



Repositórios de Dados Abertos de Pesquisa em plataformas de universidades brasileiras¹

Open Research Data Repositories on Brazilian university platforms

Adriana Alves Rodrigues 

Doutora em Ciência da Informação
Universidade Federal da Paraíba, Brasil
adrianacontemporanea@gmail.com

Guilherme Ataíde Dias 

Doutor em Ciências da Comunicação
Universidade Federal da Paraíba, Brasil
guilhermeataide@gmail.com

Resumo

A pesquisa teve como objetivo descrever e analisar os aspectos técnicos dos repositórios de dados de pesquisa vinculados às Instituições de Ensino Superior (IES) no Brasil. Para isso, realizou-se pesquisa exploratória e uma análise descritiva a partir do *site Re3data*, um dos principais diretórios de registro de dados globais de pesquisa. Foram analisados oito repositórios a partir das categorias: *software*, acesso aberto, licenças e padrão de metadados. Os resultados sugerem que estes aspectos podem apresentar vantagens por parte dos pesquisadores em realizar pesquisas frequentes nestes repositórios, por um lado, e por outro, podem acarretar inúmeras discrepâncias ao realizar suas coletas, compartilhamento, reutilização e encontrabilidade dos dados quando não estão bem gerenciados.

Palavras-chave: repositórios; universidades brasileiras; dados de pesquisa.

Abstract

The research aims to describe and analyze aspects of research data repositories linked to Higher Education Institutions (HEIs) in Brazil. To this end, exploratory research and a descriptive analysis were carried out using the Re3data website, one of the main directories for recording global research data. Eight repositories were analyzed based on the categories: software, open access, licenses and metadata standards. The results suggest that these aspects may present advantages for researchers in carrying out frequent searches in these repositories, on the one hand, and on the other, they can lead to numerous discrepancies when collecting, sharing, reusing and finding data when they are not well managed.

Keywords: *Repositories; Brazilian universities; research data.*

¹ Versão completamente reconfigurada do Pôster apresentado na 14ª Conferência Lusófona de Ciência Aberta (Confoa), em 2023.



1 INTRODUÇÃO

O compartilhamento de dados de pesquisa se tornou uma prática indispensável para a comunidade científica e está se tornando cada vez mais relevante na ciência contemporânea. Os desdobramentos advindos do contexto da *e-Science*, enquanto quarto paradigma da ciência (Gray, 2009), apontam para a demanda de uma ciberinfraestrutura para que os pesquisadores possam ter acesso aos dados. Nesse cenário, a Declaração da Organização para Cooperação e Desenvolvimento Econômico (OCDE)², em 2004, na qual autentica o acesso sem restrições aos dados de pesquisa e o uso destes em favor do progresso científico, abriu caminhos para a utilização de dados de pesquisa e construção dos repositórios com essa finalidade.

Diante esse cenário, as Instituições de Ensino Superior (IES), agências de fomento e pesquisadores iniciaram o caminho junto a exploração dos serviços de repositórios digitais nas mais variadas estruturas tecnológicas, bem como formas diversas de protótipos de planejamento de gestão dos dados de pesquisa (Sayão; Sales, 2016). No entanto, de um lado, há a questão da emergência de uma estrutura tecnológica adequada para o gerenciamento da confiabilidade, qualidade, arquivamento, entre outros. Por outro, os pesquisadores necessitam de letramento digital em como acessar esses conjuntos de dados, como compartilhar e como reutilizar de maneira ética e eficaz.

O desenvolvimento de repositórios de dados de pesquisa demanda o estabelecimento de novas práticas com vistas a gerenciar o fluxo por parte dos pesquisadores, envolvendo aporte teórico e tecnológico para sua efetivação. O compartilhamento dos dados primários dos investigadores está intrínseco nas convenções da ciência contemporânea pautada na disponibilização dos dados de pesquisa em ambiência digital. Esta propagação dos dados de pesquisa através dos repositórios está se tornando o constructo central da comunicação científica. Todavia, essa perspectiva deve ser empreendida de modo em que haja uma associação entre investigadores e repositórios, com a finalidade de facilitar essa disseminação (Castelli; Manghi; Thanos, 2013). De modo geral, repositórios de dados são espaços para armazenamentos e podem ser entendidos como “grandes infraestruturas de banco de dados definidas para gerenciar, compartilhar, acessar e arquivar conjuntos de dados dos pesquisadores” (Uzwyszyn, 2016, p. 18, tradução nossa)³.

Ao analisar a situação dos repositórios de dados institucionais em direção à gestão de dados científicos no Brasil, Henning *et al.* (2018) constatam que essa questão ainda precisa ser maturada e que as instituições brasileiras se encontram no caminho da aderência para a gestão de dados de pesquisa, e que, portanto, existe um longo caminho a ser percorrido para sua total efetivação. Ainda assim, neste cenário em constante evolução, há cerca de 110 repositórios indexados no *OpenDOAR*⁴, que é o diretório global de repositórios de dados com acesso aberto (Ribeiro, 2019). Tal fato demonstra avanços significativos no país rumo à uma implementação mais efetiva de uma cultura de compartilhamento e reuso de dados científicos.

² OECD. Declaration on Access to Research Data From Public Funding, Paris, 2004. Disponível em: <https://bit.ly/3FvMRcL> Acesso e: 17 mar. 2023

³ No original: “Online research data repositories are large database infrastructures set up to manage, share, access, and archive researchers’ datasets”.

⁴ Disponível em: <http://v2.sherpa.ac.uk/pendoar/> Acesso em: 8 nov. 2023

Nessa direção, os repositórios de dados de pesquisa contribuem para o compartilhamento e reutilização dos dados associados às investigações científicas. As principais recomendações para a gestão de dados de maneira adequada sugerem que os dados de pesquisa devem estar disponibilizados em um repositório apropriado que garanta a integridade da pesquisa (Uzwyszyn, 2016) e que possa trazer benefícios acadêmicos como visibilidade da pesquisa científica, maximização do compartilhamento, potencialização do reuso dos dados, estimule o debate científico, dentre outros aspectos centrais. Entretanto, ainda se vivencia cenários desafiantes para a implementação dos repositórios de dados no Brasil para que estes possam atuar em favorecimento do progresso da ciência. Em face desse contexto, a proposta dessa pesquisa é analisar os repositórios de dados de pesquisa no Brasil e apontar os aspectos técnicos para sua implantação e gestão dos dados de pesquisa.

2 DADOS DE PESQUISA E REPOSITÓRIOS DE DADOS

O ano de 2004 pode ser considerado como um marco teórico e temporal para os estudos dos dados. Uma reunião em Paris dos ministros de Ciência e Tecnologia vinculados aos países da OECD teve como objetivo a urgência de um documento padrão no âmbito internacional para o acesso dos dados de pesquisa (Henning *et al.*, 2018). O resultado desse deu origem à *Declaration on access to research data from public funding*⁵, na qual determina as políticas mandatórias que estão vinculadas ao acesso aos dados de pesquisa.

Alguns preceitos, como proteção da propriedade intelectual, transparência, interoperabilidade, qualidade, eficiência e segurança dos dados, estavam neste documento como o início da formalização dos dados de pesquisa para acessos e usos. (Organization For Economic Cooperation and Development, 2004). Mais adiante, a mesma instituição estabeleceu novas diretrizes atualizadas integrando quatro aspectos a serem seguidos, quais sejam: flexibilidade, sustentabilidade, qualidade e segurança. Neste período desde o lançamento destas guias, outras iniciativas foram arregimentando o terreno para os dados de pesquisa. A *The Royal Society*, por exemplo, em 2012, divulgou um relatório na qual empreende uma análise e aponta os desafios advindos pelos novos usos de dados científicos, como manipular, armazenar, acessar, dentre outros aspectos desse universo (The Royal Society, 2012).

O fato é que tantos os pesquisadores, quanto às IES, demais agências de fomento e Centros de Pesquisa nacionais e internacionais despertam para essa realidade contemporânea de importância dos dados durante todo a processualidade da pesquisa acadêmica e implantam essa prática como uma prerrogativa do fazer científico dentro de seus ambientes de estudos. Coloque-se nesse contexto o empoderamento das Tecnologias Digitais da Informação e Comunicação (TDICs) que auxiliam nos processos de acesso, coleta, manipulação, poder analítico e de visualização de dados. Assim sendo gerados pelas mais diversas disciplinas e áreas de conhecimento, os dados de pesquisa são frutos de estudos científicos sistemáticos. Estes estudos utilizam ferramentas tecnológicas em várias funções, incluindo procedimentos de experimentação, observação, análise, simulação e interpretação da pesquisa acadêmica. A OECD (2007), define o termo dado de pesquisa como “registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias de pesquisa científica e que são comumente aceitos na comunidade científica

⁵ Disponível em: <https://legalinstruments.oecd.org/en/instruments/157>. Acesso em: 7 nov. 2023

como necessários para validar resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação parcial e sistemática do tema que está sendo pesquisado” (OECD, 2007, p. 13, tradução nossa)⁶. Um outro ponto de vista de definição de dados de pesquisa expõe a categorização desses dados no universo científico. Para Sayão e Sales (2015, p. 36), dados de pesquisa “é todo e qualquer tipo de registro coletado, observado, gerado ou usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica”. Além dessa definição, os autores propõem uma categorização no Guia de Gestão de Dados de Pesquisa (2015), quais sejam:

[...] segundo sua natureza: (a) observacionais: gerados por meio de observações diretas que podem ser associadas a lugares e/ou tempo específico, por sua natureza instantânea qualificam-se como registros históricos, pois não podem ser coletados uma segunda vez e, portanto, devem ser submetidos a processos de curadoria que os preservem para sempre; (b) computacionais, são o resultado da execução de modelos computacionais ou de simulações; (c) experimentais, são provenientes de situações controladas em laboratórios. [...] Podem ser números, imagens, vídeos, áudio, software, algoritmos, equações, animações, modelos e simulações. Por fim, conforme o fluxo da pesquisa, (a) dados brutos, coletados diretamente de instrumentos científicos de pesquisa; (b) derivados, resultados do processamento ou combinação de dados brutos ou de outros dados; (c) canônicos ou referenciais, coleções de dados consolidados e arquivados geralmente em grandes centros de dados (Sayão; Sales, 2015, p. 8-9).

A proposta dos autores visa a especificação dos dados de pesquisa, considerando a complexidade e heterogeneidade que estes se apresentam. O potencial informacional trazido pelos dados no contexto da ciência contemporânea muda o cenário científico mundial. Os dados que eram antes considerados como a última etapa de uma investigação científica assumem uma posição de centralidade considerando o seu ciclo de vida. No entanto, apesar de ser o sustentáculo da pesquisa científica, os dados não se mostram decifráveis de lidar, seja no aspecto de manipulação, coleta, organização, estruturação e disponibilização (Dudziak, 2016).

Uma das iniciativas que contribuiu para o compartilhamento de dados de pesquisa foi o Movimento do Acesso Aberto (*Open Access*), na década de 1990, na qual tem como bandeira de luta defender o acesso livre das pesquisas científicas através da produção acadêmica, meios de divulgação e publicação em ambiência digital, sem custos ou restrições. Nesta época, em 1999, foi criada a *Open Archives Initiative* (OAI), no Novo México, durante a Convenção de Santa Fé. Neste encontro, foram definidos os princípios norteadores do acesso aberto em direção à pesquisa científica (Cavalcanti; Sales, 2017).

A partir desse marco histórico, outras iniciativas neste sentido foram empreendidas em diversos lugares, a exemplo de Bethesda, Berlim e Budapeste. Em nível nacional, em 2004, durante o 2º Simpósio Internacional de Bibliotecas Digitais, em Campinas - SP, foi estabelecida uma declaração realizada pelos integrantes do evento acadêmico na qual direcionava apoio ao movimento do Acesso Aberto, ocasião em que gerou outras três empreendimentos neste sentido, quais sejam: 1) “Declaração de Salvador Compromisso com a Equidade”, prerrogativa que determina o acesso equitativo e aberto à produção acadêmica, em setembro de 2005; 2) a “Declaração de Salvador sobre Acesso Aberto:

⁶No original: “factual records (numerical scores, textual records, images and sounds) used as primary sources of scientific research and that are commonly accepted in the scientific community as necessary to validate research results”.

Perspectiva do Mundo em Desenvolvimento”, que solicita ao governo a condição de acesso aberto às produções científicas que estejam vinculadas ao dinheiro público, também no mesmo ano e 3) “Declaração de São Paulo em apoio ao Acesso Aberto”, elaborada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia/IBICT (Fausto, 2015).

Nesta direção, este cenário do acesso aberto encontra ressonância no contexto do *Big Data*, entram em cena os repositórios de dados com a finalidade de promover a visibilidade das investigações científicas. Os repositórios são como robustos bancos de dados formatados para compartilhar, gerenciar, arquivar conjuntos de dados e acessar, e podem integrar em sua composição dados de pesquisas locais ou globais, seja em Universidades, Centros de Pesquisas e Agências de Fomento. A ideia simples é que o compartilhamento de dados melhora os resultados e impulsiona a investigação. A partilha de dados de investigação é cada vez mais reconhecida como uma componente essencial da investigação acadêmica e científica.

Neste sentido, Monteiro et al. (2017, p. 35-36) trazem que “os repositórios de dados científicos são ambientes implementados nas universidades com infraestrutura para dar suporte aos pesquisadores no gerenciamento e na disponibilização de dados científicos e, dessa forma, ampliar o acesso para que outros pesquisadores possam reutilizá-los”. Os autores chamam a atenção para que a finalidade maior dos repositórios se revela de forma ampla com a intenção de assegurar que estes dados “sejam armazenados em um suporte que permita sua recuperação, ampliando seu potencial de compartilhamento” (Monteiro et al., 2017, p 36).

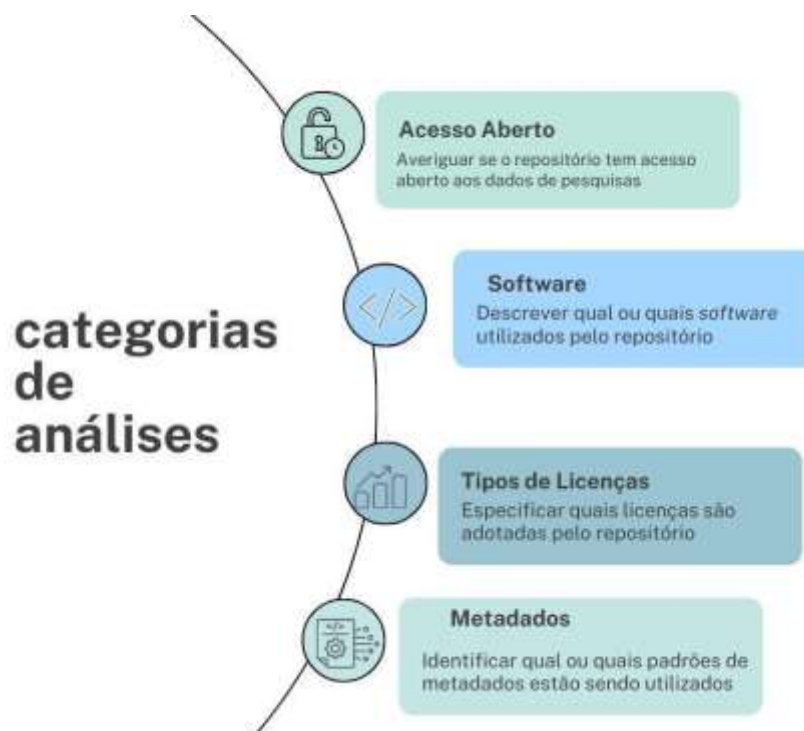
Na visão de Sayão e Sales (2016, p. 96), a missão primordial dos repositórios de dados de pesquisa frente ao contexto da *e-Science* é possibilitar o acesso de modo contínuo e aberto a esses dados de pesquisas, e listam alguns benefícios: Visibilidade dos dados, compartilhamento dos dados, crédito ao autor dos dados, preservação digital, memória científica e transparência, segurança, disponibilidade e curadoria dos dados, serviços inovadores, reuso dos dados, redes de repositórios e indicador de qualidade e produtividade da instituição.

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa tem caráter descritivo, a partir do levantamento de dados e análise descritiva. De acordo com Gil (2011, p. 28), pesquisas com esse aspecto “têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis”. Nesta direção, foi realizado o levantamento de dados em março de 2023 na plataforma *Re3data*⁷. A escolha dessa base de dados em questão se deve pela sua relevância acadêmica dentro da Ciência da Informação. Em seguida, elaboramos as categorias de análises para o nosso *corpus* empírico, conforme mostra o Gráfico 1, quais sejam: a) *Acesso Aberto*: averiguar se o repositório tem acesso aberto aos dados de pesquisa; b) *Software*: Descrever qual ou quais *software* utilizados pelos repositórios; c) *Tipos de Licença*: Especificar quais licenças são adotadas pelo repositório; e d) *Metadados*: Identificar qual ou quais padrões de metadados estão sendo utilizados.

⁷Disponível em: <https://www.re3data.org/> Acesso em: 17 mar. 2023.

Gráfico 1 - Definição das categorias de análises



Fonte: Dados de Pesquisa (2023).

Ao todo, até o momento da finalização desta pesquisa, estão registrados 18 repositórios de dados de pesquisas associados ao *Re3data*. Contudo, para especificar mais ainda nossa análise e focar na ambiência acadêmica, selecionamos os repositórios que estejam vinculados às IES. Deste modo, dos 18 repositórios disponíveis no Brasil, apenas 8 têm essa característica, são eles: 1) Base de Dados Científicos da Universidade Federal do Paraná⁸, 2) *IBICT Dataverse Network*⁹, 3) Repositório Institucional Unesp¹⁰; 4) CEDAP-(Centro de Documentação e Acervo Digital da Pesquisa - UFRGS)¹¹, 5) Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí¹², 6) Maenduar (IBICT)¹³, 7) REDU - Repositório de Dados de Pesquisas Institucionais da Unicamp¹⁴ e 8) Dados Abertos de Pesquisa @PUC-Rio¹⁵.

4 RESULTADOS

A partir das categorias de análises pré-estabelecidas, os dados foram tabulados no Quadro 1 como um panorama de compreensão a partir dos resultados encontrados (quadro 1).

⁸Disponível em: <https://bit.ly/3yQOPk5>. Acesso em: 17 mar. 2023.

⁹Disponível em: <https://bit.ly/3LyaISj>. Acesso em: 17 mar. 2023.

¹⁰Disponível em: <https://repositorio.unesp.br/>. Acesso em: 17 mar. 2023.

¹¹Disponível em: <https://www.ufrgs.br/cedap/>. Acesso em: 17 mar. 2023.

¹²Disponível em: <https://bit.ly/3n1Obh0>. Acesso em: 17 mar. 2023.

¹³Disponível em: <https://bit.ly/3TrJSaS>. Acesso em: 17 mar. 2023.

¹⁴Disponível em: <https://redu.unicamp.br/>. Acesso em: 17 mar. 2023.

¹⁵Disponível em: <https://bit.ly/3ltwrKU>. Acesso em: 17 mar. 2023.

Quadro 1 - Resultados da aplicação das Categorias de Análise

REPOSITÓRIOS	ACESSO ABERTO	SOFTWARE	TIPOS DE LICENÇA	PADRÃO DE METADADOS
Base de Dados Científicos da Universidade Federal do Paraná	✓	DSPACE	CREATIVE COMMONS	DUBLIN CORE
IBICT Dataverse Network	✓	DATAVERSE	COPYRIGHT & TERMS OF USE	DDI - DATA DOCUMENTATION INITIATIVE (FORMATO SWORD E DATA CITE)
Repositório Institucional Unesp	✓	DSPACE	CREATIVE COMMONS	DUBLIN CORE
CEDAP (Centro de Documentação e Acervo Digital da Pesquisa)	✓	DSPACE	CREATIVE COMMONS	DUBLIN CORE
Repositório de Dados de Pesquisas do IFG Urutai	✓	DATAVERSE	CREATIVE COMMONS	DDI - DATA DOCUMENTATION INITIATIVE (FORMATO SWORD)
Maenduar (IBICT)	✓	NÃO INFORMADO	CREATIVE COMMONS	DUBLIN CORE OPEN AIRE
REDU - Repositório de Dados de Pesquisas Institucionais da Unicamp	✓	DATAVERSE	COPYRIGHT & TERMS OF USE	NÃO INFORMADO
Dados Abertos de Pesquisa @PUC-Rio	✓	MAXWELL SYSTEM	CREATIVE COMMONS	MTD-BR - MODELO BRASILEIRO DE METADADOS PARA TESES E DISSERTAÇÕES, ETD-MS - AN INTEROPERABILITY METADATA STANDARD FOR ELECTRONIC THESES AND DISSERTATIONS, O PADRÃO INTERNACIONAL.

Fonte: Dados de pesquisa (2023).

Em uma análise geral, todos os repositórios digitais mantêm acesso aberto em que os dados de pesquisa são disponibilizados livremente de maneira gratuita para a comunidade acadêmica. Esse fato está alinhado ao movimento dos dados abertos (*Open Data*), movimento de grande relevância para o compartilhamento e acesso aos dados científicos. O primeiro passo rumo a esse movimento aconteceu ainda em 1995, nos Estados Unidos (EUA), momento em que o *Global Change Data and Information System (GCDIS)*¹⁶ levantou a discussão e colocou em pauta o processo de abertura dos dados de pesquisa de um modo mais amplo e global (Rodrigues *et al*, 2010). Desta forma, “a abertura deve permitir que a produção científica seja constante e eficaz a partir da reutilização dos conjuntos informacionais dispostos em pesquisas anteriores” (Gama; Cianconi; González de Gómez, 2023, p. 34), integrando as ciberinfraestrutura das plataformas digitais como os repositórios, por exemplo, bem como a “valorização de ativos digitais”.

Na categoria *Software*, foram apontados os mais utilizados pelos repositórios digitais em atividade. O *Dataverse* e *Dspace*¹⁷ são os mais recorrentes em nossa análise, ambos pelas funcionalidades, usos e segurança. O primeiro, sendo licenciado sob a Licença *Apache*,

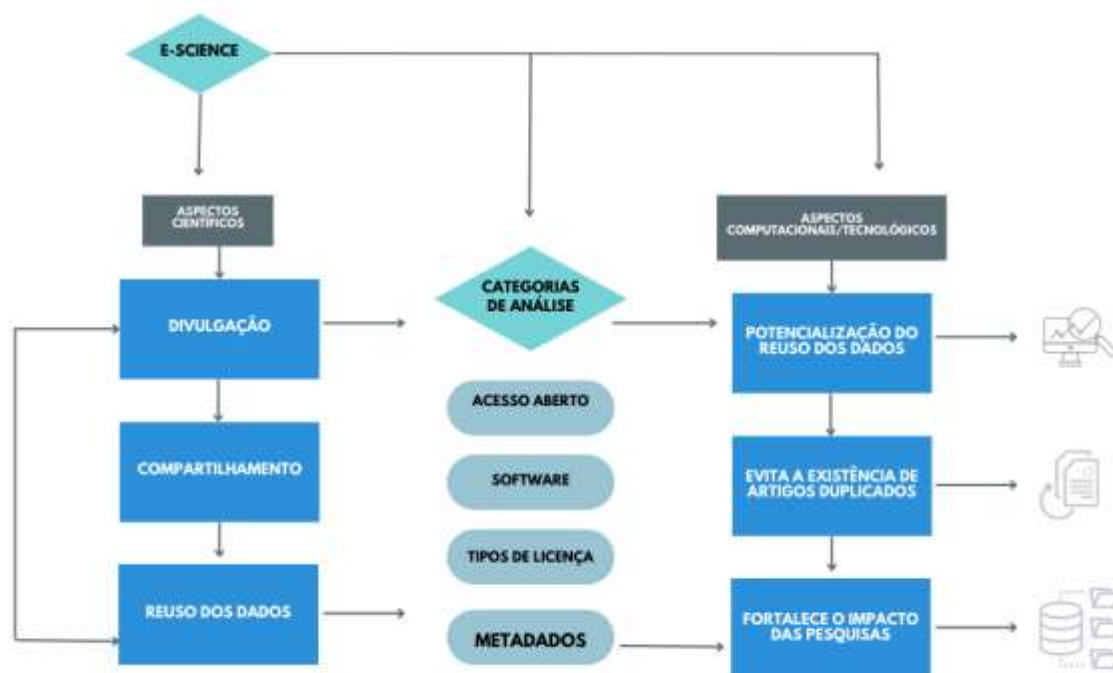
¹⁶ Disponível em: <https://bit.ly/3Z2PKIG>. Acesso em: 20 mar. 2023.

¹⁷ Disponível em: <https://en.wikipedia.org/wiki/Dspace>. Acesso em: 20 mar. 2023.

Versão 2.0 e por ser de código-fonte aberto armazenando de forma segura os dados, abrigando outros *dataverses*¹⁸; e o segundo, também de código-fonte aberto e é específico como um sistema de arquivos digitais e segue a licença de distribuição BSD. Merece destaque o *Maxwell System*, da Puc-Rio, que desenvolveu um repositório próprio para seus dados de pesquisa, o que denota uma cautela na curadoria dos dados e a personalização do repositório próprio da instituição. Em relação aos tipos de licenças utilizadas pelas instituições de ensino superior, a investigação constatou as preferências pela *Creative Commons*, na qual demonstra mais flexibilidade e modos de facilitar o compartilhamento dos dados e reutilização sob a filosofia do *copyleft*. No que se refere aos padrões de metadados analisados, constatamos que o mais recorrente utilizado é o *Dublin Core*, padrão criado especificamente para informações bibliográficas, e por essa razão maior, o mais utilizado em repositórios acadêmicos.

Os resultados desta investigação sugerem que estes aspectos constatados podem apresentar vantagens por parte dos pesquisadores em realizar pesquisas frequentes nestes repositórios, por um lado, e, por outro, podem acarretar inúmeras discrepâncias ao realizar suas coletas, compartilhamento, reutilização e encontrabilidade dos dados quando não estão bem gerenciados. Durante o processo de análise, observou-se que parte dos dados que estavam sendo disponibilizados em alguns repositórios apresentavam inconsistências ao acessá-los, por exemplo, quando o *input* não atendia ao solicitado. Os repositórios analisados apresentam o potencial para dinamizar o acesso, uso, reuso e compartilhamento dos dados de pesquisa em favorecimento ao cenário da *e-Science*, porém, é necessário salientar a implementação de políticas para estimular os depósitos de dados oriundos das pesquisas científicas, e, conseqüentemente, ampliar o raio de alcance tanto para a comunidade acadêmica quanto para as esferas sociais.

Figura 1 - Repositórios analisados estão alinhados às características da *e-Science*



Fonte: Dados de pesquisa (2023)

¹⁸ Disponível em: <https://en.wikipedia.org/wiki/Dataverse> Acesso em: 20 mar. 2023.

Neste sentido e considerando os resultados, a discussão proposta está inserida no quarto paradigma da ciência, a partir das contribuições de Gray (2009), a *e-Science*, em que o mesmo assinala ser um ponto de encontro entre as tecnologias digitais e pesquisadores. Para Jankowski, (2007), *e-Science* se refere ao conjunto de instrumentos e aparatos tecnológicos adequados para auxiliar a ciência contemporânea diante da produção dos grandes volumes de dados produzidos diariamente, com características colaborativas e multidisciplinares. Os repositórios de dados analisados apresentam algumas das prerrogativas da *e-Science* (figura 1), sobretudo, quando estão em alinhamento os aspectos científicos e aspectos computacionais/tecnológicos em favorecimento da ciência contemporânea pautada pelo gerenciamento de grandes volumes de dados.

Nesse contexto, Albagli, Appel e Maciel (2013, p. 10) estabelecem três mecanismos como as principais que compõem o cenário da *e-Science* como: distribuição, volume e a heterogeneidade. Para os autores, “A questão, em particular, desempenha um papel chave devido à grande variedade de plataformas, fontes de dados, sistemas, formatos, interfaces, políticas, abordagens etc.”. Deste modo, ao estimular as práticas de publicação de dados e como forma de potencializar o seu uso, os resultados nos mostram que os repositórios devem considerar estes aspectos importantes: a) Superar de barreiras tecnológicas, operacionais e normativas para que os dados estejam qualificados e assegurados para os usos; b) Lidar com a comunidade acadêmica multidisciplinar em termos de proprietários de dados e consumidores de dados e c) Desenvolver um Plano de Gestão de Dados específico para os repositórios e práticas de edição de cópias para garantir que os documentos de dados que publicam contenham uma citação inteligível para o conjunto de dados. Além disso, um aspecto chave para o aprimoramento é melhorar a organicidade da heterogeneidade dos *datasets* e os formatos que os suportam.

Assim, as práticas da *e-Science* buscam a disponibilização dos dados abertos públicos a partir do conjunto de ações que favoreçam as investigações científicas tanto no que concerne ao trabalho individual quanto nas pesquisas colaborativas. E os repositórios, neste sentido, exercem a função de serem os faróis científicos para o acesso, compartilhamento, uso e reuso dos dados de pesquisa. Evidentemente, faz-se necessário a implementação de Políticas de Gestão de Dados nas IES para um melhoramento na gestão de dados dada a heterogeneidade dos dados, fontes, sistemas, metadados, interfaces, entre outros aspectos que reúnam esforços que otimizem o acesso aos dados de pesquisa.

5 CONSIDERAÇÕES FINAIS

A pesquisa levantou os principais aspectos para a publicação de dados de pesquisa promovidos por oito repositórios de dados científicos selecionados a partir de quatro categorias (acesso aberto, *software*, tipos de licenças e padrão de metadados). Observamos que os repositórios analisados adotam diferentes características em conformidade à realidade daquela Instituição, havendo, portanto, distintas configurações e formas de usos, compartilhamento e reuso dos dados científicos. Pode trazer benefícios de conforto por parte dos pesquisadores em realizar pesquisas frequentes nestes repositórios e, portanto, já conhece o caminho dos dados, por um lado, mas por outro, pode acarretar inúmeras discrepâncias ao realizar suas coletas, compartilhamento, reutilização e encontrabilidade dos dados.

A prática da existência dos repositórios não devem ser enxergada como substituta das publicações em revistas científicas, teses e dissertações, mas como uma ponte para o

acesso aos dados de toda a investigação científica, tendo em vista que, ao finalizar pesquisas densas, muitos dados podem ficar de fora por inúmeras questões, sejam de ordem de limitação de espaço por exemplo, é o caminho natural é que esses dados fiquem armazenados nos ambientes de arquivos virtuais dos pesquisadores (*Google Drive, Dropbox, etc*) do pesquisador. Ou seja, ficam na invisibilidade. Os repositórios de dados, quando bem estruturados, resguardam um lugar de muita relevância durante toda a processualidade da pesquisa científica, tanto no sentido de depositar, organizar, compartilhar e reusar, quanto também na perspectiva documental em se tornar um artefato de memória digital, retroalimentando novos ciclos de vida dos dados para futuras gerações, e ampliando o fazer científico em larga escala.

No entanto, a implementação dos repositórios nas IES, por exemplo, aponta para novos desafios que extrapolam a dimensão tecnológica de operacionalizar os dados em ambiência digital, considerando a diversidade dos dados advindos em vários formatos, modelos e áreas. Barreiras de ordem financeira por parte da Instituição, de ordem da manutenção, capacitação para que esses pesquisadores possam gerenciar os seus próprios dados de pesquisa mostram que ainda falta um longo caminho a ser percorrido para sua real efetivação, e assim, instaurar uma cultura de compartilhamento de dados de pesquisa nas Universidades, e, desta forma, a interdisciplinaridade fará mais sentido com mais impacto social e científico.

Assim, conclui-se que os repositórios devem considerar estes aspectos importantes para sua eficácia, tais quais: a) Superar de barreiras tecnológicas, operacionais e normativas para que os dados estejam qualificados e assegurados para os usos; b) Lidar com a comunidade acadêmica multidisciplinar em termos de proprietários de dados e consumidores de dados e c) Desenvolver um Plano de Gestão de Dados específico para os conjuntos de dados a serem depositados nos repositórios e práticas de edição de cópias para garantir que os documentos de dados que publicam contenham uma citação inteligível para a comunidade acadêmica. A pesquisa aponta ainda para conteúdos que merecem maior aprofundamento como investigar critérios no que se refere à propriedade intelectual, plano de gestão de dados, qualidade e políticas de compartilhamento de dados, cruzando os dados com repositórios estrangeiros para ser ter um termômetro das práticas adotadas nesses países e quais aspectos podem ser mais potencializados. Ou, ainda, quais os impactos reais desses gerenciamentos para a comunidade científica.

REFERÊNCIAS

ALBAGLI, S.; APPEL, A. L.; MACIEL, M. L. e-Science e ciência aberta: questões em debate. In: XIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais**[...] Florianópolis: PPGCI/UFSC, 2013 Disponível em: <https://ridi.ibict.br/handle/123456789/465>. Acesso em: 2 nov. 2023.

CASTELLI, D., MANGHI, P. E THANOS, C. Uma visão para infra-estruturas de comunicação científica: unindo os domínios das bibliotecas digitais de investigação e dos centros de dados científicos. **Revista Internacional de Bibliotecas Digitais**, 2013. Disponível em: <https://shre.ink/Tdnt>. Acesso em: 13 nov. 2023.

CAVALCANTI, M. T; SALES, L. Gestão De Dados De Pesquisa: um panorama da atuação da União Europeia. **Biblos**: revista do Instituto de Ciências Humanas e da Informação, Rio

Grande, v. 31, n. 1, p. 73-98, jan./jun.2017. Disponível em: <https://periodicos.furg.br/biblos/article/view/5789/4616>. Acesso em: 13 nov. 2023.

DUDZIAK, Elisabeth. Dados de Pesquisa agora devem ser armazenados e citados. **SIBIUSP**, 2016. Disponível em: <https://n9.cl/cpk2p>. Acesso em: 2 nov. 2023.

FAUSTO, S. Evolução do Acesso Aberto: breve histórico. **SciELO em Perspectiva**, [Internet], 2013. Disponível em: <https://blog.scielo.org/blog/2013/10/21/evolucao-do-acesso-aberto-breve-historico/>. Acesso em: 12 nov. 2023.

GAMA, I. O.; CIANCONI, R. B.; GONZÁLEZ DE GOMEZ, M. N. G. A abertura científica: o processo de ressignificação a partir dos movimentos Open Access e Open Science. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 27, n. 4, p. 28-53, out/dez 2022. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/29247/32172>. Acesso em: 4 mar. 2023.

GIL, A.C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GRAY, J. 2009 Jim Gray on e-science: a transformed scientific method. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. (Eds.). **The fourth paradigm: data intensive scientific discovery**. Redmond, Washington: MicroSoft Research, 2009.

HENNING, Patrícia Corrêa; RIBEIRO, Cláudio José Silva; SALES, Luana Faria; MOREIRA, J. L. R.; SANTOS, L. O. B. S. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no Ecosistema dos dados FAIR. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais[...]**. Londrina: PPGCI/UEL, 2018. Disponível em: <https://brapci.inf.br/index.php/res/v/150613> . Acesso em: 12 nov. 2023.

JANKOWSKI, N.W. Exploring e-Science: An Introduction. **Journal of Computer-Mediated Communication**, v. 2, n. 2 jan. 2007. Disponível em: <https://shre.ink/TdH5>. Acesso em: 12 nov. 2023.

MONTEIRO, E. C. S. A.; AFFONSO, E. P; BORBA, V. U. ; SANTANA, R. C. G. . Repositório de Dados Científicos: aspectos sobre privacidade de dados. In: DUTRA, M. L.; MACEDO, D. D. J. (Org.). **Repositório de Dados Científicos: aspectos sobre privacidade de dados**. 1 ed. Florianópolis: [s.n.], 2017. p. 100-106.

OECD. DECLARATION ON ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING 2004 **Declaration on Access to Research Data From Public Funding**. 2004. Disponível em: <https://bit.ly/403v3hjm>. Acesso em: 2 fev. 2023.

RE3DATA, REGISTRY OF RESEARCH DATA REPOSITORIES 2023. **Re3data Registry of Research Data Repositories**. 2023. <https://www.re3data.org/>. Acesso em: 2 fev. 2023.

RIBEIRO, C. J. S.. Modelo de Maturidade para Repositórios Digitais: um caminho para sua adoção na gestão de dados de pesquisa. **Liinc em revista**, Rio de Janeiro, v. 15, p. 224, 2019. Disponível em: <https://revista.ibict.br/liinc/article/view/4816/4317>. Acesso em: 3 nov. 2023.

RODRIGUES, E.; et al. **Os repositórios de dados científicos: estado da arte**. Porto: UP: 2010. Disponível em: <https://repositorium.sdum.uminho.pt/bitstream/1822/10830/1/Relat%C3%B3rio%20Estado%20da%20Arte%20dos%20Reposit%C3%B3rios%20de%20Dados%20Cient%C3%ADficos%20Vers%C3%A3o%201.0%20Final.pdf>. Acesso em: 4 mar. 2023.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários de pesquisadores**. Rio de Janeiro: CNEN, 2015. D. Disponível em: <https://www.aben.com.br/Arquivos/420/420.pdf>. Acesso em: 7 nov. 2023.

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: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27939/20122>. Acesso em: 4 out. 2023.

THE ROYAL SOCIETY. **Science as an open enterprise**, 2012. Disponível em: <https://shre.ink/Used>. Acesso em: 12 nov. 2023.

UZWYSHYN, Ray. Research data repositories: the what, when, why, and how. **Computers in Libraries, Westport**, v. 36, n. 3, Apr. 2016. Disponível em: <https://shre.ink/Tdn5>. Acesso em: 4 nov. 2023.

AGRADECIMENTOS

À Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ/PB), Termo no 1969/2022, referente a pesquisa de Pós-Doutorado desenvolvida no Programa de Pós-Graduação em Ciência da Informação na Universidade Federal da Paraíba (PPGCI/UFPB), assim como ao Edital FAPESQ/PB 09/2021, demanda universal.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo 311563/2018-0.