

INDEXAÇÃO DE DADOS CIENTÍFICOS: uma análise a partir das políticas da Rede de Repositórios de Dados Científicos do Estado de São Paulo

INDEXING OF SCIENTIFIC DATA: an analysis based on the policies of the Network of Scientific Data Repositories of the State of São Paulo

Joyce Mirella dos Anjos Viana | Paula Regina Dal'Evedove

<https://doi.org/10.21747/21836671/pagnesppk26>

Resumo: Os repositórios de dados científicos são uma realidade vivenciada no mundo todo, contribuindo com o armazenamento, preservação e acesso aos dados provenientes de pesquisa científica. Em atenção ao importante papel que esses sistemas contemporâneos desempenham, objetiva-se investigar a indexação de dados científicos no âmbito da Rede de Repositórios de Dados Científicos do Estado de São Paulo, plataforma que permite acesso aos dados científicos e amplia a visibilidade das pesquisas conduzidas no âmbito das instituições participantes. Para tanto, são analisadas as políticas de informação estabelecidas por cada instituição integrante, no intuito de contribuir com os estudos de representação da informação em repositórios digitais. Trata-se de um estudo exploratório e documental, com coleta de dados realizada a partir de consulta aos sítios eletrônicos dos repositórios de dados científicos e instituições integrantes. A análise das políticas de informação identificadas revela que os repositórios de dados vinculados à Rede atendem parcialmente os princípios FAIR, utilizam *softwares* e padrões que permitem a interoperabilidade, possuem sistemas de identificação dos dados e protocolo para a coleta de metadados. A indexação dos dados científicos nos repositórios integrantes ocorre pelo próprio autor-pesquisador ou pela equipe responsável pelo sistema. Constata-se a necessidade de estudos mais aprofundados na Rede de Repositórios de Dados Científicos do Estado de São Paulo, com destaque para a qualidade do metadado assunto e as especificidades das políticas de indexação em repositórios de dados científicos.

Palavras-chave: Dados científicos; Indexação; Rede de Repositórios de Dados Científicos do Estado de São Paulo; Repositórios de dados científicos.

Abstract: The repositories of scientific data are a reality experienced worldwide, contributing to the storage, preservation and access to data from scientific research. In view of the important role that these contemporary systems play, the objective is to investigate the indexing of scientific data within the scope of the Scientific Data Repositories Network of the State of São Paulo, a platform that allows access to scientific data and increases the visibility of research conducted in the participating institutions. To this end, the information policies established by each member institution are analyzed in order to contribute to studies of information representation in digital repositories. This is an exploratory and documentary study, with data collection carried out by consulting the electronic sites of the scientific data repositories and member institutions. The analysis of the identified information policies reveals that the data repositories linked to the Network partially meet the FAIR principles, use software and standards that allow interoperability, have data identification systems and a protocol for collecting metadata. The indexing of scientific data in the member repositories occurs by the author-researcher himself or by the team responsible for the system. There is a need for more in-depth studies on the Network of Scientific Data Repositories of the State of São Paulo, with emphasis on the quality of the subject metadata and the specificities of the indexing policies in scientific data repositories.

Keywords: Scientific data; Indexing; Network of Scientific Data Repositories of the State of São Paulo; Scientific data repositories.

Introdução

Os dados científicos, também denominados de dados de pesquisa ou dados de investigação abertos, recebem atenção especial nos últimos tempos por atuarem como recursos informacionais estratégicos para o avanço da pesquisa científica. Dados científicos “muito rapidamente deixam de ser meros subprodutos das atividades de pesquisa e se tornam um foco de grande interesse para todo o mundo científico”, especialmente para as universidades e agências de fomento à pesquisa que passam a atuar como protagonistas na produção e na disseminação do conhecimento, além de grandes produtoras e consumidoras desses recursos informacionais (SAYÃO e SALES, 2016:91).

A fim de pesquisadores e instituições terem acesso aberto aos dados associados aos resultados científicos e usarem as informações de ciência com maior eficiência, plataformas especializadas estão em funcionamento e/ou desenvolvimento em todo o mundo. Os repositórios de dados estão sincronizados com o movimento de Acesso Aberto, atuando na gestão e na definição de políticas que estabeleçam as melhores práticas para a curadoria digital de dados científicos, contribuindo com uma pesquisa científica cada vez mais colaborativa, aberta e transparente. Walport e Brest (2011) compreendem que os repositórios de dados científicos configuram-se como uma estratégia eficiente para a organização, preservação e compartilhamento dos dados científicos.

Nesta perspectiva, investiga-se a indexação de dados científicos no âmbito da Rede de Repositórios de Dados Científicos do Estado de São Paulo mediante análise das políticas de informação estabelecidas por suas instituições integrantes, no intuito de contribuir com a representação da informação em repositórios digitais. A Rede disponibiliza em uma plataforma aberta os dados provenientes de pesquisas desenvolvidas nas mais diversas áreas do conhecimento. Por meio da plataforma é possível ter acesso aos dados usados em pesquisas científicas, independentemente de sua publicação em artigos científicos.

Procedimentos metodológicos

Pesquisa de caráter documental, empregando abordagem qualitativa para análise dos dados e apresentação dos resultados. A coleta das políticas de informação disponíveis para consulta nos *Websites* dos repositórios de dados associados à Rede e no sítio eletrônico da própria instituição foi realizada em março de 2020.

A iniciativa da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) com a plataforma foi forjada a partir de um movimento internacional de disponibilização de dados científicos de agências de fomento públicas e privadas da América do Norte, Austrália, Grã-Bretanha, Holanda, Alemanha e países escandinavos. Trata-se, portanto, de uma iniciativa prática para a gestão dos dados científicos no Brasil, sintonizada com as práticas de Acesso Aberto. Lançada em 16 de dezembro de 2019, a plataforma reúne dados científicos de oito instituições do Estado de São Paulo, a saber: Embrapa Informática Agropecuária (CNPTIA/EMBRAPA); Instituto Tecnológico de Aeronáutica (ITA); Universidade de São Paulo (USP); Universidade Estadual de Campinas (UNICAMP); Universidade Estadual Paulista (UNESP); Universidade Federal de São Carlos (UFSCar); Universidade Federal de São Paulo (UNIFESP); e Universidade Federal do ABC (UFABC).

Segundo Medeiros (2019), cada instituição integrante da rede desenvolveu seu próprio repositório de dados científicos e grupos permanentes internos para o gerenciamento e compartilhamento dos dados. A integração dos repositórios é viabilizada por um portal único que busca e disponibiliza informações de forma integrada. O portal de acesso, um buscador de metadados, foi desenvolvido pela Universidade de São Paulo a partir do *software* livre DSpace. A plataforma de acesso e os repositórios de dados associados foram criados para atender a Política de Gestão de Dados da FAPESP, compreendendo a relevância do gerenciamento e compartilhamento de dados científicos para o avanço da ciência e da tecnologia, racionalização de recursos, facilidade da reprodutibilidade da pesquisa, além do treinamento de novos pesquisadores e exploração de aspectos não previstos no projeto original.

Resultados e discussão

Em atenção ao objetivo delineado, constata-se que a maior parte das políticas analisadas trazem considerações sobre os dados científicos e a figura do autor-pesquisador. Ainda, que os repositórios de dados vinculados à Rede atendem, mesmo que parcialmente, os princípios FAIR ao fazerem uso de *softwares* e padrões que permitem a interoperabilidade, sistemas de identificação dos dados e protocolo para a coleta de metadados. A UFSCar, a UNESP e a UNIFESP são as únicas instituições que permitem o autoarquivamento dos dados científicos.

Na USP, a Resolução 7.900, de 11 de dezembro de 2019, estabelece normas para a gestão de dados científicos da instituição, considerando os princípios da Ciência Aberta e a necessidade de coleta, armazenamento, gerenciamento, compartilhamento e preservação desses valiosos recursos informacionais, em atenção às exigências de agências de fomento e periódicos para o acesso aberto aos dados e visibilidade da produção científica. A resolução estabelece como condições para o armazenamento de dados científicos o preenchimento dos metadados de maneira adequada pelo autor-pesquisador responsável e adequação aos aspectos éticos e legais envolvidos na questão. Neste caso o autor-pesquisador faz a solicitação de inserção dos dados científicos e metadados no repositório de dados ao Grupo Gestor de Dados Científicos da instituição, o qual fará a análise segundo critérios pré-estabelecidos. Após a análise do grupo gestor a biblioteca realiza uma curadoria dos metadados propostos, seguidos de inclusão e publicação no repositório de dados da instituição. A indexação dos dados científicos é realizada no momento em que o pesquisador envia seu material para o grupo, por meio do preenchimento dos metadados. Assim, o responsável pela atribuição dos termos no metadado assunto é o próprio autor-pesquisador do documento.

A UFSCar adota o Manual de Autodepósito de Dados de Pesquisa que orienta o autoarquivamento de pesquisas científicas produzidas na instituição, sendo dividido em duas seções que contemplam o depósito de dados de pesquisa com e sem embargo. Os dados disponibilizados no repositório institucional compreendem todos os tipos de recursos informacionais produzidos em qualquer fase do ciclo de vida da pesquisa, viabilizando o acesso, preservação e o uso em longo prazo. O autor-pesquisador é quem insere os dados científicos e metadados no repositório institucional. A representação por assunto no manual é contemplada nos metadados de descrição e palavras-chave, nos quais o autor-pesquisador tem a obrigatoriedade de informar a finalidade, natureza e escopo do

conjunto de dados científicos, assim como as principais palavras-chaves representativas. Outro metadado contemplado, porém não obrigatório, é a descrição do arquivo no qual o autor-pesquisador deve fornecer uma contextualização geral do mesmo.

No momento da coleta dos dados, a UFABC indicou estar na fase de elaboração da política de uso do repositório de dados. A universidade está se mobilizando para promover a abertura dos seus dados institucionais, sendo instituído o Plano de Dados Abertos 2020-2022 pela Portaria nº 451. O documento se configura como orientador para as ações de transparência ativa das bases de dados institucionais de natureza administrativa, não contemplando diretamente os dados científicos.

Seguindo na mesma linha, a UNIFESP desenvolveu o seu Plano de Dados Abertos 2018-2019, que estabelece as suas estratégias para abertura de dados científicos, definindo um conjunto de ações para viabilizar a prática de dados abertos na instituição. A mesma, pela Resolução nº 178, dispõe sobre a Política de Gestão de Dados Estratégicos Institucionais que estabelece objetivos, princípios, diretrizes e governança dos dados, ainda em fase de elaboração. Como instrumento para auxiliar os usuários do repositório de dados, a instituição criou os Termos Gerais de Uso do Repositório de Dados de Pesquisa UNIFESP. O termo prevê regras de conduta, *upload* pelo usuário, licenças de uso, citações e um vocabulário de termos utilizados no texto. Ficam concedidas ao repositório de dados todas as permissões e licenças necessárias para o arquivamento, preservação e acesso ao conteúdo depositado, sem restrição ou permissão para disseminar cópias, promover e divulgar o conteúdo, armazenar, traduzir, copiar ou reformatar, incorporar metadados ou documentação. O depositante não precisa ser o autor-pesquisador, devendo fornecer informações sobre os dados científicos, incluindo, mas não limitado, ao nome do autor, data de publicação, título do conteúdo de dados, descrição do conteúdo e outras informações relacionadas.

A Política de Governança de Dados, Informação e Conhecimento da EMBRAPA e o Termo de Uso da Base de Imagens de Sintomas de Doenças de Plantas dispõem sobre os princípios, diretrizes, atribuições e responsabilidades para a gestão dos dados científicos e divulgação de informações relevantes na empresa. A política contempla os dados científicos, dados abertos, dados administrativos e dados pessoais, estes produzidos no exercício das funções da instituição, sendo gerenciados como ativos corporativos. A gestão dos dados científicos está alinhada com os objetivos da alta administração da empresa. O usuário tem a garantia da confidencialidade de seus dados de navegação no sistema. Na versão atual do repositório de dados da EMBRAPA, a responsabilidade pela inserção dos dados científicos, gestão e curadoria dos metadados é da instituição.

As instituições UNICAMP e ITA não disponibilizam informações e políticas de gestão de dados científicos em seus respectivos repositórios. Por seu turno, a UNESP está atuando no desenvolvimento de guias, termos, materiais de apoio e políticas de acesso, conforme informações presentes em seu repositório de dados.

Considerações finais

Os repositórios digitais tem se consolidado como bases de dados de alto valor científico, confiáveis, estáveis e acessíveis. A Rede de Repositórios de Dados Científicos do Estado de São Paulo é uma iniciativa pioneira para o compartilhamento e preservação de dados

científicos, oriundos das pesquisas conduzidas no Brasil. Os dados científicos disponíveis nesta plataforma necessitam de um ecossistema que possibilite o gerenciamento adequado e que favoreça as melhores práticas de representação e recuperação por assunto.

Nos repositórios de dados analisados, a indexação dos dados científicos ocorre com a inserção dos itens pelo próprio autor-pesquisador, em que a relação com o sistema é direta (UFSCar, UNESP e UNIFESP); ou mediante o envio dos dados para a equipe responsável, a qual realiza a inserção nos repositórios de dados (EMBRAPA, ITA, UFABC, UNICAMP e USP). No primeiro caso, alguns repositórios de dados solicitam contribuições de seus pesquisadores para armazenar e representar, por meio de metadados, os assuntos dos dados científicos. A contribuição permitida é apenas do pesquisador responsável. No segundo, existe a confluência de duas percepções – a do autor-pesquisador com os metadados da pesquisa e a do profissional responsável pela inserção, representação e curadoria dos dados no sistema. Nesta o processo de indexação não se configura exclusivamente como unidirecional, mas parte do princípio de colaboração entre os envolvidos, contribuindo com a construção de metadados mais representativos.

Para o avanço do tema, considera-se oportuno que os esforços iniciais conduzidos neste estudo sejam ampliados com pesquisas que observem a qualidade do metadado assunto no âmbito da Rede de Repositórios de Dados Científicos do Estado de São Paulo. Além disso, observações direcionadas para as políticas de indexação e a exploração das iniciativas de representação colaborativa em repositórios de dados científicos são oportunas.

Referências bibliográficas

MEDEIROS, C. B.

2019 *FAPESP lança Rede de Repositórios de Dados Científicos do Estado de São Paulo: Agência FAPESP: 20 dez. 2019: Entrevista concedida a Maria Fernanda Ziegler.* [Em linha]. 2019. [Consult. 20 dez. 2019]. Disponível em: <http://agencia.fapesp.br/fapesp-lanca-rede-de-repositorios-de-dados-cientificos-do-estado-de-sao-paulo/32251/>.

SAYÃO, L. F.; SALES, L. F.

2016 Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação.* [Em linha]. 21:2 (2016) 90-115. [Consult. 20 dez. 2019]. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>.

WALPORT, M.; BREST, P.

2011 Sharing research data to improve public health. *The Lancet.* 377:9.765 (2011) 537-539.

Joyce Mirella dos Anjos Viana | joyce@estudante.ufscar.br

Universidade Federal de São Carlos (UFSCar), Brasil

Paula Regina Dal'Evedove | dalevedove@ufscar.br

Universidade Federal de São Carlos (UFSCar), Brasil