

Planejamento e incubação de repositórios de dados em ICTs brasileiras

CATERINA GROPOSO PAVÃO

Universidade Federal do Rio Grande do Sul (UFRGS) (Brasil)

caterina@ufrgs.br

CAROLINA HOWARD FELICISSIMO

RNP

carolina.felicissimo@rnp.br

RAFAEL PORT DA ROCHA

Universidade Federal do Rio Grande do Sul (UFRGS) (Brasil)

rafael.rocha@ufrgs.br

RENE FAUSTINO GABRIEL JUNIOR

Universidade Federal do Rio Grande do Sul (UFRGS) (Brasil)

rene.gabriel@ufrgs.br

SAMILE ANDREA DE SOUZA VANZ

Universidade Federal do Rio Grande do Sul (UFRGS) (Brasil)

samile.vanz@ufrgs.br

SÔNIA ELISA CAREGNATO

Universidade Federal do Rio Grande do Sul (UFRGS) (Brasil)

sonia.caregnato@ufrgs.br

RESUMO

O texto relata as atividades do GT Rede de Dados de Pesquisa (GT-RDP) no âmbito do Edital de 1/06/2021, lançado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), o

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a Rede Nacional de Ensino e Pesquisa (RNP). O edital abrange a implementação de um repositório de dados com por quatro instituições, Centro Brasileiro de pesquisas Físicas (CBPF), Universidade Federal do Ceará (UFC), Universidade Federal de Goiás (UFG) e Universidade Federal Rural do Amazonas (UFRA). O objetivo geral da equipe do GT-RDP, responsável pela execução do projeto, foi atender as quatro Instituições de Ciência, Tecnologia e Inovação (ICTs) nas atividades desenvolvidas nas vertentes tecnológica e conceitual, assim como desenvolver atividades de capacitação. A prática de atendimento das quatro ICTs aconteceu por meio de reuniões semanais, onde foram realizadas orientações e direcionamento para a implementação dos repositórios. Utilizou-se o software Dataverse, com base no modelo OAIS e princípios FAIR. O momento de finalização das atividades é dedicado à definição de políticas para funcionamento do repositório. O projeto encontra-se em fase final dos testes e em breve passando para a fase de produção e disponibilização pública.

PALABRAS CLAVE

Ciência aberta; Dados de Pesquisa; Repositórios de dados de pesquisa.

Open Science; Research Data; Research Data Repositories.

Introdução

Com base nos princípios essenciais de liberdade acadêmica, integridade nas pesquisas e excelência científica, a ciência aberta estabelece um novo paradigma que integra práticas científicas de reprodutibilidade, transparência, intercâmbio e colaboração resultante da maior abertura de conteúdos científicos, ferramentas e processos, com o objetivo de tornar o conhecimento científico disponível, acessível e reutilizável, para aumentar a colaboração

científica e o compartilhamento de informações e para abrir os processos de criação, avaliação e comunicação do conhecimento científico (UNESCO, 2021). Os dados de pesquisa abertos são um dos pilares da ciência aberta, usados como fontes primárias para apoiar as pesquisas e a produção do conhecimento, são utilizados como evidências no processo de pesquisa e são comumente aceitos na comunidade científica como necessários para validar as descobertas e resultados da pesquisa (RESEARCH, 2022).

A necessidade de gestão dos dados de pesquisa contribuiu para a criação dos repositórios de dados que, na sua maioria, estão vinculados às universidades e instituições de pesquisa e contribuem para assegurar que os dados sejam publicados e disponibilizados para a comunidade científica com o mínimo de restrições possíveis. Os repositórios de dados de pesquisa são ambientes digitais implementados com infraestrutura para dar suporte aos pesquisadores na gestão e na disponibilização de dados o que potencializa a reutilização por outros pesquisadores.

As universidades internacionais têm implementando repositórios de dados para armazenar e disponibilizar os dados de pesquisa da sua comunidade. As universidades brasileiras, alinhadas com essa tendência, seguem nessa direção. Conforme levantamento realizado por GABRIEL JUNIOR et al. (2019), foram identificados no Brasil 15 repositórios de dados de pesquisa, 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. As temáticas atendidas por esses repositórios são voltadas às Ciências Sociais Aplicadas, às Ciências Agrárias, às Ciências Exatas e da Terra, às Ciências Biológicas e à Engenharia. Atualmente no RE3data¹ estão registrados 14 repositórios brasileiros.

¹ [https://www.re3data.org/search?query=&countries\[\]=BRA](https://www.re3data.org/search?query=&countries[]=BRA)

Baseados nesse contexto o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a Rede Nacional de Ensino e Pesquisa (RNP), por meio do Edital de 1/06/2021, convidaram Instituições de Ciência, Tecnologia e Inovação (ICTs) a colaborar para o fomento à Ciência, Tecnologia e Inovação por meio da produção, socialização e integração do conhecimento científico e tecnológico. A ideia foi realizar a colaboração por meio do planejamento e implementação de ações de pesquisa e desenvolvimento (P&D) e transferência de conhecimentos relacionados à criação de repositórios de dados de pesquisa. A equipe de execução do projeto foi composta pelos pesquisadores do GT-RDP.

O GT-RDP é responsável pelo projeto Rede de Dados de Pesquisa Brasileira (RDP Brasil), que 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). Em 2017 o grupo iniciou suas atividades, com o objetivo inicial de realizar um levantamento de informações que subsidiassem as primeiras ações no âmbito do Acesso Aberto da Dados de Pesquisa (AADP) no país, visando à identificação de práticas de AADP e o desenvolvimento de um futuro serviço de acesso aberto. Até 2019 o GT- RDP contou com a contribuição do Centro de Ciências Computacionais da Universidade do Rio Grande (C3/FURG).

A execução dos objetivos iniciais do projeto permitiu reunir informações acerca do contexto brasileiro. O grupo desenvolveu uma survey com 4.676 respondentes (CAREGNATO *et al.*, 2019), revelando que, a pesar do grande interesse pelo tema, evidenciado pela prevalência de variáveis relacionadas ao compartilhamento e ao uso de dados e aos repositórios institucionais, não há clareza por parte dos pesquisadores sobre os principais tópicos relacionados. Em média, 49,36% dos respondentes nunca utilizou 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. Entende-se que a gestão e a disponibilização de dados de pesquisa é a chave para que esses possam ser utilizados por pesquisadores brasileiros ou estrangeiros em futuras pesquisas. Apesar de 57,79 % dos respondentes armazenarem parte dos seus dados em arquivos em papel e 66,80 % guardarem todos os dados em seu computador pessoal, há um entendimento geral de que os dados devam ser compartilhados. Fica evidente o problema de infraestrutura para o armazenamento e o compartilhamento dos dados de pesquisa nas instituições brasileiras ao analisar que 71,31 % dos respondentes apontaram que a instituição de trabalho nunca oferece orientações sobre o tema e que 78,56 % deles responderam também que a instituição não oferece pessoal especializado para preparar os dados a serem compartilhados.

Munidos de diversas informações, o grupo passou a dedicar-se ao estudo de estratégias de planejamento para implantação de repositórios, preservação digital e certificação. Este trabalho relata as atividades de planejamento e incubação de repositórios em quatro ICTs atendidas pelo Edital de 1/06/2021, conforme detalhado a seguir.

O projeto

Considerando a crescente demanda de compartilhamento de dados de pesquisa motivada pelo novo paradigma da Ciência Aberta, as diretrizes produzidas pelo Grupo de Trabalho formado para discutir uma proposta de Política Nacional para Ciência Aberta, as recentes recomendações da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) sobre acesso a dados de pesquisas científicas com financiamento público e o trabalho realizado pelo IBICT e RNP no âmbito do Compromisso três do 4º Plano de Ação Nacional para Governo Aberto (OGP), o Plano de Trabalho executado teve como objetivo principal o apoio à implantação de repositórios de dados de pesquisa nas quatro ICTs selecionadas pelo Edital. Como compromisso, cada partícipe deveria aplicar recursos humanos e financeiros,

computacionais, assinaturas ou licenças de software, dentre outros, para a execução do plano de trabalho.

Para pôr em prática o plano de trabalho as ICTs se comprometeram a implementar, até o final do projeto piloto, um repositório de dados com, pelo menos, três conjuntos de dados de pesquisadores que estivessem dispostos a colaborar com a iniciativa. As instituições selecionadas e a área dos datasets foram:

- 1. Centro Brasileiro de pesquisas Físicas (CBPF)**, o piloto do repositório de dados de pesquisa incubado conterá datasets, previamente selecionados, dos laboratórios: a) Laboratório Multiusuário de Nanociência e Nanotecnologia – LABNANO; b) Laboratório de Ressonância Magnética Nuclear – RMN; c) Laboratório de Raios X - RX e Nanomagnetismo e d) Espectroscopia Mössbauer. Laboratório de Magnetismo Aplicado. A equipe é formada por um físico e pesquisador do Centro, juntamente com uma bibliotecária e um responsável pelo setor de TI.
- 2. Universidade Federal do Ceará (UFC)**, o piloto do repositório de dados conterá datasets das seguintes áreas: a) Engenharia de Produção, Inteligência Computacional e Otimização; b) Fisioterapia, Tecnologias assistivas, Bioengenharia, Engenharia Biomédica; c) Letras, Estudos da Língua Inglesa, Literatura, Tradução, Fonética e Multilinguismo e d) Finanças, Economia, Engenharia de Transportes. A equipe é formada por uma bibliotecária.
- 3. Universidade Federal de Goiás (UFG)**, o piloto do repositório de dados de pesquisa conterá datasets, das seguintes áreas: a) Ciências Biológicas-Botânica / Ciências Agrárias-Fitotecnia-Tecnologia de Sementes / Ciências Agrárias-Engenharia Florestal e Recursos Florestais-Silvicultura; b) Ciências da Saúde, Cardiologia; c) Ciências Sociais, Sociologia e d) Ciências Humanas, Antropologia, Arqueologia e Museologia. A equipe é formada por professores do Departamento de Ciência da Informação,

gerentes de TI, e professores vinculados a administração central da Universidade.

- 4. Universidade Federal Rural do Amazonas (UFRA)** o piloto do repositório de dados de pesquisa conterà datasets das seguintes áreas: a) Fisiologia Animal - Reprodução de Primata na Amazônia legal Fisiologia Reprodutiva de Animais Selvagens e Domésticos na Amazônia; b) Recuperação de Áreas Degradadas, Manejo florestal em concessões federais no Pará, Ecologia, manejo e produção florestal das várzeas do Marajó, Avaliação de componentes químicos de diferentes espécies florestais armazenadas, passíveis de contaminação no solo e c) Reprodução animal- Desenvolvimento de um novo diluente crioprotetor a base de castanha-do-pará /brasil, Suplementação para reduzir a idade a puberdade de novilhas zebuínas em pastagens tropicais. A equipe é formada por uma bibliotecária, por um responsável pela TI e pela superintendente a rede da instituição.

Execução do projeto

O objetivo geral da equipe do GT-RDP, responsável pela execução do projeto, foi atender as quatro ICTs nas atividades desenvolvidas nas vertentes tecnológica e conceitual, assim como desenvolver atividades de capacitação. Mais especificamente, proporcionar suporte tecnológico e conceitual para execução das atividades; capacitar as ICTs na vertente conceitual e tecnológica e participar em atividades de disseminação dos resultados para criação de repositórios de dados. Ainda, como objetivos secundários, apoiar instituições de ensino e pesquisa brasileiras na abertura de dados de pesquisa; fomentar o debate acerca da importância do desenvolvimento de repositórios de dados de pesquisa em instituições brasileiras; apoiar atividades de P&D e experimentação em repositório de dados de pesquisa a partir de estudos de caso com datasets de diferentes áreas do conhecimento; apoiar ações da RNP no processo de descoberta e estruturação de serviços para a implantação de

repositórios de dados de pesquisa; apoiar ações nacionais voltadas para a elaboração de uma política para ciência aberta; ampliar as instituições participantes do Consórcio Nacional para Ciência Aberta (CoNCienciA) com a inclusão das ICTs selecionadas, consórcio que fornecerá identificadores persistentes associados a seus conjuntos de dados (DOI); e finalmente, planejar a estruturação de uma Rede Federada Nacional de Repositórios de Dados de Pesquisa, com a participação dos repositórios de dados do IBICT, CNPq, RNP e de um conjunto de instituições de pesquisa convidadas, entre elas as ICTs selecionadas por meio do edital.

Para alcançar esses objetivos foi necessário capacitar as equipes das ICTs selecionadas em diversos aspectos relacionados à gestão de repositórios de dados de pesquisa, sobre a instalação, manutenção, operação, utilização e governança de repositórios de dados de pesquisa. Relatam-se a seguir alguns dos princípios que guiaram esse processo. A implantação de um repositório de dados é orientada por uma série de modelos de referência e princípios, como o Modelo de Referência para Repositórios, Open Archival Information System (OAIS) e os princípios de compartilhamento FAIR e de citação. Além disso, para o desenvolvimento de um repositório, é importante considerar critérios para construção de repositórios digitais confiáveis.

OAIS é um modelo de referência que proporciona uma linguagem comum para facilitar a discussão entre comunidades interessadas no desenvolvimento de repositórios com responsabilidade em preservação digital (Consultive..., 2012). É composto pelas especificações do ambiente do repositório, do modelo de dados das informações a serem preservadas e do modelo funcional do repositório. Um Repositório confiável é aquele cuja missão é fornecer acesso de longo prazo a recursos digitais gerenciados. Um repositório com essas características 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 de sistemas 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. Critérios para verificação e certificação de repositórios digitais confiáveis têm como base e medem a proximidade do repositório com o Modelo de Referência OAIS. Os principais instrumentos para certificação de repositórios digitais são Core Trusted Seal (Becker,2009) e ISO 16363 27.

A garantia de armazenamento e acessibilidade dos dados a longo prazo é uma exigência importante e nesse sentido, adoção dos princípios TRUST (Transparency, Responsibility, User focus, Sustainability and Technology) foi trabalhada com as quatro ICTs atendidas pelo projeto, conforme mostra o quadro 1:

PRINCÍPIO	ORIENTAÇÃO PARA OS REPOSITÓRIOS
T (transparência)	Ser transparente sobre os serviços de repositório específicos e acervos de dados
R (responsabilidade)	Ser responsável por garantir a autenticidade e integridade dos acervos de dados e pela confiabilidade e persistência de seu serviço.
U (foco no usuário)	Para garantir que as normas de gerenciamento de dados e expectativas das comunidades de usuários sejam atendidas.
S (sustentabilidade)	Para sustentar serviços e preservar acervos de dados para o longo prazo.
T (tecnologia)	Para fornecer infraestrutura e recursos para suporte a serviços seguros, persistentes e confiáveis.

QUADRO 1. Princípios TRUST para repositórios digitais
Fonte: Lin et al. (2020)

Os Princípios TRUST se constituem como um meio para agilizar a comunicação com todas as partes interessadas e servem como um guia para que os repositórios demonstrem transparência, responsabilidade, foco no usuário, sustentabilidade e tecnologia (LIN et al., 2020). Aliados a eles, os princípios FAIR e de citação orientam no sentido de promover o compartilhamento dos dados (FORCE11, 2017). Os princípios FAIR indicam que

dados devem ser localizáveis (Findable), acessíveis (Accessible), interoperáveis (Interoperable) e reusáveis (Reusable). Para dados serem FAIR, necessitam identificadores únicos, persistentes e globais (F). Devem ser descritos por metadados indexáveis e ricos (F), representados em linguagens formais (I), aceitos pela comunidade (R), com atributos relevantes, precisos e úteis ao contexto (R), incluindo metadados de proveniência (R) e usando vocabulários controlados que seguem princípios FAIR (I). Estes dados devem ser recuperáveis pelo seu identificador através de um protocolo de comunicação padronizado, aberto e gratuito (A). Também devem ser acompanhados de licenças claras e acessíveis (R), e referências qualificadas devem ligar (meta)dados para enriquecer o conhecimento sobre os mesmos (I).

Identificadores persistentes viabilizam a referência de longa duração aos objetos digitais, à medida que permitem que estes objetos sejam identificados de forma independente de suas localizações. O Digital Object Identifier (DOI) e Handle System são os principais serviços de identificadores persistentes usados em repositórios digitais. DOI é um sistema que agrega metadados ao Handle System, permitindo a adição de semântica aos identificadores. Especifica um conjunto nuclear de metadados, que podem ser expandidos com novos elementos para atender domínios específicos. O DOI possui um modelo de negócio em que o registro é realizado por meio de uma federação de agências. Essas agências atendem a domínios específicos, e usam esquemas de metadados específicos para atender às necessidades desses domínios.

O software Dataverse foi utilizado como solução tecnológica para repositório de dados de pesquisa em uma prova de conceito de instanciação do modelo OAIS para o objeto de estudo proposto. A escolha do software Dataverse baseia-se, principalmente: no estudo comparativo de soluções tecnológica para compartilhamento de dados realizado pelo GT-RDP (Rocha, 2018); na aceitação e direcionamento do IBICT pelo uso do software Dataverse; e, pela experiência adquirida pelas equipes da RNP e IBICT na instalação e operação do Dataverse, que será usada para o suporte das atividades previstas do projeto.

De acordo com ROCHA (2018), o Dataverse possui recursos para configuração de vários tipos de ambientes de repositório, incluindo hierarquias organizacionais e políticas de gestão distintas para unidades ou grupos, incluindo esquemas de metadados e licenças e por isso apresentou-se como a melhor opção entre os softwares analisados no estudo para a implementação de repositórios de dados de pesquisa.

Alguns esquemas de metadados já vêm pré-definidos na instalação no Dataverse, compostos por metadados de citação e outras variáveis. O Dataverse disponibiliza os esquemas para as áreas de: Ciências Sociais e Humanidades, Geoespaciais, Astronomia e Astrofísica, Ciências da Vida, Revistas (Journals). Esses esquemas foram criados pensando no Data Documentation Initiative (DDI) e DataCite. O Esquema de Citação é obrigatório, embora nem todos os metadados dentro desse esquema sejam obrigatórios. É possível manter um conjunto mínimo mandatório do esquema e citação, composto pelo Identificador, Criador (Autor), Título, Publicador, Ano de publicação e tipo de recurso. É possível mesclar o Esquema de Citação com os específicos das áreas do conhecimento previstas, criando um perfil de aplicação. Por exemplo: Esquema de Citação + Esquema de Ciências Sociais e Humanidades. Como os campos do Esquema de Citação são pré-definidos, eles já possuem mapeamento com o DDI (crosswalks). Os metadados do Esquema de Citação no Dataverse incluem aqueles que descrevem os dados mais gerais de um documento. Dentre vários outros possíveis, elencou-se a seguir 25 elementos para exemplificar:

METADADOS NO DATAVERSE	DESCRIÇÃO EM PORTUGUÊS
title	Título
otherId	Outra identificação
author	Autor
datasetContact	Contato do conjunto de dados
dsDescription	Descrição
subject	Assunto
keyword	Palavra-chave
topicClassification	Classificação de tópicos
publication	Publicação
notesText	Notas de texto
Language	Idioma
producer	Produtor
contributor	Contribuidor
grantNumber	Número de concessão
distributorName	Nome do distribuidor
depositor	Depositante
dateOfDeposit	Data do depósito
timePeriodCovered	Cobertura temporal
dateOfCollection	Data de coleta
kindOfData	Tipo de dados
series	Série
software	Programa
relatedMaterial	Material relacionado
dataSources	Fontes de dados
accessToSources	Acesso a fontes

QUADRO 2. Esquema de citação do Dataverse. Fonte: Criação do GT-RDP

A capacitação das quatro ICTs é realizada por meio de reuniões semanais, onde acontece a apresentação de boas práticas e fundamentos teóricos, além da discussão das atividades práticas e de experimentação desenvolvidas pelo grupo. Os atendimentos seguem uma vertente conceitual, onde foram discutidos, entre outros temas, a estrutura de organização do repositório, os metadados de descrição, definição de papéis e grupos, fluxos de submissão, políticas para o repositório, identificadores persistentes e requisitos de certificação.

As questões técnicas de instalação e configuração do software Dataverse foram contempladas em atendimentos da vertente tecnológica. Ao longo deste processo, desenvolveu-se um checklist dos itens básicos a serem operacionalizados na instanciação de um repositório:

<input type="checkbox"/>	Ambiente de instalação (min. 4 núcleos, 8Giga RAM, 1Tera disco)
<input type="checkbox"/>	Pré-requisitos (banco de dados PostGres, Java 11, plataforma de pesquisa SOLR, Servidor de aplicativo PAYARA, JQ para Json, ImageMagick para imagens e RServer para dados estatísticos)
<input type="checkbox"/>	Definição da URL de acesso (http)
<input type="checkbox"/>	Serviço de monitoramento (Ex: Google Analytics)
<input type="checkbox"/>	Instalação de uma chave de segurança
<input type="checkbox"/>	Servidor para envio de e-mail (SMTP) – podendo ser servidor externo como o Google
<input type="checkbox"/>	Serviço de Identificador persistente (DOI do DataCite ou Handle), deve-se definir se a atribuição do DOI será com conjunto de Datasets ou por Arquivos (gera maior número de DOIs)
<input type="checkbox"/>	Definição da forma de autenticação (bultin: do próprio Dataverse ou Shibboleth – federada)
<input type="checkbox"/>	Rotinas de backup e restauração do Dataverse (Postgres e Payara – arquivos)
<input type="checkbox"/>	Customizações

<input type="checkbox"/>	Customização da Página inicial (opcional)
<input type="checkbox"/>	Página de políticas
<input type="checkbox"/>	Página do Guia para usuários
<input type="checkbox"/>	Logo da instância (altura recomendada até 45pixel)
<input type="checkbox"/>	Criação da página sobre
<input type="checkbox"/>	Implementação dos idiomas de navegação
<input type="checkbox"/>	Customizações (adicionais)
<input type="checkbox"/>	Perfil de aplicação – perfil de metadados personalizado (<i>Schema</i>)
<input type="checkbox"/>	Configurações para armazenamento elástico em nuvem (AWS, Azure, Google)

QUADRO 3. Checklist para implementação tecnológica

Fonte: Criação do GT-RDP

Nas configurações do Dataverse ressalta-se que o registro de um Dataset pode ocorrer de duas maneiras, uma com a identificação de cada arquivo e sua versão, o que na prática gera uma quantidade grande de registros, por consequência aumentando o seu custo², ou seja, ao registrar um Dataset com 1.000 arquivos, serão gerados 1.001 DOIs diferentes, um para cada arquivo e outro para o Dataset. A outra forma é atribuição para o Dataset, tendo os arquivos herdados o DOI do Dataset, gerando apenas um DOI, menor custo. Essa opção deve ser avaliada pelos gestores na implementação do repositório.

A estratégia de instanciamento do Dataverse foi o uso de dois servidores, uma para os testes denominado de desenvolvimento e outro de produção com acesso público, no servidor de desenvolvimento foi utilizado do DOI FAKE³. Observou-se que a se atender os requisitos e a obtenção de das informações previamente, a instalação pelas ICTs ocorreu em três semanas. A maior

² A cobrança do DOI pelo DataCite é realizada previamente com a contratação de pacotes anuais, com valor próximo de 0,80 euros por registro de DOI. <https://datacite.org/feemodel.html>

³ DOI registrado no ambiente de testes do DATACITE

dificuldade foi na definição das pessoas que compõem a equipe tecnológica, pois no início houve uma grande rotatividade. Outro ponto a destacar está na documentação dos problemas enfrentados na instalação e principalmente nos registros de senhas que deve estar disponível para os gestores do projeto.

Ressalta-se também a vertente tecnológica depende muito da vertente teórica para as definições da instalação e customização, e que reuniões em conjunto são essenciais para o bom desenvolvimento dos repositórios, com as definições das políticas, pois o Dataverse permite muitas configurações. Destaca-se também que as customizações realizadas são realizadas pelas APIs e BASH, sem a necessidade de alteração do código fonte, ou seja, em novas atualizações todas as configurações permanecem compatíveis.

Além dos atendimentos realizados nas reuniões semanais, a produção de vídeos mostrou-se uma aliada para disseminar conhecimentos, tanto para atendimento às ICTs como para fomentar o debate e o entendimento a acerca da importância e do desenvolvimento de repositórios de dados de pesquisa em instituições brasileiras. O grupo RDP Brasil produziu 7 vídeos relativos aos conceitos básicos de ciência aberta, metadados e operacionalização do Dataverse, disponíveis em

<https://www.youtube.com/playlist?list=PLSk7nwl2wEh9CBEEqFRB6E5BqXiiC7HNP>

Conclusão

O projeto encontra-se em fase de finalização dos testes e as quatro ICTs já estão com repositórios instalados, operando a admissão (ingest) de datasets e descrição dos mesmos, os próximos passos estão relacionados a implementação dos repositórios em produção. Paralelamente ao desenvolvimento das questões técnicas, o grupo vem trabalhando nas políticas do repositório e fluxos de submissão e uso. Pretende-se, ao longo do segundo semestre, incentivar uma Rede Brasileira de Repositórios a partir das quatro ICTs, fomentando maior colaboração e apoio entre as instituições parceiras e o

lançamento do LattesData pelo CNPq e o Aléia pelo Ibict. As quatro ICTs participantes do projeto automaticamente tem prioridade de entrada no Consórcio Nacional para Ciência Aberta (CoNCienciA), o que possibilitará ter identificadores persistentes DOI associados a seus conjuntos de dados, com o custo inicial da emissão de DOIs patrocinado pelo CNPq.

Da experiência do GT-RDP na implementação e capacitação de gestores dos repositórios não enfrentou dificuldades no instanciamento do Dataverse e em suas definições, como a política do repositório, sendo fruto de muitas discussões. Entretanto o maior problema está na alteração da cultura da instituição com o pesquisador, principalmente em disponibilizar seus dados e compreensão do ciclo de vida dos dados de pesquisa atendendo os princípios FAIR.

Bibliografia

- BECKER, C., KULOVITS, H., GUTTENBRUNNER, M., STRODL, S., RAUBER, A., HOFMAN, H. (2009). Systematic planning for Digital. Preservation: evaluating potential strategies and building preservation plans. *International Journal on Digital Libraries*, 10 (4), 133-157. DOI:10.1007/s00799-009-0057-1
- CAREGNATO, S. E.; VANZ, S. A.; PAVÃO, C. M. G.; PASSOS, P. C. S. J.; BORGES, E. N.; GABRIEL JUNIOR, R. F.; AZAMBUJA, L. A. B.; ROCHA, R. P. (2019). Práticas e percepções dos pesquisadores brasileiros sobre serviços de acesso aberto a dados de pesquisa. *LIINC em Revista*, 15, 121-141.
- CONSULTIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS) (2012). Reference Model for an Open Archival Information System (OAIS). // CCSDS 650.0-M-2. Magenta Book, Recommended Practice, Issue 2. <https://public.ccsds.org/Pubs/650x0m2.pdf>
- LIN, D. ET AL. The TRUST Principles for digital repositories. *Scientific Data*, 7(144). <https://www.nature.com/articles/s41597-020-0486-7>
- FORCE11 (2017). Guiding principles for findable, accessible, interoperable and re-usable data publishing version b1.0. 2017. <https://force11.org/info/guiding-principles-for-findable-accessible-interoperable-and-re-usable-data-publishing-version-b1-0/>

GABRIEL JUNIOR, R. F.; ROCHA, R. P.; CAREGNATO, S. E.; PAVÃO, C.G.; PASSOS, P. C. S. J.; BORGES, E. N.; VANZ, S. A. S.; AZAMBUJA, L. A. B. (2019). Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias. *Ciência da Informação*, 48, 87-101.

RESEARCH DATA CANADA (2022). Original RDC Glossary. <https://www.rdc-drc.ca/glossary/original-rdc-glossary/>

ROCHA, R. P. DA; ET AL. (2018). Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas: relatório 2018. Porto Alegre, UFRGS. 75 p. <http://hdl.handle.net/10183/185126>

UNESCO. Conferência General (2021). 41ª Reunión – París. Proyecto de recomendación sobre la ciencia abierta. https://unesdoc.unesco.org/ark:/48223/pf0000378841_spa