

DADOS DE PESQUISA DA CIÊNCIA DA INFORMAÇÃO PRESENTES NO REPOSITÓRIO ZENODO

Marcia Regina da Silva

Doutora em Educação
Faculdade de Filosofia, Ciências e Letras
Universidade de São Paulo (USP)
E-mail: marciaregina@usp.br

Edneia Silva Santos-Rocha

Doutora em Política Científica e Tecnológica
Faculdade de Direito
Universidade de São Paulo (USP)
E-mail: edneia@usp.br

Fernanda Cristina da Silva Rosa Martins

Graduação em Biblioteconomia e Ciência da Informação
Universidade de São Paulo (USP)
E-mail: fernanda.cristina.martins@usp.br

INTRODUÇÃO

A produção e editoração de dados de pesquisas tomaram outros direcionamentos com os princípios FAIR – Findability, Accessibility, Interoperability, and Reusability –, publicados em 2016 na revista *Scientific Data*, da Nature. (WILKINSON et al., 2016) Tratam-se de orientações que contribuem para melhorar a capacidade das máquinas encontrarem dados de pesquisas, apoiando a reutilização desses dados por outros indivíduos, além disso, contribuem para aumentar o valor agregado das publicações digitais.

Dados de pesquisa são registros que subsidiam resultados de pesquisas publicados, esses dados de pesquisas em sua forma tangível incluem, por exemplo, “fatos, observações, entrevistas, gravações, medições, experimentos, simulações e softwares; dados numéricos, descritivos e visuais; brutos, limpos e processados”. (VAN BERCHUM; GROOTVELD, 2017, p. 5) Esses dados são difíceis de serem organizados e estruturados para a reutilização, justamente por isso, a gestão estratégica de dados passou a ser uma prática importante na rotina do pesquisador. Além da publicação de dados em apêndices de periódicos, os pesquisadores podem publicar seus dados em repositório específicos de uma área ou em repositórios multidisciplinares como Figshare, Dryad e Zenodo, que seguem os Princípios FAIR.

Algumas pesquisas a respeito de dados abertos de pesquisa no escopo da Ciência da Informação já foram desenvolvidas (ESTEVÃO; ARNS; STRAUHS, 2019; SEMELER; PINTO, 2019), no entanto, a caracterização dos dados indexados em repositórios que seguem os princípios FAIR ainda precisa ser mais bem explorada. Esses repositórios se tornaram importantes fontes de informação para a consulta e para o monitoramento das áreas e disciplinas em relação ao compartilhamento e reutilização de dados de pesquisa. Embora haja maior conscientização dos pesquisadores quanto à publicação em periódicos de acesso aberto e disponibilização de dados de pesquisa, nem toda comunidade acadêmica tem incorporado essas práticas no processo de produção e divulgação científica. Nesse sentido, o objetivo desta pesquisa é caracterizar os dados de pesquisa da Ciência da Informação indexados no repositório Zenodo, como vistas a buscar subsídios para futuras discussões no que se refere ao comportamento da área da Ciência da Informação em relação ao compartilhamento de dados de pesquisa.

O repositório Zenodo foi lançado em 2013 com a proposta de compartilhar resultados de pesquisas de instituições menores publicados em diversos formatos, tornando-se uma ferramenta interessante de curadoria de dados de pesquisa. Apesar de não possuir mecanismos de controle dos dados enviados, permite a formação de comunidades que podem criar e selecionar conteúdos, fomentando a própria validação do conteúdo disponível. (ZENODO, 2020) A escolha do Zenodo, que é um repositório multidisciplinar e sem custo para depósito, deve-se justamente ao interesse em verificar a aderência da Ciência da Informação no compartilhamento de dados fora dos repositórios da própria área.

COMPARTILHAMENTO DE DADOS DE PESQUISA

O movimento Open Science, ou Ciência Aberta, tem como cerne a adoção de mecanismos que possibilitem que o processo de pesquisa e de sua divulgação se torne mais transparente e acessível. Dentro do escopo de discussões que permeiam esse movimento encontram-se conceitos, ferramentas, plataformas e meios que promovam a transparência do processo de produção e divulgação científica. Transparência no sentido de acesso tanto aos dados de pesquisa como aos resultados da pesquisa, de forma livre, aberta e inclusiva, possibilitando a obtenção de maiores benefícios. (RIGOLIN; SILVA, 2019)

Nos limitaremos aqui a tratar dos dados abertos ou *open data*, prática que vem ganhando maior atenção entre os pesquisadores. Os dados abertos de pesquisa significam que devem estar disponíveis para todas as pessoas acessarem, usarem e até compartilharem. Neste seguimento, não há entraves para dados sem licenças ou direitos autorais ou mesmo patentes. Já os dados FAIR são aqueles que atendem os padrões de Encontrabilidade, Acessibilidade, Interoperabilidade e Reusabilidade.

Os repositórios de dados de pesquisas são responsáveis pelo arquivamento e pela curadoria de conjuntos de dados, resultados legítimos de pesquisas. Contribuem para a

preservação de diversos tipos de dados, juntamente com suas metadescrições e citações de suporte. (SANCHEZ-ALONSO, 2017; SICILIA; GARCIA-BARRIOCANAL) Essas fontes de dados são importantes porque corroboram para maior preocupação dos pesquisadores quanto à qualidade dos dados e resultados das pesquisas divulgadas.

METODOLOGIA

Este trabalho caracteriza-se como descritivo e exploratório. A busca foi realizada no mês de janeiro de 2020. O Zenodo não permite a realização da busca avançada, o que prejudica o refinamento, por isso, o termo de busca foi *Information Science*, com a utilização das aspas para a restrição da busca. A escolha dessa expressão pode ter limitado os resultados. No entanto, entende-se que a caracterização deles poderá embasar estudos sobre essa temática no campo da Ciência da Informação. Foram recuperados 2.173 registros, importados individualmente para uma planilha do Excel. Após checagem dos registros, tomou-se a decisão de considerar todos para a análise. Para essa comunicação buscou-se caracterizar a tipologia desses recursos, a tipologia das publicações, o formato do arquivo e as palavras-chave.

RESULTADOS

Aproximadamente 99% dos documentos são de acesso aberto, ou seja, podem ser acessados sem restrição. Os dados revelam que publicações representam a maioria dos recursos informacionais sobre Ciência da Informação disponibilizados no Zenodo (n=1.928). Também são disponibilizados softwares (n=65), apresentações (n=60), dataset (n=59), pôster (n=23), imagens (n=22), anotações (08), vídeo (n=01) e outros (n=07). Observa-se que apenas 11,2% dos registros podem ser caracterizados especificamente como dados de pesquisa, revelando a necessidade de explorar e considerar a natureza das pesquisas da área.

A maioria dos arquivos encontra-se no formato PDF. Esse fato possivelmente justifica-se pela maioria dos documentos ser formada por publicações (n=1.722). Mesmo com baixa frequência, pontua-se o aparecimento de outras tipologias como relatórios, trabalho em eventos (n=133), pré-publicações (n=15), partes de documentos (n=12), teses (n=12), produto ou resultado de um processo (entregável) (n= 8), documentos preliminares científicos ou técnicos (n=8), relatórios (n=6) e outros (n=5). Observa-se que o compartilhamento de documentos que subsidiam a pesquisa nas diversas fases denotam possível engajamento de alguns pesquisados na prática de compartilhamento de dados.

Embora o conteúdo desses dados de pesquisa não tenha sido objeto da caracterização dos dados, destacam-se as palavras-chave que tiveram maior frequência: Biodiversidade (n=92), Taxonomia (n=39), Digitalização (n=35), Sistema Global de Informação sobre a Biodiversidade (GBIF) (n=29), Informática em Biodiversidade (n=27), Coleções (n=27),

Ontologia (n=27), Darwin Core (n=25) e Ciência Cidadã (n=25). Vale ressaltar a necessidade de investigar os registros que foram representados com as palavras Taxomia e Ontologia, temas específicos da Ciência da Informação como forma de identificar quais dados de pesquisas subsidiaram o desenvolvimento dessas pesquisas.

CONCLUSÃO

O compartilhamento de dados de pesquisa é uma prática que vem sendo estimulada pela comunidade científica. Algumas áreas estão mais engajadas dentro do sistema de gestão de dados científicos, outras como a Ciência da Informação, ainda carecem de maiores discussões e iniciativas. Os resultados apresentados aqui denotam que os artigos científicos ainda prevalecem entre os recursos informacionais da área que são indexados no Zenodo. No entanto, o fato de dados de pesquisas da Ciência da Informação estarem presentes nesse repositório pode apontar para a atenção dessa comunidade com as práticas oriundas da Ciência Aberta. O repositório Zenodo ainda precisa ser melhor estruturado para tornar-se uma fonte importante para os estudos métricos, tendo em vista o processo de busca e exportação dos dados.

REFERÊNCIAS

- ESTEVÃO, J. S. B.; ARNS, E. M.; STRAUHS, F. R. Gestão de dados de pesquisa. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 17, p. 1-26, 2019. Disponível em: <https://doi.org/10.20396/rdbci.v17i0.8656239>. Acesso em: 10 maio 2020.
- RIGOLIN, C. C. D.; SILVA, M. R. Sistema de revisão por pares na ciência. In: CARNEIRO, F. F. B.; FERREIRA NETO, A.; SANTOS, W. (org.). *A comunicação científica em periódicos*. Curitiba: Appris, 2019. p. 95-118.
- SEMLER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. *Ciência da Informação*, Brasília, DF, v. 48, n. 1, p. 113-129, 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 10 maio 2020.
- SICILIA, M. A.; GARCIA-BARRIOCANAL, E.; SANCHEZ-ALONSO, S. Community curation in open dataset repositories: insights from zenodo. *Procedia Computer Science*, Red Hook, v. 106, p. 54-60, 2017. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050917302776>. Acesso em: 10 jan. 2020.
- VAN BERCHUM, M.; GROOTVELD, M. Research data management: an overview of recent developments in the Netherlands. *Data Archiving and Networked Services*, Amsterdam, p. 1-27, 2017. Disponível em: https://dans.knaw.nl/en/about/organisation-and-policy/information-material/Whitepaper_ResearchdatamanagementAnoverview_DEF.pdf. Acesso em: 1 jun. 2020.
- WILKINSON, M.; DUMONTIER, M.; MONS, B. et al. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, London, v. 3, n. 160018, 2016. Não Paginado. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 10 mar. 2020.
- ZENODO. [Geneva], 2020. Disponível em: <https://zenodo.org/>. Acesso em: 10 mar. 2020.