

**XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB****ISSN 2177-3688****GT-7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação****Caracterização de Repositórios de Dados de Periódicos Científicos*****Characterization of Scientific Journal Data Repositories*****Laura Vilela Rodrigues Rezende. UFG.****Larissa Barbara Borges Drumond. UFG.****Modalidade: Resumo Expandido**

Resumo: Este trabalho é um dos desdobramentos da pesquisa de mestrado em um Programa de Pós-Graduação em Comunicação, linha de pesquisa Mídia e Informação. O estudo em andamento trata, especificamente, da ampliação da informação veiculada nos periódicos científicos geridos por Instituições de Ensino Superior, sob a ótica da Ciência Aberta. A Ciência Aberta é um tema amplo que, de maneira geral, preocupa-se com um fazer científico mais transparente e colaborativo. Dentre as vertentes possíveis de serem abordadas, tanto a pesquisa de mestrado da autora, quanto este trabalho, estão focados na abertura dos dados científicos. Objetiva-se no presente trabalho apresentar as principais características de repositórios de dados de periódicos científicos, especificamente no Dataverse de Harvard. Busca-se conhecer as normativas e definições que pautam a abertura de dados científicos de dois periódicos científicos hospedados no repositório Dataverse da Universidade de Harvard, com base nos requisitos de repositórios confiáveis da certificação *Core Trust Seal*. Para que isso fosse possível, realizou-se um levantamento documental visitando as páginas web e respectivas coleções de dados dos periódicos presentes no Repositório supracitado. Como resultados desses procedimentos metodológicos, obtiveram-se as caracterizações propostas que poderão encorajar a implementação de repositórios de dados de periódicos científicos brasileiros visando alinhar este processo de abertura ao contexto da Ciência Aberta, em suas práticas de disseminação científica.

Palavras-Chave: Comunicação Científica. Ciência Aberta. Repositórios de Dados Científicos.

Abstract: This study is one of the developments of the master's research in Communication Postgraduate Program, specifically at the Media and Information research group. The master research in progress specifically deals with the expansion of information published in scientific journals managed by Higher Education Institutions, from the perspective of Open Science. Open Science is a broad topic that, in general, is concerned with a more transparent and collaborative scientific work. Among the possible aspects to be addressed, both the author's master's research and this preliminary study are focused on the openness of scientific data. The objective of this study is to present the main characteristics of scientific journals data repositories, specifically in Harvard Dataverse. It seeks to understand the regulations and definitions that guide the opening of scientific data from two scientific journals hosted in this repository, based on the requirements of trusted repositories characterized by Core Trust Seal certification. To make this possible, a documentary survey was carried out visiting the web pages of the journals hosted in Harvard Repository. As a result, the characterizations obtained



could encourage the implementation of data repositories of Brazilian scientific journals in order to align this process of opening with the context of Open Science, in its scientific dissemination practices.

Keywords: Scientific Communication. Open Science. Scientific Data Repositories.

1 INTRODUÇÃO

A comunicação do conhecimento científico visa, basicamente, a disseminação de informações especializadas. Nesse contexto, os periódicos científicos são os mais comumente utilizados para publicar os resultados das pesquisas realizadas individual ou coletivamente por pesquisadores vinculados a laboratórios, Institutos de pesquisa ou projetos científicos das Universidades. A maioria dos dados gerados para a execução dessas pesquisas acaba sendo “esquecida” em computadores pessoais dos pesquisadores ou das instituições, sendo fornecidos de maneira privada, quando solicitados e concedidos pelos seus “donos”. O compartilhamento adequado desses dados faz parte do que vem a ser o cerne da Ciência Aberta com vistas a possibilitar a reprodutibilidade da Ciência e a economia de recursos.

A abertura dos dados científicos é apenas uma das vertentes do Movimento da Ciência Aberta que ao promover a inovação e a disseminação do conhecimento, afeta diretamente os periódicos científicos (MIRANDA; DAMASIO; FIRME, 2020).

As autoras Oliveira e Silva (2016) afirmam que no Brasil as práticas voltadas para Ciência Aberta e, por conseguinte, abertura de dados científicos, ainda estão em seus estágios iniciais. Embora o artigo dessas autoras seja datado de 2016, essa realidade ainda não mudou muito. Um exemplo que corrobora com essa afirmação é o fato de que poucas Universidades, Institutos de Pesquisas e até mesmo periódicos científicos já possuem repositórios para organizar, armazenar, preservar e divulgar sua produção intelectual.

A partir do exposto, este trabalho objetiva apresentar as principais características de repositórios de dados de periódicos científicos, especificamente no Dataverse de Harvard. Busca-se conhecer as normativas e definições que pautam a abertura de dados científicos de dois periódicos científicos hospedados no repositório mencionado, com base nos requisitos de repositórios confiáveis da certificação *Core Trust Seal*. Para que isso fosse possível, realizou-se um levantamento documental visitando as páginas web e respectivas coleções de dados dos periódicos selecionados. O Dataverse de Harvard é um repositório de dados científicos



abertos ao público interno e externo à esta universidade, podendo ser utilizado por qualquer pesquisador, periódico científico ou instituição.

Espera-se que, com os resultados desse levantamento inicial, que faz parte da referida pesquisa de mestrado, seja possível obter referenciais necessários para elaborar a proposta de implementação de um repositório de dados científicos para um periódico científico de uma universidade brasileira, utilizando-se o software Dataverse, que poderá servir de iniciativa piloto para esta instituição contemplando sua implementação, povoamento e uso, além de ações de formação para autores, editores e avaliadores do referido periódico científico que será utilizado como piloto.

2 REFERENCIAL TEÓRICO

Por muito tempo, os meios pelos quais se comunicava a ciência eram, basicamente, por meio dos livros, periódicos impressos ou, informalmente, em encontros acadêmicos. Graças ao advento das Tecnologias da Informação e Comunicação (TICs) esses canais de comunicação foram ampliados, passando a coexistir, por exemplo, os periódicos científicos eletrônicos e, um pouco mais tarde, os repositórios de *pre-prints* e as bases de dados.

As TIC e, sobretudo a internet, trouxeram muitos benefícios à comunicação científica, como a ampliação de armazenamento, distribuição e recuperação dos resultados das pesquisas (VANZ; SILVA FILHO, 2019).

Com o aumento exponencial do número de periódicos científicos online, muitas preocupações surgiram, como o controle bibliográfico e as questões ligadas à preservação dos registros efetuados em ambiente virtual. No sentido do controle das publicações em meio online foi criado o Identificador de Objeto Digital (DOI) ou, como também é conhecido, Identificador persistente dos dados, uma sequência numérica aceita internacionalmente para identificar conteúdos digitais (documentos, imagens, artigos, e-books, *datasets* etc.) o objetivo principal é localizar e acessar mais facilmente, conteúdos na internet (FACHIN; HILLESHEIM, 2006).

Se por um lado, têm-se vários órgãos preocupados com essas normalizações e meios para facilitar o acesso aos arquivos disponíveis online, por outro há a discussão sobre o fazer científico mais transparente e colaborativo, com foco na Ciência Aberta.



A Ciência Aberta surge, então, para atender às necessidades de novos modelos de publicação, nos quais não se tenha acesso apenas aos resultados finais das pesquisas científicas, por meio de artigos científicos publicados, mas também modelos que evidenciem como a pesquisa foi desenvolvida em sua totalidade, seus métodos, suas anotações, os dados obtidos nos levantamentos etc., a fim de que essas informações “possam servir de base para avaliações mais confiáveis, testes de veracidade e replicação de experimentos, além de estarem disponíveis para serem reutilizadas mais eficazmente em novas pesquisas ou que permitam ao pesquisador interagir com outros pesquisadores” (SHINTAKU; SALES, 2019, p.11).

O tema da Ciência Aberta é amplo, permeado por muitas dimensões, porém ao se falar em periódicos científicos geridos por Instituições de Ensino Superior (IES), é comum que já se pratique o Acesso Aberto, no qual pesquisadores não precisam pagar nenhuma taxa para publicarem seus artigos científicos e os leitores não têm nenhum tipo de barreira financeira, de cadastro ou login para acessá-los. Esses periódicos estão agora se preparando para a divulgação dos dados científicos, estabelecendo políticas e infraestrutura específica para essa finalidade com base em padrões e princípios internacionalmente difundidos.

Nesse contexto dos periódicos científicos é possível afirmar que, “os repositórios de dados desempenham uma função vital para a preservação, a integridade e a difusão dos dados científicos” (SILVA, 2019, p. 89). Além disso, permitem que a literatura científica seja reconhecida, utilizada e que o conhecimento científico possa ser explorado em sua complexidade. Existe uma infinidade de tipos de repositórios, podendo ser gratuitos ou que requerem pagamento de taxas. Dentre as tipologias estão os repositórios institucionais, temáticos, de propósitos gerais ou ainda, multidisciplinares.

Além disso, esses repositórios possuem políticas que orientam a forma em que os arquivos, antes de serem depositados, devem ser preparados para que atendam aos princípios de qualidade de dados garantindo que sejam localizáveis, acessíveis, interoperáveis e reutilizáveis, ou seja, que estejam em consonância com os princípios FAIR (*Findable, Accessible, Interoperable e Reusable*).



3 PROCEDIMENTOS METODOLÓGICOS

Para atender ao objetivo norteador desse trabalho, que é identificar as principais características de dois repositórios de periódicos científicos conforme requisitos da certificação *Core Trust Seal*, procedeu-se com um levantamento bibliográfico, seguido de um levantamento documental nas páginas web de duas coleções de periódicos presentes no repositório Dataverse da Universidade de Harvard: o *Review of Economics and Statistics* e o *Brazilian Political Science Review*. A escolha desses periódicos se deu a partir do levantamento dos periódicos científicos presentes no repositório Dataverse de Harvard. Foi adotado o seguinte procedimento: acessou-se a página web do Dataverse de Harvard e na caixa de busca da página inicial, utilizou-se a palavra “*journal*” e, em seguida, na lateral esquerda, em “*Dataverse Category*”, selecionou-se a opção de categoria “*journal*”. Com essas especificações de busca, na data do levantamento (06 de junho de 2022), foram retornados 91 resultados. Optou-se pelo uso da expressão “*journal*”, tradução em inglês para periódico, dado o recorte desejado para a pesquisa, somente coleções categorizadas como sendo pertencentes a periódicos científicos.

Das 91 coleções encontradas, 45 correspondiam a periódicos científicos geridos por IES, as demais pertenciam a Organizações Científicas ou Editoras Autônomas. Das 45 coleções, para este estudo inicial da pesquisa de mestrado, optou-se por selecionar duas delas, considerando os seguintes critérios: a que possuía maior quantidade de *datasets* neste escopo e uma vinculada a um periódico brasileiro. Foi selecionada a coleção *Review of Economics and Statistics (REST)*, periódico científico da *Harvard Kennedy School*, por ser essa a que possuía, na data do levantamento, a maior quantidade de *datasets* disponibilizados, 1.134. A outra escolha foi do periódico brasileiro *Brazilian Political Science Review (BPSR)*, vinculado à Universidade de São Paulo. Destaca-se aqui que o Repositório Dataverse de Harvard possui apenas dois repositórios de periódicos científicos brasileiros, vinculados a IES, o *Brazilian Political Science Review (USP)* e o *Brazilian Journal of Urban Management (PUCPR)*, este último, ainda não disponibilizou nenhum *dataset*, justificando a escolha do primeiro que possui 69 *datasets* disponibilizados.

Os resultados obtidos no levantamento documental dos dois casos específicos serão explanados na próxima seção.



4 RESULTADOS

O *International Science Council* (WDS) é uma organização que oferece a certificação *Core Trust Seal* (CTS) que visa avaliar repositórios digitais de dados com base em uma lista de requisitos de repositórios de dados considerados confiáveis (CORETRUSTSEAL, 2022, tradução nossa). Trata-se de uma auto avaliação realizada pela equipe gestora dos repositórios a serem certificados. Os requisitos são categorizados seguindo as seguintes vertentes:

- A. Infraestrutura organizacional: referente à dimensão organizacional do repositório. Abrange questões relacionadas à definição, governança, processos, políticas, licenças e recursos financeiros e humanos;
- B. Gestão de objetos digitais: modo como ocorre a criação e gestão dos objetos e informações digitais. Abrange a forma como ocorre a criação, preservação e gestão dos pacotes de informação que serão submetidos, armazenados e recuperados;
- C. Tecnologia: Gestão de riscos em infraestrutura e segurança abrangendo a gestão de segurança e controle dos riscos e ameaças à infraestrutura do repositório.

Uma das finalidades da referida certificação é apoiar pesquisadores de todo o mundo no gerenciamento, compartilhamento e preservação de seus dados científicos por meio do depósito em repositórios confiáveis, favorecendo assim a abertura do fazer científico, ou seja, a Ciência Aberta.

Para esse trabalho foram considerados os Requisitos de Repositórios de Dados Confiáveis *Core Trust Seal: Orientação Estendida 2020–2022*¹. A partir das orientações do CTS, foi elaborada a Tabela 1 como instrumento de verificação de conformidade das informações descritivas que estão divulgadas nas páginas web dos repositórios e nos sites dos periódicos selecionados.

¹ <https://zenodo.org/record/3632533#.YqdMAnbMJPY>



Tabela 1 – Caracterização dos Periódicos Científicos Seleccionados com base nos Requisitos

Core Trust Seal

REPOSITÓRIOS	BPSR	REST
INFORMAÇÕES DESCRITIVAS GERAIS		
Finalidade do repositório / apresentação do periódico	Sim	Sim
Comunidade atendida; Parceiros;	Não	Sim
Governança do repositório (Nomes/contato/qualificação da equipe); dotação orçamentária	Sim	Sim
Vínculo com alguma rede, consórcio, comunidade de prática, etc.(Equipe IQSS)	Sim	Sim
Recursos Adicionais (estatísticas / fator de impacto, etc.)	Sim	Sim
POLÍTICAS ADOTADAS DIVULGADAS		
Glossário (termos utilizados nas políticas)	Não	Sim
Depósito de dados: Guia de preparação de dados;	Não	Sim
Depósito de dados: termo de depósito	Não	Sim
Uso de dados: Política de acesso e uso	Não	Sim
Uso de dados: orientações sobre citações corretas	Não	Sim
Política de Curadoria de dados / Workflow detalhado:	Não	Não
Política de Preservação Digital: <ul style="list-style-type: none">● Conformidade OAIS;● Segurança do sistema;● Integridade e autenticidade dos dados;● Explicações sobre as Funções: "deaccessioning" (tornar o arquivo indisponível mantendo os metadados), versionamento dos dados em caso de upgrade ou modificação;● Controle de acesso, autenticação;● Infraestrutura adequada;	Não	Não

Fonte: Elaborado pelas autoras (2022) com base nos requisitos da certificação *Core Trust Seal* (2022)

Review of Economics and Statistics (REST) é um periódico científico da *Harvard Kennedy School* que publica artigos científicos correlatos ao tema "Economia". Os conselhos editorial e científico contam com pesquisadores de universidades de diferentes estados norte-americanos. O periódico publica dois formatos: artigos em Acesso Aberto e Artigos Restritos,



cujo acesso está disponível apenas aos pagantes de taxas específicas ao *MIT Press Direct*², plataforma que disponibiliza o acesso a vários periódicos científicos de praticamente todas as áreas do conhecimento.

Quanto às informações gerais, o REST não apresenta no Dataverse de Harvard a finalidade ou missão do repositório, sendo que consta apenas no próprio site do periódico a sua finalidade. Quanto à governança, são informados os nomes e contatos dos editores do periódico que em geral realizam a curadoria dos dados. Quanto às informações estatísticas, o site do periódico divulga estatísticas sobre os artigos mais lidos e os mais citados.

Quanto às políticas de abertura de dados, no REST faltam informações mais claras sobre curadoria e preservação dos dados. Em contrapartida, existe um glossário dos termos utilizados nas políticas e um link para acesso ao guia de depósito de dados, elaborado pela *Crossref*³ com a exigência de preenchimento de uma Declaração de Disponibilidade de Dados (DAS), na qual devem constar informações claras sobre tipos e formato de dados e políticas de acesso e uso dos dados.

O periódico *Brazilian Political Science Review* (BPSR) é publicado pela Associação Brasileira de Ciência Política, sediada na Universidade de São Paulo. As temáticas dos artigos publicados contemplam abordagens teóricas e metodológicas da Ciência Política e suas relações internacionais. O conselho editorial é formado por pesquisadores brasileiros e de vários outros países. Os artigos são publicados somente em Acesso Aberto, característica de grande parte dos periódicos brasileiros.

O BPSR apresenta de maneira bem clara, tanto na página do Dataverse de Harvard, quanto em seu próprio site, a finalidade e apresentação do repositório de dados, entretanto faltam informações mais precisas sobre os parceiros e qual a comunidade atendida. Quando se fala em governança, é citado o nome e contato de um editor de dados e origem do financiamento mantenedor. A página do periódico apresenta dados estatísticos sobre acesso aos artigos publicados.

Sobre os requisitos específicos da CTS vinculados às políticas de abertura de dados, na BPSR observou-se a ausência de informações relevantes, tais como um glossário orientador

² <https://direct.mit.edu/>

³ <https://www.crossref.org/documentation/reference-linking/data-and-software-citation-deposit-guide/>



para entendimento das políticas, guia e termo de depósito de dados, política de acesso e uso, orientações sobre citações e política de curadoria e preservação dos dados. O periódico divulga informações sobre a obrigatoriedade de disponibilização de *datasets* utilizados nas pesquisas acompanhados de arquivos descritores dos procedimentos de coleta, análise e etapas de pesquisa. Além disso, existem informações que orientam sobre os formatos possíveis nos quais os *datasets* devem ser disponibilizados (arquivos dos softwares R, Stata, SPSS, Excel).

5 CONSIDERAÇÕES FINAIS

Obviamente, o trabalho realizado não teve a pretensão de esgotar análises sobre o tema, principalmente porque utilizou-se como escopo investigativo somente o repositório Dataverse de Harvard. Na categoria de “*journals*” deste repositório, existiam ainda outros 44 periódicos científicos, até o momento do levantamento, abrindo possibilidades para outros estudos sobre as especificidades de diferentes políticas de abertura dos dados científicos de pesquisas publicadas em periódicos científicos geridos por Instituições de Ensino Superior.

Chega-se à conclusão de que, embora existam organizações preocupadas com a padronização dos requisitos de abertura de dados, seguindo orientações como a certificação CTS, os repositórios dos periódicos científicos analisados, hospedados no Dataverse de Harvard, publicam suas políticas sem considerar a integralidade desses requisitos.

Um ponto relevante a ser considerado são os diferentes desenhos dos periódicos no tocante à disponibilização dos dados científicos dos seus autores. Os periódicos oferecem auto depósito para que os autores possam depositar os dados científicos em algum repositório a escolher, além daquele oferecido pelo periódico, como é o caso do REST. Além disto, os próprios editores dos periódicos realizam os depósitos dos dados, necessariamente em seus respectivos repositórios de dados científicos, como no BPSR. Estas duas configurações alteram sobremaneira as informações descritivas presentes nas páginas web dos periódicos. Isto porque o fluxo de submissão e depósito dos *datasets* é diferente.

Com base nesse entendimento, considera-se relevante a investigação preliminar realizada, a qual servirá de embasamento para que a pesquisa de mestrado da autora possibilite elaborar uma proposta de implementação de um repositório de dados científicos para um periódico científico de uma universidade brasileira, utilizando-se o software Dataverse, considerando o conjunto de requisitos de repositórios confiáveis, a partir da lista



elaborada pela *Core Trust Seal* e que possa servir de iniciativa piloto para esta instituição contemplando sua implementação, povoamento e uso, além de ações de formação para autores, editores e avaliadores do referido periódico científico.

Ações como essas, são muito importantes para a consolidação da Ciência Aberta no âmbito dos periódicos científicos geridos por IES e favorecem, sobretudo, a preservação dos dados científicos, a reprodutibilidade da Ciência e o impacto dos periódicos uma vez que facilita a citação de suas publicações.

REFERÊNCIAS

CORETRUSTSEAL. Disponível em: <https://www.coretrustseal.org/>. Acesso em 12 de junho de 2022.

FACHIN, Gleisy Regina Bories; Hillesheim, Araci Isaltina de Andrade. **Periódico científico: padronização e organização**. Florianópolis: Editora da UFSC, 2006.

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Organizadores). **O quarto paradigma: descobertas científicas na era da eScience**. Tradução de Leda Maria Marques Dias Beck. São Paulo: Oficina de Textos, 2011.

MIRANDA, Angélica C. D.; DAMASIO, Edilson; FIRME, Simone Machado (Organizadores). **Ciência Aberta: visão e contribuição a partir dos Periódicos Científicos**. Rio Grande: Editora da Furg, 2020.

MIT PRESS DIRECT. Disponível em <https://direct.mit.edu/>. Acesso em 06 de junho de 2022.

OLIVEIRA, Adriana Carla Silva de; SILVA, Edilene Maria da. Ciência Aberta: dimensões para um novo fazer científico. **Informação & Informação**, Londrina, v. 21, n. 2, p. 5 – 39, mai./ago., 2016.

SILVA, Fabiano Couto Corrêa da. **Gestão de dados científicos**. Rio de Janeiro: Editora Interciência, 2019.

SHINTAKU, Milton; SALES, Luana Farias (Orgs.) **Ciência aberta para editores científicos**. Botucatu, SP: ABEC, 2019. 108 p. DOI: <http://dx.doi.org/10.21452/978-85-93910-02-9>

VANZ, Samile Andréa de Souza; SILVA FILHO, Rubens da costa. O Protagonismo das Revistas na Comunicação Científica: Histórico e Evolução. In: Carneiro, Felipe Ferreira Barros; Ferreira Neto, Amarílio; Santos, Wagner dos. (Orgs.) **A Comunicação Científica em Periódicos**. 1. Ed. Curitiba: Appris, 2019. p. 20-57 (Versão Kindle).

WOELFLE, Michael; OLLIARO, Piero; TODD, Matthew H. Open science is a research accelerator. **Nature Chemistry**, v. 3, n. 10, p. 745–748, 23 set. 2011. Doi: <https://doi.org/10.1038/nchem.1149>. Acesso em: 12 nov. 2021.