
PROPOSTA DE CRIAÇÃO DE UMA REDE DE DADOS ABERTOS DA PESQUISA BRASILEIRA

PROPOSAL FOR THE CREATION OF AN OPEN DATA NETWORK IN
BRAZILIAN RESEARCH

PROPUESTA DE CREACIÓN DE UNA RED DE DATOS ABIERTOS
DE LA INVESTIGACIÓN BRASILEÑA

¹Caterina Groposo Pavão

¹Universidade Federal do Rio Grande do Sul

Correspondência

¹Caterina Groposo Pavão
Universidade Federal do Rio Grande do Sul
Porto Alegre, RS
Email: caterina@cpd.ufrgs.br
ORCID: <http://orcid.org/0000-0003-3712-7200>

Submetido em: 11/12/2018

Aceito em: 28/02/2018

Publicado em: 25/04/2018



JITA: IM. Open data

RESUMO: Aborda os dados de pesquisa resultantes de investigações realizadas em instituições de ensino e pesquisa brasileiras e a necessidade de torná-los públicos, por meio do acesso aberto em repositórios, a fim de validar os resultados obtidos e publicados, como também para impulsionar novas pesquisas e socializar o conhecimento. Tem como objetivo explorar o cenário nacional e internacional e apresentar o planejamento de uma investigação que busca uma solução tecnológica para efetivar o acesso aberto a dados de pesquisa (AADP) dentro de uma perspectiva nacional. Propõe uma metodologia dividida em cinco etapas: identificação de práticas de AADP em instituições brasileiras; mapeamento dos usuários de AADP e suas necessidades; proposta de um portal web para reunir a comunidade nacional em AADP; levantamento dos serviços e soluções tecnológicas para o compartilhamento de dados de pesquisa existentes no cenário internacional; proposição de recomendações para o apoio a criação de repositórios de dados de pesquisa em instituições nacionais e a sua agregação a uma rede de pesquisa em AADP. Como resultado, traz iniciativas e estratégias internacionais para levar a cabo a criação de repositórios de dados de pesquisa, assim como para criar comunidades de práticas em torno do assunto. Recomenda que a proposta seja constituída de três vertentes, a fim de apoiar a criação de repositórios de dados de pesquisa em instituições nacionais e a sua agregação a uma rede de pesquisa em AADP.

PALAVRAS-CHAVE: Acesso aberto. Dados de pesquisa. Repositórios.

ABSTRACT: It addresses research data resulting from research carried out in Brazilian teaching and research institutions and the need to make it public, through open access in repositories, in order to validate the results obtained and published, as well as to foster new research and socialize the knowledge. It aims to explore the national and international scenario and present the planning of an investigation that seeks a technological solution to effect open access to research data (AADP) from a national perspective. It proposes a methodology divided into five stages: identification of AADP practices in Brazilian institutions; mapping of AADP users and their needs; proposal of a web portal to bring together the national community in AADP; survey of services and technological solutions for the sharing of research data on the international scene; proposing recommendations to support the creation of repositories of research data in national institutions and their aggregation into a research network in AADP. As a result, it brings international initiatives and strategies to undertake the creation of research data repositories, as well as to create communities of practice around the subject. It recommends that the proposal be made up of three strands, in order to support the creation of research data repositories in national institutions and their aggregation into a research network in AADP.

KEYWORDS: Open access. Research data. Repositories.

RESUMEN: Aborda los datos científicos resultantes de investigaciones realizadas en instituciones de enseñanza e investigación brasileñas y la necesidad de hacerlos públicos, a través del acceso abierto en repositórios, a fin de validar los resultados obtenidos y publicados, así como para estimular nuevas investigaciones y socializar el conocimiento. Tiene como objetivo explorar el escenario nacional e internacional y presentar un plan de investigación que busca una solución tecnológica para el acceso abierto a datos de investigación (AADP) dentro de una perspectiva nacional. Propone una metodología dividida en cinco etapas: identificación de prácticas de AADP en instituciones brasileñas; mapeo de los usuarios de AADP y sus necesidades; propuesta de un portal web para reunir la comunidad nacional en AADP; el levantamiento de los servicios y soluciones tecnológicas para el intercambio de datos de investigación existentes en el escenario internacional; propuesta de recomendaciones para el apoyo a la creación de repositórios de datos de investigación en instituciones nacionales y su agregación a una red de investigación en AADP. Como resultado, presenta iniciativas y estrategias internacionales para llevar a cabo la creación de repositórios de datos de investigación, así como para crear comunidades de prácticas alrededor del tema. Recomienda que la propuesta esté constituída por tres vertientes con el fin de apoyar la creación de repositórios de datos de investigación en instituciones nacionales y su agregación a una red de investigación en AADP.

PALABRAS CLAVE: Acceso abierto. Datos de investigación. Repositórios.

1 INTRODUÇÃO

A divulgação dos dados de pesquisa vem sendo reconhecida como de grande importância, não só para validar os resultados obtidos e publicados, mas propiciar seu reuso, como também para impulsionar novas pesquisas e socializar o conhecimento. Neste contexto, os repositórios de dados de pesquisa assumem o papel de fornecer mecanismos de busca eficientes e serviços de valor agregado para a produção de conhecimento.

Se os repositórios desempenham uma função vital na preservação, integridade e divulgação de dados de pesquisa, uma rede de repositórios pode gerar conexões entre as comunidades, aumentando assim a interface entre fontes de dados de diferentes disciplinas e em repositórios específicos ou multidisciplinares.

O Brasil é atuante na construção de repositórios institucionais, para armazenamento da produção científica produzida na instituição, dentre vários repositórios o LUME da UFRGS destaca-se como referência em âmbito nacional, e junto com o IBICT tornaram-se formadores de comunidades de apoio ao desenvolvimento de repositórios institucionais. O país, no entanto, ainda não conta com uma estrutura de apoio à construção de repositórios de dados de pesquisa, nem há um planejamento em nível nacional para a efetivação de uma rede que reúna indivíduos e instituições, paralelamente aos seus conhecimentos e práticas, envolvidos com a promoção de dados abertos da pesquisa brasileira.

Assim, o objetivo deste artigo é explorar o cenário nacional e internacional e apresentar o planejamento de uma investigação que busca uma solução tecnológica para efetivar o Acesso Aberto a Dados de Pesquisa (AADP) dentro de uma perspectiva nacional. Dentro da proposta do planejamento da investigação, relacionam-se os objetivos específicos: a) identificar de práticas de AADP em instituições brasileiras; b) mapear os usuários de AADP e suas necessidades; c) propor um portal web para reunir a comunidade nacional em AADP; d) levantar os serviços e soluções tecnológicas para o compartilhamento de dados de pesquisa existentes no cenário internacional; e) propor recomendações para o apoio a criação de repositórios de dados de pesquisa em instituições nacionais e a sua agregação a uma rede de pesquisa em AADP.

Destaca-se que o trabalho está em andamento e seus resultados não são conclusivos.

2. CONTEXTO E ANTECEDENTES DO ESTUDO

O trabalho aqui apresentado é fruto de uma parceria entre a Universidade Federal do Rio Grande do Sul (UFRGS) e a Universidade Federal do Rio Grande (FURG), envolvendo o Centro de Documentação e Acervo Digital de Pesquisa (CEDAP), órgão auxiliar da Faculdade de Biblioteconomia e Comunicação (FABICO) da UFRGS, o Centro de

Processamento de Dados (CPD) da UFRGS, e o Grupo de Pesquisa em Gerenciamento de Informações do Centro de Ciências Computacionais (C3) da FURG, descritos a seguir.

O CEDAP é um órgão auxiliar da UFRGS, instituído em 2012 e vinculado à FABICO, conforme consta no Art. 96 do Estatuto da UFRGS. Desde sua concepção, o Centro tem por objetivo dar suporte à pesquisa científica, tecnológica, artística e cultural realizada na UFRGS, e fora dela, através da curadoria de dados de pesquisa, natos ou provenientes da digitalização, a fim de permitir o seu reuso, validar resultados, preservar a memória e aplicar essas experiências no ensino e para o bem público. Oferece assessoria em gestão e curadoria de dados da pesquisa, apoiando pesquisadores durante os estágios ativos do ciclo de vida dos dados (planos de gestão de dados, coleta e organização dos dados), assim como oferecendo recursos para disseminação e para armazenamento a longo prazo (repositório de dados da pesquisa). Dispõem de serviços para conversão de documentos para o formato digital, por meio de uma abordagem interdisciplinar das áreas da Ciência da Informação, Comunicação e Computação, desenvolvendo projetos de investigação, participando de redes com temáticas afins, estimulando a produção científica e promovendo o compartilhamento e a disseminação de novos conhecimentos.

O CEDAP tem como filosofia o processo de inovação aberta, visto que explora tanto ideias internas como externas, trata a Pesquisa e Desenvolvimento (P&D) como um sistema aberto e assume que o conhecimento útil deva ser amplamente distribuído. Para tal desenvolveu um ambiente fundamentado em trabalho colaborativo, em que o conhecimento e as práticas em dados da pesquisa são adquiridos e socializados por meio de um sistema wiki, baseado em corpos de conhecimento, processos de inovação e modelos conceituais, conforme relatado por Rocha e Caregnato (2015), Rocha (2016), Pavão, Caregnato e Rocha (2016) e Rocha, Caregnato e Gabriel (2017).

Atualmente, o Centro está em fase de implantação e operacionalização de seu repositório de dados em DSpace, empregando uma estratégia que passa a considerar o repositório de dados como parte final de uma cadeia de ações a ser explorada. Como decorrência, estuda-se todo o ciclo de vida do dado da pesquisa, que envolve, desde o apoio aos pesquisadores na elaboração de seus planos de gestão de dados de pesquisa, na coleta e organização dos dados, assim como no armazenamento a curto e longo prazo desses dados em um repositório confiável.

A constituição do CEDAP reforçou o pioneirismo da UFRGS entre as universidades brasileiras no compartilhamento de dados de pesquisa, e acompanha uma tendência mundial encontrada em universidades que estão bem posicionadas no cenário internacional. Ela também representou mais uma parceria da FABICO com o CPD da UFRGS.

O CPD é um Órgão Suplementar da UFRGS, que tem como objetivo prover os serviços de Tecnologia da Informação e Comunicação que apoiem a UFRGS no

desenvolvimento do ensino, pesquisa, extensão, gestão e serviços à comunidade, de acordo com as diretrizes do Plano de Desenvolvimento Institucional (PDI) e planos específicos para a área de Tecnologia da Informação (TI). O CPD é referência na gestão de repositórios institucionais, com o LUME, e tem na gestão de dados de pesquisa o próximo desafio, dando continuidade aos esforços e investimentos já realizados. Ao contar com a equipe do CPD responsável pela implantação e operação do LUME, o projeto é contemplado com a experiência na gestão e no compartilhamento de informações em repositórios, de extrema relevância para a definição de estratégias semelhantes para dados da pesquisa.

A equipe gestora do LUME possui experiência na definição de uma infraestrutura organizacional para institucionalização e gestão do repositório, no estabelecimento de políticas para a criação e gerenciamento de coleções digitais, na implantação e customização do software DSpace, na sua operação com alta disponibilidade e desempenho, assim como no desenvolvimento de extensões para o compartilhamento das teses e dissertações e outros documentos do LUME com a Biblioteca Brasileira de Teses e Dissertações (BDTD¹), o Portal brasileiro de publicações científicas em acesso aberto (OASISBr²) e o Repositório Científico de acesso Aberto de Portugal (RCAAP³), entre outros.

O LUME é um importante instrumento promotor da visibilidade institucional e tem contribuído, também, para o posicionamento da UFRGS no *Webometrics Ranking of World Universities*, no qual ocupa, atualmente, a 2ª posição no Brasil, a 2ª na América Latina, a 42ª posição na classificação geral (num total de 2.284 repositórios temáticos e institucionais) e a 35ª posição na classificação geral dentre os repositórios institucionais. Desde a sua implantação, em janeiro de 2008, foram realizados, na comunidade de produção Científica 7.461.253 *downloads*, dos quais quase cinco milhões no triênio 2014-2016. Os dados sobre o número de acessos e *downloads*, por ano e país, permitem visualizar o impacto e a extensão de uso dos documentos nele depositados.

O LUME é, ainda, uma ferramenta gerencial de apoio à tomada de decisão para as instâncias superiores da Universidade; para as agências de fomento permite verificar os resultados de seus investimentos em projetos e programas de pesquisa, para os usuários finais é uma fonte de informação imprescindível para o desenvolvimento da pesquisa e realimentação do ciclo de geração de conhecimento e para os autores oferece dados sobre o uso da sua produção.

O Centro de Ciências Computacionais (C3) é uma das treze unidades acadêmicas que constituem a FURG, atendendo as demandas por ensino, pesquisa e extensão nas áreas da Ciência da Computação, da Automação, da Tecnologia da Informação e da Computação Científica. Atualmente, o C3 é responsável por três cursos de graduação: Engenharia de

¹ <http://bdt.d.ibict.br/vufind/>

² <http://oasisbr.ibict.br/vufind/>

³ <https://www.rcaap.pt/>

Computação, Engenharia de Automação e Bacharelado em Sistemas de Informação. No que tange a pós-graduação lato sensu, o C3 oferece quatro cursos de especialização: Aplicações para Web, Tecnologias de Informação e Comunicação na Educação, Engenharia de Automação e Instrumentação e Engenharia de Elétrica.

Em relação a pós-graduação stricto sensu, é responsável pelo Programa de Pós-Graduação em Computação, com curso de Mestrado em Engenharia de Computação e coopera com outras unidades/universidades no Programa de Pós-Graduação em Modelagem Computacional e no Programa de Pós-Graduação em Educação em Ciências. O Grupo de Pesquisa em Gerenciamento de Informações é um dos que atuam junto ao Programa de Pós-Graduação em Computação. O grupo é coordenado pelo professor Eduardo Nunes Borges desenvolvendo atividades de pesquisa relacionadas à coleta, extração, visualização, recuperação e recomendação de informações, o que permite trazer conhecimentos, técnicas e pesquisas que envolvem informações heterogêneas e distribuídas, no que diz respeito a similaridades, consultas vagas ou aproximadas, metadados, recuperação de informações e descoberta de conhecimento.

Em síntese, a equipe que trabalhou na proposta aqui apresentada é composta de participantes da Ciência da Informação e da Computação, tem experiência em lidar com todas as áreas de conhecimento, conhecendo suas características e especificidades no que diz respeito à publicação de conteúdos digitais decorrentes de pesquisa e estudo do compartilhamento de dados de pesquisa.

3. MATERIAS E MÉTODOS

O trabalho foi planejado para ser realizado em cinco etapas, uma para cada objetivo específico, empregando métodos e técnicas específicos para cada uma delas.

A **primeira etapa**, de identificação de práticas de AADP em instituições brasileiras, baseia-se na pesquisa documental no diretório *Research Data Repositories Information (Re3data)*⁴, de abrangência internacional, que inclui uma descrição detalhada baseada na identificação dos tipos de repositórios, seu contexto, conteúdo e características. Igualmente, é necessário realizar um levantamento junto à comunidade científica nacional nas diversas áreas do conhecimento, incluindo universidades e centros de pesquisa, de forma a identificar as experiências em AADP e as soluções e práticas alternativas em armazenar e disseminar dados da pesquisa, assim como as demandas para a incorporação de dados em repositórios deste tipo e potenciais interessados em participar de uma iniciativa de construção de repositório de AADP.

⁴ Re3Data - Registry of Research Data Repositories - <http://www.re3data.org/>

A **segunda etapa**, de mapeamento dos usuários de AADP e suas necessidades, consiste na realização de uma consulta à comunidade científica brasileira (*survey*) por meio de questionário a ser respondido online por pesquisadores de todas as áreas, cujo objetivo é levantar as percepções, atitudes, experiências, práticas informais e necessidades dos cientistas quanto à disponibilização, gestão, acesso e reutilização aberta dos dados de pesquisa.

A coleta de dados para esta segunda etapa é constituída também de entrevistas com pesquisadores identificados na *survey* e que tenham disponibilidade e interesse de continuar participando da pesquisa. Ela compreende um aprofundamento dos temas, buscando identificar qualitativamente as necessidades dos usuários sobre o tema a partir das especificidades das grandes áreas do conhecimento, já que reconhecidamente práticas culturais e epistemológicas variam conforme as diferentes comunidades científicas.

A **terceira etapa**, que consiste em propor um portal web para reunir a comunidade nacional em AADP, apoia-se tecnologicamente na existência de softwares estáveis, seguros e com possibilidade de gerenciamento remoto, conhecidos como *Content Management System* (CMS), ou Sistema de Gerenciamento de Conteúdo.

Além de realizar o levantamento de requisitos do portal web, de forma a estabelecer qual CMS é mais adequado para a tarefa, a equipe do projeto deve trabalhar na arquitetura da informação do site para encontrar a melhor forma de distribuição dos conteúdos, além de propor formas de estimular a participação efetiva de pessoas chave na rede.

A **quarta etapa** consiste no levantamento dos serviços e soluções tecnológicas para o compartilhamento de dados de pesquisa existentes no cenário internacional. Tendo em vista esse cenário, propõe-se o levantamento comparativo dos serviços e soluções tecnológicas existentes para compartilhamento de dados conforme demonstrado a seguir.

O estudo se dá por meio da pesquisa bibliográfica e documental, envolvendo publicações científicas presentes nos principais periódicos e anais da área, assim como documentações técnicas, recomendações, padrões e estratégias presentes nos sites de agentes relacionados ao compartilhamento de dados da pesquisa. Também envolver a experimentação de serviços automatizados oferecidos em sites.

A **quinta etapa**, que consiste na formulação de recomendações para o apoio a criação de repositórios de dados de pesquisa em instituições nacionais e a sua agregação a uma rede de pesquisa em AADP, será realizada partir dos requisitos, funcionalidades e tecnológicas experimentadas internacionalmente, identificados na primeira, segunda e quarta etapas do trabalho, justapostos a criação de uma rede de participantes e de conhecimentos em AADP (realizada na etapa 3).

4 RESULTADOS PARCIAIS E DISCUSSÃO

Algumas iniciativas internacionais buscam mapear e identificar repositórios de AADP. O *Research Data Repositories Information* (Re3data) é um diretório internacional que inclui uma descrição detalhada dos principais repositórios e pode auxiliar a identificação dos tipos de repositórios, seu contexto, conteúdo e características. Portanto, o Re3data é utilizado como primeira alternativa para identificar repositórios brasileiros de dados de pesquisa.

Numa busca inicial, identificou-se o registro de apenas sete repositórios brasileiros, sendo que a metade deles são repositórios resultantes de cooperação internacional com diversos países. Com esses baixos resultados se faz necessário implementar outros mecanismos de coleta de informações para identificar iniciativas existentes no Brasil e que não foram registradas em diretórios internacionais.

Uma saída é a busca de informações diretamente com o público interessado. Os usuários potenciais de serviços de AADP no país são todos os pesquisadores brasileiros, tanto aqueles que já atentaram para a necessidade de se estabelecer processos de gestão de dados de pesquisa para atingir a finalidade de disponibilizar, acessar e reutilizar os resultados da ciência nacional, como os que ainda não conhecem as dimensões da questão e suas possíveis consequências. Da mesma forma, é necessário considerar os pesquisadores de todas as grandes áreas do conhecimento, que, segundo tabela do CNPq, são: Ciências Exatas e da Terra; Ciências Biológicas; Engenharias; Ciências da Saúde; Ciências Sociais Aplicadas; Ciências Humanas; Linguística, Letras e Artes.

Assim, o levantamento por meio de questionário dirigido a pesquisadores de todas as áreas e instituições permite identificar as experiências em repositórios de AADP nas universidades e centros de pesquisa, assim como desvelar soluções e práticas alternativas em armazenamento e disseminação de dados da pesquisa. Adicionalmente, permite constatar as demandas para a incorporação de dados em repositórios deste tipo e potenciais interessados em participar de uma iniciativa de construção de repositório de AADP.

As seguintes características das práticas e dos repositório de AADP levantadas são importantes nesta etapa: a) âmbito (nacional ou cooperação internacional), b) as áreas temáticas de abrangência, c) tipo de repositório (temático, institucional, outro), d) número de instituições envolvidas na coleta e disponibilização de dados, e) políticas do repositório (institucional, direitos autorais, licenças de uso, etc), f) formato dos dados, g) interoperabilidade e padrões de metadados, h) software utilizado, aplicações e interfaces. I) existência de um ciclo de vida da pesquisa.

Em relação à reunião da comunidade interessada em AADP e à disponibilização do corpo de conhecimento gerado por ela, constata-se que a utilização das tecnologias da informação e comunicação torna o empreendimento viável, simples e eficaz. A construção de

um portal web baseado em Sistema de Gerenciamento de Conteúdo (CMS) é a solução tecnológica proposta.

Estes sistemas permitem que o gestor tenha total autonomia sobre o conteúdo, acompanhamento da evolução da sua presença na internet, além de dispensar a assistência de terceiros para sua manutenção (GALIANO; et. al. 2007). A utilização de CMS torna mais ágil a publicação de conteúdos, além de delegar perfis diferentes para cada usuário. Requer pouca experiência do usuário, além dos conhecimentos de um simples editor de texto.

Também, permitem a customização de seu visual e a incorporação de *plugins* conforme identificação de demanda do portal. Outra grande vantagem no uso dos CMS está na possibilidade de colaboração de comunidades desses aplicativos para o desenvolvimento de *plugins* específicos a serem incorporados no portal, como por exemplo, a gestão e organização dos Grupos de Trabalhos (GTs), e sistemas de adesão de membros.

Entre os sistemas mais conhecidos disponíveis, desse tipo, e com a maior comunidade de colaboradores e desenvolvedores, estão o WordPress, Joomla ou Drupal. Cada um deles com centenas de milhares de *templates* (temas) gratuitos disponíveis, bem com inúmeros *plugins* passíveis de incorporação gratuitamente, o que agiliza o processo de construção do portal.

No que diz respeito aos requisitos técnicos mínimos estabelecidos por esses sistemas destacam-se: Servidor Apache (com mod_mysql, mod_xml e mod_zlib) com 4 ou 8 Gigabytes de memória, com interpretação de códigos PHP e Sistema de Gerenciamento de Banco de Dados (SGBD) MySQL (recomendado) ou PostGreSQL, espaço disponível em disco de 500 MBytes, e acesso à Internet com um domínio ou subdomínio.

Além da tecnologia, é necessário considerar o detalhamento da arquitetura da informação e a melhor forma de distribuição dos conteúdos. Os requisitos básicos já identificados são: a) informações sobre a governança do portal; b) possibilidade de adesão ao portal pela comunidade científica, profissional, estudantil e institucional; c) informações sobre a estruturação, organização e publicações dos grupos que forem constituídos; d) área de documentação com apresentação das publicações, documentos normativos, relatórios entre outros; e e) uma área destinada a comunicações, novidades e eventos relacionadas a área. Esses requisitos estão focados em um portal com características semelhantes às do portal da Research Data Alliance (RDA) (THE RESEARCH DATA ALLIANCE, 2017).

O portal deve ter instrumentos de monitoramento de visitantes e número de acessos de forma a mapear as maiores demandas, assim como a produção de relatórios periódicos de acompanhamento para a equipe gestora. O portal está disponível em seu endereço eletrônico <https://dadosdepesquisa.rnp.br> hospedados dentro dos servidores da RNP.

Nos primeiros levantamentos no cenário internacional, já foi possível identificar a existência de serviços nacionais de dados de pesquisa, com destaque para o serviço holandês Data Archiving and Networked Services (DANS⁵), que incentiva os pesquisadores a tornarem seus dados de pesquisa digital e resultados relacionados viáveis, acessíveis, interoperáveis e reutilizáveis, oferecendo consultoria especializada e serviços certificados desde 2005. O DANS disponibiliza o National Academic Research and Collaborations Information System (NARCIS⁶), sendo o principal portal nacional para quem procura informações sobre pesquisadores e seu trabalho.

Destaca-se também a colaboração da Monash University e da Australian National University com o Commonwealth Scientific and Industrial Research Organisation (CSIRO) que desenvolvem o Australian National Data Service (ANDS⁷), que tem como proposta tornar os ativos de dados de pesquisa mais valiosos para pesquisadores e institutos de pesquisa. A ANDS foi estabelecida em 2008, sendo seu principal produto o Research Data Australia (RDA), onde pode-se encontrar, acessar e reutilizar dados de pesquisa.

Na Inglaterra, a UK Research Data Discovery Service⁸ (Alpha) desenvolve um serviço piloto de busca envolvendo repositórios do Reino Unido, patrocinado pelo JIST, órgão que provém a universidades britânicas infraestrutura digital compartilhada (BEAGRIE, 2004).

Essas iniciativas estão embasadas em estudos que auxiliam governos na definição de políticas e estratégicas para compartilhamento e gestão de dados da pesquisa. Os panoramas são elaborados por grupos de trabalhos ligados a órgãos governamentais, como no caso do Canadá pelo Research Data Strategy Working Group (2011); nos Estados Unidos pela Interagency Working Group on Digital Data (2009); e pela Comunidade Européia (European Union, 2010).

Em relação as iniciativas de acesso aberto a dados de pesquisa, observam-se diversas propostas, porém bastantes singulares no que tange às formas de promovê-las e de realizá-las. Para identificar os serviços e soluções tecnológicas, parte-se dos modelos já conhecidos pela comunidade, com destaque para o modelo para descrição e disseminação de dados desenvolvido pela Data Documentation Initiative (DDI⁹); o modelo Data Description Registry Interoperability Model (DDRI¹⁰), desenvolvido pela Research Data Alliance (RDA¹¹) para interconectar conjuntos de dados; e o modelo de ciclo de vida de dados do Digital Curation

⁵ <https://dans.knaw.nl/en>

⁶ <http://www.narcis.nl/>

⁷ <http://www.ands.org.au/>

⁸ <http://ckan.data.alpha.jisc.ac.uk/dataset>

⁹ DDI - Data Documentation Initiative DDI-Lifecycle - <http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/>

¹⁰ DDRI - Data Description Registry Interoperability Model - <http://dx.doi.org/10.15497/RDA00003>

¹¹ RDA - Research Data Alliance - <https://www.rd-alliance.org/>

Centre Life Cycle (DCC¹²), dos repositórios de dados UK Data Archive¹³ e do Institute for Social Research (ICPSR¹⁴), que são referência internacional; e do serviço distribuído para dados observacionais da terra Data Observation Network for Earth (DataONE¹⁵).

No que diz respeito às soluções adotadas internacionalmente, observa-se a presença majoritária dos softwares, projetos pilotos e serviços com base no DSpace¹⁶, Dataverse¹⁷, DataCite¹⁸, EUDAT¹⁹ e RD-Switchboard²⁰, sendo que um dos objetivos desse projeto é estudá-los individualmente.

Em termos de guias e recomendações de melhores práticas para implantação de serviços de dados da pesquisa, considera-se relevantes os desenvolvidos pelas organizações OCLC (ERWAY et al., 2016), EDUCASE (FARY, 2013), JISC, pelo projeto Data Share (GREEN; MACDONALD; RICE, 2009) e pelo serviço e repositório de dados Institute for Social Research (AUSTIN, 2010).

Para desenho da solução proposta, primeiramente serão caracterizados os agentes envolvidos na promoção do compartilhamento de dados da pesquisa, considerando objetivos (promover, padronizar, experimentar, compartilhar, armazenar, entre outros), natureza (projeto, órgão, associação, aliança), estrutura de funcionamento (grupos, comitês), financiadores e estratégias usadas para promoção do compartilhamento de dados da pesquisa (arquiteturas, modelos, padrões, guias, projetos pilotos, oferta de serviços, apoio ao desenvolvimento de serviços de dados, etc.).

A seguir serão identificados os principais serviços para compartilhamento de dados da pesquisa, caracterizados considerando seus objetivos e soluções tecnológicas, arquiteturas, padrões, softwares, e a forma com que estes se relacionam com agentes que promovem o compartilhamento de dados.

¹² Digital Curation Centre Life Cycle - <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

¹³ <http://www.data-archive.ac.uk/create-manage/life-cycle>

¹⁴ ICPSR Institute for Social Research - <http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/#cycle>

¹⁵ DataONE - Data Observation Network for Earth - <https://www.dataone.org/data-life-cycle>

¹⁶ DSpace - <http://www.dspace.org/>

¹⁷ Dataverse - Ambiente computacional aberto para armazenamento, preservação, compartilhamento e análise de dados - <http://dataverse.org/>

¹⁸ DataCite - Organização não lucrativa que provém identificadores persistentes para dados da pesquisa. <https://www.datacite.org/index.html>

¹⁹ EUDAT - Projeto Europeu patrocinado pelo programa Horizon 2020 para prover infraestrutura de serviços computacionais para gerenciar dados de pesquisa dispersos em centros de dados e repositórios europeus. <https://www.eudat.eu/>

²⁰ RD-Switchboard - Projeto que visa criar infraestrutura conectar várias fontes de dados, composto pela Australian National Data Service (ANDS) e administrado pelo grupo de trabalho Data Description Registry Interoperability (DDRI) da Aliança RDA. Switchboard. <http://www.rd-switchboard.org/>

Então, os serviços serão analisados comparativamente para identificar aqueles que atendem às necessidades, requisitos e funcionalidades levantadas para usuários nacionais de serviços de AADP, observando também serviços praticados em nível internacional.

Também, serão analisadas, comparativamente, as soluções tecnológicas para a implementação dos serviços identificados, considerando inovação, proximidade com plataformas e práticas usadas nacionalmente em repositórios, uso de padrões, possibilidade de formação de redes internacionais de compartilhamento de dados, possibilidade de apoio e participação de atores e iniciativas internacionais que promovem o compartilhamento de dados, etc.

Finalmente, deverão ser detectadas a viabilidade, estratégias e adaptações necessárias no sentido de utilizar soluções tecnológicas identificadas para implementação de serviços de AADP para usuários nacionais.

5 CONSIDERAÇÕES FINAIS

Entende-se que as recomendações para uma proposta de apoio à criação de repositórios de dados de pesquisa em instituições nacionais e a sua agregação a uma rede de pesquisa em AADP deva ser constituída em três vertentes.

A primeira, contemplada na etapa 3 deste trabalho, baseia-se na implementação de um portal de forma a reunir a comunidade interessada no tema, abrindo espaço para discussões das necessidades individuais e coletivas para armazenamento de dados de pesquisa. Ao mesmo tempo, pretende incentivar a criação de grupos de trabalho para formulação de políticas, diretrizes e padrões para construção, implementação dos repositórios ADP. Como premissa dos repositórios está a possibilidade de interoperabilidade de dados, com o uso de boas práticas no compartilhamento de dados de pesquisa. Vê-se então no portal, parte fundamental na consolidação da proposta tecnológica com a interação com a comunidade científica.

A segunda vertente propõe desenvolver uma metodologia que possibilite a implementação de um ou mais protótipos de repositórios AADP em instituições, abrangendo todas as áreas do conhecimento, ou repositórios temáticos. A solução compreende a escolha do melhor software, conforme análise dos resultados obtidos na etapa 4, conforme proposta. Todas as soluções devem ser documentadas, tanto em manuais operacionais como técnicos e estes disponibilizados no portal para colaboração da comunidade. Dos requisitos de escolha do software, está a possibilidade de desenvolvimento e incorporação de novos padrões de metadados, de forma a melhor representar os dados de pesquisa, bem como a implementação de classes e métodos para o preenchimento de lacunas e adaptações necessárias. Ainda nessa etapa devem ser elegíveis, por uma amostragem de conveniência, uma ou mais instituições para implementar e testar o modelo de repositório AADP, com capacitações *in loco* ou

virtuais, bem como coletar qualitativamente as percepções por meio de entrevista, identificando as facilidades e dificuldades de implementação e uso do ambiente. Os resultados contribuem para o ajuste da metodologia, podendo ocorrer em mais de uma instituição.

A terceira e última vertente comporta o desenvolvimento de um protótipo de agregador de dados dos repositórios AADP, com implementação de protocolos OAI-PMH ou similares, que propiciem a coleta e reunião dos dados em uma única base de dados. É importante que se empreguem técnicas da Biblioteconomia, como controle de autoridade e cabeçalhos de assunto, de forma a melhorar a precisão na recuperação da informação.

REFERÊNCIAS

AUSTIN, Erik. et al. Guide to Social Science Data Preparation and Archiving Best Practice Throughout the Data Life Cycle. ICPSR Institute for Social Research University of Michigan 2010. <<http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>>. Acesso em: 14 abr. 2017.

BEAGRIE, Neil. The Continuing Access and Digital Preservation Strategy for the UK Joint Information Systems Committee (JISC). **D-Lib Magazine**, v.10, n. 7/8, Jul/Aug 2004. Disponível em: ><http://www.dlib.org/dlib/july04/beagrie/07beagrie.html>>. Acesso em: 22 nov. 2017.

ERWAY, Ricky. et al. Building Blocks: Laying the Foundation for a Research Data Management Program. Dublin, Ohio: OCLC Research. Disponível em: <<http://www.oclc.org/research/publications/2016/oclcresearch-data-management-building-blocks-2016.html>>. Acesso em: 14 abr. 2017.

EUROPEAN UNION. **Riding the wave. How Europe can gain from the rising tide of scientific data**. Final report of the High level Expert Group on Scientific Data, European Union. 2010. Disponível em: <http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707>. Acesso em: 14 abr. 2017.

FARY, Michael; OWEN, Kim PIENAAR, H. Developing an institutional research data management planservice. EDUCAUSE, 2013. <<https://www.educause.edu/ir/library/pdf/ACTI1301.pdf>>. Acesso em: 21 NOV. 2017.

GALIANO, V.; PEREZ-CAPARROS, David; MIGALLÓN, Hector; PALOMINO-BENITO, Juan Alejandro. Universal WCMS: An open source web content management system and customer relationship management for e-business. In: CONFERENCE ICE-B 2007, Barcelona. **Proceedings...**Barcelona. Disponível em:

<https://www.researchgate.net/publication/220917626_Universal_WCMS_An_Open_Source_Web_Content_Management_System_and_Customer_Relationship_Management_for_e-Business>. Acesso em: 21 nov. 2017.

GREEN, Ann.; MACDONALD, Stuart.; RICE, Robin. Policy-making for Research Data in Repositories: A Guide. EDINA, 2009. <<http://www.disc-uk.org/docs/guide.pdf>>. Acesso em: 14 abr. 2017.

INTERAGENCY WORKING GROUP ON DIGITAL DATA. **Harnessing the power of digital data for science and society**. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. 2009. Disponível em: <https://www.nitrd.gov/About/Harnessing_Power_Web.pdf>. Acesso em: 14 abr. 2017.

NEUROTH, Heike.; STRATHMANN, Stefan.; OßWALD, Achim.; LUDWIG, Jens. (Org.). **Digital Curation of Research Data: Experiences of a Baseline Study in Germany**. Glückstadt, Verlag Werner Hülsbusch, 2013. Disponível em: <http://www.nestor.sub.uni-goettingen.de/bestandsaufnahme/Digital_Curation.pdf>. Acesso em: 21 nov. 2017.

PAVÃO, Caterina Marta. Groposo; CAREGNATO, Sônia Elisa; ROCHA, Rafael Port da. Implementação da preservação digital em repositórios: conhecimento e práticas. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 14, n.3, p. 407-425, 2016. Disponível em: <<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646326/pdf>>. Acesso em: 21 nov. 2017.

RESEARCH DATA STRATEGY WORKING GROUP. **Mapping the Data Landscape: Report of the 2011 Canadian Research Data Summit**. Research Data Strategy Working Group. 2011 Disponível em: <https://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/obj/doc/2011_data_summit-sommet_donnees/Data_Summit_Report.pdf>. Acesso em: 14 abr. 2017.

ROCHA, Rafael. Um sistema baseado na construção coletiva de em corpo de conhecimento para apoiar as atividades de um centro de digitalização. In ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., Salvador. **Proceedings...** Salvador, 2016. Disponível em: <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/viewFile/3662/2567>. Acesso em: 21 nov. 2017.

ROCHA, Rafael Port da; CAREGNATO, Sônia Elisa. Corpo de Conhecimento em Digitalização na Implantação do Centro de Documentação e Acervo Digital da Pesquisa -

CEDAP. In: CTCM 2015 - CONFERÊNCIA SOBRE TECNOLOGIA, CULTURA E MEMÓRIA, 3., 2015, Recife. **Anais...** Recife: Liber, 2015. p. 1-14. Disponível em: <http://www.liber.ufpe.br/home/wp-content/uploads/2016/09/13-Corpo-de-Conhecimento_Rocha-Caregnato.pdf>. Acesso em 21 nov. 2017.

ROCHA, Rafael; CAREGNATO, Sônia Elisa; GABRIEL JUNIOR, Rene Faustino. Aspectos de inovação na implantação de um centro de digitalização e dados da pesquisa. In: CONGRESSO DE GESTÃO ESTRATÉGICA DA INFORMAÇÃO: EMPREENDEDORISMO E INOVAÇÃO, 2017, Florianópolis. **Anais...** Florianópolis: UFSC, 2017.

THE RESEARCH DATA ALLIANCE. **Research data sharing without barriers.** 2017. Disponível em: <<https://www.rd-alliance.org/>>.. Acesso em: 14 abr. 2017.

