

Sarah Rúbia de Oliveira Santos

Universidade Federal de Minas Gerais
sarahrubia@ufmg.br

Dalgiza Andrade Oliveira

Universidade Federal de Minas Gerais
dalgizamg@gmail.com

Universidade Federal de Minas
Gerais

Correspondência/Contato
Av. Antônio Carlos, 6627
Pampulha: 31270-901
BELO HORIZONTE - MG

Escola de Ciência da Informação da UFMG

AUTOARQUIVAMENTO NA CIÊNCIA DA INFORMAÇÃO

Uma análise dos documentos depositados no repositório digital e-LIS

RESUMO

Discute as práticas de autoarquivamento na Ciência da Informação e as contribuições do repositório *E-prints in Library and Information Science* para a comunicação científica aberta na área. O estudo é quantitativo-descritivo, de caráter exploratório, e utiliza-se a bibliometria como aporte metodológico para mensuração dos indicadores científicos. Os instrumentos de coleta e análise de dados foram construídos utilizando a linguagem de programação *Python*. Cerca de 34 mil documentos foram arquivados no repositório entre 2002 e 2019. As tipologias de documentos mais frequente são artigo de periódico (47%), artigo de conferência (18%), apresentação (8%), capítulo de livro (6%). Os assuntos mais abordados são “Uso da informação e sociologia da informação”, “Tecnologia da informação e tecnologia de biblioteca” e “Tratamento da informação para serviços de informação”. Os resultados são parciais e representam uma visão geral dos documentos depositados no repositório.

Palavras-Chave: Comunicação Científica, Movimento de Acesso Aberto, Autoarquivamento, e-LIS, Bibliometria.

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 bibliometrics 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 “Information Use and Sociology of Information”, “Information Technology and Library Technology” and “Information Treatment for Information Services”. The results are partial and represent an overview of documents deposited in the repository.

Keywords: Scientific communication, Open Access Movement, Self-archiving, e-LIS, Bibliometrics.

1. INTRODUÇÃO

A crise dos periódicos na década de 1980, resultado dos altos preços cobrados para manutenção de assinaturas por bibliotecas universitárias, e os canais eletrônicos, como um canal alternativo para a comunicação científica, propiciaram a emergência de um movimento que busca disponibilizar o conhecimento científico sem restrições para aquele que dele precisar.

O Movimento de Acesso Aberto emerge diante dessas inquietações a respeito do modelo de negócios que vigora no mercado editorial de periódicos científicos. Dessa forma, ele propõe a disponibilização da produção científica de forma gratuita, on-line e sem restrições. Para isso, propõe duas estratégias: a Via Dourada e a Via Verde (SUBER *et al.*, 2003).

A Via Dourada decorre da publicação de um artigo em um periódico de acesso aberto, assim, o artigo estará disponível desta forma imediatamente após a publicação. A Via Verde, por sua vez, acontece por meio do autoarquivamento realizado pelo autor em um repositório digital (BAPTISTA *et al.*, 2007). O autoarquivamento pode ser feito em repositórios digitais institucionais e temáticos. O foco deste trabalho está na Via Verde e na prática de autoarquivamento em repositórios digitais temáticos.

A aceitação de repositórios temáticos como o *arXiv*¹, com mais de um milhão e meio de publicações, e a relativa ausência de trabalhos sobre repositórios temáticos (BERNIUS, 2010; ADAMICK; REZNIK-ZELLEN, 2007; BJÖRK, 2014), especificamente sobre o *E-prints in Library and Information Science (e-LIS)*, motivaram o interesse em analisar as características do autoarquivamento nesse repositório e as consequências dessa atividade para a comunicação científica aberta em Ciência da Informação.

Na medida em que mais de 34 mil documentos foram depositados nesse repositório no decorrer de 16 anos, e essas publicações juntas acumulam mais de 16 milhões de *downloads*, pressupõe-se que esse material tem alguma representação na área de Ciência da Informação.

Em vista disso, o presente estudo busca contribuir para os debates acerca do Movimento de Acesso Aberto, a partir da discussão sobre o que é e como funciona a estratégia de autoarquivamento em repositórios digitais temáticos. Logo, pergunta-se:

¹ O *ArXiv*, primeiro repositório temático na Internet, foi criado em 1991 por Paul Ginsparg. Atualmente, ele é hospedado pela Universidade Cornell, em Nova Iorque nos Estados Unidos, e fornece mais de um milhão e meio de documentos eletrônicos em acesso aberto nos campos de Física, Matemática, Ciência da Computação, Biologia Quantitativa, Estatística, entre outros.

como se dão as práticas de autoarquivamento e quais as contribuições do repositório *e-LIS* para a comunicação científica aberta em Ciência da Informação? Tal questionamento envolve não só o estudo do repositório *e-LIS*, em si, mas também da coleção de materiais disponíveis nele e do comportamento dos pesquisadores frente a esse espaço.

2. REPOSITÓRIOS DIGITAIS TEMÁTICOS

Nos espaços em que os pesquisadores lidam de forma mais direta com os obstáculos da publicação em periódicos científicos pagos e com a cobrança para acesso a publicações, nota-se que há uma busca por maneiras de possibilitar o compartilhamento do saber produzido com recursos públicos. Entende-se que disponibilizar esse conhecimento pode ser considerado um dever para o pesquisador e, para a sociedade, o acesso a esse saber é um direito.

Apesar das estratégias do Movimento de Acesso Aberto, de acordo com Khabsa e Giles (2014), apenas cerca de 2% de todos os documentos científicos disponibilizados na *Web* estão em acesso aberto. Tennant *et al.* (2016) apontam que, mesmo com todos esses documentos disponíveis abertamente, um problema central permanece: cerca de 75% dos artigos restantes não estão diretamente disponíveis a menos que o indivíduo esteja em uma posição privilegiada e possua acesso por meio de sua instituição, isto é, que tenha a assinatura do periódico ou de uma base de dados. De toda forma, é improvável que alguém, alguma universidade ou instituto de pesquisa tenha acesso às assinaturas de todos os periódicos revisados por pares existentes.

Neste contexto, os repositórios temáticos ou disciplinares se caracterizam por fornecerem os meios para o armazenamento e disseminação de informação digital. Faz-se necessário frisar que estes se diferenciam dos repositórios institucionais por possuírem coleções de documentos pertencentes a uma disciplina (ou conjunto de disciplinas) científica específica.

No entendimento de Björk (2014), os repositórios mais interessantes fornecem acesso livre ao texto completo de publicações acadêmicas e permitem que estas estejam localizáveis ou pesquisáveis na *Web*, o que é o caso do *e-LIS*. Bernius (2010), sugere que os repositórios temáticos fornecem as melhores condições para recuperação do conhecimento científico. 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, os pesquisadores.

3. O AUTOARQUIVAMENTO NA CIÊNCIA DA INFORMAÇÃO

Mas como funciona esse autoarquivamento? Mais do que apenas depositar o trabalho em um repositório, o autor deve estar ciente de seus direitos sobre o manuscrito e o artigo, uma vez estes poderão ser negociados com as editoras.

Algumas editoras, por exemplo, propõem um acordo para a transferência dos direitos autorais do autor para a editora e, então, proibem 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 determinado tempo, o autor disponibilize seu artigo em um serviço de sua preferência.

Esses acordos têm a função de definir o que o autor pode ou não fazer com o manuscrito no qual o artigo é baseado. 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, alguns autores afirmam que na área de Biblioteconomia e Ciência da Informação (BCI) os atritos entre autor e editora não parecem ser tão fortes como em outras áreas (COLEMAN, 2007; EMERY, 2018; CHILIMO; ONYANCHA, 2018). Em consonância com Coleman (2007), em seu estudo sobre o autoarquivamento e a existência de acordos de transferência de direitos autorais em periódicos de BCI, 90% de 52 periódicos classificados na Thomson-Scientific (ISI)², não proíbe, de nenhuma maneira, o autoarquivamento.

Em outro recorte, Emery (2018) analisa periódicos de BCI da editora Taylor e Francis, que permite o auto-depósito de artigos publicados em suas revistas. Nele, a autora descobriu que a taxa de depósito em BCI é de aproximadamente 22%. Para Emery (2018), esse é um valor abaixo do esperado (50%), visto que a própria editora incentiva esse depósito e até realizou campanhas nesse sentido.

² 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.

Chilimo e Onyancha (2018), ao investigarem o grau de abertura da pesquisa em Acesso Aberto na área de BCI, 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 BCI disponível e, se cada autor depositasse seu artigo nele, os resultados poderiam chegar próximo à marca de 50%. De acordo com Emery (2018), o campo de BCI 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 se pese o exposto acima, entende-se que um repositório aberto de BCI já existe.

4. METODOLOGIA

O estudo é quantitativo-descritivo, de caráter exploratório, e a pesquisa bibliométrica será utilizada como aporte metodológico para mensuração dos indicadores científicos. Segundo Oliveira (2018, p. 54), entre os diversos aspectos que podem ser analisados, os indicadores científicos “evidenciam os pesquisadores, as instituições, as temáticas, as áreas do conhecimento, assim 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”.

O ambiente de pesquisa é o *e-LIS*, um repositório digital de documentos técnico-científicos para as disciplinas relacionadas à Ciência da Informação (Figura 1). Para Santillan-Aldana (2009), o *e-LIS* é o primeiro servidor eletrônico especializado em BCI e está estabelecido como o maior repositório aberto da área. A amostra constitui-se de metadados de 34.534 documentos, subdivididos entre os sete continentes: África, América do Norte (e Central), América do Sul, Antártica, Ásia, Europa e Oceania.

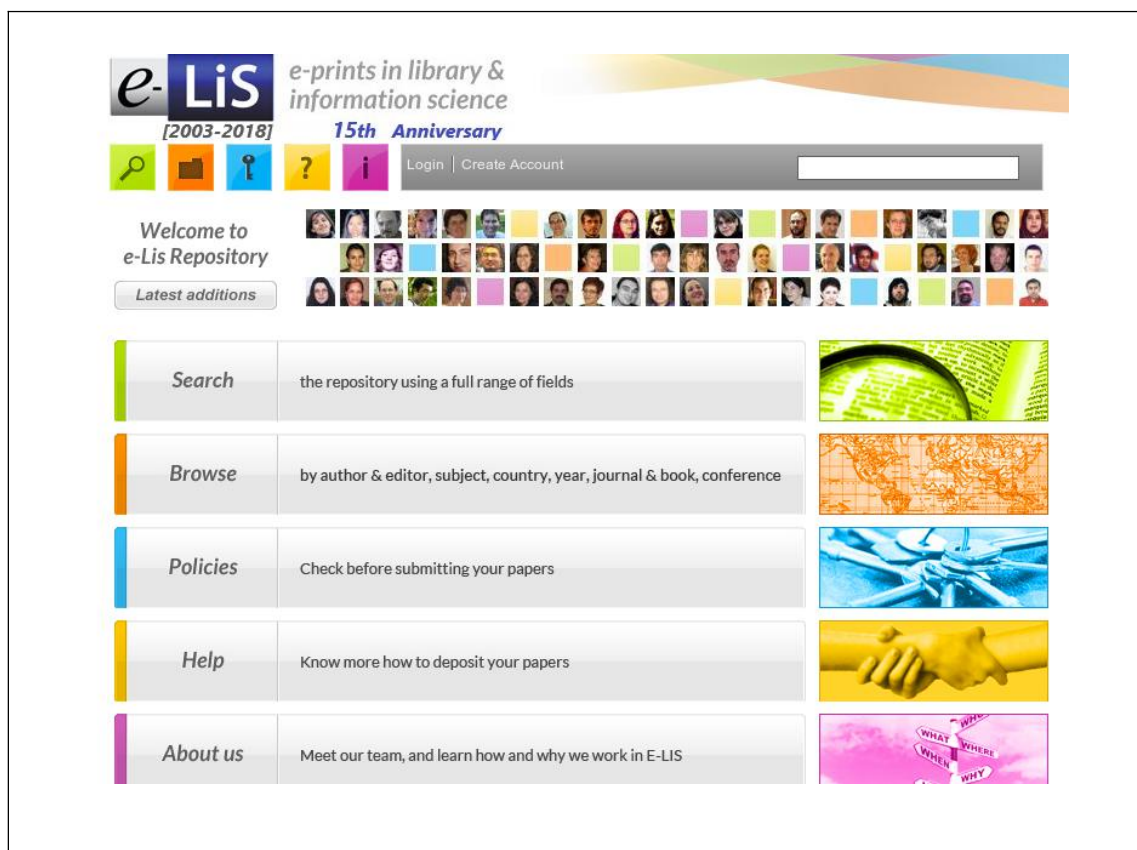


Figura 1. Página inicial do repositório e-LIS

A coleta de dados foi realizada por meio da técnica denominada *Scrapping*, que significa extrair de forma automática dados da *Web*. Com a linguagem de programação *Python 3*, os metadados foram exportados no formato de planilhas CSV (arquivos de texto delimitado que usam vírgula para separar valores (em inglês: *comma-separated values* - CSV). Como ambiente computacional para a programação, foi utilizado o *Jupyter Notebook*, uma ferramenta de código aberto que possui uma lista ordenada de células de entrada/saída, podendo conter código de programação, texto (usando *Markdown*, linguagem simples de marcação), matemática, gráficos e texto enriquecido, o que possibilitou a leitura dos dados bem como sua representação.

Dessa forma, foi possível prosseguir para a análise, ler o conteúdo das planilhas, extrair as informações relevantes e processá-las de forma a construir um entendimento acerca do material coletado.

5. RESULTADOS PRELIMINARES

Nesta seção apresenta-se alguns resultados preliminares referentes à análise geral da amostra coletada. Tenha-se em vista que os resultados aqui, expostos, são parte de uma pesquisa de mestrado em andamento.

O período registrado em que houve depósito de documentos no repositório compreende os anos de 2002 (ano de criação do repositório) até 2019. 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 uma análise quantitativa (Figura 2).

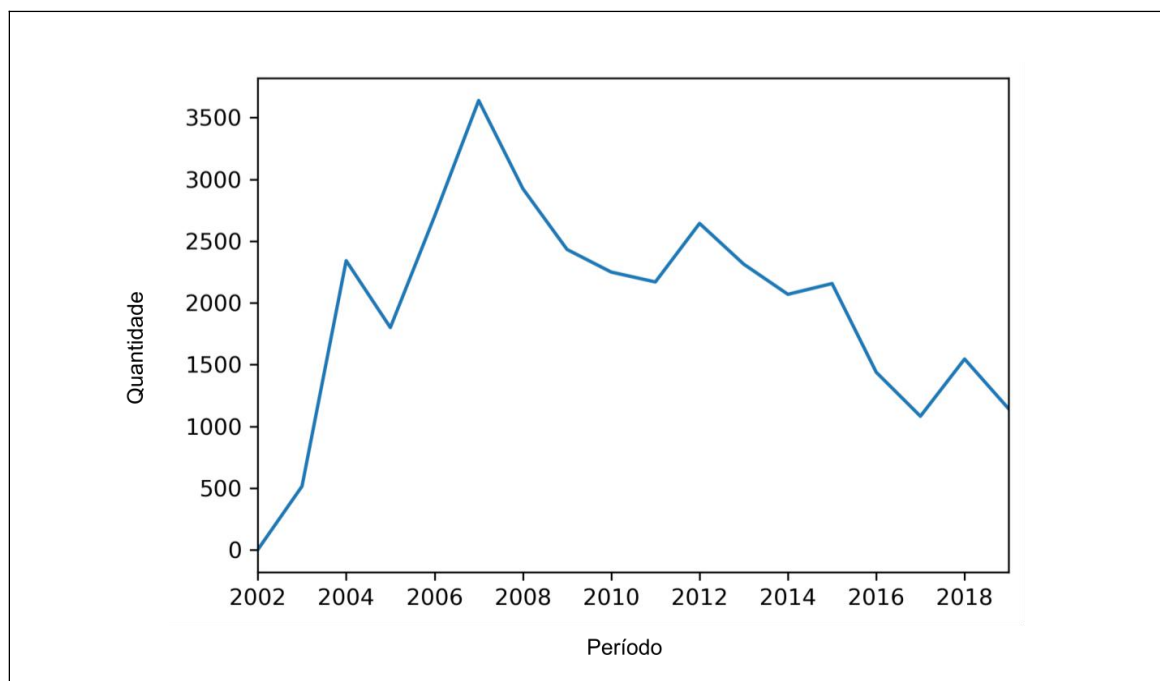


Figura 2. Quantidade de depósitos por ano.

Em 2007, há um pico na quantidade de depósitos chegando a ultrapassar 3,5 mil documentos autoarquivados no mesmo ano. A partir desse período, no entanto, há um declínio na quantidade de depósitos, com picos em 2012, 2015 e 2018. No começo do *e-LIS*, os editores que eram responsáveis pela manutenção e acompanhamento do repositório realizavam, também, atividades de divulgação, convidando os pesquisadores em eventos a fazerem depósitos e utilizarem a ferramenta. Sendo assim, nota-se que o crescimento no número de documentos depositados não foi orgânico e isso pode ter desencadeado o declínio nos anos subsequentes.

Apesar de “artigo” ser o tipo mais comum de documento a ser depositado, diferentes tipologias de documentos podem ser encontradas no *e-LIS*. Na Figura 3 são apresentados os quantitativos de cada tipologia que constam no repositório:

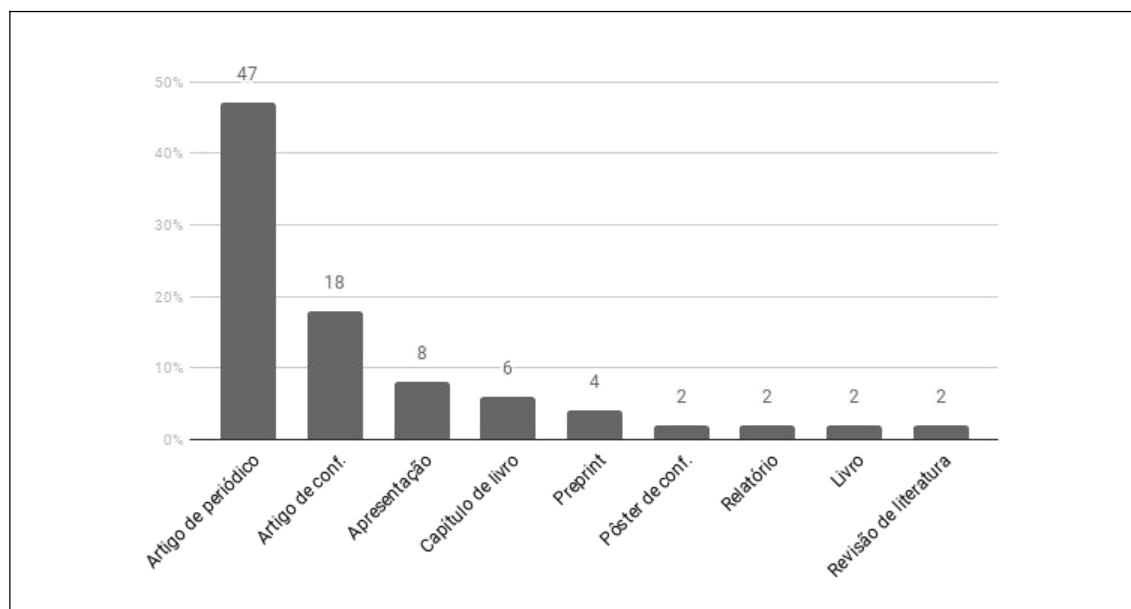


Figura 3. Tipologia dos documentos depositados no *e-LIS*.

De maneira geral, observa-se que dos 34.534 documentos, os tipos mais depositados são: artigo de periódico (47%), artigo de conferência (18,4%), apresentação/*slides* (8,3%), capítulo de livro (5,7%), manuscrito pré-publicação/*preprint* (4,2%), tese e dissertação (3,1%), pôster de conferência (2,4%), relatório (2%), livro (1,7%) e revisão de literatura (1,6%). Os outros tipos de documento, a saber: bibliografias, conjunto de dados (*dataset*), relatórios técnico-departamentais, guias e manuais, material instrucional de biblioteca, matérias de jornal ou revista, *anais*, planos de negócios, relatórios técnicos e tutoriais tiveram menor representação, com menos de 1,5% cada.

Quanto ao conteúdo dos documentos, no *e-LIS* estes têm seus assuntos categorizados de acordo com a classificação JITA, uma classificação simples, alfabética, com dois níveis hierárquicos. Seu nome 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). Os assuntos quando utilizados para classificar não são exclusivos, isto é, um mesmo documento pode ser classificado em mais de um assunto. Por meio desse esquema de classificação, foi possível observar quais os assuntos dos documentos depositados no repositório (Tabela 1). As letras do alfabeto antes do assunto são as notações da classificação.

Tabela 1. Tipologia dos documentos depositados no *e-LIS*

ASSUNTO	QUANTIDADE
B. Uso da informação e sociologia da informação	2329
L. Tecnologia da informação e tecnologia de biblioteca	2294
I. Tratamento da informação para serviços de informação	1801
H. Fontes de informação, suportes e canais	1745
A. Aspectos teóricos e gerais de bibliotecas e informação	1611
C. Usuários, nível de instrução e leitura	1124
F. Gerenciamento	1044
G. Indústria, profissão e formação	1015
E. Publicação e questões legais	780
J. Serviços técnicos em bibliotecas, arquivos e museus	758
D. Bibliotecas como coleções físicas	684
K. Técnicas de Armazenamento	74

Assim, os assuntos mais abordados são “Uso da informação e sociologia da informação” (2.329 documentos), “Tecnologia da informação e tecnologia de biblioteca” (2.294 documentos) e “Tratamento da informação para serviços de informação” (1.801 documentos). A partir dessas temáticas observa-se 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, também se faz presente nos documentos depositados.

6. CONSIDERAÇÕES FINAIS

As práticas de autoarquivamento em repositórios digitais possibilitam o armazenamento e a preservação do corpo de conhecimento de uma área. O *e-LIS* pode se caracterizar, nesse sentido, como uma ferramenta para preservação do conhecimento produzido na área de Biblioteconomia e Ciência da Informação. Todavia, suas características e contribuições para a área ainda estão sendo investigadas como parte de uma pesquisa de mestrado em andamento.

Os resultados apresentados neste trabalho são parciais e representam uma visão geral do autoarquivamento de trabalhos por pesquisadores da Ciência da Informação, bem como dos documentos contidos no repositório. Como passos futuros no desenvolvimento da dissertação, os dados de cada continente serão analisados de

forma separada, bem como de maneira comparativa. Outros indicadores bibliométricos também serão considerados, tais como os países de origem dos autores, as instituições às quais eles estão filiados, o tipo de autoria (única ou co-autoria), palavras-chave utilizadas, os assuntos dos documentos de maneira mais específica, entre outros. Dessa forma, busca-se compreender as práticas de autoarquivamento na área, assim como o que constitui o acervo do repositório *e-LIS*.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- ADAMICK, J.; REZNIK-ZELLEN, R. Representation and recognition of subject repositories. **D-Lib Magazine**, 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.
- BAPTISTA, A. A. et al. Comunicação científica: o papel da Open Archives Initiative no contexto do acesso livre. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, 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.
- BERNIUS, S. The impact of open access on the management of scientific knowledge. **Online Information Review**, v. 34, n. 4, p. 583-603, 2010. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/14684521011072990>. Acesso em: 21 maio 2019.
- BJÖRK, B. Open Access Subject Repositories: An Overview. **Journal of the American Society for Information Science**, v. 65, n. 4, p. 698-706, 2014. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23021>. Acesso em: 27 maio 2019.
- CHILIMO, W. L.; ONYANCHA, O. B. How open is open access research in Library and Information Science? **SA Jnl Libs & Info Sci**, v. 84, n. 1, p. 11-19, 2018. DOI: 10.7553/84-1-1710. Disponível em: <http://sajlis.journals.ac.za/pub/article/view/1710/1488>. Acesso em: 27 maio 2019.
- COLEMAN, A. Self-archiving and the copyright transfer agreements of ISI-ranked Library and Information Science journals. **JASIST**, v. 58, n. 2, p. 286-296, 2007. DOI: 10.1629/uksg.444. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20494>. Acesso em: 27 maio 2019.
- DE ROBBIO, A; KATZMAYR, M. The management of an international open access repository: the case of E-LIS. **GMS Medizin-Bibliothek-Information**, v. 9, n. 1, 2009. Disponível em: <http://hdl.handle.net/10760/13220>. Acesso em: 24 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**, v. 31, p. 1-9, 2018.

Disponível em: <https://insights.uksg.org/articles/10.1629/uksg.406/>. Acesso em: 27 maio. 2019.

KHABSA, M.; GILES, C. L. The number of scholarly documents on the public web. **PLoS one**, v. 9, n. 5, p. 1-6, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4015892/>. Acesso em: 19 ago. 2019.

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

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

NEVES, B.; FERREIRA, C. Caracterização da produção científica portuguesa em Ciência da Informação disponibilizada em acesso aberto no e-LIS. **Cadernos BAD**, 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.

OLIVEIRA, E. F. T. **Estudos Métricos da Informação no Brasil**. São Paulo: Cultura Acadêmica, 2018.

SANTILLAN-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**, v. 25, n. 2, p. 135-147, maio. 2009. Disponível em: <https://www.emeraldinsight.com/doi/full/10.1108/10650750910961938>. Acesso em: 24 maio. 2019.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, 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.

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.

TENNANT, J. P. *et al.* The academic, economic and societal impacts of Open Access: an evidence-based review. **F1000Research**, v. 5, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4837983/>. Acesso em: 19 ago. 2019.

Sarah Rúbia de Oliveira Santos

Mestranda no Programa de Pós-graduação em Gestão e Organização do Conhecimento, Universidade Federal de Minas Gerais (PPGGOC/UFMG), Belo Horizonte, MG, Brasil. Graduada em Biblioteconomia pela Universidade Federal de Alagoas (UFAL), Maceió, AL, Brasil.
Lattes: <http://lattes.cnpq.br/2150772913464368>

Dalgiza Andrade Oliveira

Professora Adjunta na Escola de Ciência da Informação e no Programa de Pós-Graduação em Gestão da Organização e do Conhecimento da Universidade Federal de Minas Gerais (PPGGOC/UFMG), Belo Horizonte, MG, Brasil. Doutora em Ciência da Informação, Universidade

Federal de Minas Gerais (PPGCI/UFMG), Belo Horizonte, MG, Brasil.

Lattes: <http://lattes.cnpq.br/2803696275869287>