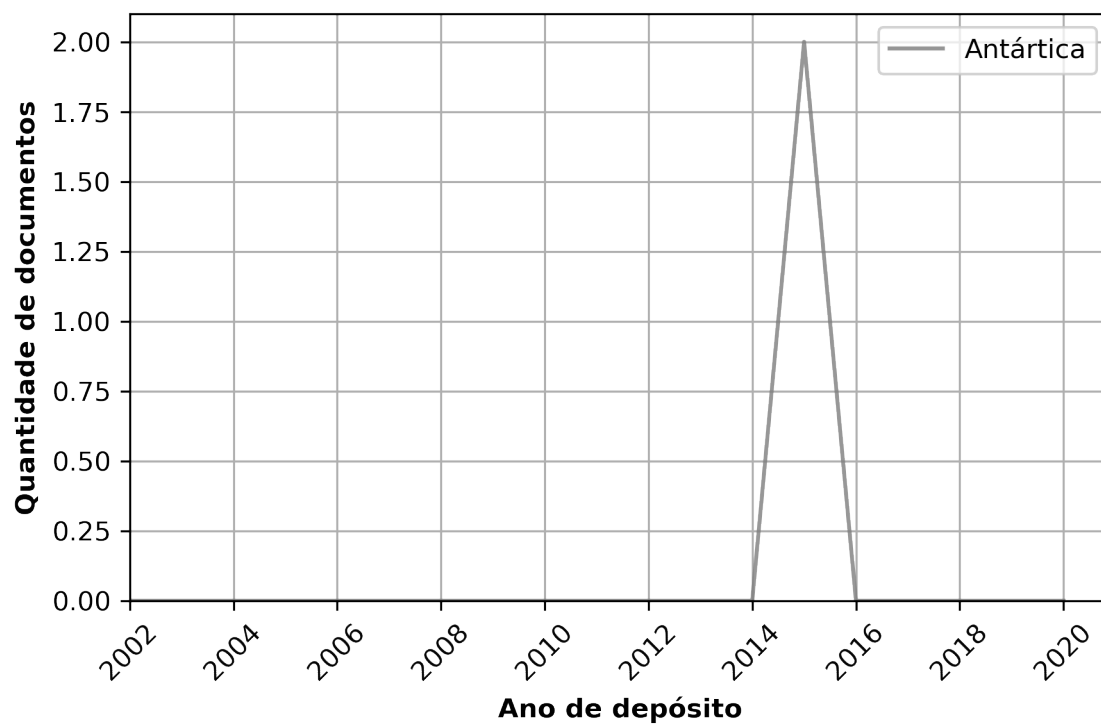
Fonte: dados da pesquisa (2020).

Gráfico 17 – Distribuição por ano de depósitos realizados na América do Sul



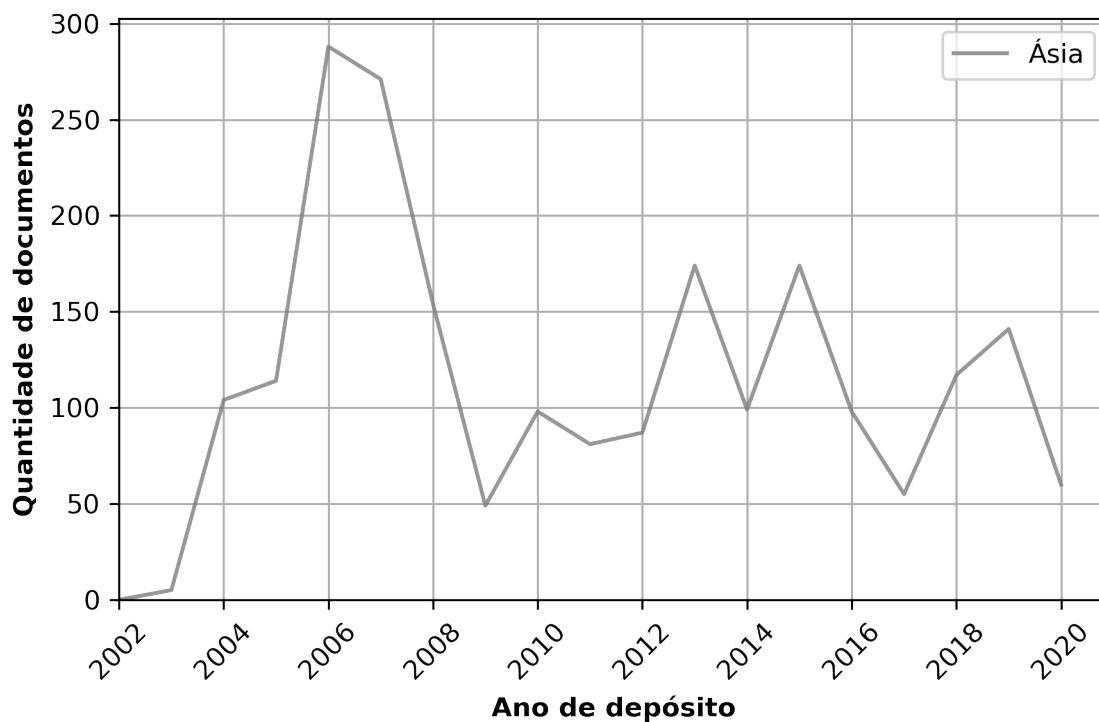
Fonte: dados da pesquisa (2020).

Gráfico 18 – Distribuição por ano de depósitos realizados na Antártica



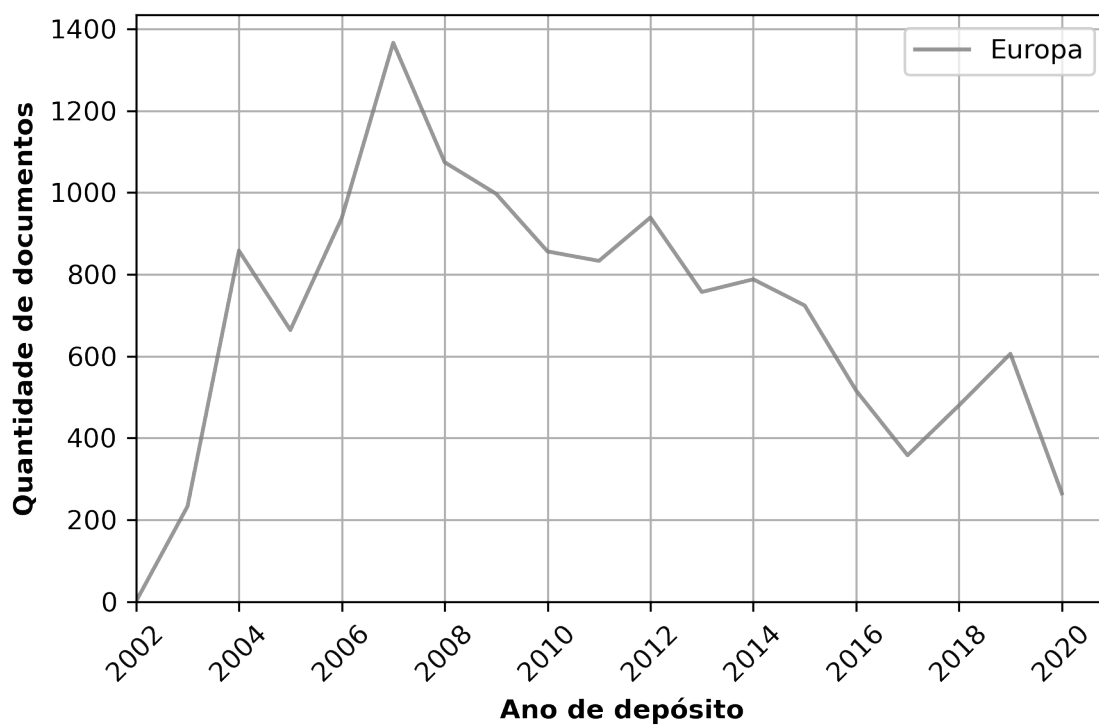
Fonte: dados da pesquisa (2020).

Gráfico 19 – Distribuição por ano de depósitos realizados na Ásia



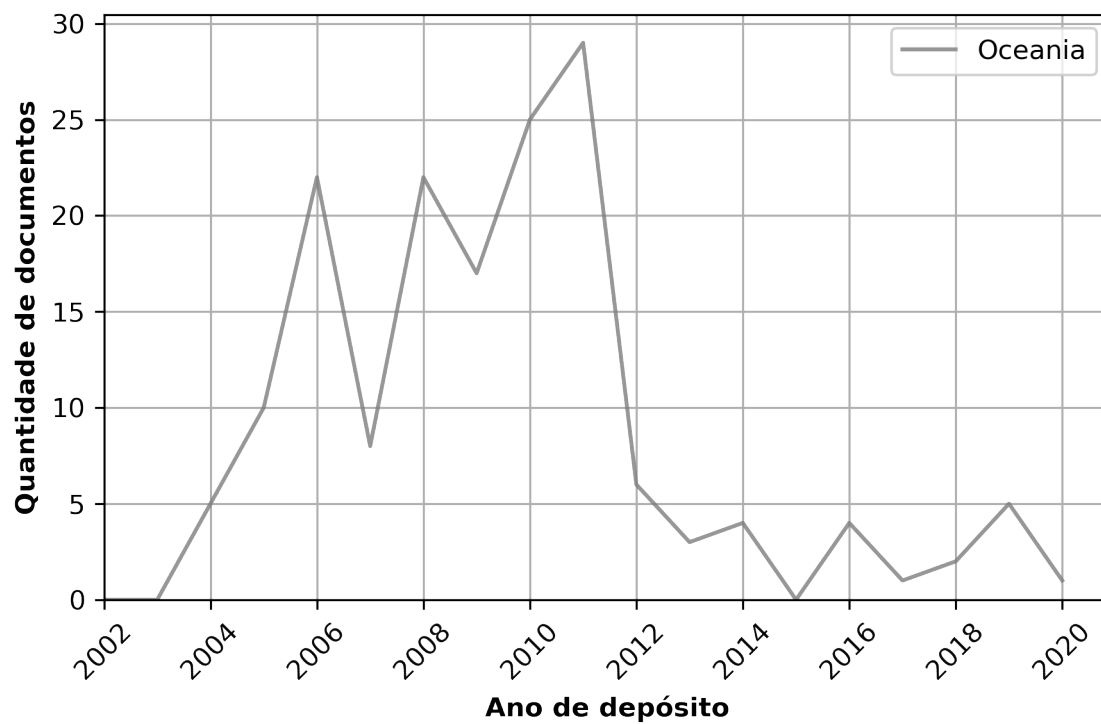
Fonte: dados da pesquisa (2020).

Gráfico 20 – Distribuição por ano de depósitos realizados na Europa



Fonte: dados da pesquisa (2020).

Gráfico 21 – Distribuição por ano de depósitos realizados na Oceania

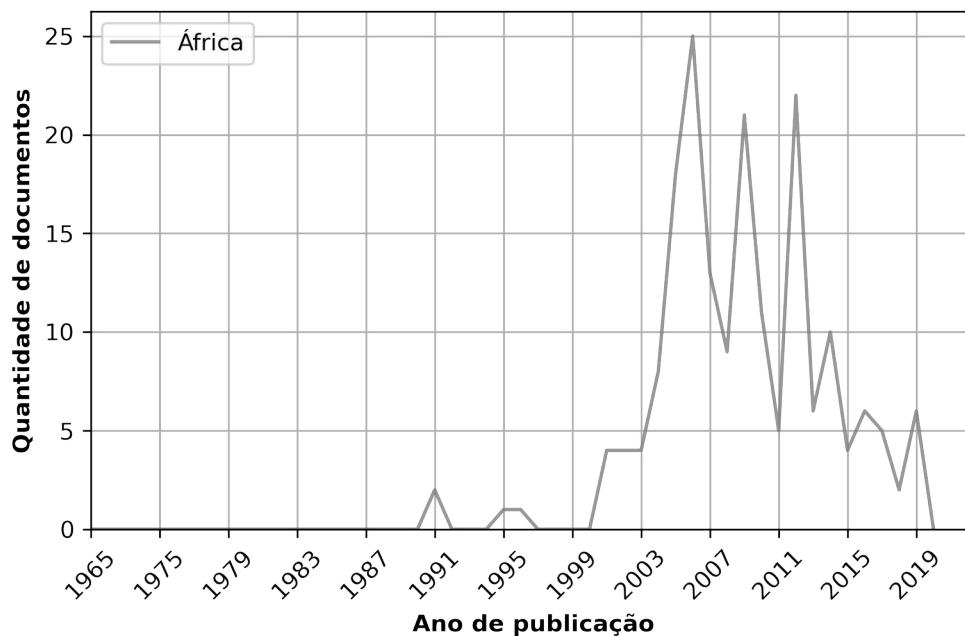


Fonte: dados da pesquisa (2020).

APÊNDICE C – Distribuição por ano de publicação dividido por continente

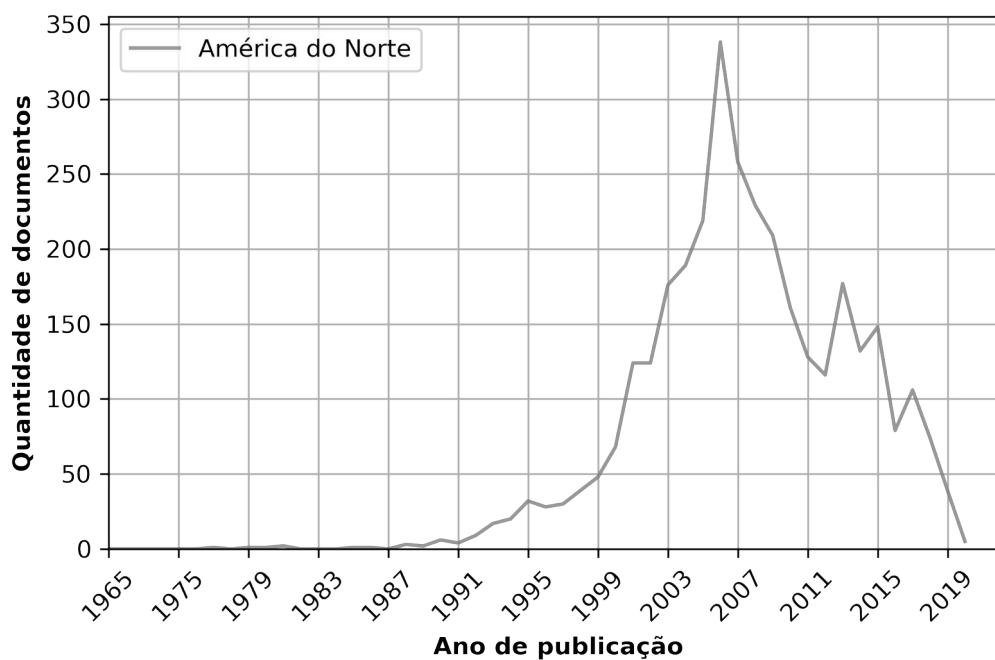
Aqui são apresentados os gráficos gerados pelo ELIScript referentes à distribuição por ano de publicação para cada continente.

Gráfico 22 – Distribuição por ano de publicação na África



Fonte: dados da pesquisa (2020).

Gráfico 23 – Distribuição por ano de publicação na América do Norte e Central



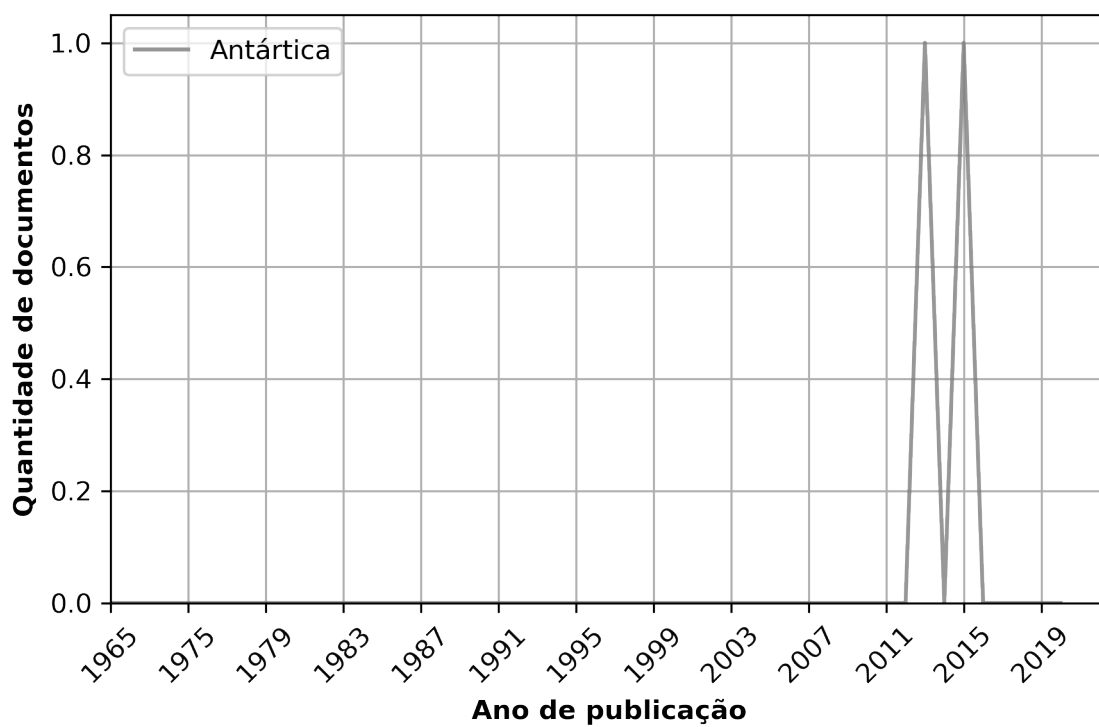
Fonte: dados da pesquisa (2020).

Gráfico 24 – Distribuição por ano de publicação na América do Sul



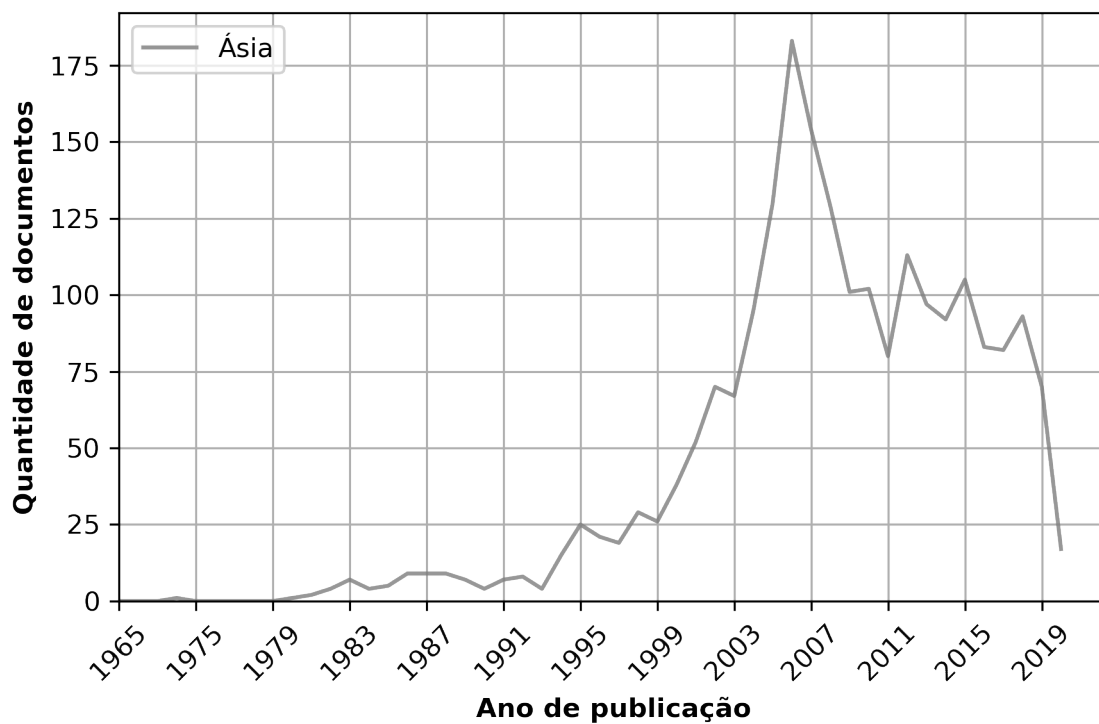
Fonte: dados da pesquisa (2020).

Gráfico 25 – Distribuição por ano de publicação na Antártica



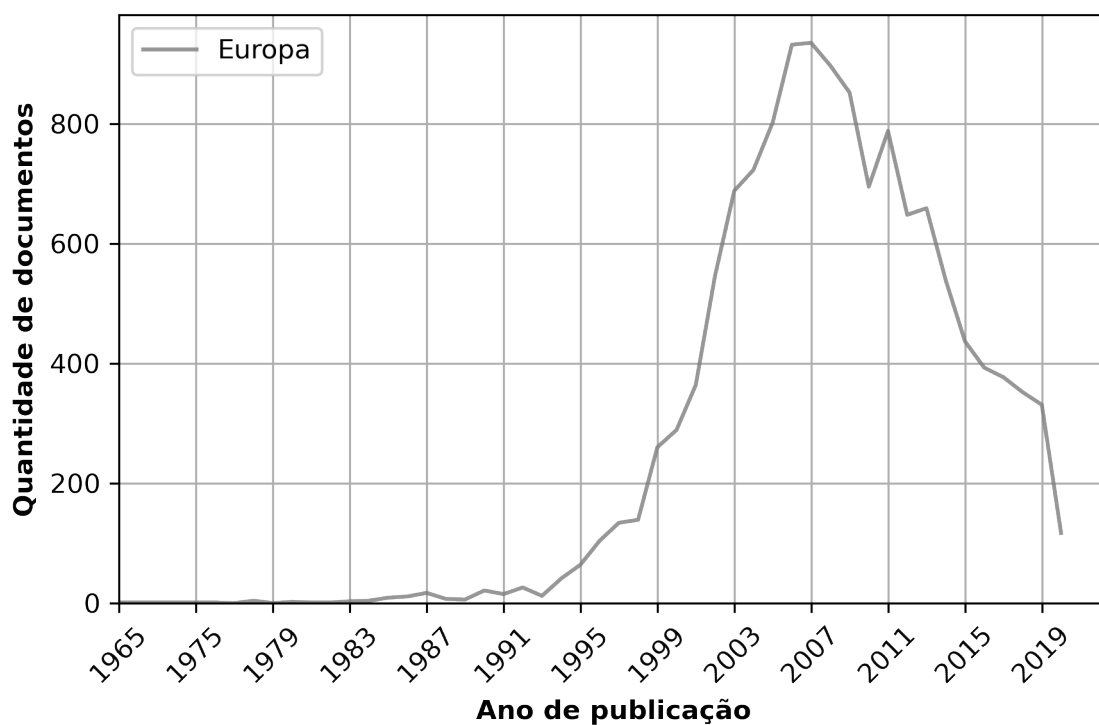
Fonte: dados da pesquisa (2020).

Gráfico 26 – Distribuição por ano de publicação na Ásia



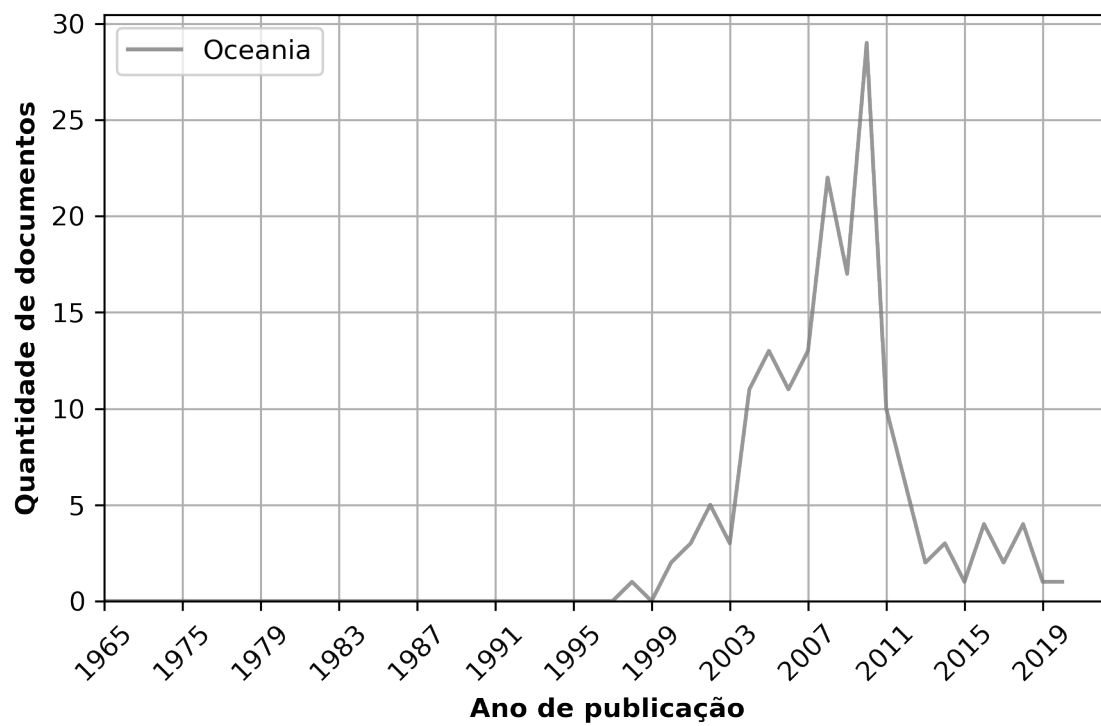
Fonte: dados da pesquisa (2020).

Gráfico 27 – Distribuição por ano de publicação na Europa



Fonte: dados da pesquisa (2020).

Gráfico 28 – Distribuição por ano de publicação na Oceania



Fonte: dados da pesquisa (2020).

APÊNDICE D – Quantidade de documentos depositados no E-LIS dividido por assunto e continente

A partir das funções criadas para análise dos assuntos dos documentos depositados no repositório E-LIS, foi possível extrair um quadro geral com informações comparativas entre continente para cada assunto (Quadro 13).

Legenda do quadro:

- ÁF: África
- AN: Antártica
- ÁS: Ásia
- EU: Europa
- ANC: América do Norte e Central
- OC: Oceania
- AMS: América do Sul.

Quadro 13 – Quantidade de documentos depositados no E-LIS por assunto e continente

ASSUNTOS	ÁF	AN	ÁS	EU	ANC	OC	AMS
Agentes inteligentes	2	0	5	37	6	0	16
Análise de conteúdo (indexação, resumos, classificação)	0	0	19	166	27	2	58
Apresentação da informação: hipertexto, hipermedia	3	0	12	112	29	0	24
Aquisição	0	0	37	89	28	0	20
Arquitetura	0	0	0	54	1	1	11
Arquivos	1	0	23	250	29	1	87
Aspectos curriculares	0	0	7	171	39	0	114
Aspectos teóricos e gerais de bibliotecas e informação	6	0	143	505	216	1	348
Audiovisual, Multimídia	0	0	8	189	16	3	41
Autenticação e controle de acesso	4	0	11	48	4	0	3
Autoridades Centrais	0	0	0	16	4	0	1
Bases de Dados e Redes	3	0	47	412	65	1	58
Bibliotecas científicas	3	0	19	87	15	0	23
Bibliotecas como coleções físicas	4	0	26	244	67	0	135
Bibliotecas de Instituições de Ensino Superior	15	1	66	882	155	9	213
Bibliotecas de Saúde e Medicina	0	0	18	282	28	1	37
Bibliotecas escolares	0	0	8	66	27	0	44
Bibliotecas especializadas	0	0	16	184	23	2	51
Bibliotecas governamentais	1	0	8	47	13	1	19
Bibliotecas Mundiais	1	0	5	27	7	0	7

ASSUNTOS	ÁF	AN	ÁS	EU	ANC	OC	AMS
Bibliotecas Nacionais	4	0	28	135	12	6	24
Bibliotecas Particulares	0	0	2	35	5	0	4
Bibliotecas Públicas	1	0	46	374	90	0	106
Bibliotecas técnicas	3	0	5	32	6	0	10
Biblioteconomia e Ciência da Informação (CI) como um campo	11	1	77	361	112	3	214
Biografias	0	0	1	65	12	0	13
Câmeras digitais	0	0	0	7	0	0	8
Catálogo, controle bibliográfico	2	0	106	397	110	11	116
Catálogos on-line de acesso público (OPAC's)	0	0	14	161	20	0	26
Categorias de usuários: crianças, jovens, grupos sociais	1	0	6	150	60	0	66
CD-ROM	0	0	9	23	1	0	2
Censura	1	0	4	53	29	3	15
Centros de recursos	0	0	3	13	4	1	3
Circulação	1	0	7	26	5	1	12
Computadores	0	0	25	25	9	1	10
Cooperação	4	0	36	273	46	0	36
Descarte	0	0	1	4	3	0	0
Digitalização	2	0	61	284	38	36	40
Disseminação de Documentos	1	0	9	72	7	0	14
Disseminação e difusão da informação	21	0	103	740	252	3	227
Documentos de arquivo	1	0	15	142	10	0	29
Economia da informação	4	0	19	107	73	0	16
Edifícios de bibliotecas, arquivos e museus	1	0	2	89	7	2	8
Empregados	1	0	14	142	21	2	23
Empréstimo interbibliotecário	1	0	11	79	8	0	3
Estruturas de dados e de metadados	1	0	41	361	80	2	98
Estudos de uso	2	0	22	146	38	0	50
Estudos de usuário	2	0	46	297	110	8	112
Filtros e sistemas de filtragem	0	0	3	46	8	0	3
Finanças	0	0	4	77	14	0	13
Fontes de informação, suportes e canais	4	0	78	672	239	2	166
Formação	7	0	61	695	189	4	273

ASSUNTOS	ÁF	AN	ÁS	EU	ANC	OC	AMS
Fotocopiadoras	0	0	1	4	3	0	1
Gerenciamento	3	0	60	384	100	1	152
Gestão de Recursos Humanos	3	0	27	96	35	1	34
Governo Local	0	0	4	81	12	1	18
Hiperlinks de referência	1	1	16	120	76	1	35
Impressão, publicação eletrônica, transmissão	1	0	68	418	143	8	123
Impressos em geral	0	0	7	72	20	1	12
Indústria da informação	1	0	18	221	38	3	22
Indústria de computadores e telecomunicações	0	0	8	33	7	1	5
Indústria de software	0	0	5	37	15	0	6
Indústria, profissão e formação	4	0	21	427	109	0	167
Informação na sociedade	7	0	50	470	250	7	237
Instrução literária	3	0	37	253	103	5	97
Interfaces de usuário, usabilidade	2	0	24	218	71	3	57
Internet, inclusive WWW	15	0	124	668	128	16	148
Inventário	0	0	3	6	1	0	1
Leitura e narração de estórias	1	0	8	73	31	0	48
Liberdade intelectual	2	0	9	86	45	2	26
Linguagens de indexação, processos e esquemas	0	0	30	314	35	3	80
Literatura cinzenta	0	0	25	66	13	0	18
Livros eletrônicos	0	0	17	186	43	1	27
Livros raros e manuscritos	0	0	10	139	14	0	21
Manutenção de registros (cadastros)	0	0	6	64	10	0	11
Marketing	2	0	26	181	31	1	33
Material não impresso, mimeografado, datilografado, manuscrito etc.	0	0	6	42	11	0	10
Mecanismos de busca	1	0	43	312	35	9	33
Meios de comunicação de massa	1	0	6	165	9	2	12
Metodologia e engenharia de softwares	0	0	10	79	30	5	48
Métodos bibliométricos	18	0	232	696	176	9	209
Microformas	0	0	6	3	0	2	5
Mídia eletrônica	0	0	28	192	32	2	30
Mobiliário	0	0	0	9	1	0	0

ASSUNTOS	ÁF	AN	ÁS	EU	ANC	OC	AMS
Museus	0	0	3	64	2	0	21
Necessidades de informação e análise de requisitos de informação	3	0	44	303	118	1	106
Nenhum destes, mas na seção Aspectos teóricos e gerais de bibliotecas e informação	1	0	24	122	37	1	58
Nenhum destes, mas na seção Uso da informação e sociologia da informação	0	0	12	128	30	9	53
Nenhum destes, mas na seção Usuários, nível de instrução e leitura	0	0	6	67	14	2	40
Nenhum destes, mas na seção Bibliotecas como coleções físicas	2	0	2	71	11	0	14
Nenhum destes, mas na seção Publicação e questões legais	1	0	7	127	23	0	19
Nenhum destes, mas na seção Gerenciamento	1	0	12	161	32	1	60
Nenhum destes, mas na seção Indústria, profissão e formação	13	0	6	79	23	2	27
Nenhum destes, mas na seção Fontes de informação, suportes e canais	2	0	13	194	47	3	30
Nenhum destes, mas na seção Tratamento da informação para serviços de informação	1	0	9	111	17	4	28
Nenhum destes, mas na seção Serviços técnicos em bibliotecas, arquivos e museus	0	0	7	101	7	7	34
Nenhum destes, mas na seção Técnicas de Armazenamento	0	0	1	5	1	0	3
Nenhum destes, mas na seção Tecnologia da informação e tecnologia de biblioteca	5	0	41	175	52	22	53
Organizações	0	0	22	167	28	1	26
Páginas da Web	0	0	29	322	46	1	56
Periódicos eletrônicos	5	0	55	490	155	1	112
Periódicos, Publicações	0	0	21	350	66	8	103
Planejamento, projeto, mudança	0	0	4	32	6	0	11
Política da informação	7	0	35	261	178	0	116
Portais	2	0	16	301	48	1	45
Preservação de documentos em papel	0	0	6	25	14	1	40
Preservação digital	7	0	45	310	47	9	80
Processamento automatizado de idiomas	0	0	9	75	10	0	13
Projeto, desenvolvimento, implementação e manutenção	1	0	19	182	33	0	43
Propriedade intelectual: os direitos de autor, propriedade, direitos autorais e copyleft	17	0	51	517	169	9	89
Publicação e questões legais	4	0	36	281	121	3	87
Recuperação automática de texto	0	0	9	98	6	0	15

ASSUNTOS	ÁF	AN	ÁS	EU	ANC	OC	AMS
Recursos eletrônicos	6	0	78	533	108	2	58
Recursos Financeiros	0	0	6	44	8	1	9
Rede de computadores	5	0	59	140	15	0	29
Relacionamento da Biblioteconomia e CI com outros campos	5	0	41	281	108	2	105
Relações Públicas	0	0	4	59	14	0	5
Reorganização	1	0	5	78	20	0	32
Repositórios (baseado ou não em OAI)	24	0	90	834	180	9	191
Representação do conhecimento	4	0	32	441	62	5	86
Scanners	0	0	2	4	3	0	2
Segurança	0	0	0	24	2	0	1
Segurança de computadores e redes	6	0	21	63	3	1	6
Seriados (gerenciamento)	0	0	6	50	26	1	25
Serviços técnicos em bibliotecas, arquivos e museus	3	0	40	283	92	2	101
Servidores on-line	0	0	8	57	8	0	5
Sistemas de automação de bibliotecas	5	0	73	159	60	0	57
Sistemas de Gerenciamento de Bancos de Dados orientados a objetos	0	0	3	3	0	0	6
Sistemas de gerenciamento de bases de dados	0	0	27	108	19	2	37
Sistemas de imagens	1	0	6	56	3	0	9
Sistemas de OPACs	3	0	23	163	24	1	25
Situações de calamidade	0	0	0	17	3	0	8
Sociedade da informação	2	0	32	295	116	2	194
Software	1	0	39	237	69	3	100
Técnicas de Armazenamento	0	0	12	27	13	0	5
Tecnologia da informação e tecnologia de biblioteca	20	0	231	803	264	7	269
Telecomunicações	4	0	32	93	29	0	14
Teoria da informação e teoria de bibliotecas	2	1	42	197	99	1	124
Transferência da informação: protocolos, formatos, técnicas	0	0	16	186	23	0	46
Tratamento da informação para serviços de informação	11	0	105	689	203	1	155
Treinamento	6	0	19	183	94	1	102
Treinamento de usuários, promoções, atividades, educação	4	0	31	417	135	8	108
Uso da informação e sociologia da informação	22	0	253	728	312	9	440

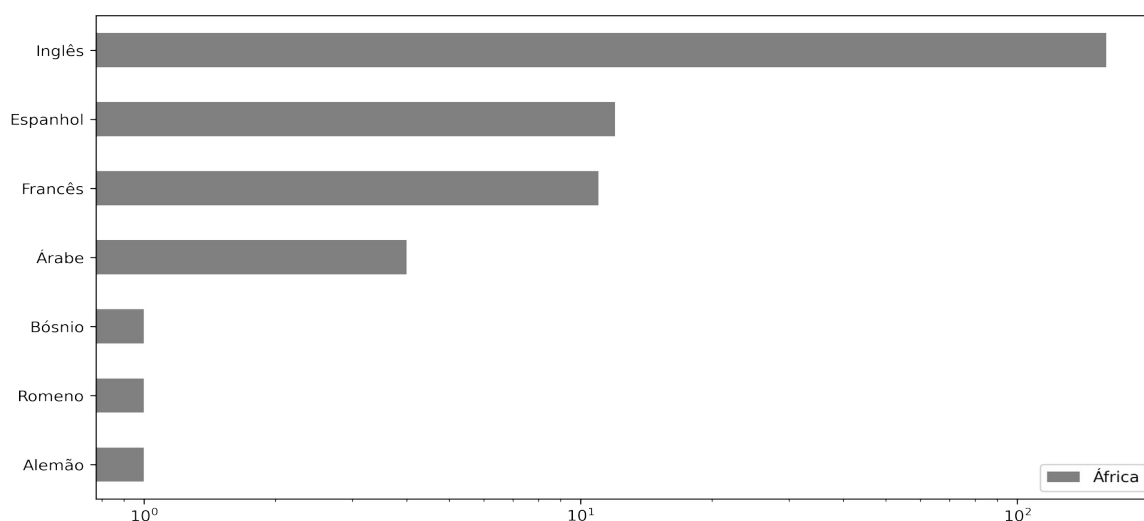
ASSUNTOS	ÁF	AN	ÁS	EU	ANC	OC	AMS
Uso e impacto da informação	5	0	106	609	199	9	227
Usuários, nível de instrução e leitura	7	0	67	377	111	0	242
Veículos	0	0	0	4	0	0	1
Venda de livros	1	0	9	47	10	0	3
Web Semântica	0	0	0	0	0	0	1

Fonte: dados da pesquisa (2020).

APÊNDICE E – Idioma das publicações por continente

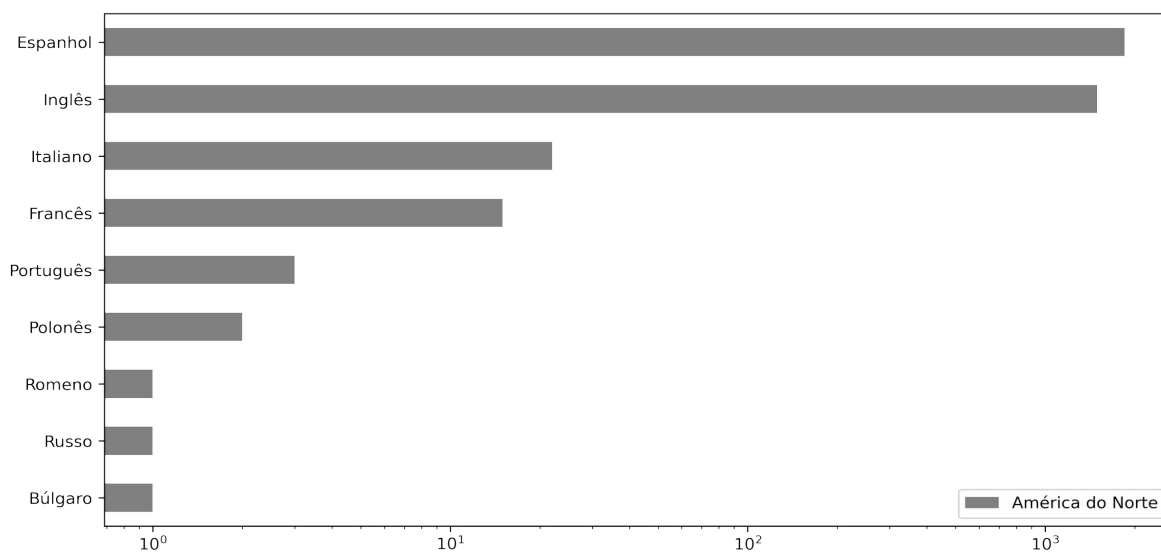
Aqui, são apresentados os gráficos sobre os idiomas das publicações de cada continente. Embora os dados já tenham sido apresentados no capítulo **Análise e discussão dos resultados**, os gráficos também são evidenciados em apêndice como exemplo de produto gerado pelo ELIScript.

Gráfico 29 – Principais idiomas de publicação da África



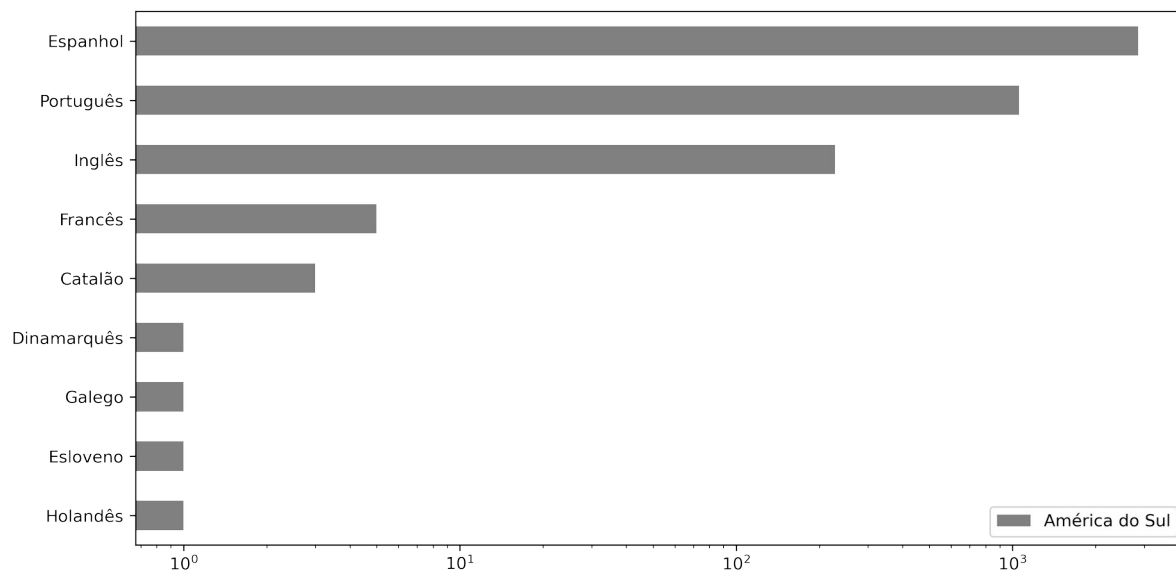
Fonte: dados da pesquisa (2020).

Gráfico 30 – Principais idiomas de publicação da América do Norte e Central



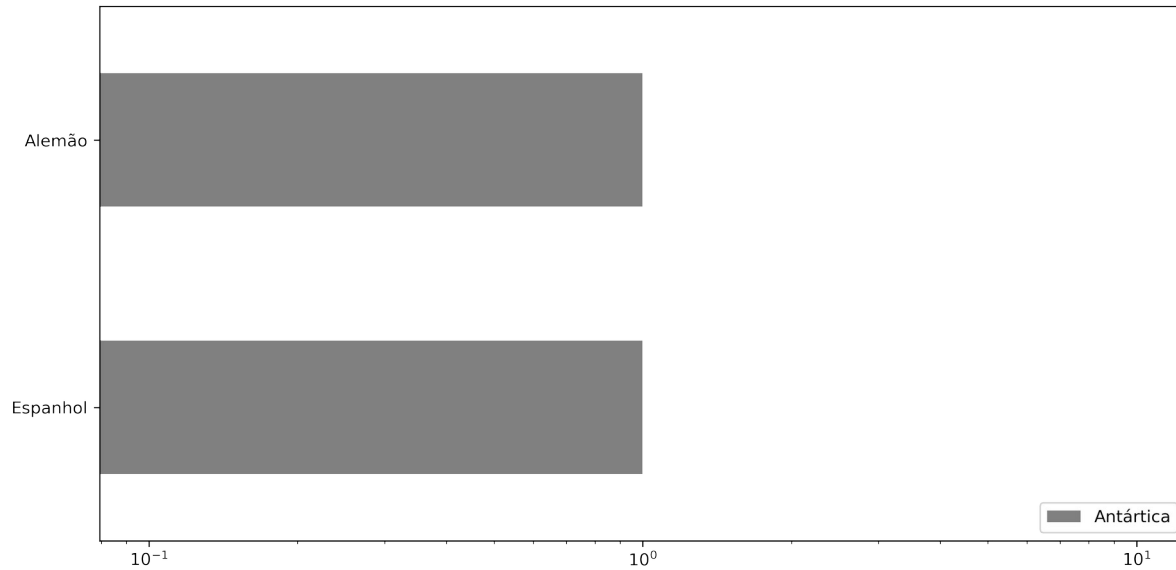
Fonte: dados da pesquisa (2020).

Gráfico 31 – Principais idiomas de publicação da América do Sul



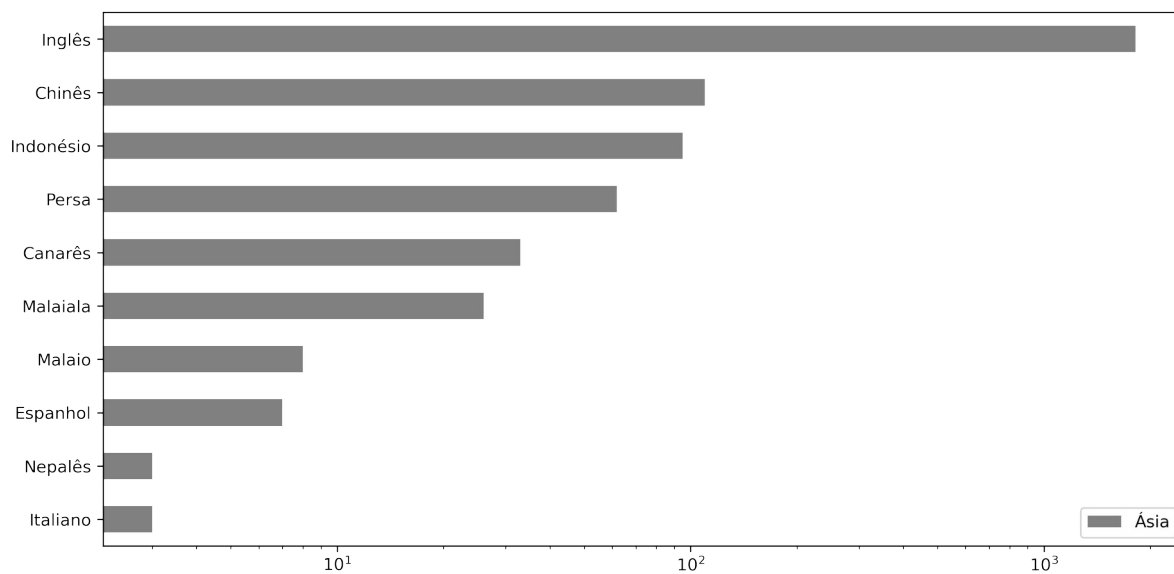
Fonte: dados da pesquisa (2020).

Gráfico 32 – Principais idiomas de publicação da Antártica



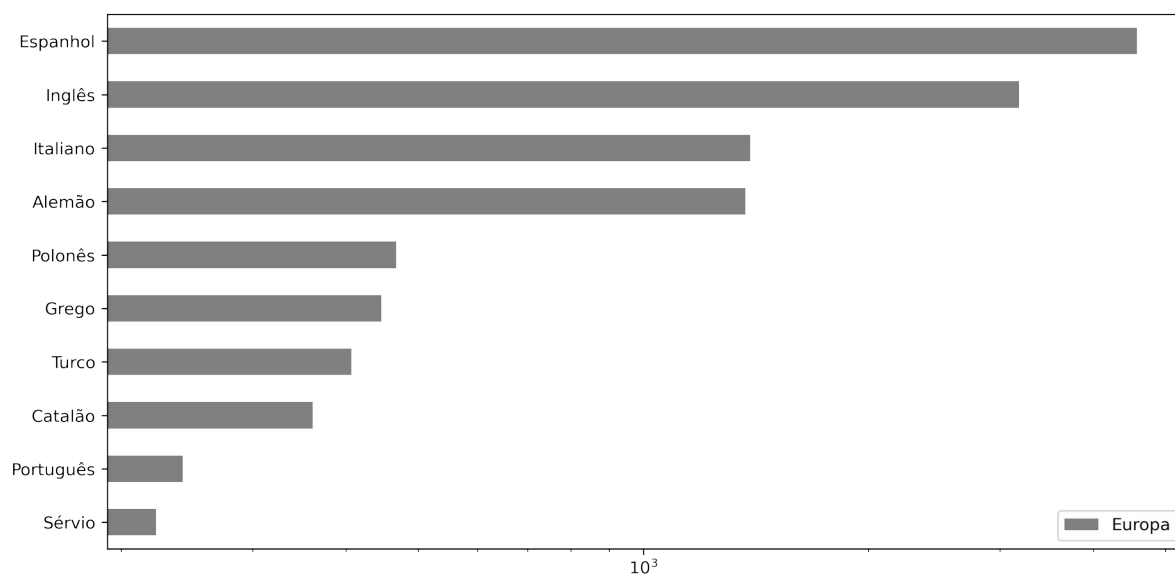
Fonte: dados da pesquisa (2020).

Gráfico 33 – Principais idiomas de publicação da Ásia



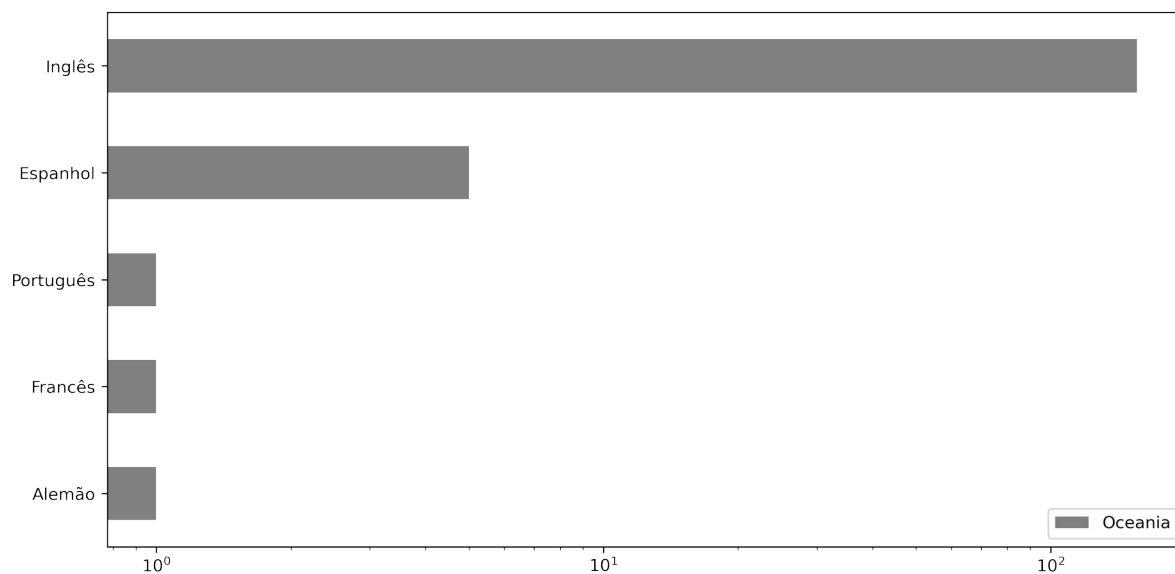
Fonte: dados da pesquisa (2020).

Gráfico 34 – Principais idiomas de publicação da Europa



Fonte: dados da pesquisa (2020).

Gráfico 35 – Principais idiomas de publicação da Oceania



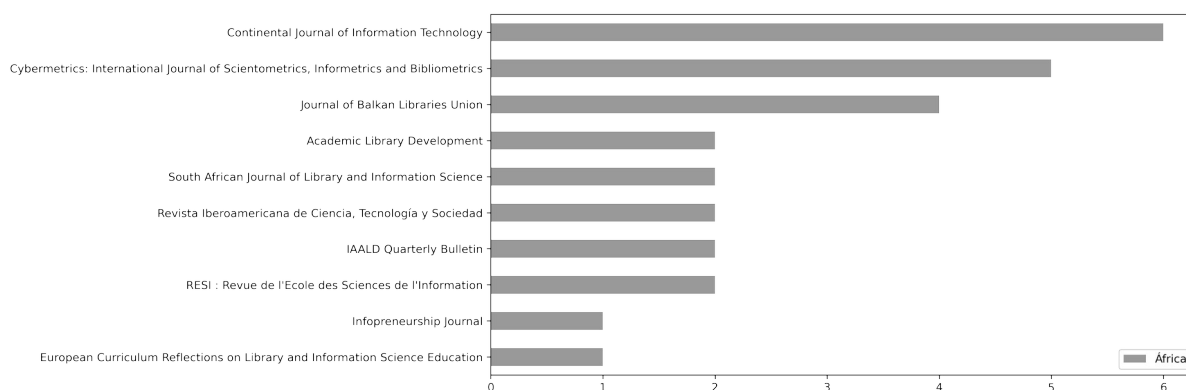
Fonte: dados da pesquisa (2020).

Optou-se por apresentar os dados em escala logarítmica para tornar os idiomas menos comuns mais perceptíveis nos gráficos. No entanto, no código do ELIScript é possível mudar essa opção e apresentar os valores numéricos inteiros. Outras opções de customização também são possíveis, como: mudar as cores das barras ou apresentá-las na disposição vertical ao invés de horizontal.

APÊNDICE F – Periódicos com mais publicações depositadas

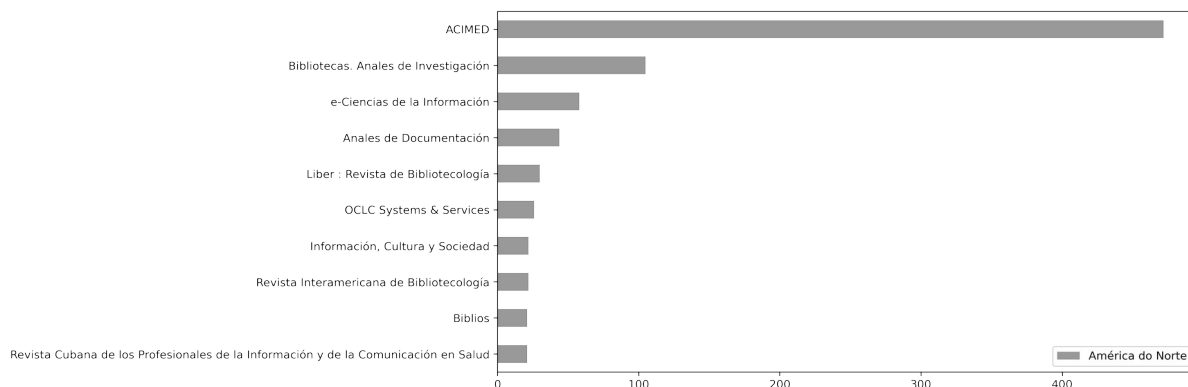
Conforme apontado anteriormente, os gráficos com os títulos dos periódicos foram gerados, mas não foram apresentados na discussão dos resultados. Optou-se por apresentar os dados em tabela. No entanto, como os gráficos também são produtos gerados pelo ELIScript, são apresentados aqui.

Gráfico 36 – Periódicos com mais documentos depositados na África



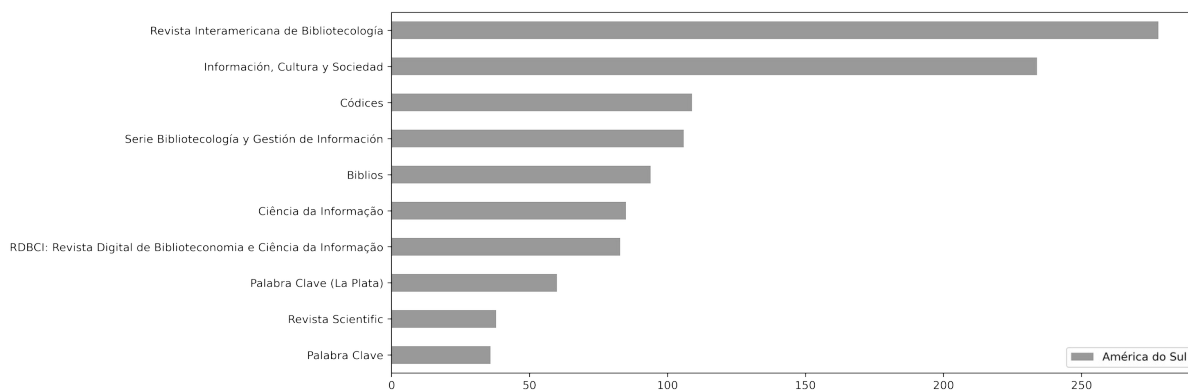
Fonte: dados da pesquisa (2020).

Gráfico 37 – Periódicos com mais documentos depositados na América do Norte e Central



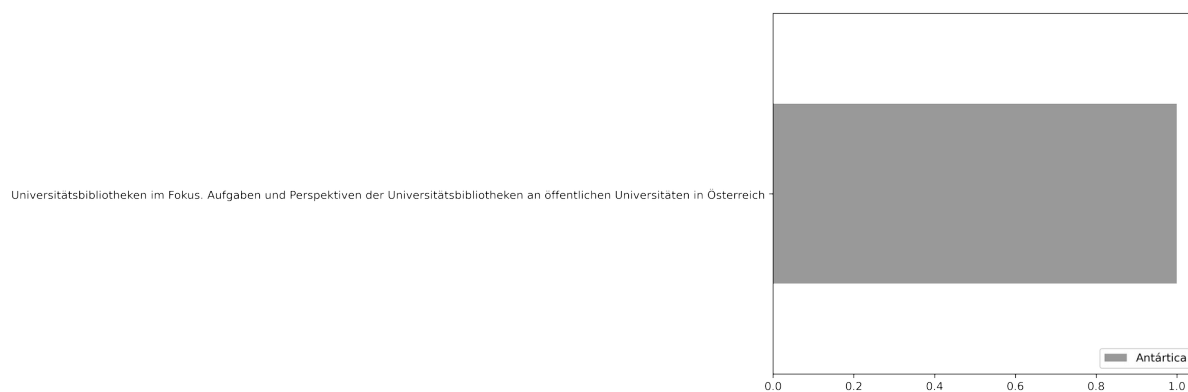
Fonte: dados da pesquisa (2020).

Gráfico 38 – Periódicos com mais documentos depositados na América do Sul



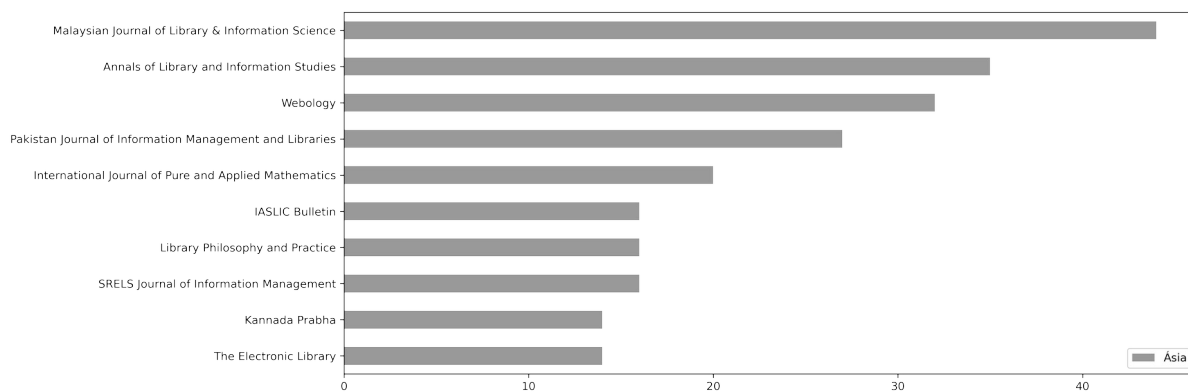
Fonte: dados da pesquisa (2020).

Gráfico 39 – Periódicos com mais documentos depositados na Antártica



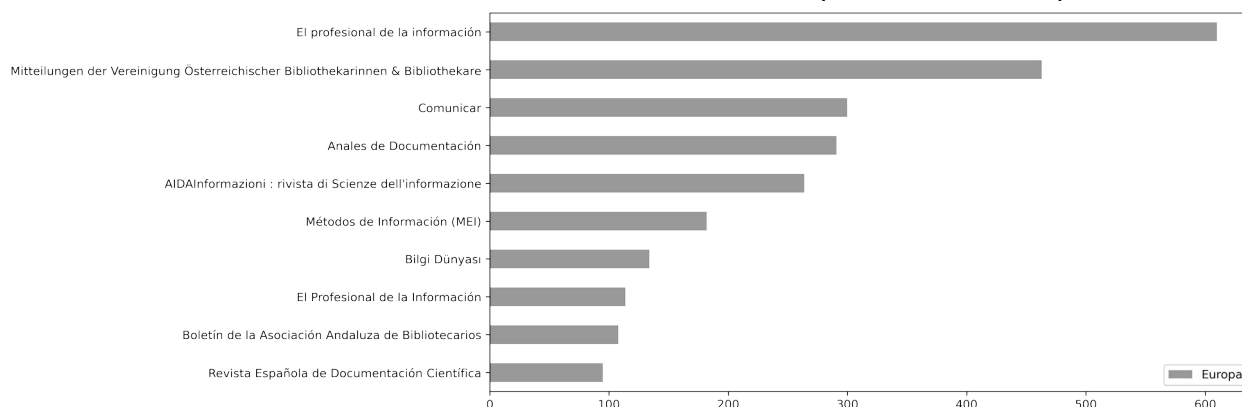
Fonte: dados da pesquisa (2020).

Gráfico 40 – Periódicos com mais documentos depositados na Ásia



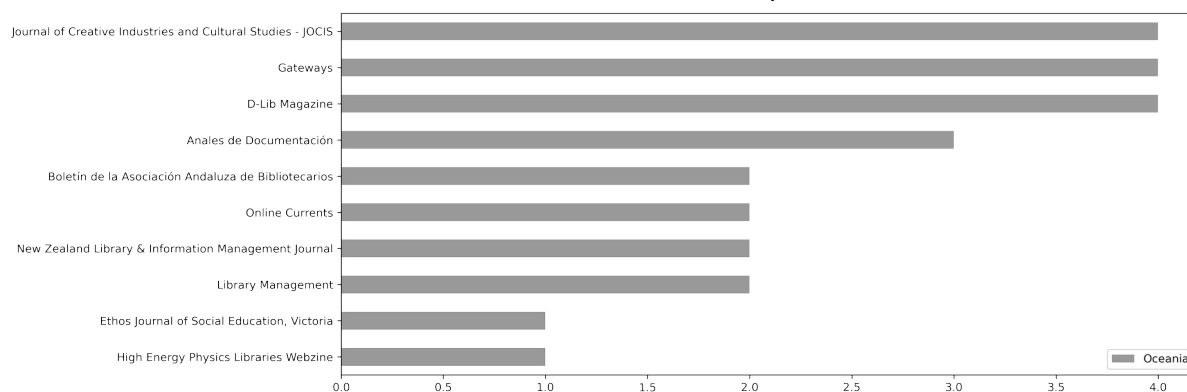
Fonte: dados da pesquisa (2020).

Gráfico 41 – Periódicos com mais documentos depositados na Europa



Fonte: dados da pesquisa (2020).

Gráfico 42 – Periódicos com mais documentos depositados na Oceania



Fonte: dados da pesquisa (2020).

Como os gráficos de idioma, estes também podem ser personalizados. Nota-se que as quantidades de publicações por periódicos são representadas em valores numéricos inteiros, pois as diferenças são perceptíveis nos gráficos.

ANEXO A – Esquema de classificação JITA, versão em português – Brasil

Cassandra Lúcia de Maya Viana

Tradução e adaptação

1. [Teorética e Generalidades]: nível geral

A. Aspectos teóricos e gerais de bibliotecas e informação

- AA. Biblioteconomia e Ciência da Informação (CI) como um campo.
- AB. Teoria da informação e teoria de bibliotecas.
- AC. Relacionamento da Biblioteconomia e CI com outros campos.
- AZ. Nenhum destes, mas nesta seção.

B. Uso da informação e sociologia da informação

- BA. Uso e impacto da informação.
- BB. Métodos bibliométricos.
- BC. Informação na sociedade.
- BD. Sociedade da informação.
- BE. Economia da informação.
- BF. Política da informação.
- BG. Disseminação e difusão da informação.
- BH. Necessidades de informação e análise de requisitos de informação.
- BI. Interfaces de usuário, usabilidade.
- BZ. Nenhum destes, mas nesta seção.

2. [Funcionalidades orientadas ao usuário, de direção e de gerenciamento]: nível intermediário (inclui questões sócio-econômicas e legais).

C. Usuários, nível de instrução e leitura

- CA. Estudos de uso.
- CB. Estudos de usuário.
- CC. Categorias de usuários: crianças, jovens, grupos sociais.
- CD. Treinamento de usuários, promoções, atividades, educação.
- CE. Instrução literária.
- CF. Leitura e narração de histórias.
- CZ. Nenhum destes, mas nesta seção.

D. Bibliotecas como coleções físicas

- DA. Bibliotecas Mundiais.
- DB. Bibliotecas Nacionais.

- **DC. Bibliotecas Públicas**⁷⁴.
- DD. Bibliotecas de Instituições de Ensino Superior.
- DE. Bibliotecas escolares.
- DF. Bibliotecas governamentais.
- DG. Bibliotecas Particulares.
- DH. Bibliotecas especializadas.
- DI. Bibliotecas científicas.
- DJ. Bibliotecas técnicas.
- DK. Bibliotecas de Saúde e Medicina.
- DL. Arquivos.
- DM. Museus.
- DZ. Nenhum destes, mas nesta seção.

E. Publicação e questões legais

- EA Meios de comunicação de massa.
- EB. Impressão, publicação eletrônica, transmissão.
- EC. Venda de livros.
- ED. Propriedade intelectual: os direitos de autor, propriedade, direitos autorais e *copyright*.
- ⁷⁵
- EE. Liberdade intelectual.
- EF. Censura.
- EZ. Nenhum destes, mas nesta seção.

F. Gerenciamento

- FA. Cooperação.
- FB. Marketing.
- FC. Finanças.
- FD. Relações Públicas.
- FE. Gestão de Recursos Humanos.
- FF. Recursos Financeiros.
- FG. Governo Local.
- FH. Reorganização.
- FI. Autoridades Centrais.
- **FJ. Gestão do Conhecimento**
- FZ. Nenhum destes, mas nesta seção.

G. Indústria, profissão e formação

- GA. Indústria da informação.
- GB. Indústria de software.
- GC. Indústria de computadores e telecomunicações.

74 Para completez dos dados, as notações que estão em negrito foram adicionadas pela autora da dissertação com base nas notações apresentadas no próprio repositório E-LIS, pois observou-se que a versão de Viana (2006) está desatualizada.

75 Direitos autorais, porém com termos que permitem a distribuição do produto.

- GD. Organizações.
- GE. Empregados.
- GF. Biografias.
- GG. Aspectos curriculares.
- GH. Formação.
- GI. Treinamento.
- GZ. Nenhum destes, mas nesta seção.

3. [Objetos, Aspectos Pragmáticos, Aspectos Técnicos]: nível específico

H. Fontes de informação, suportes e canais

- HA. Periódicos, Publicações.
- HB. Literatura cinzenta.
- HC. Documentos de arquivo.
- HD. Livros raros e manuscritos.
- HE. Impressos em geral.
- HF. Microformas.
- HG. Material não impresso, mimeografado, datilografado, manuscrito etc.
- HH. Audiovisual, Multimídia.
- HI. Mídia eletrônica.
- HJ. CD-ROM.
- HK. Servidores on-line
- HL. Bases de Dados e Redes
- HM. Catálogos on-line de acesso público (OPAC's).
- HN. Periódicos eletrônicos.
- HO. Livros eletrônicos.
- HP. Recursos eletrônicos.
- HQ. Páginas da Web.
- HR. Portais.
- HS. Repositórios (baseado ou não em OAI).
- **HT. Web 2.0, Redes Sociais**
- HZ. Nenhum destes, mas nesta seção.

I. Tratamento da informação para serviços de informação

- IA. Catalogação, controle bibliográfico.
- IB. Análise de conteúdo (indexação, resumos, classificação).
- IC. Linguagens de indexação, processos e esquemas.
- ID. Estruturas de dados e de metadados.
- IE. Representação do conhecimento.
- IF. Transferência da informação: protocolos, formatos, técnicas.
- IG. Apresentação da informação: hipertexto, hipermídia.
- IH. Sistemas de imagens.
- II. Filtros e sistemas de filtragem.
- IJ. Hiperlinks de referência.
- IK. Projeto, desenvolvimento, implementação e manutenção.
- **IL. Web Semântica**
- IZ. Nenhum destes, mas nesta seção.

J. Serviços técnicos em bibliotecas, arquivos e museus

- JA. Aquisição.
- JB. Seriados (gerenciamento)
- JC. Descarte.
- JD. Inventário.
- JE. Manutenção de registros (cadastros).
- JF. Preservação de documentos em papel.
- JG. Digitalização.
- JH. Preservação digital.
- JI. Circulação.
- JJ. Disseminação de Documentos.
- JK. Empréstimo interbibliotecário.
- JZ. Nenhum destes, mas nesta seção.

K. Técnicas de Armazenamento

- KA. Centros de recursos.
- KB. Edifícios de bibliotecas, arquivos e museus.
- KC. Mobiliário.
- KD. Veículos.
- KE. Arquitetura.
- KF. Planejamento, projeto, mudança.
- KG. Segurança.
- KH. Situações de calamidade.
- KZ. Nenhum destes, mas nesta seção.

L. Tecnologia da informação e tecnologia de biblioteca

- LA. Telecomunicações.
- LB. Rede de computadores.
- LC. Internet, inclusive WWW.
- LD. Computadores.
- LE. Scanners.
- LF. Câmeras digitais.
- LG. Fotocopiadoras.
- LH. Segurança de computadores e redes.
- LI. Autenticação e controle de acesso.
- LJ. Software.
- LK. Metodologia e engenharia de softwares.
- LL. Processamento automatizado de idiomas.
- LM. Recuperação automática de texto.
- LN. Sistemas de gerenciamento de bases de dados.
- LO. Sistemas de Gerenciamento de Bancos de Dados orientados a objetos.
- LP. Agentes inteligentes.
- LQ. Sistemas de automação de bibliotecas.
- LR. Sistemas de OPACs.
- LS. Mecanismos de busca.
- **LT. Dispositivos Móveis**
- LZ. Nenhum destes, mas nesta seção.