

ANÁLISE DE CARACTERÍSTICAS DE REPOSITÓRIOS MULTIDISCIPLINARES DE DADOS CIENTÍFICOS ABERTOS¹

Vitória Maria Hartmann Caitano², Elaine Rosangela de Oliveira Lucas³.

¹ Vinculado ao projeto “Curatore Brasil: custódia e gestão de dados abertos de pesquisa”

² Acadêmica do Curso de Biblioteconomia FAED/UDESC – bolsista PROBIC/UDESC

³ Orientadora, PPGInfo, Depto. de Biblioteconomia FAED/UDESC – lani@udesc.br

Esse resumo é resultado parcial da pesquisa realizada até o final de julho de 2021, vinculada ao projeto de pesquisa intitulado ‘Curatore Brasil: custódia e gestão de dados abertos de pesquisa’. Teve como objetivo desenvolver um estudo com o mapeamento e descrição de características de alguns repositórios de dados científicos abertos e multidisciplinares.

Nele, foi destacada a importância do compartilhamento de dados científicos, que pode resumir-se na possibilidade de acelerar o desenvolvimento de novos conhecimentos, aumentar a eficiência dos recursos, e ampliar as possibilidades de reuso e de reproduzibilidade de pesquisas. Além disso, apresentou-se um conjunto de informações com características de repositórios de dados científicos abertos, com o objetivo de auxiliar os pesquisadores na escolha por um repositório para depósito dos dados resultantes das suas pesquisas.

Acerca da metodologia aplicada, o estudo teve caráter exploratório e foi realizado por meio de pesquisa bibliográfica e documental. Durante o período, foram analisados os repositórios: DANS¹, DataHub², Dataverse³, Dryad⁴, Figshare⁵ e Zenodo⁶. A definição destes foi realizada a partir de um artigo publicado pela Agência USP de Gestão da Informação Acadêmica (AGUIA) que aborda a mesma temática, citando estes seis repositórios como alternativas para o depósito de dados de pesquisa.

Partindo de categorias de análise previamente definidas, o estudo foi realizado a partir de informações coletadas nos sites oficiais de cada um desses repositórios. São elas: fundação e responsabilidade pelo manutenção do repositório; idiomas da plataforma e dos *datasets* aceitos, mídias sociais e possibilidade de compartilhamento direto de conteúdo; tipos e formatos de arquivos aceitos, limites de uso, curadoria de conteúdo; identificadores persistentes de conteúdo, de autores e informações acerca de financiamento de pesquisa; controle de versões; métricas de uso; integração com outros aplicativos e custos.

Quanto aos métodos de busca, as análises foram realizadas majoritariamente a partir das informações disponíveis nas abas ‘Sobre’ e ‘Perguntas Frequentes’. Na inexistência dessas abas ou na inviabilidade de respostas a partir delas, foram feitos levantamentos no conteúdo do site como um todo. Apesar de não terem sido feitos testes de preenchimento com *datasets* experimentais, em alguns repositórios houve a necessidade de realização de cadastro, com a finalidade de obter mais informações acerca da visualização dos campos de preenchimento.

Sobre as dificuldades encontradas, além das barreiras linguísticas (todos os repositórios têm o inglês como idioma primário), houve dificuldades na obtenção de alguns dados,

¹ <https://dans.knaw.nl/en>

² <https://datahub.io/>

³ <https://dataverse.org/>

⁴ <https://datadryad.org/stash>

⁵ <https://figshare.com/>

⁶ <https://zenodo.org/>

principalmente no que se refere a ano de criação do repositório, idiomas de publicação aceitos e identificação de financiamento de pesquisa.

Após o mapeamento e caracterização dos repositórios analisados neste estudo, foi possível constatar que as possibilidades de escolha no momento de optar por um repositório para depositar os dados científicos de uma pesquisa são muito amplas, além da variedade de dados já disponíveis ser bastante diversificada. Com isso, o projeto ainda continuará a ser desenvolvido pelo próximo ano, com o desenvolvimento de mais estudos até a conclusão da pesquisa.

Palavras-chave: Dados Abertos. Repositórios de dados. Ciência Aberta.