



Pecha Kucha

DOI: [10.21680/2447-7842.2023v9n2ID33622](https://doi.org/10.21680/2447-7842.2023v9n2ID33622)

Panorama dos repositórios de dados de pesquisa em Geociências

Overview of research data repositories in Geosciences

Roberta Pereira da Silva de Paula ¹

Submetido em: 17/04/2023	Aprovado na ConfOA: 14/06/2023	Publicado em: 04/12/2023
--------------------------	--------------------------------	--------------------------

Resumo: Os repositórios de dados de pesquisa são uma ferramenta para disseminação e preservação da produção institucional, cuja premissa é permitir o acesso livre à informação científica impulsionando assim a geração de novos conhecimentos. Dessa maneira, este estudo realizou um mapeamento descritivo dos repositórios de dados de pesquisa em Geociências no diretório Re3data, com a finalidade de analisar sua organização, através das categorias: ranking de países, tipos de dados em relação ao conteúdo, sistemas de identificadores de autoria, tipos de acesso, licenças de uso, upload de dados, palavras-chave mais usadas, metadados utilizados, softwares utilizados e versionamento do dado. De acordo com o levantamento realizado foram identificados 789 repositórios de geociências. Os resultados alcançados apontam que há inconformidade com padrões internacionais estabelecidos entre os repositórios analisados e que a distribuição geográfica é heterogênea, com predomínio quantitativo de repositórios de dados no contexto norte-americano. A pouca quantidade de repositórios de dados em geociências analisados na América Latina e Caribe aponta para o fato de que o estado da arte neste domínio de conhecimento, nessa localidade, ainda se encontra em estágio inicial.

¹ Doutoranda em Ciência da Informação pelo convênio IBICT/UFRJ. Bibliotecária do Serviço Geológico do Brasil – CPRM.



Palavras-chave: repositórios digitais; dados de pesquisa; Geociências.

Abstract: Research data repositories are a tool for disseminating and preserving institutional production whose premise is to allow free access to scientific information, thus boosting the generation of new knowledge. Thus, this study carried out a descriptive mapping of research data repositories in Geosciences in the Re3data directory, in order to analyze their organization, through categories: ranking of countries, types of data in relation to content, systems of authorship identifiers, types of access, use licenses, data upload, most used keywords, metadata used, software used and data versioning. According to the survey carried out, 789 geoscience repositories were identified. The results achieved indicate that there is non-compliance with established international standards among the analyzed repositories and that the geographic distribution is heterogeneous, with a quantitative predominance of data repositories in the North American context. The small amount of geoscience data repositories analyzed in Latin America and the Caribbean points to the fact that the state of the art in this field of knowledge in this location is still in its initial stage.

Keywords: digital repositories; search data; Geosciences.

1 INTRODUÇÃO

O acesso aberto a dados de pesquisa tem um papel relevante no progresso da ciência em escala mundial, pois contribui para a geração de novas pesquisas a partir da eliminação de etapas que já foram realizadas por outros pesquisadores. Há impacto também na esfera econômica com a diminuição de pesquisas duplicadas.

A enorme quantidade de dados de pesquisa gerados pelas instituições sem o devido controle e organização traz inúmeros problemas de ordem econômica, social e ambiental.

Assim, a proposta deste estudo é realizar um mapeamento da situação atual da gestão de dados no domínio das geociências com a finalidade de identificar tipos



de dados, ferramentas e ações importantes já desenvolvidas no âmbito da gestão dos dados geocientíficos.

2 DESENVOLVIMENTO

Dados de pesquisa fazem parte da ciência há bastante tempo; mas apenas recentemente com o uso de novas tecnologias da informação é que esses dados passaram a ser produzidos em uma larga escala, tornaram-se públicos e disseminados para reuso em outras pesquisas científicas.

De acordo com Sayão e Sales (2016), para apoiar a execução dos processos de gestão é necessário um arcabouço tecnológico e gerencial que compreenda todo o ciclo de vida dos dados. No centro desse arcabouço estão os repositórios digitais de dados de pesquisa que, por muitas razões e demandas, rapidamente se tornaram parte essencial da infraestrutura mundial de pesquisa. Entretanto, para que um determinado dado possa ser reutilizado é necessária uma gestão de dados eficiente.

Os tipos de dados gerados no domínio das geociências, variam de acordo com determinados propósitos, e são de origem **observacionais**, como por exemplo os dados de fotointerpretação geológica de erosões, **experimentais**, tais como as análises de água mineral realizadas em laboratório, e **computacionais**, como os dados de simulação computacional. Quanto à natureza dos dados é bem ampla e inclui multimídia, software, visualização, textual e artefato.

Entretanto, há de se ressaltar que os dados em geociências, quanto à sua origem, são considerados em grande parte como observacionais. São dados que não podem ser coletados em uma segunda vez ou que são caros demais para serem repetidos e, portanto, devem ser preservados de maneira permanente.

Para Pinnick (2017, p. 180), uma característica específica de todos os dados de geociências é sua longa validade, que estende consideravelmente a vida útil dos dados digitais, assumindo que ações sejam tomadas para preservá-los.

A necessidade das instituições e pesquisadores depositarem seus dados brutos ou dados primários levou a criação de repositórios de dados, cuja finalidade maior é a preservação dos dados gerados pelas pesquisas científicas.



Partindo da premissa que os dados devem estar em um local acessível ao público, muitas instituições e pesquisadores estão depositando seus dados em repositórios de dados, os quais são indexados no diretório Re3data. Assim sendo, a abordagem metodológica desenvolvida foi realizada através de mapeamento descritivo dos repositórios de dados constantes no diretório Re3data no domínio das Geociências.

3 CONSIDERAÇÕES FINAIS

No contexto do domínio em geociências, constatou-se que os dados geocientíficos além de serem heterogêneos têm um valor muitas vezes insubstituíveis. Além disso, os volumes de dados geoespaciais são enormes, necessitando de uma grande quantidade de espaço de armazenamento, no qual os sistemas tradicionais não podem satisfazer adequadamente.

Em relação aos tipos de dados depositados nos repositórios há predomínio de dados científicos, estatísticos e imagens. Em relação ao padrão de metadados adotado prevalece a norma ISO 19115, seguido do Dublin Core. Sobre o sistema identificador de autoria o mais adotado é o ORCID. A maioria dos repositórios afirma utilizar o versionamento de dados e a maioria declararam disponibilizar seus dados no formato aberto.

Pode-se ainda constatar que os repositórios em geociências não estão distribuídos geograficamente de maneira homogênea. Há um predomínio norte-americano e há uma grande lacuna em repositórios cadastrados no Re3data no âmbito da América Latina e Caribe.



REFERÊNCIAS

Pinnick, J. (2017). Exploring digital preservation requirements. *Records Management Journal*, 27(2), 175–191. <https://doi.org/10.1108/rmj-04-2017-0009>

Re3data.org - Registry of Research Data Repositories. ([2023]). Site.

<https://doi.org/10.17616/R3D>. recuperado de:

[https://www.re3data.org/search?subjects\[\]=34%20Geosciences%20\(including%20Geography\)](https://www.re3data.org/search?subjects[]=34%20Geosciences%20(including%20Geography)).

Sayão, L. F., & Sales, L. F. (2016). Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, 21(2), 90.

<https://doi.org/10.5433/1981-8920.2016v21n2p90>