

Ricardo César Gonçalves Sant'Ana

Moisés Lima Dutra

Guilherme Ataíde Dias

Organizadores

WIDaT 2018

II WORKSHOP DE INFORMAÇÃO,
DADOS E TECNOLOGIA

ANAIS
WIDaT 2018

Organização do WIDaT 2018

- **Organização Geral:**

Guilherme Ataíde Dias (PPGCI-UFPB) - Coordenador geral do evento
Moisés Lima Dutra (PPGCIN-UFSC) - Vice-coordenador

- **Coordenador da Comissão Científica:**

Ricardo César Gonçalves Sant'Ana (PPGCI-UNESP)

- **Comissão científica**

Adilson Luiz Pinto (PPGCIN-UFSC)
Ana Alice Baptista (Universidade do Minho, Portugal)
Ana Carolina Simionato (PPGCI-UFSCar)
Angela Maria Grossi de Carvalho (PPGCI-UNESP)
Bernardina Maria Juvenal Freire de Oliveira (PPGCI-UFPB)
Cristian Berrío-Zapata (PPGCI-UFPA)
Dalton Lopes Martins (FCI-UnB)
Denysson Axel Ribeiro Mota (PPGB-UFCA)
Douglas Dyllon Jeronimo de Macedo (PPGCIN-UFSC)
Ed Porto Bezerra (PPGI-UFPB)
Edgar Bisset Alvarez (PPGCIN-UFSC)
Edna Gusmão de Goés Brennand (MPGOA-UFPB)
Edna Gomes Pinheiro (DCI-UFPB)
Elaine Parra Affonso (FATEC-SP)
Elvis Fusco (UNIVEM-Marília)
Enrique Muriel Torrado (PPGCIN-UFSC)
Evandro de Barros Costa (IC-UFAL)
Fábio Paraguaçu (IC-UFAL)
Fernando de Assis Rodrigues (PPGCI-UNESP)
Gustavo Medeiros de Araújo (PPGCIN-UFSC)
Henry Pôncio Cruz de Oliveira (PPGCI-UFPB)
Joana Coeli Ribeiro Garcia (PPGCI-UFPB)
José Eduardo Santarém Segundo (USP-FFCLRP)
Leonardo Castro Botega (UNIVEM-Marília)
Luana Farias Sales Marques (PPGCI-IBICT-UFRJ)
Marckson Roberto Ferreira de Sousa (PPGCI-UFPB)
Luís Fernando Sayão (CNEN)
Marcelo Morandini (EACH-USP)
Márcio Matias (PPGCIN-UFSC)
Marcos Mucheroni (CBD-USP)
Marynice de Medeiros Matos Autran (PPGCI-UFPB)

Maurício Barcellos Almeida (PPGGOC-UFMG)
Moisés Lima Dutra (PPGCIN-UFSC)
Plácida Leopoldina V. da Costa Santos (PPGCI-UNESP)
Pedro Luiz Pizzigatti Corrêa (POLI-USP)
Renata Baracho (PPGGOC-UFMG)
Ricardo César Gonçalves Sant'Ana (PPGCI-UNESP)
Robson Rodrigues Lemos (UFSC-Araranguá)
Rogério Ramalho (PPGCI-UFSCar)
Ryan Ribeiro de Azevedo (UFRPE-UAG)
Sandra de Albuquerque Siebra (PPGCI-UFPE)
Sandro Rautenberg (DECOMP-UNICENTRO)
Silvana Aparecida Borsetti G. Vidotti (PPGCI-UNESP)
Virginia Bentes Pinto (PPGCI-UFC)
Wagner Junqueira de Araújo (PPGCI-UFPB)
Zaira Regina Zafalon (PPGCI-UFSCar)

- **Coordenador do Cerimonial:**

André Luiz Dias de França (PPGCI-UFPB)

- **Coordenador da Equipe Técnica Local:**

Laerte Pereira da Silva Júnior (CCHLA-UFPB)

- **Equipe Técnica Local:**

Adriana Alves Rodrigues (PPGCI-UFPB)
Antonio Felipe dos Santos (MPGOA-UFPB)
Débora Gomes de Araújo (PPGCI-UFPB)
Pedro Augusto de Lima Barroso (PPGCI-UFPB)
Pollianna Marys de Souza e Silva (PPGCI-UFPB)
Renata Lemos dos Anjos (PPGCI-UFPB)

REPOSITÓRIOS DE DADOS DE PESQUISA:

investigando sua adoção nos institutos brasileiros de pesquisa

RESEARCH DATA REPOSITORIES:

investigating its adoption in the brazilian research institutes

André Luiz França de Madeiro

Universidade Federal da Paraíba, Departamento de Ciência da Informação, andreluizmadeiro@gmail.com

Resumo:

A ciência atual é baseada no uso e compartilhamento de dados de pesquisa por meio das Tecnologias de Informação e da Comunicação. Neste contexto é possível acessar estes dados em bancos de dados *online* conhecidos como repositórios digitais. Os repositórios de dados são um dos tipos desta categoria. Eles podem estar associados a instituições de pesquisa e destacam-se cada vez mais no cenário nacional e internacional. A pesquisa possui como objetivo geral investigar os repositórios de instituições brasileiras de pesquisa no que diz respeito ao suporte para a gestão de dados de pesquisa. A pesquisa apresentou um percurso metodológico bibliográfico, exploratório e quanti-qualitativo. Após a análise dos repositórios encontrados, percebemos uma incipiente presença de repositórios de dados de pesquisa no contexto das instituições de pesquisa brasileiras. Estima-se que este contexto tende a mudar devido à necessidade global de gerenciar novas produções de dados de pesquisa que aumenta constantemente. Novos repositórios de pesquisa poderão surgir à medida que a comunidade de pesquisadores e as instituições estiverem cientes dos benefícios proporcionados por eles e que o tema possa contribuir para novas investigações científicas.

Palavras-chave: Repositório de dados científicos. Instituições brasileiras de pesquisa. Gestão de dados de pesquisa.

Abstract:

Nowadays science is based on the use and sharing of research data through Information and Communication Technologies. In this context it is possible to access this data in online databases known as digital repositories. Data repositories are one of a kind in this category. They can be associated with research institutions and are becoming more relevant in the national and international scenario. The research general objective is to investigate the repositories of Brazilian research institutions regarding the support of research data management. This research presents a bibliographic, exploratory and quanti-qualitative methodology. After the analysis of the repositories found, we noticed an incipient presence of repositories of research data in the context of Brazilian research institutions. It is estimated that this context tends to change due to the global need to manage new and constantly increasing research data productions. New research repositories may emerge as the research community and institutions are aware of the benefits they provide and the subject can contribute to new scientific research.

Keywords: Research Data Repositories. Brazilian Research Institutions. Research Data Management.

I INTRODUÇÃO

A ciência atual está contextualizada no âmbito de um fenômeno conhecido como *e-Science*¹. Segundo Oliveira (2016) este termo representa o quarto paradigma científico e o cientista Jim Gray foi seu precursor. No contexto histórico dos paradigmas da ciência, o primeiro paradigma foi marcado pela experimentação, ocorrido há mais de mil anos. O segundo fundamentou-se na ciência teórica com o surgimento de modelos e generalizações, no século

XVII. O terceiro é representado na ciência computacional com base na simulação de fenômenos complexos. O paradigma atual caracteriza-se pela presença da geração intensiva de grande quantidade de dados que surgem do processo de investigação científica (OLIVEIRA, 2016).

Na ciência contemporânea, os dados de pesquisa não são insumos exclusivamente destinados para os pesquisadores que os produziram. Estes dados também devem beneficiar toda a comunidade científica através de estratégias que possibilitem o seu uso e reuso. Os dados podem estar disponibilizados através de repositórios de dados científicos com o acesso livre de barreiras geográficas e de restrições temporais

A relevância da temática em tela é de interesse dos pesquisadores, curadores de dados, comunidades científicas, órgãos de fomento à pesquisa, editores de periódicos científicos, gestores e executores de programas e projetos de dados de pesquisa, profissionais da computação e da Ciência da Informação em geral.

Devido a crescente expansão e compartilhamento do volume de dados de pesquisa, é necessário o estabelecimento e a consolidação de uma política de gestão de dados científicos, visando a garantia dos mesmos por meio de processos de curadoria, preservação, arquivamento e compartilhamento de coleções de dados de pesquisa, em sustentabilidade contínua e permanente (IBICT, 2016).

O valor dos dados de pesquisa em nossa contemporaneidade motivou o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) a lançar o “Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã”. Esta ação foi efetivada na data de 28 de setembro de 2016, configurando-se como um posicionamento brasileiro diante do movimento mundial ao acesso aberto à informação científica. O objetivo desta iniciativa foi incentivar o acesso livre, abrangente e ilimitado às fontes primárias de dados de pesquisa que tornam possível o uso, reuso, compartilhamento e reprodutibilidade em novos contextos de investigação científica (IBICT, 2016).

Dentre os mais variados tipos de repositórios existentes no Brasil, os associados com as instituições de pesquisa têm presença cada vez mais marcante na Internet. A pesquisa de Sayão e Sales (2016) é um exemplo deste assunto: os referidos pesquisadores trazem informações sobre repositórios de dados de pesquisa, sobre sua definição, objetivos, benefícios, apresentação de fluxos de gestão de dados de pesquisa em repositório de dados, tipos de repositórios de dados de pesquisa, valor e permanência de dados e análise dos repositórios *Dryad* e *R3DATA*.

Repositórios de dados associados a instituições de pesquisa são novas entidades que despontam na seara da Tecnologia da Informação. Eles favorecem a propagação das informações científicas e tecnológicas. De acordo com Sayão e Sales (2015, p.53):

¹ O termo *data intensive science* é equivalente ao termo *e-Science*. Doravante, adotaremos o termo *e-Science*.

Os repositórios de dados de pesquisa são banco de dados que recebem, gerenciam e disseminam dados e conjuntos de dados de pesquisa. Eles constituem o lugar mais apropriado para que seus dados sejam preservados e possam ser recuperados, acessados e citados por outros pesquisadores, ou seja, tenham visibilidade em escala mundial.

2 OBJETIVOS

No que se refere a problematização, esta pesquisa questiona de que forma se configuram os repositórios de pesquisa das instituições brasileiras de pesquisa no que diz respeito a gestão de dados de pesquisa.

A pesquisa possui como objetivo geral investigar os repositórios de instituições brasileiras de pesquisa no que diz respeito ao suporte para a gestão de dados de pesquisa. Quanto aos objetivos específicos elencamos os seguintes: identificar os repositórios de dados de institutos brasileiros de pesquisa; identificar a tipologia dos repositórios em institutos brasileiros de pesquisa; levantar características específicas dos repositórios de dados de institutos brasileiros de pesquisa; disponibilizar os dados coletados e o material de apoio usado na pesquisa no repositório de dados no repositório *Dataverse*² da UFPB.

3 PROCEDIMENTOS METODOLÓGICOS

A trilha metodológica usada para a condução desta investigação baseou-se numa abordagem bibliográfica, de cunho exploratório, quanti-qualitativo.

A busca por textos para a revisão de literatura e para a identificação dos repositórios de dados foi realizada através das ferramentas de busca Google e Google Acadêmico no período compreendido entre agosto de 2017 a julho de 2018. Os descritores selecionados para a pesquisa foram: “repositórios brasileiros”; “repositórios de dados”; “*brazilian repositories*” e “*data repositories*”.

Os sites *R3DATA*, *ROAR* e *IBICT* foram usados como fontes de referência para identificar os repositórios de dados existentes. O processo investigativo foi bastante exaustivo. Todos os repositórios levantados foram acessados para serem identificados com relação a sua tipologia e as características dos objetos digitais ora depositados. Para auxiliar na identificação de repositórios de dados de pesquisa, foram utilizados os estudos de Leite et al. (2012), Pampel et al. (2013) e Sales (2014).

A partir dos repositórios brasileiros identificados, verificou-se a existência de muitos repositórios institucionais. Após uma análise precisa, consultando os tipos de dados presentes, a finalidade da instituição, informações referentes dos repositórios, verificamos que alguns poucos destes repositórios poderiam ser classificados como repositórios de dados de pesquisa.

2 <https://dataverse.ufpb.br/>

4 RESULTADOS

Foram encontrados seis repositórios de dados em instituições brasileiras de pesquisa. Dentre os repositórios identificados como repositórios de dados, listamos os seguintes: *PPBiodata*, *IBICT Dataverse*, *carpeDIEN*, Portal GEOINFO de infraestrutura de dados espaciais da EMBRAPA, Portal da Biodiversidade e o repositório SinBiota (Sistema de Informação Ambiental do Programa Biota/Fapesp). Todos apresentam diversidade na tipologia dos dados em e variados formatos.

O repositório de dados científicos *IBICT Dataverse* encontra-se no site <https://repositorioesquisas.ibict.br/>. Ele apresenta uma tipologia orientada para várias áreas do conhecimento, de acordo com as informações em seu site:

O IBICT está disponibilizando um repositório de dados de pesquisa que cuida da preservação a longo prazo e de boas práticas de arquivamento, para que os pesquisadores possam compartilhar, manter o controle e obter o reconhecimento de seus dados. O repositório suporta o compartilhamento de dados de pesquisa com citação de dados persistentes, permitindo que eles sejam reproduzidos. O Dataverse é um grande repositório aberto de dados de todas as disciplinas, criado pelo Institute for Quantitative Ciências Sociais da Universidade de Harvard (*IBICT Dataverse*, 2018, *online*).

O repositório IBICT usa o *software Dataverse*. O *Dataverse* (2018, *online*, tradução nossa) é definido como “um aplicativo da Web de código aberto para compartilhar, preservar, citar, explorar e analisar dados de pesquisa³”.

O *CarpeDIEN*, pode ser acessado através do site <http://carpedien.ien.gov.br/>, é o nome do repositório do IEN (*Instituto de Energia Nuclear*). Apresenta uma tipologia voltada para “dados e informações em energia nuclear” (*carpeDIEN*, 2018, *online*) e faz uso do *software DSpace*. O *DSpace* (*DSPACE*, 2018, *online*, tradução nossa) é conhecido como o “software preferido das organizações acadêmicas, sem fins lucrativos e comerciais que criam repositórios digitais abertos⁴”.

O *PPBiodata* pode ser acessado através do link <https://ppbiodata.inpa.gov.br/metacatui/>. Ele faz parte do *PPBio* (*Programa de Pesquisa em Biodiversidade*). *PPBio* foi:

criado em 2004 com o objetivo de intensificar estudos sobre biodiversidade no Brasil, descentralizar a produção científica dos centros desenvolvidos academicamente, integrar atividades de pesquisa e divulgar os resultados para diferentes finalidades, como gestão ambiental e educação (*PPBio*, 2018, *online*).

Ainda, com relação ao *PPBiodata*, é importante mencionar que o uso faz uso do *metacat*. Este *software* é um “repositório de dados e metadados (documentação sobre dados), que ajuda os cientistas a encontrar, entender e usar efetivamente os conjuntos de dados que gerenciam ou que foram criados por outros⁵” (*KNB*, 2018, *online*, tradução nossa).

3 Fonte original em língua inglesa: *is an open source web application to share, preserve, cite, explore, and analyze research data*. (*DATAVERSE*, 2018, *online*).

4 Fonte original em língua inglesa: *DSpace is the software of choice for academic, non-profit, and commercial organizations building open digital repositories* (*DSpace*, 2018, *online*).

5 Fonte original em língua inglesa: *is a repository for data and metadata (documentation about data), that helps scientists find, understand and effectively use the data sets they manage or that have been created by others* (*KNB*, 2018, *online*).

Costa (2017) menciona o Portal GEOINFO de infraestrutura de dados espaciais da EMBRAPA e o Portal da Biodiversidade. Peixoto (2012) indica o repositório SinBiota (Sistema de Informação Ambiental do Programa Biota/Fapesp).

Conforme Pampel et al. (2013) e Sales (2014), respectivamente, os repositórios encontrados estão de acordo com a tipologia repositórios de dados de pesquisa institucional e repositórios de dados. Verifica-se uma consonância entre a fundamentação teoria e a investigação realizada.

5 CONSIDERAÇÕES FINAIS

A partir da pesquisa realizada, concluímos que o número de repositórios de dados em instituições de pesquisa brasileira ainda é insipiente. Concordamos com Sales (2014, p.62) sobre este contexto quando afirma que: “este tipo de repositório é mais raro no contexto brasileiro”. Contudo, a partir do crescimento exponencial na geração de dados e considerando o valor deste recurso no fazer científico, acreditamos que esta situação poderá modificar-se. Novos repositórios de dados de pesquisa poderão surgir, sejam a partir de iniciativas isoladas das instituições ou mesmo através da geração de consórcios.

REFERÊNCIAS

COSTA, M.M. **Diretrizes para uma política de gestão de dados científicos no Brasil**. Brasília: 2017.288p. Tese (Doutorado em Ciências da Informação). Faculdade de Ciência da Informação. Universidade de Brasília. Disponível em:<<http://repositorio.unb.br/handle/10482/24895>> Acesso em: 24 abr. 2018.

DATAVERSE. **About**. Disponível em:< <https://dataverse.org/about>> Acesso em: 03 jun. 2018.

IBICT. IBICT lança Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã.

Disponível em:<<http://www.ibict.br/Sala-de-Imprensa/noticias/2016/ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>> Acesso em: 29 jun. 2018.

KNB. **The Knowledge Network for Biocomplexity**. Disponível em:<https://knb.ecoinformatics.org/knb/do%20cs/intro.html#what-s-in-this-guide>. Acesso em: 08 out. 2018.

LEITE, F. et al. **Repositórios Institucionais: boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: IBICT, 2012. Disponível em:< <http://livroaberto.ibict.br/bitstream/1/703/1/Boas%20pr%C3%A1ticas%20para%20a%20constru%C3%A7%C3%A3o%20de%20reposit%C3%B3rios%20institucionais%20da%20produ%C3%A7%C3%A3o%20cient%C3%ADfica.pdf>> Acesso em: 10 jun. 2018.

OLIVEIRA, A. C. S. **Desvendando a autorialidade colaborativa na e- Science sob a ótica dos direitos da propriedade intelectual**. João Pessoa: 2016. 297p. Tese (Doutorado em Ciências

da Informação). Centro de Ciências Sociais Aplicadas. Universidade Federal da Paraíba. Disponível em: <<http://tede.biblioteca.ufpb.br/handle/tede/8849>>. Acesso em: 7 jun 2018.

PAMPEL, H. et al. **Making research data repositories visible**: the re3data.orgRegistry. PLoSOne, San Francisco, v. 8, n. 11, 2013. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078080>>. Acesso em: 19 maio. 2018.

PEIXOTO, N.M. **Arquitetura de informação em repositório de dados científicos**: análise da interface do repositório do PELD: Programa de Pesquisa Ecológicas de Longa Duração. 2012. Porto Alegre: 2012. 78p. Trabalho de Conclusão de Curso (Graduação) Universidade Federal do Rio Grande do Sul. Faculdade de Biblioteconomia e Comunicação. Curso de Biblioteconomia. Disponível em: <<https://lume.ufrgs.br/handle/10183/69727>>. Acesso em: 29 jul. 2018.

SALES, L.F. **Integração semântica de publicações científicas e dados de pesquisa**: proposta de modelo de publicação ampliada para a área de Ciências Nucleares. Rio de Janeiro: 2014. 265p. Tese apresentada ao Programa de Pós-graduação em Ciência da Informação da Universidade Federal do Rio de Janeiro/. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf>>. Acesso em: 05 dez. 2017.

SAYÃO, L.F; SALES, L.F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90 – 115, maio/ago 2016. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939/20122>>. Acesso em: 18 abr. 2018.