

# Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias

## **Rene Faustino Gabriel Junior**

Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) – Marília, SP – Brasil. Professor da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.

<http://lattes.cnpq.br/5900345665779424>

<http://orcid.org/0000-0003-1021-3360>

E-mail: [renefgj@gmail.com](mailto:renefgj@gmail.com)

## **Rafael Port da Rocha**

Doutor em Ciência da Computação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil. Professor da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.

<http://lattes.cnpq.br/5118387541734094>

<http://orcid.org/0000-0003-4131-0534>

E-mail: [rafael.rocha@ufrgs.br](mailto:rafael.rocha@ufrgs.br)

## **Sônia Elisa Caregnato**

Doutora em Information Studies pela University of Sheffield (SHEFFIELD) - Inglaterra.

Professora da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.

<http://lattes.cnpq.br/5627209208288722>

<https://orcid.org/0000-0002-5676-2763>

E-mail: [sonia.caregnato@ufrgs.br](mailto:sonia.caregnato@ufrgs.br)

## **Caterina Marta Groposo Pavão**

Doutora em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS – Brasil, com período sanduíche em Universidad Complutense de Madrid (UCM) - Espanha.

Professora da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.

<http://lattes.cnpq.br/4834791532698069>

<http://orcid.org/0000-0003-3712-7200>

E-mail: [caterina@cpd.ufrgs.br](mailto:caterina@cpd.ufrgs.br)

## **Paula Caroline Schifino Jardim Passos**

Pós-Doutorado pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.

Doutora em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.

<http://lattes.cnpq.br/1766582596876949>

<http://orcid.org/0000-0002-2354-847X>

E-mail: [paulacarolinejardim@gmail.com](mailto:paulacarolinejardim@gmail.com)

## **Eduardo Nunes Borges**

Pós-Doutorado pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.

Doutor em Computação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil. Professor da Universidade Federal do Rio Grande (FURG) - Rio Grande, RS – Brasil.

<http://lattes.cnpq.br/5851601274050374>

<http://orcid.org/0000-0003-1595-7676>

E-mail: [eduardoborges@furg.br](mailto:eduardoborges@furg.br)

### **Samile Andréa de Souza Vanz**

Pós-Doutorado pela Universidad Carlos III de Madrid (UC3M) – Espanha. Doutora em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS – Brasil, com período sanduíche em Dalian University of Technology (DUT) – China. Professora da Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, RS - Brasil.

<http://lattes.cnpq.br/5243732207004083>

<https://orcid.org/0000-0003-0549-4567>

E-mail: [samilevanz@terra.com.br](mailto:samilevanz@terra.com.br)

### **Luís Alberto Barbosa Azambuja**

Mestre em Engenharia de Computação pela Universidade Federal do Rio Grande (FURG) – RS - Brasil. Servidor Técnico em Tecnologia da Informação da Universidade Federal do Rio Grande (FURG) - Rio Grande, RS - Brasil.

<http://lattes.cnpq.br/2167942665157676>

<http://orcid.org/0000-0003-3383-0513>

E-mail: [l.azambuja@gmail.com](mailto:l.azambuja@gmail.com)

Data de submissão: 19/09/2019. Data de aprovação no ConfOA: 12/06/2019. Data de publicação:

## **RESUMO**

O Acesso Aberto a Dados de Pesquisa (AADP) tem crescido significativamente em todo o mundo, nos últimos anos; entretanto, ainda é incipiente no Brasil. Este artigo apresenta o projeto RDP Brasil, uma parceria entre a Rede Nacional de Ensino e Pesquisa (RNP), o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), a Universidade Federal do Rio Grande do Sul (UFRGS) e a Universidade Federal do Rio Grande (FURG). O projeto foi dividido em cinco etapas: mapeamento de serviços e potenciais usuários de AADP em instituições brasileiras; desenvolvimento de portal web para reunião da comunidade nacional interessada em AADP; levantamento dos serviços e das soluções tecnológicas existentes para compartilhamento de dados; e proposta de solução tecnológica que atenda aos usuários identificados no estudo. Os resultados demonstraram que iniciativas institucionais de AADP ainda são escassas no Brasil; que há receptividade dos pesquisadores brasileiros à ideia do compartilhamento e do reúso de dados, contudo há desconhecimento e falta infraestrutura nas instituições; que as soluções tecnológicas existentes oferecem grande número de opções e, portanto, devem ser constantemente monitoradas; e que a implementação de pilotos de repositórios de dados de pesquisa é viável, no curto prazo. Conclui-se que o projeto foi pioneiro em investigar pesquisadores brasileiros de todas as áreas do conhecimento sobre AADP, trazendo contribuições relevantes para iniciativas governamentais, como o compromisso assumido entre o Brasil e a Open Government Partnership (OGP). No entanto, evidencia-se a necessidade de ampliação das investigações sobre o tema e de esforços públicos para promover e consolidar iniciativas.

**Palavras-chave:** Acesso aberto a dados de pesquisa. Software. Repositório de dados de pesquisa. Compartilhamento de dados. Reuso de dados.

## **Open access to research data in Brazil: mapping repositories, practices and perceptions of researchers and technologies**

### **ABSTRACT**

*Initiatives to promote Open Access to Research Data have grown significantly worldwide, although in Brazil there are still few studies on the subject. The aim of this paper is to present the development and results of the project RDP Brasil, developed in a partnership between RNP, IBICT, UFRGS, and FURG. The project, which has been under development since March 2018, has been divided into five stages: mapping of practices and of services in Brazilian institutions; assembling of the community to discuss the topic through a web portal; survey of existing technology for sharing data; and development of a technology solution to meet the requirements of a selected group of users. The results show that institutional initiatives for open research data sharing in Brazil are still few; that Brazilian researchers are motivated by the idea of data sharing and reuse, but there is a lack of knowledge and poor institutional infrastructure; that existing technologies offer a large number of options and should therefore be constantly monitored; and that developing data repository pilots is a viable solution in the short term. It is concluded that the project was a pioneer in investigating Brazilian researchers from all areas of knowledge about Open Access to Research Data in Brazil, bringing relevant contributions to emerging government initiatives, such as Brazil's commitment to the Open Government Partnership initiative. However, there is still a need to expand research on the subject and make public efforts to promote new and consolidate existing initiatives.*

**Keywords:** *Open access to research data. Software. Research data repository. Data sharing. Data reuse.*

## **Acceso abierto a datos de investigación en Brasil: mapeo de repositorios, prácticas y percepciones de investigadores y tecnologías**

### **RESUMEN**

*En los últimos años, las iniciativas para promover el Acceso Abierto a los Datos de Investigación (AADP) crecieron significativamente en todo el mundo, pero todavía es insipiente en Brasil. Este artículo presenta el desarrollo del proyecto RDP Brasil, en colaboración entre RNP, IBICT, UFRGS e FURG. Se dividió en cinco etapas: mapeo de servicios y potenciales usuarios nacionales de AADP; desarrollo de un portal web; levantamiento de servicios y soluciones tecnología existentes para intercambio y el desarrollo de una propuesta de solución tecnológica que satisfaga a los usuarios identificados en el estudio. Los resultados mostraron que las iniciativas para intercambio abierto de datos de investigación, en Brasil, son pocas; que existe la receptividad a la idea del intercambio y la reutilización de datos por parte de los investigadores, pero hay una falta de conocimiento y de infraestructura por parte de las instituciones; que las soluciones tecnológicas existentes ofrecen una gran cantidad de opciones y, por lo tanto, debe haber un monitoreo constantemente y que implementar pilotos de repositorios de datos de investigación es una solución viable a corto plazo. Se concluye que el proyecto fue pionero en la identificación de investigadores brasileños de todas las áreas de conocimiento, aportando contribuciones relevantes a iniciativas gubernamentales de Brasil con la Open Government Partnership (OGP). Sin embargo, existe la necesidad de ampliar la investigación sobre el tema y los esfuerzos públicos para promover y consolidar iniciativas.*

**Palabras clave:** *Acceso abierto a los datos de investigación. Software. Repositorio de datos de investigación. Intercambio de datos. Reutilización de datos.*

## INTRODUÇÃO

O compartilhamento de dados de pesquisa consiste no ato de disponibilizar dados gerados ou coletados por uma pesquisa para reuso de outros pesquisadores, prática que também permite a formação da memória digital das instituições de pesquisa a partir da preservação por longo prazo (SAYÃO; SALES, 2015). Os benefícios mais evidentes são a possibilidade da reprodução ou da verificação da pesquisa; a disponibilização dos resultados de pesquisas financiadas com fundos públicos; a continuidade das pesquisas e dos questionamentos a respeito dos dados existentes e, conseqüentemente, a viabilização de avanços no estado da pesquisa e na inovação (KIM; ADLER, 2015; KIM, 2017; HENNING *et al.*, 2019). Governos, comunidades e instituições nacionais e regionais passaram a identificar o valor estratégico desse compartilhamento, fomentando o desenvolvimento de infraestruturas e tecnologias que estimulem a disponibilização dos dados de pesquisa. Nesse sentido, diversos são os estudos desenvolvidos com o objetivo de entender as práticas nacionais e por áreas do conhecimento (BEAGRIE; BEAGRIE; ROWLANDS, 2017; TENOPIR *et al.*, 2011; 2015; KIM; ADLER, 2015; KIM, 2017; WALLIS; ROLANDO; BORGMAN, 2013; CARVALHO; LEITE, 2019).

Nesse contexto, a Rede Nacional de Ensino e Pesquisa (RNP), em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), lançou a Carta-Convite 01/2017, com o objetivo principal de identificação de práticas e prototipação de serviço de acesso aberto a dados de pesquisa. Dentre o total de sete submissões, o projeto proposto por pesquisadores do Estado do Rio Grande do Sul foi escolhido no início de maio daquele ano.

O projeto Rede de Dados de Pesquisa Brasileira (RDP Brasil) vem sendo desenvolvido por pesquisadores ligados ao Centro de Documentação e Acervo Digital da Pesquisa, unidade da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul

(Cedap/Fabico/UFRGS) e ao Centro de Ciências Computacionais da Universidade do Rio Grande (C3/Furg). A primeira etapa do projeto teve como objetivo geral realizar um levantamento de informações que subsidiassem as ações iniciais no âmbito do Acesso Aberto da Dados de Pesquisa (AADP) no país, visando a identificação de práticas de AADP e o desenvolvimento de um futuro serviço de acesso aberto. Seus objetivos específicos foram: identificação das atuais práticas de AADP em instituições brasileiras; mapeamento de potenciais usuários nacionais de serviços de AADP e dos principais requisitos e funcionalidades para o atendimento dos mesmos; elaboração de um portal web para reunir a comunidade nacional de interessados em AADP para discussão sobre o tema; levantamento comparativo dos serviços e soluções tecnológicas existentes para compartilhamento de dados, com identificação de lacunas e adaptações necessárias para atender aos usuários mapeados; desenvolvimento de uma proposta de solução tecnológica para atender a um conjunto de usuários selecionados após o mapeamento dos potenciais usuários nacionais.

Este artigo tem como objetivo apresentar uma síntese dos principais resultados de cada um dos objetivos do projeto, com o intuito de contribuir com a discussão sobre o tema no Brasil. Os relatórios completos podem também ser acessados no Lume, Repositório institucional da UFRGS. A seguir, apresentam-se os resultados principais, e ao final do trabalho, a síntese das atividades e propostas para a segunda fase do projeto.

## RESULTADOS E DISCUSSÃO

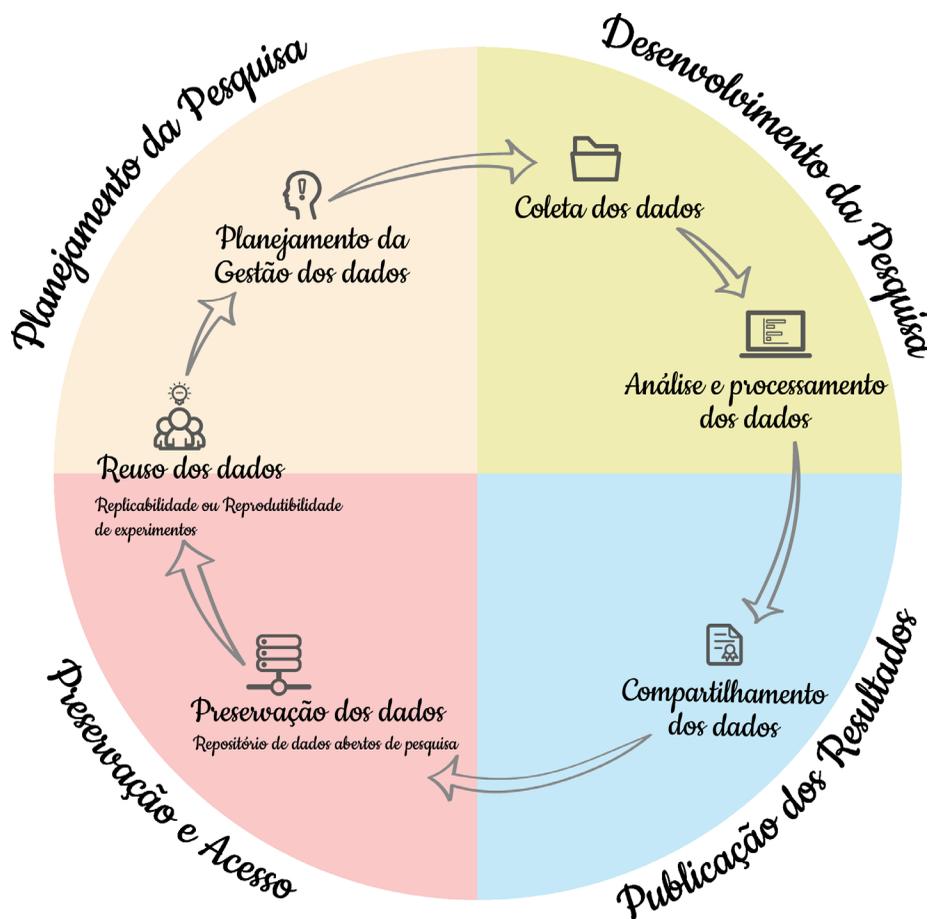
Como já destacado, o objetivo central do projeto foi o desenvolvimento de uma proposta de solução tecnológica que atendesse a um conjunto de usuários selecionados após o mapeamento dos potenciais usuários nacionais. Visando alcançar esse objetivo, buscou-se desenvolver um protótipo de repositório para dados de pesquisa, com conjuntos reais de dados. Essa iniciativa está fundamentada em três pilares: o mapeamento de práticas de

AADP em instituições brasileiras (PAVÃO *et al.*, 2018); o estudo de potenciais usuários nacionais de serviços de AADP (VANZ *et al.* 2018); e o levantamento de serviços e soluções tecnológicas para compartilhamento de dados realizado a partir da avaliação dos *softwares* mais usados para este fim (ROCHA *et al.*, 2018).

Além do conteúdo expresso nesses relatórios, o planejamento de um repositório necessita considerar temas relacionados como o ciclo de vida do dado da pesquisa, os princípios FAIR (*Findable, Accessible,*

*Interoperable, and Reusable*) (WILKINSON *et al.*, 2016; HENNING *et al.*, 2019), os padrões de citação, o modelo OAIS (Reference Model for an Open archive Information System) (CCSDS, 2012) e os critérios para repositórios confiáveis (INTERNATIONAL SCIENCE COUNCIL, 2016; JANTS; GIARLO, 2006). A respeito do ciclo de vida do dado da pesquisa, considerou-se necessário o desenvolvimento de um modelo específico, orientado às peculiaridades dessa pesquisa, conforme mostra a figura 1.

Figura 1 – Modelo de Ciclo de Vida do Dado Aberto de Pesquisa



Fonte: Elaborado pelos autores

O Modelo de Ciclo de Vida do Dado Aberto de Pesquisa. A proposta relaciona as ações inerentes ao ciclo de vida do dado às etapas de desenvolvimento da pesquisa científica. A figura é dividida em quadrantes e tem base em um círculo, indicando que se trata de um processo iterativo. A divisão em quadrantes representa etapas indispensáveis da pesquisa científica, que consistem no planejamento e no desenvolvimento da pesquisa, na publicação dos resultados e na preservação para o acesso.

Conforme o modelo da figura 1, a etapa de planejamento da pesquisa inicia com o reuso de dados, considerando-se a replicabilidade ou a reprodutibilidade dos experimentos científicos. Em seguida, na mesma etapa, tem-se o planejamento da gestão dos dados. Na etapa de desenvolvimento da pesquisa, o modelo destaca a coleta, a análise e o processamento dos dados obtidos. Segue-se a essa, a fase de publicação dos resultados, quando acontece o compartilhamento dos dados. Fechando o ciclo, tem-se a etapa de preservação e acesso, na qual ocorre a preservação dos dados por meio dos repositórios de dados abertos de pesquisa. Assim, possibilita-se o reuso de dados, ação compartilhada com a etapa de planejamento da pesquisa, reiniciando-se o ciclo de vida do dado da pesquisa.

Com base nos pressupostos elencados anteriormente, o trabalho foi dividido em cinco etapas, correspondendo aos objetivos específicos do projeto. A seguir, apresentam-se os resultados divididos nas cinco etapas, cada uma contendo também objetivo, metodologia utilizada e conclusões.

## PRINCIPAIS PRÁTICAS EM AADP EM INSTITUIÇÕES BRASILEIRAS

A fim de identificar as principais práticas em AADP em instituições brasileiras (primeiro objetivo do projeto), buscou-se mapear todos os repositórios brasileiros que disponibilizam dados de pesquisa. Isso ocorreu em três etapas, entre março e junho de 2018. Na primeira delas, os dados foram coletados a partir do diretório *Research Data Repositories Information* (Re3data), as buscas foram realizadas na opção *Browser by country* do menu principal. Em seguida, foi selecionada a opção Brasil.

O Re3data é um registro global de repositórios de dados de pesquisa, lançado em 2012, que abrange diferentes áreas do conhecimento e que busca promover uma cultura de compartilhamento e maior acesso e visibilidade dos dados da pesquisa. Para registrar-se no Re3data, um repositório de dados de pesquisa deve obedecer a certos requisitos, a saber: ser administrado por uma entidade legal como, por exemplo, uma biblioteca ou uma universidade; esclarecer as condições de acesso aos dados e ao repositório, bem como os termos de uso; e ter foco em dados de pesquisa. Ainda, o Re3data considera um repositório de dados de pesquisa como uma das partes de uma infraestrutura de informação sustentável que fornece armazenamento em longo prazo e acesso a dados de pesquisa, considerados a base para uma publicação científica.

Na segunda etapa, buscou-se identificar repositórios de dados em publicações na *web* sobre o tema e em *sites* de instituições de pesquisa reconhecidas nacionalmente pelo seu trabalho. Na terceira e última etapa, foram levantadas as respostas a uma das questões da *survey*, realizada principalmente para contemplar o objetivo 2 do projeto RDP Brasil, que deixava espaço livre para que o respondente especificasse os nomes dos repositórios de dados por ele utilizados. Porém, nenhum repositório diferente dos já identificados nas etapas 1 e 2 foi mencionado.

Os repositórios de dados, mantidos ou nos quais há participação de instituições brasileiras, identificados no Re3Data foram: IBICT Dataverse Network<sup>1</sup>, BDEP - Banco de Dados de Exploração e Produção<sup>2</sup>, GLOBE - Global Collaboration Engine<sup>3</sup>, International Ocean Discovery Program (IODP)<sup>4</sup>, PPBio Data

<sup>1</sup> <https://repositoriopesquisas.ibict.br>. Acesso em: 10 maio 2019.

<sup>2</sup> <http://www.anp.gov.br/exploracao-e-producao-de-oleo-e-gas/dados-tecnicos/acesso-aos-dados-tecnicos>. Acesso em: 10 maio 2019.

<sup>3</sup> <http://globe.umbc.edu>. Acesso em: 10 maio 2019.

<sup>4</sup> <http://www.iodp.org>. Acesso em: 10 maio 2019.

Repository<sup>5</sup>, WorldClim - Global Climate Data<sup>6</sup> e a Base de Dados Científicos da Universidade Federal do Paraná<sup>7</sup>.

Os repositórios identificados em universidades e/ou institutos de pesquisa do país e que não estavam registrados no Re3data foram: Base Tuiuiu<sup>8</sup>, Consórcio de Informações Sociais (CIS)<sup>9</sup>, Instituto Nacional de Meteorologia (INMET)<sup>10</sup>, Instituto Brasileiro de Geografia e Estatística (IBGE)<sup>11</sup>, Sistema Maxwell<sup>12</sup>, Repositório de dados PELD<sup>13</sup>, Projeto speciesLink<sup>14</sup> e IPAgriDados<sup>15</sup>.

Em cada um dos repositórios encontrados procurou-se identificar as seguintes características: a) âmbito (nacional ou cooperação internacional); b) áreas temáticas de abrangência; c) número de instituições envolvidas na coleta e disponibilização de dados; d) políticas do repositório (institucional, direitos autorais, licenças de uso); e) formato dos dados; f) interoperabilidade e padrões de metadados; g) software utilizado, aplicações e interfaces; e h) existência de informações quanto às orientações para elaboração do plano de gestão de dados e ao ciclo de vida da pesquisa. Em alguns casos, não foi possível identificar todas as características elencadas acima, pois as páginas dos repositórios nem sempre fornecem todas as informações; em outros casos, as informações estavam em notícias ou em documentos veiculados pela instituição.

Foram identificados 15 repositórios, sendo quatro de abrangência internacional, ou seja, envolvendo várias instituições, com pelo menos uma do Brasil, e 11 de abrangência nacional, sendo que dois desses são do tipo multi-institucional. Predominam as temáticas voltadas às Ciências Sociais Aplicadas,

às Ciências Agrárias, às Ciências Exatas e da Terra, às Ciências Biológicas e à Engenharia. Os padrões de metadados encontrados foram o Dublin Core, o DataCite Metadata Schema, a ISO 19115 e o DarwinCore. O Digital Object Identifier (DOI) é empregado em quatro repositórios, e o padrão de interoperabilidade Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) é citado em cinco repositórios. O Dspace é utilizado em dois repositórios, assim como o Metacat. O Dataverse é utilizado em um repositório. Em três repositórios são utilizados softwares desenvolvidos localmente. Grande parte dos dados depositados são dos tipos texto, bases de dados, imagens, planilhas e gráficos. As políticas dos repositórios foram encontradas em apenas 8 dos 15 repositórios analisados.

## MAPEAMENTO DE POTENCIAIS USUÁRIOS NACIONAIS DE SERVIÇOS DE AADP

A investigação de práticas e percepções dos usuários nacionais de serviços de AADP foi realizada por meio de uma pesquisa survey enviada em março de 2018 a 71.712 e-mails válidos de 48.335 líderes e 27.869 vice-líderes dos grupos de pesquisa cadastrados no Diretório de Grupos de Pesquisa CNPq; 4.677 e-mails institucionais e de coordenadores dos PPGs cadastrados na Capes; e 249 e-mails de coordenadores de Institutos Nacionais de Ciência e Tecnologia. Após um mês de coleta, o número total de respondentes foi 4.703.

O questionário intitulado Práticas e percepções sobre acesso aberto a dados de pesquisa foi criado com base em levantamentos internacionais sobre compartilhamento e reuso de dados de pesquisa, especialmente os trabalhos de Tenopir e colaboradores (2011; 2015), Kim e Adler (2015), Kim e Zhang (2015) e Digital Curation Center (2009). O questionário foi organizado no software Survey Monkey em uma estrutura composta por oito questões demográficas e mais dois eixos principais, contendo 13 questões acerca das práticas dos pesquisadores sobre dados de pesquisa e seis questões sobre as percepções em relação ao compartilhamento e uso de dados.

<sup>5</sup> <https://ppbiodata.inpa.gov.br/metacatui/>. Acesso em: 10 maio 2019.

<sup>6</sup> <http://worldclim.org>. Acesso em: 10 maio 2019.

<sup>7</sup> <https://bdc.c3sl.ufpr.br>. Acesso em: 10 maio 2019.

<sup>8</sup> <http://tuiuiu.cpap.embrapa.br>. Acesso em: 10 maio 2019.

<sup>9</sup> <http://www.cis.org.br>. Acesso em: 10 maio 2019.

<sup>10</sup> <http://www.inmet.gov.br/portal>. Acesso em: 10 maio 2019.

<sup>9</sup> <https://ww2.ibge.gov.br>. Acesso em: 10 maio 2019.

<sup>12</sup> [https://www.maxwell.vrac.puc-rio.br/research\\_data.php?tipBusca=inicio](https://www.maxwell.vrac.puc-rio.br/research_data.php?tipBusca=inicio). Acesso em: 10 maio 2019.

<sup>13</sup> <http://peld.inpa.gov.br>. Acesso em: 10 maio 2019.

<sup>14</sup> <http://www.splink.org.br/index?setlang=pt>. Acesso em: 10 maio 2019.

<sup>15</sup> <https://www.dados.cnpqia.embrapa.br>. Acesso em: 10 maio 2019.

No total, foram formuladas 27 questões, que tomaram, em média, 14 minutos para serem respondidas por participante.

Os resultados revelaram grande interesse dos pesquisadores brasileiros em questões relacionadas ao AADP. Para tal afirmação, considera-se o alto número de respondentes da pesquisa, que foi de 4.703 pesquisadores. Desses, 1.061 (22,56%) aceitaram participar de entrevista que daria continuidade ao levantamento e 1.667 (35,44%) declararam interesse em participar de lista de discussão a respeito do tema. Dentre os principais resultados, destacam-se que 57,79% dos respondentes armazenam parte dos seus dados de pesquisa em papel; 66,80% guardam todos os dados em seu computador pessoal; 55,51% alegam não armazenar nenhuma parte de seus dados em servidor/data center da instituição. A nuvem também está sendo muito utilizada pelos respondentes, como opção para armazenamento de todos os seus dados (33,64%) ou parte deles (52,53%).

Em média, 49,36% dos respondentes nunca utilizaram dados compartilhados por outros grupos; e quase um quarto dos respondentes (23,49%) nunca compartilhou dados de pesquisa, com percentuais que variam de 14,18% na Linguística, Letras e Artes até 35,29% nas Ciências da Saúde. Dentre os principais motivos apontados pelos respondentes para compartilhar dados de pesquisa em acesso aberto, destaca-se a obrigatoriedade para publicação em periódico ou em anais de evento, indicada por 36,60% dos respondentes – com variações relevantes por áreas específicas, como as Ciências Agrárias, na qual o percentual é de 42,43%.

A prática de compartilhamento de dados, de fato, não é tão comum na pesquisa brasileira. Tal efeito fica evidente no baixo percentual de respondentes que indicaram compartilhar todos os seus dados de pesquisa (9,19%), e o alto percentual de respondentes que não compartilham dado algum (37,02%). O compartilhamento de parte dos dados é uma prática apontada por 53,79% dos respondentes. A necessidade de publicar os resultados da pesquisa antes do compartilhamento

foi apontada por 52,73% dos respondentes como motivo para não compartilhar os dados ou compartilhar parcialmente. Outras razões apontadas são a falta de exigência (31,94%), de infraestrutura (31,14%) e financiamento (24,86%).

O uso de repositório para acessar dados de pesquisa de outros pesquisadores foi confirmado por 31,71% dos respondentes, enquanto 68,29% disseram que não os utilizam. Ao serem solicitados a especificar o nome do repositório utilizado, os respondentes, na sua maioria, informaram bases de dados, portais, redes sociais acadêmicas, repositórios institucionais e revistas científicas, o que pode indicar certo desconhecimento por parte dos pesquisadores brasileiros do que seja um repositório de dados de pesquisa.

O compartilhamento total e irrestrito não é aceito pela maioria dos respondentes, tendo em vista que apenas 21,86% concordaram com a afirmativa de que compartilhariam todos os seus dados de pesquisa em um repositório de acesso aberto sem restrições, 42,68% dos respondentes concordaram parcialmente, e 31,71% discordaram dela. Grande parte dos respondentes considera que o compartilhamento de dados pode ampliar o reconhecimento acadêmico, por exemplo, através da citação (69,12% concordam e 24,04% concordam parcialmente com a afirmativa). Estes resultados sem dúvida demonstram que há receptividade à ideia do compartilhamento e reuso de dados por parte dos pesquisadores brasileiros.

A falta de oferta de infraestrutura por parte das instituições brasileiras é evidenciada quando 58,41% dos respondentes afirmaram que não têm à disposição um repositório institucional para compartilhamento de dados de pesquisa. No que diz respeito à oferta de orientação e pessoal especializado no tema, a situação é ainda mais crítica: 71,31% dos respondentes apontaram que a instituição de trabalho nunca oferece orientações, e 78,56% ressaltam que a instituição não oferece pessoal especializado para preparar os dados a serem compartilhados.

## O PORTAL DADOS DE PESQUISA ABERTOS

A elaboração de um portal web para reunir a comunidade nacional de interessados em AADP para discussão sobre o tema iniciou com a provisão de informações institucionais do projeto e divulgação do survey. A segunda fase compreendeu a produção de conteúdo e publicação no site de textos produzidos pela equipe e reprodução de notícias obtidas em sites especializados. A última fase envolveu a criação e a utilização de ferramentas que possibilitam aprimorar as discussões sobre o AADP, por meio de *newsletter* e de perfis no Twitter e Facebook.

A Portal cumpriu seu papel de divulgação durante o primeiro ano do projeto. Embora seu conteúdo continue disponível<sup>16</sup>, suas atividades foram descontinuadas com a criação do Grupo RDA in Brazil, que busca reunir a comunidade brasileira interessada para discutir e promover ações sobre dados de pesquisa.

O RDA in Brazil é um dos 24 grupos nacionais que se organizam no escopo da Research Data Alliance (RDA), iniciativa internacional da Comissão Europeia e de órgãos governamentais dos Estados Unidos e da Austrália, formada em 2013, para fomentar a construção de uma infraestrutura técnica para o compartilhamento e reuso de dados abertos (RESEARCH DATA ALLIANCE, 2016). Para tornar-se membro do RDA e participar das discussões, basta registrar-se no portal e seguir as normas de conduta.

## SOLUÇÕES TECNOLÓGICAS

O quarto objetivo do projeto - levantamento de soluções tecnológicas para compartilhamento de dados - visou identificar soluções tecnológicas (*softwares*) para a construção de repositório para AADP. A busca enfatizou as funcionalidades e recursos que o software oferece para dar apoio ao desenvolvimento de repositórios que venham a prover o compartilhamento de dados segundo os princípios FAIR (GO FAIR, 2019) e que atendam a critérios estabelecidos para repositórios digitais

confiáveis, isto é, repositórios que assumem compromisso e adotam soluções técnicas e de gestão para garantir o acesso em longo prazo aos objetos digitais (preservação digital).

A literatura científica que compara diferentes soluções tecnológicas para repositórios de dados de pesquisa ainda é incipiente. AMORIM *et al.* (2015) traçam um comparativo entre algumas soluções tecnológicas abordando múltiplas características. Os autores estenderam o trabalho totalizando 15 softwares ou serviços, os quais incluem CKAN, DSpace e Dataverse, mas o estudo é restrito a poucos critérios (AMORIM *et al.*, 2017). Mais recentemente, Mahato e Gajbe (2018) comparam Dataverse e CKAN.

Este projeto investigou em profundidade os manuais técnicos e a documentação das soluções DSpace e Dataverse, por serem as mais usadas e por serem adotadas por repositórios que obtiveram certificação de repositório digital confiável, segundo o diretório de repositórios Re3Data. Ambas as soluções são adequadas para o desenvolvimento de repositórios de dados de pesquisa. O Dataverse é um software integrado para publicação, compartilhamento e armazenamento de dados. O DSpace é um software desenvolvido para repositório institucional. O Dataverse possui recursos para configuração de vários tipos de ambientes de repositório. Isso é possível no DSpace, entretanto exige adaptações. Como o Dataverse foi desenvolvido para repositório de dados, a representação e a gestão automatizada dos conjuntos de dados são estruturadas por meio do conceito dataset, que inclui dados, metadados de citação, metadados específicos, documentação adicional, citação, gerenciamento de versões, entre outros. Já o DSpace está estruturado no conceito de coleção de itens. Para gerenciar dados no DSpace, é necessário configurar metadados, fluxos e interfaces de usuário para conduzir a submissão de dados.

Os dois sistemas permitem publicação, compartilhamento e armazenamento de conteúdo digital e possuem recursos bem similares. Eles permitem uma estrutura hierárquica representativa de comunidades diversas, entidades

<sup>16</sup> <https://dadosdepesquisa.rnp.br/>. Acesso em: 10 maio 2019.

autônomas com customizações próprias, chegando à personalização do acesso ao nível do documento ou arquivo digital individual, inclusive licenças de uso e identificadores únicos universais e persistentes: ambos aceitam DOI e Handle System. Apesar de o DSpace ser originalmente destinado à publicação de documentos e o Dataverse à disponibilização de dados de pesquisa, ambos permitem armazenar e gerenciar praticamente qualquer tipo de material digital.

O PostgreSQL é a opção de código aberto para o Sistema de Gerenciamento de Banco de Dados em ambos os softwares. Ainda com relação ao armazenamento, tanto DSpace quanto Dataverse suportam a submissão por máquinas pelo protocolo SWORD e a coleta de registros de metadados pelo protocolo OAI-PMH.

As configurações de submissão também são similares, oferecendo desde autoarquivamento até níveis suplementares de conferência e aprovação, e permitindo que cada comunidade estabeleça suas regras e fluxos. Os sistemas de permissões são bastante flexíveis. Os usuários recebem autorizações que os habilitam a realizar certas funções dentro dos ambientes. As permissões são geralmente agrupadas em papéis que identificam o nível de autoridade do indivíduo ou grupo de usuários. É possível ainda, se assim desejado, liberar acesso de leitura e download para usuários sem cadastro (anônimo). Um recurso imprescindível ao projeto foi a integração com a comunidade acadêmica federada (CAFe) da RNP, usando as credenciais do usuário na instituição de origem. Essa importante funcionalidade é possível em ambos os sistemas de repositório através da configuração do sistema Shibboleth<sup>17</sup>.

A interface gráfica é formada por temas configuráveis podendo ser diferentes para cada comunidade. Também é possível alterar a língua padrão de mensagens do sistema, sendo essa tarefa bem mais fácil na interface XMLUI do DSpace, na qual basta ajustar um parâmetro nos arquivos de configuração e o arquivo de tradução incluído na distribuição ficará disponível no ambiente Web.

Já o Dataverse oferece um arquivo de mensagens que deve ser traduzido na língua desejada. Algumas traduções para determinados idiomas são disponibilizadas pela comunidade de usuários, mas funcionam apenas em versões específicas do sistema.

Os dois softwares suportam Dublin Core para descrição de metadados, sendo possível utilizar outros padrões e até estender padrões existentes. Apesar de ambos gerarem citações, elas podem incluir subconjuntos de observações e/ou variáveis específicas. O DSpace não oferece esse suporte até a versão analisada neste comparativo. Também não oferece versionamento automático dos conjuntos de dados publicados de forma incremental. No requisito embargo de dados é o contrário, o Dataverse ainda não provê uma funcionalidade automatizada, mas informa essa funcionalidade como meta estratégica para o ano de 2020.

Dataverse e DSpace suportam o sistema Lockss<sup>18</sup> para distribuir cópias de conteúdo de forma compartilhada com outros repositórios, aumentando assim a segurança na preservação destes materiais. No quesito recuperação de informações, ambos utilizam o software Solr<sup>19</sup>, uma popular plataforma de código aberto de indexação e busca, que é baseada no Apache Lucene.

Para analisar por completo as soluções tecnológicas, foram elaborados 56 critérios, classificados em Representação do Ambiente do Repositório (AMB1-6), Representação dos Conjuntos de Dados (PAC1-6), Descrição e Documentação dos Conjuntos de Dados (DOC1-11), Produção dos Conjuntos de Dados (SUB1-7), Armazenamento a Longo Prazo e Planejamento da Preservação (PD1-5), Acesso e Uso dos Conjuntos de Dados (AC1-15) e Desenvolvimento, Manutenção e Uso do Software (SW1-6). Esses critérios foram estruturados com base no modelo OAI (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS, 2012) adicionados de critérios relativos ao desenvolvimento e uso de software.

<sup>17</sup> <https://www.shibboleth.net> . Acesso em: 15 set. 2019

<sup>18</sup> <https://www.lockss.org/> . Acesso em: 10 maio 2019.

<sup>19</sup> <http://lucene.apache.org/solr> . Acesso em: 10 maio 2019.

O quadro 1 apresenta comparativamente Dspace e Dataverse com relação aos critérios estabelecidos. Para quantificar o atendimento aos critérios, a seguinte escala foi utilizada: 3 = plenamente ou suficiente; 2 = suficiente com restrições; 1 = parcialmente ou muito parcialmente; 0 = Insuficiente ou não atende. Ao analisar o somatório das avaliações,

confirma-se o equilíbrio entre os softwares para repositório de dados de pesquisa analisados. Resultados completos do comparativo estão publicados em relatório de pesquisa (ROCHA *et al.*, 2018), onde são detalhados os motivos pelos quais cada ferramenta foi classificada em determinado nível na escala de atendimento a cada critério.

Quadro 1 – Comparação entre os principais softwares para implantação de repositórios de dados de pesquisa de acordo com os critérios estabelecidos

Critério de comparação		DS	DV
AMB1	Representação do ambiente	2	3
AMB2	Políticas de funcionamento do ambiente	3	3
AMB3	Políticas descentralizadas	2	3
AMB4	Sistemas de Informação de Pesquisa	1	0
AMB5	Dados Abertos e Ligados	2	0
AMB6	Transparência e feedback	3	3
PAC1	Natureza dos conjuntos de dados	3	3
PAC2	Representar os conjuntos de dados em pacotes	2	3
PAC3	Formatos de arquivos aceitos	2	1
PAC4	Versionamento de conjuntos de dados	2	3
PAC5	Uso de padrões	2	1
PAC6	Dados 5 estrelas	2	2
DOC1	Informação de proveniência	2	3
DOC2	Metadados administrativos	2	2
DOC3	Metadados descritivos	2	3
DOC4	Metadados estruturais	2	1
DOC5	Metadados técnicos	1	1
DOC6	Vocabulários controlados	1	1
DOC7	Definir e estender esquemas de metadados	3	3
DOC8	Recursos para realizar mapeamentos (crosswalks)	3	2
DOC9	Descrever em Linked Data	2	0
DOC10	Descr. integrado com Sistemas de Informação da Pesquisa	2	1
DOC11	Documentação dos dados	1	2
SUB1	Gerenciamento de Acordos de Submissão e Licença	1	1
SUB2	Fluxos de submissão	3	3
SUB3	Transferência legal da custódia dos dados ao repositório	3	3
SUB4	Validação e verificação de pacotes de submissão	2	2
SUB5	Extração, verificação, produção e registro de metadados	1	2
SUB6	Submissão em lote por máquina	3	3
SUB7	Uso de padrões	2	1

(Continua)

Quadro 1 – Comparação entre os principais softwares para implantação de repositórios de dados de pesquisa de acordo com os critérios estabelecidos (Conclusão)

Critério de comparação		DS	DV
PD1	Serviços / microsserviços para garantir acesso em longo prazo	2	1
PD2	Planejamento e ações de preservação digital	1	0
PD3	Integração com serviços de preservação digital	2	2
PD4	Exclusão de dados	3	3
PD5	Usado em Repositório Digital Confiável certificado	3	3
AC1	Recuperação de informação	3	3
AC2	Informações sobre direitos de uso, licenças	3	3
AC3	Informações de proveniência e para uso dos dados	1	3
AC4	Informações de citação	2	3
AC5	Identificadores globais e persistentes	3	3
AC6	Identificadores globais e persistentes para recursos relacionados	2	2
AC7	Acesso aos (meta)dados via identificador e protocolo aberto	3	3
AC8	Restrições de acesso, embargos	3	3
AC9	Gerenciamento e autenticação de usuários	3	3
AC10	Entrega de dados ao consumidor	2	2
AC11	Entrega dos metadados ao consumidor	3	3
AC12	Ferramentas para visualização e análise de dados	0	2
AC13	Estatísticas e relatórios de uso	2	2
AC14	Acesso às descrições em formatos para Linked Data	2	0
AC15	Recuperação em ambiente Linked Data	2	0
SW1	Tecnologia e Plataforma	3	3
SW2	Distribuição e Versionamento	3	3
SW3	Estratégia de Desenvolvimento do Software	2	2
SW4	Licença de Uso	3	3
SW5	Desempenho e Escalabilidade	2	2
SW6	Presença – Usuários, Uso no Brasil	3	1
<b>SOMATÓRIO</b>		<b>123</b>	<b>117</b>

Fonte: Elaborada pelos autores.

Legenda das pontuações

3 = plenamente suficiente;

2 = suficiente com restrições;

1 = parcialmente ou muito parcialmente;

0 = Insuficiente ou não atende.

Ainda foram considerados na análise outras soluções tecnológicas. CKAN é um software desenvolvido para abertura de dados governamentais. Ele permite representar ambientes em que organizações possuem e gerenciam conjuntos de dados, na qual usuários, por meio dos papéis de administrador, editor e membro, são autorizados a acessar, tornar público, editar e remover conjuntos de dados. CKAN possui um pequeno conjunto próprio de metadados e permite a inclusão de campos (pares propriedade:valor) para descrever um conjunto de dados.

Possibilita também a seleção de licenças, controla versões de conjuntos de dados, mantém histórico das modificações e gerencia formatos. O sistema não possui identificador global persistente, nem permite o uso de esquemas de metadados, mas oferece flexibilidade para a criação de extensões. CKAN é uma boa solução quando usado como ambiente de publicação e acesso a dados, em que a submissão é feita por outro sistema. Entretanto, é mais limitado que Dataverse e DSpace no que diz respeito a recursos para organizar unidades e seus dados, para definir políticas de gestão e submissão de dados específicas para cada unidade, assim como no que diz respeito a metadados.

Por fim, Zenodo é um serviço de repositório de dados de pesquisa do CERN, que permite aos pesquisadores compartilhar e preservar resultados de pesquisa em qualquer tamanho, formato e ciência. Zenodo é uma camada de software distribuída sob licença GPLv2 que roda sobre o Invenio<sup>20</sup>, um software livre para a construção de bibliotecas digitais e repositórios de dados e documentos. Em sua documentação, o Zenodo recomenda considerá-lo como inspiração para construção de sistemas de repositório, sendo que a marca e o logo não são cobertas pela licença.

### PROPOSTA DE SOLUÇÃO TECNOLÓGICA

O desenvolvimento de uma proposta de solução tecnológica que atenda a um conjunto de usuários selecionados foi realizado com a construção de um modelo de planejamento com base na metodologia OAIS para identificar a interface entre o produtor e o repositório, nos requisitos FAIR e nos critérios para obtenção de certificação de repositório digital confiável, Core Trust Seal. Foram identificados: comunidade de interesse, produtores, consumidores, gestores, unidades organizacionais e políticas; tipos e estruturas de conjuntos de dados; descrição e documentação dos conjuntos de dados, metadados e padrões; detalhes da produção dos conjuntos de dados e das políticas de submissão; políticas de acesso e uso dos dados; e o armazenamento e critérios de preservação digital, com destaque para o armazenamento em longo prazo, planejamento e execução da preservação digital.

Após a identificação dos detalhes mencionados e dos critérios de comparação elencados na seção anterior, optou-se por testar DSpace e Dataverse. As soluções foram instaladas e configuradas em máquinas virtuais. Uma bateria de testes foi executada visando confirmar a análise dos requisitos levantados e dos recursos oferecidos. Por fim, decidiu-se pela adoção do software Dataverse como base do protótipo de repositório de dados de pesquisa com acesso aberto da RNP, por atender à maior parte dos requisitos identificados no planejamento do repositório.

Foram prototipados Dataverses para armazenar dados de pesquisa produzidos pelo próprio projeto RDP Brasil durante a execução do projeto, dados da comunidade de pesquisadores em Redes de Computadores, e dados da Rede Ipê, que apresentam medições de qualidade de 2012 a 2019 realizadas nos 27 Pontos de Presença (PoPs) da RNP destinados à rede acadêmica brasileira, a qual interliga 1522 câmpus e unidades de instituições de ensino, pesquisa e saúde em todo o país. Atualmente, o repositório está implantado em versão beta<sup>21</sup>. Estão sendo implementados pilotos de repositórios de dados de pesquisa para as seguintes organizações: a RedeRio<sup>22</sup>, rede que interconecta instituições acadêmicas, centros de ensino e pesquisa e órgãos públicos fluminenses; e a Fundação Oswaldo Cruz (FIOCRUZ)<sup>23</sup>, instituição de ciência e tecnologia em saúde, vinculada ao Ministério da Saúde.

### CONSIDERAÇÕES FINAIS

O estudo sobre o acesso aberto a dados de pesquisa no Brasil demonstrou não somente que iniciativas institucionais de compartilhamento de dados abertos de pesquisa no Brasil ainda são poucas, mas também que esforços públicos ainda são necessários para consolidar tais iniciativas. Este estudo possibilitou identificar a necessidade de investigação do AADP no cenário nacional, bem como um interesse da comunidade científica sobre o tema, principalmente pelo excelente retorno do survey da pesquisa.

<sup>20</sup> <https://invenio-software.org/> . Acesso em: 10 maio 2019.

<sup>21</sup> <https://dadosabertos.rnp.br> . Acesso em: 15 set. 2019.

<sup>22</sup> <http://www.rederio.br/> . Acesso em: 15 set. 2019.

<sup>23</sup> <https://portal.fiocruz.br/> . Acesso em: 15 set. 2019.

O estudo foi abrangente na investigação do AADP no Brasil, com contribuições relevantes para iniciativas governamentais que estão surgindo para promover o AADP, como o compromisso de “Estabelecer mecanismos de governança de dados científicos para o avanço da ciência aberta no Brasil”, assumido pelo Brasil com a Parceria de Governo Aberto (OGP), por meio do seu Quarto Plano de Ação.

A aceitação do projeto pela comunidade nacional é comprovada através do elevado número de respondentes do survey, pelas parcerias que vêm se desenvolvendo para construção do protótipo de repositório e pela proposição de um projeto para a segunda fase das atividades do grupo de pesquisa junto à RNP, o qual envolve atividades voltadas à ampliação da comunidade brasileira reunida em torno da temática AADP, implantação de um piloto de repositório de dados abertos (em parceria com outras instituições) e implantação de um piloto de federação para esses repositórios.

## REFERÊNCIAS

- AMORIM, R. C. *et al.* A Comparative study of platforms for research data management: interoperability, metadata capabilities and integration potential. *In: ROCHA, A. et al. (ed.). New contributions in information systems and technologies. Advances in Intelligent Systems and Computin. [S.l.]: Springer, Cham, 2015. p. 101-111. v. 353. DOI: 10.1007/978-3-319-16486-1\_10.*
- AMORIM, R. C. *et al.* A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access in the Information Society, [S.l.]*, v. 16, n. 4, p. 851-862, 2017. DOI: 10.1007/s10209-016-0475-y.
- BEAGRIE, N.; BEAGRIE, R.; ROWLANDS, I. Research data preservation and access: the views of researchers. *Ariadne, [S.l.]*, n. 60, 2009. Disponível em: <http://www.ariadne.ac.uk/issue/60/beagrie-et-al/>.
- CARVALHO, E. R. S.; LEITE, F. C. L. Diferenças na produção, no compartilhamento e no (re)uso de dados de pesquisa: a percepção de pesquisadores de Química, Antropologia e Educação. *Em Questão, Porto Alegre*, v. 25, n. 3, p. 321-347, set./dez. 2019.
- CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. *Reference Model for an Open Archival Information System (OAIS)*. Washington, DC: CCSDS Secretariat, 2012. Disponível em: <https://public.ccsds.org/pubs/650x0m2.pdf>.
- DIGITAL CURATION CENTRE. *Data Asset Framework: Implementation guide. [S.l.:s.n.]*, 2009. Disponível em: [https://www.data-audit.eu/docs/DAF\\_Implementation\\_Guide.pdf](https://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf).
- GO FAIR. *FAIR principles. [S.l.]*, 2019. Disponível em: <https://www.go-fair.org/fair-principles>.
- HENNING, P. C. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. *Em Questão, Porto Alegre*, v. 25, n. 2, p. 389-412, maio/ago. 2019.
- INTERNATIONAL SCIENCE COUNCIL. *Core Trustworthy Data Repositories Requirements v01*. Tokyo: WDS, 2016. Disponível em: [https://www.coretrustseal.org/wp-content/uploads/2017/01/Core\\_Trustworthy\\_Data\\_Repositories\\_Requirements\\_01\\_00.pdf](https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf) Acesso em: 10 ago. 2019
- JANTZ, R.; GIARLO, M. Digital Preservation - Architecture and Technology for Trusted Digital Repositories. *Dlib Magazine, [S.l.]*, v. 11, n. 6, 2006. Disponível em: <http://www.dlib.org/dlib/june05/jantz/06jantz.html>.
- KIM, Y.; ADLER, M. Social scientists' data sharing behaviors: investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management, [S.l.]*, v. 35, p. 408-418, 2015.
- KIM, Y.; ZHANG, P. Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research, [S.l.]*, v.37, n.3, p. 189-200, 2015.

KIM, Y. Fostering scientists' data sharing behaviors via data repositories, journal supplements, and personal communication methods. *Information Processing and Management*, [S.l.], v. 53, p. 871-885, 2017.

MAHATO, S. S.; GAJBE, S. A. Comparative study of open source data repository software: Dataverse and CKAN. *Library Herald*, [S.l.], v. 56, n. 1, p. 36-54, 2018. DOI:10.5958/0976-2469.2018.00005.2.

RESEARCH DATA ALLIANCE. *About RDA*. [S.l.]: RDA, 2016. Disponível em: <https://www.rd-alliance.org/about-rda>.

ROCHA, R. P. *et al.* *Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas: relatório 2018*. Disponível em: <http://hdl.handle.net/10183/185126>.

SAYÃO, L. F.; SALES, L. F. *Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores*. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: [http://www.cnen.gov.br/images/CIN/PDFs/guia\\_de\\_dados\\_de\\_pesquisa.pdf](http://www.cnen.gov.br/images/CIN/PDFs/guia_de_dados_de_pesquisa.pdf).

TENOPIR, C. *et al.* Data sharing by scientists: practices and perceptions. *PLoS ONE*, [S.l.], v. 6, n. 6, e21101, 2011. DOI:10.1371/journal.pone.0021101.

TENOPIR, C. *et al.* Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE*, [S.l.], 2015. DOI:10.1371/journal.pone.0134826.

WALLIS, J. C.; ROLANDO, E.; BORGMAN, C. L. If we share data, will anyone use them? data sharing and reuse in the long tail of science and technology. *PLoS ONE*, [S.l.], v. 8, n. 7, e673322013, 2013. DOI: [doi.org/10.1371/journal.pone.0067332](https://doi.org/10.1371/journal.pone.0067332).

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, [S.l.], v. 3, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>.