

ACESSO ABERTO A DADOS DE PESQUISA NO BRASIL

RELATÓRIO FINAL – 2018



Coordenador

Rafael Port da Rocha (Universidade Federal do Rio Grande do Sul)

Coordenadora adjunta

Sônia Elisa Caregnato (Universidade Federal do Rio Grande do Sul)

Pesquisadores**Universidade Federal do Rio Grande do Sul**

Caterina Groposo Pavão
Paula Caroline Schifino Jardim Passos
Rene Faustino Gabriel Junior
Samile Andréa de Souza Vanz

Universidade Federal de Rio Grande

Eduardo Nunes Borges
Luís Alberto Barbosa Azambuja

Bolsista**Universidade Federal do Rio Grande do Sul**

Iván Andrés Fornos Angues
Victor Andrews Garcia Lima

Contato

e-mail: dadosdepesquisa@rnp.br

site: <https://dadosdepesquisa.rnp.br>

Rua Ramiro Barcelos, 2705 - Campus Saúde

Sala: 106 - Anexo 1

Brasil - Porto Alegre - RS - CEP 90.035-007

Telefone: +55(51)3308.5942



Resumo

O projeto atingiu seus objetivos, obtendo resultados que irão subsidiar ações e políticas de AADP no Brasil. Foi pioneiro em investigar AADP no cenário nacional, trazendo contribuições relevantes para iniciativas governamentais que estão surgindo, como o compromisso assumido pelo Brasil com a iniciativa OGP. Os resultados alcançados levaram a RNP a propor a prorrogação do projeto por um período de mais um ano. A prorrogação justifica-se pela possibilidade de desenvolver novos estudos a partir dos resultados já alcançados. Os resultados dos marcos 1 e 2, que buscaram a identificação das atuais práticas de AADP em instituições brasileiras e o mapeamento de potenciais usuários nacionais de serviços de AADP subsidiando o trabalho desenvolvido nos marcos 4 e 5. Na próxima fase deste projeto, o Marco 3, elaboração de um portal web para reunir a comunidade nacional em AADP, envolverá a ampliação da comunidade de AADP, para acompanhar a evolução da discussão e das iniciativas que vêm ocorrendo no Brasil. Os marcos 4 e 5, que na fase 1, respectivamente, analisaram e prototiparam soluções tecnológicas, ampliam o escopo da experimentação através da implantação de um piloto de repositório AADP e agregador.

Sumário

| | |
|------------------------------------|----|
| 1 Introdução..... | 5 |
| 2 Atividades desenvolvidas | 6 |
| 3 Próxima fase do RDP Braisl | 10 |
| 4 Considerações finais | 11 |

1 Introdução

O compartilhamento de dados da pesquisa é o ato de disponibilizar dados para reuso. Os benefícios mais evidentes são a possibilidade da reprodução ou da verificação da pesquisa; da disponibilização dos resultados de pesquisas financiadas com fundos públicos; e dos questionamentos a respeito dos dados existentes e consequentemente, viabilização de avanços no estado da pesquisa e na inovação. 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.

O projeto RDP-Brasil foi selecionado pela Rede Nacional de Ensino e Pesquisa (RNP) em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) por atender a carta-convite RNP/IBICT de 01/2017, que buscou a seleção de um Grupo de Trabalho (GT) para atuar em parceria na identificação de práticas de Acesso Aberto a Dados de Pesquisa no Brasil e no desenvolvimento de experimentos tecnológicos visando um futuro serviço de AADP.

O projeto foi desenvolvido por pesquisadores ligados ao Centro de Documentação e Acervo Digital da Pesquisa (CEDAP/FABICO) e ao Centro de Ciências Computacionais (C3) da FURG. Teve como objetivo geral realizar um levantamento de informações que subsidiassem as primeiras ações no âmbito do Acesso Aberto a Dados de Pesquisa (AADP) no país, visando à identificação de práticas de AADP e para o desenvolvimento de um futuro serviço de acesso aberto. Seus objetivos específicos foram:

- 1) Identificação das atuais práticas de AADP em instituições brasileiras;
- 2) mapeamento de potenciais usuários nacionais de serviços de AADP e dos principais requisitos e funcionalidades para o atendimento dos mesmos;
- 3) elaboração de um portal web para reunir a comunidade nacional de interessados em AADP para discussão sobre o tema;
- 4) Levantamento comparativo dos serviços e soluções tecnológicas existentes para compartilhamento de dados, com identificação de lacunas e adaptações necessárias para atender aos usuários mapeados;
- 5) Desenvolvimento de uma proposta de solução tecnológica que atenda a um conjunto de usuários selecionados após o mapeamento do objetivo 2.

2 Atividades Desenvolvidas

O projeto teve início em março de 2017, a partir da assinatura do contrato entre RNP, UFRGS e FURG. As atividades previstas foram desenvolvidas pela equipe das duas universidades, em parceria com RNP e IBICT, o que permitiu alcançar os resultados esperados e contribuir para a definição de serviços e estratégias para AADP no Brasil. Ao final desse primeiro ano, a RNP propôs a prorrogação do projeto, a fim de ampliar aos avanços obtidos. Cada um dos cinco objetivos da proposta se transformou em um marco do projeto. As principais entregas de cada marco são apresentadas no Quadro 1. Uma descrição detalhada sobre cada marco está disponível nos respectivos relatórios e no ambiente Wiki da RNP¹.

Quadro 1 – Objetivos e Principais Entregas

| Objetivos | Entregas |
|--|---|
| Identificação das atuais práticas de AADP em instituições brasileiras | Relatório: PAVÃO, C. et al. "Acesso aberto a dados de pesquisa no Brasil: repositórios brasileiros de dados de pesquisa", 2018. Handle: 2050011959/127 Dados da Pesquisa: GT-RDP Brasil, 2018, "Acesso aberto a dados de pesquisa no Brasil : repositórios brasileiros de dados de pesquisa", https://hdl.handle.net/20.500.11959/4 , RNP - Dados Abertos Versão BETA, V2 |
| Mapeamento de potenciais usuários nacionais de serviços de AADP | Relatório: VANZ, S. et al. Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório 2018. Disponível em: handle: 20.500.11959/128 Dados da Pesquisa: GT-RDP Brasil , 2018, "Acesso aberto a dados de pesquisa no Brasil : práticas e percepções dos pesquisadores", https://hdl.handle.net/20.500.11959/5 , RNP - Dados Abertos Versão BETA, V1, UNF:6:8Zp8DHrva8G4nRSM0MOevw== |
| Elaboração de um portal web para reunir a comunidade nacional de interessados em discutir AADP | Portal AADP: Site: https://dadosdepesquisa.rnp.br/ |
| Levantamento de soluções tecnológicas para compartilhamento de dados | Relatório: ROCHA, R. et al. Acesso aberto a dados de pesquisa no Brasil: Soluções Tecnológicas para Compartilhamento de Dados no Brasil repositórios brasileiros de dados de pesquisa. 2018. pdf |
| Desenvolvimento de uma proposta de solução tecnológica que atenda a um conjunto de usuários selecionados | Protótipo: Repositório de Dados de Pesquisa para a Área de Redes de Computadores Site: https://dadosabertos.rnp.br/ |

A seguir, segue um breve relato sobre o desenvolvimento de cada marco e suas entregas.

Marco 1: Identificação das atuais práticas de AADP em instituições brasileiras

Nesta etapa, buscou-se identificar as principais práticas em AADP em instituições brasileiras. Para tal, foram mapeados os repositórios brasileiros que disponibilizam dados de

¹ Diário de Bordo – GT RDP Brasil <https://wiki.rnp.br/display/dpd/GT-RDP+Brasil>

pesquisa, observando características relacionadas à sua abrangência e temática e aos padrões de metadados e softwares disponíveis. Para esse levantamento, foi utilizado como fonte de informação principal o diretório de repositórios Re3data, além de buscas na internet e de indicações feitas pelos respondentes da survey realizado no marco 2.

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 oito dos 15 analisados. Conclui-se 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.

Marco 2: Mapeamento de potenciais usuários nacionais de serviços de AADP

Esse marco visou investigar as práticas e as percepções dos usuários nacionais de serviços de Acesso Aberto a Dados de Pesquisa (AADP). A coleta de dados ocorreu através de uma pesquisa do tipo survey com pesquisadores de todo o país. 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)², Kim e Adler (2015)³, Kim e Zhang (2015)⁴ e do *Digital Curation Center* (2009)⁵.

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 acerca de percepções em relação ao compartilhamento e uso de dados.

Um convite para a participação foi enviado, em março de 2018, a três grupos de endereços: 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 institucio-

2 TENOPIR C. et al. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE, v.6, n.6, e21101, 2011. doi:10.1371/journal.pone.0021101

3 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, v. 35, p. 408-418, 2015.

4 KIM, Y.; ZHANG, P. Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. Library & Information Science Research, v.37, n.3, p. 189-200, 2015.

5 DIGITAL CURATION CENTRE. Data Asset Framework: Implementation guide [online], 2009. Available from: https://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf

nais e de coordenadores dos PPGs cadastrados na CAPES; 249 e-mails de coordenadores de Institutos Nacionais de Ciência e Tecnologia. O número total de respondentes foi 4.703. Os principais resultados demonstram que, apesar de haver um grande interesse pelo tema, há equívocos em relação ao que significa compartilhamento e reuso de dados de pesquisa. Observou-se que a metade dos respondentes (49,36%) nunca havia utilizado dados compartilhados por outros grupos e quase um quarto dos respondentes (23,49%) nunca compartilhou dados de pesquisa. O compartilhamento de parte dos dados é uma prática apontada por 53,79% dos respondentes. Constatou-se ainda que mais da metade deles (58,41%) não tem a sua disposição um repositório institucional para compartilhamento de dados de pesquisa. Conclui-se que a ideia de compartilhamento total e irrestrito ainda não é aceita pelos respondentes, apesar de haver receptividade à ideia do compartilhamento e reuso de dados por parte dos pesquisadores brasileiros.

Marco 3: Elaboração de portal web para reunir a comunidade nacional para discutir AADP

Este marco visou a criação e operacionalização de um portal⁶ para reunir a comunidade de interessados para discutir AADP. A elaboração do portal envolveu levantamento de requisitos, identificação da ferramenta (CMS) adequada, definição da arquitetura da informação, implementação do portal e incorporação dos conteúdos.

Foram analisadas cinco softwares CMS, Wordpress, Joomla, Drupal, TextPattern e Radiant. Wordpress foi o escolhido a partir da consideração de critérios como: curva de aprendizado; comunidade de usuários e desenvolvedores; quantidade, qualidade e segurança de *plug-ins*; facilidade de customização do site; e velocidade de carregamento das páginas.

A implantação do portal iniciou com a provisão de informações institucionais do projeto e divulgação do *survey*. A segunda fase compreendeu na produção de conteúdo e publicação no site de textos produzidos pela equipe, por meio de buscas em sites área. A última fase, envolveu a criação e a utilização de ferramentas que possibilitam aprimorar as discussões sobre o AADP, através de *newsletter* e de perfis no Twitter⁷ e Facebook⁸. Os perfis possibilitam o estabelecimento de relações com usuários desses aplicativos. A incorporação dos conteúdos teve como estratégia principal ter uma produção constante e consistente de conteúdos, a fim de que usuários passem seguir e contribuir como o portal. Nesse sentido, o portal adotou como estratégia a produção de duas postagens por semana, realizada através do acompanhamento de notícias e novidades na área.

Marco 4: Levantamento de soluções tecnológicas para compartilhamento de dados

Esse marco visou identificar soluções tecnológicas (software) 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 comparti-

⁶ Portal Dados de Pesquisa - <https://dadosdepesquisa.rnp.br/>

⁷ Portal Dados de Pesquisa no Twitter - <https://twitter.com/dadosdepesquisa>

⁸ Portal Dados de Pesquisa no Facebook - <https://www.facebook.com/dados.depesquisa>

lhamento de dados segundo os princípios FAIR⁹ e que atendam a critérios estabelecidos para repositórios digitais confiáveis¹⁰, isto é, repositórios que assumem compromisso e adotam soluções técnicas e de gestão para garantir o acesso a longo prazo aos objetos digitais (preservação digital).

Para analisar as soluções tecnológicas, foram elaborados 56 critérios, classificados em Representação do Ambiente do Repositório (6), Representação dos Conjuntos de Dados (6), Descrição e Documentação dos Conjuntos de Dados (11), Produção dos Conjuntos de Dados (7), Armazenamento a Longo Prazo e Planejamento da Preservação (5), Acesso e Uso dos Conjuntos de Dados (15) e Uso, Desenvolvimento e Manutenção do Software (6). Esses critérios foram estruturados com base no modelo OAIS¹¹, adicionados de critérios relativos ao desenvolvimento e uso de software.

Foram investigadas em profundidade as soluções DSpace e Dataverse, por estas serem as soluções mais usadas e por serem adotadas por repositórios que obtiveram certificação de repositório digital confiável, segundo o diretório de repositórios Re3Data¹². 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 através do conceito dataset, que inclui dados, metadados de citação, metadados específicos, documentação adicional, citação, gerenciamento de versões, etc. Já o DSpace está estruturado no conceito de coleção de itens. No caso do uso do DSpace para gerenciar dados, é necessário configurar metadados, fluxos e interfaces de usuário para conduzir a submissão de dados.

Marco 5: Desenvolvimento de uma proposta de solução tecnológica que atenda a um conjunto de usuários selecionados

A construção do protótipo iniciou com a elaboração de um modelo para desenvolver o planejamento do repositório. Esse modelo de planejamento foi elaborado 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.

⁹ Princípios que indicam que dados devem ser localizáveis (Findable), acessíveis (Accessible), interoperáveis (Interoperable) e reusáveis (Reusable). <https://www.force11.org/group/fairgroup/fairprinciples>

¹⁰ Repositório que aceita a responsabilidade pela manutenção a longo prazo dos recursos digitais, projeta seus sistemas de acordo com as convenções e os padrões comumente aceitos, estabelece metodologias para avaliação que atendem às expectativas de confiabilidade da comunidade, cumpre suas responsabilidades com depositantes e usuários de forma aberta e explícita, e permite que sejam auditadas e medidas suas políticas, práticas e desempenho. (Jantz, R., Giarlo, M.: Digital Preservation - Architecture and Technology for Trusted Digital Repositories. 2006 - <http://www.dlib.org/dlib/june05/jantz/06jantz.html>)

¹¹ Open Archival Information System - Modelo de Referência para Repositórios, sendo base para estudos e pesquisas em preservação digital e construção de repositórios digitais confiáveis. (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. Producer-Archive Interface Methodology Abstract Standard. CCSDS 651.0-M-1. 2004 <https://public.ccsds.org/Pubs/651x0m1.pdf>)

¹² Diretório de Repositórios de Dados Re3Data - <https://www.re3data.org/browse>

O repositório foi planejado através do modelo criado. Foram definidos o seu ambiente (comunidade, produtores, consumidores, gestores, unidades organizacionais, políticas), os conjuntos de dados (tipos e estruturas de conjuntos de dados), a descrição e a documentação dos conjuntos de dados (documentação, metadados, padrões), a produção dos conjuntos de dados (políticas de submissão), o acesso e uso dos conjuntos de dados (políticas de acesso e uso), e o armazenamento e o planejamento da preservação digital (armazenamento a longo, planejamento e execução da preservação digital).

O protótipo foi desenvolvido para a comunidade de redes de computadores e está apoiado no software Dataverse, por este atender a maior parte dos requisitos identificados no planejamento do repositório. O ambiente do repositório permite que grupos ou unidades organizacionais possam ter autonomia para criar e gerenciar as estruturas e as políticas de submissão, acesso e uso dos seus conjuntos de dados. Duas unidades foram criadas e implementadas: Rede-IPE e GT-RDP Brasil. A primeira é ligada à produção dos conjuntos de dados referentes às monitorações periódicas da Rede Ipê. O GT-RDP Brasil armazena os dados dos estudos realizados pelo GT-RDP Brasil. Para este último, inicialmente foram criados dois conjuntos de dados, referentes aos dois primeiros objetivos/marcos do projeto: “Repositórios brasileiros de dados de pesquisa” e “Práticas e percepções dos pesquisadores”. As soluções tecnológicas Dataverse e DSpace foram avaliadas operacionalmente, através de instalações e configurações em máquinas virtuais.

3 Próxima Fase do RDP Brasil

Os resultados alcançados levaram a RNP a propor a prorrogação do projeto por um período de mais um ano, denominando-a de Fase 2. Essa prorrogação justifica-se pela possibilidade de desenvolver novas pesquisas a partir dos resultados já alcançados. Planeja-se na nova fase do projeto, avançar na implementação de um piloto com mais entidades participando, ampliando a proposta. A fase 2 será dividida em três etapas, relacionadas aos marcos da fase anterior. A primeira delas, relacionada aos marcos 1, 2 e 3, envolverá a ampliação da comunidade de AADP no Portal, para acompanhar a evolução da discussão e das iniciativas que vêm ocorrendo no Brasil e no mundo. Os marcos 4 e 5, que na primeira fase analisaram e prototiparam soluções tecnológicas, ampliam o escopo da experimentação através da implantação de um piloto de repositório AADP e de um agregador, e estão representados nas duas etapas seguintes.

As atividades a serem desenvolvidas na Fase 2 serão:

| Marco | Etapas | Entregas |
|--|--|--|
| Identificação das atuais práticas de AADP em instituições brasileiras | Operação do Portal Web e Ampliação da Comunidade | Relatórios |
| Mapeamento de potenciais usuários nacionais de serviços de AADP | | |
| Elaboração de um portal web para reunir a comunidade nacional de interessados em discutir AADP | | |
| Levantamento comparativo dos serviços e soluções tecnológicas para compartilhamento de dados | Implantação de um piloto de um repositório de dados abertos <ul style="list-style-type: none"> ■ Identificação das novas entidades produtoras ■ Planejamento do Repositório para as comunidades identificadas – ambiente(s), conjuntos de dados, produção, acesso, uso e armazenamento dos dados ■ Implantação das comunidades identificadas no Repositório | Entrega da Documentação Técnica (Planejamento do Repositório e Documentação de Instalação e Uso) |
| Desenvolvimento de uma proposta de solução tecnológica que atenda a um conjunto de usuários selecionados | Implantação de um piloto de federação de repositório de dados abertos <ul style="list-style-type: none"> ■ Levantamento de requisitos para a federação ■ Desenvolvimento de agregador piloto | Entrega da Documentação Técnica (Planejamento do agregador e Documentação de Instalação e Uso) |

4 Considerações Finais

O grupo RDP realizou as atividades propostas, produziu as entregas prevista e atingiu os objetivos previstos no projeto, proporcionando resultados que poderão subsidiar ações e políticas de AADP no Brasil para repositórios de dados, interoperabilidade entre repositórios, compartilhamento de dados por pesquisadores, formação de comunidade e proposição de soluções tecnológicas.

O projeto foi pioneiro na investigação do AADP no cenário nacional, 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 frente à Parceria de Governo Aberto (OGP), através do seu Quarto Plano de Ação.

