

ALEIA: REPOSITÓRIO DE DADOS DE PESQUISA DO IBICT

Tatyane Guedes Martins da Silva
Leticia Guarany Bonetti



O repositório de dados de pesquisa Aleia é uma das iniciativas do Instituto Brasileiro de Informação em Ciência e Tecnologia para promoção da informação científica e tecnológica nacional e internacional. Sua primeira versão foi oficialmente lançada em 2023 no XXII Seminário Nacional de Bibliotecas Universitárias (SNBU), em Florianópolis. Em maio de 2025, o Aleia conquistou a certificação internacional *CoreTrustSeal*, a mais adotada para a validação de repositórios digitais confiáveis. A certificação consolida o repositório como um repositório digital de alta confiabilidade para a comunidade científica do Instituto. O processo de certificação foi iniciado em 2023 e demandou o cumprimento de 16 requisitos que abrangem áreas como infraestrutura informacional, gestão de objetos digitais, tecnologia da informação e segurança, alinhando-se aos Princípios FAIR (*Findability, Accessibility, Interoperability e Reusability*). O Aleia tornou-se o segundo repositório de dados da América Latina a obter esse selo, atrás apenas do Arca Dados, da Fiocruz, destacando-se entre os 159 repositórios certificados no mundo (Ibict, 2025).

O Aleia surgiu da necessidade de apoiar e facilitar a adoção do acesso aberto aos dados de pesquisa, fomentar novas pesquisas e promover práticas de Ciência Aberta. Sua criação foi motivada pelo Compromisso 3, do 4º Plano de Ação Nacional que visava estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil (*Open Government Partnership*, 2018).

Sanchez, Vidotti e Vechiato (2017, p. 3) descrevem os repositórios de dados como aqueles que "[...] buscam organizar, estruturar, permitir acesso, disseminar e

preservar todos os dados gerados por meio de pesquisas realizadas em sua maioria por Instituições de Ensino e Pesquisa". De acordo com Sayão e Sales (2015, p. 49), "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".

Sabe-se que os dados de pesquisa possuem um nível de abstração maior que outros objetos digitais como artigos, dissertações, teses e livros, o que pode prejudicar a interpretação, reutilização e citação adequada. Entretanto, os dados, quando reutilizados, podem trazer benefícios como: 1) transparência e prevenção de fraudes; 2) economia de recursos financeiros e humanos; 3) agilização do ciclo científico; 4) reprodutibilidade, dentre outros. Por isso é fundamental o seu armazenamento em repositórios, uma vez que:

Sem uma descrição minuciosa do contexto tecnológico dos arquivos de dados, do contexto no qual os dados foram criados ou coletados, das medidas que foram feitas, dos detalhes espaciais e temporais, dos instrumentos usados, dos parâmetros e unidades e da qualidade dos dados e da sua proveniência, é improvável que os dados possam ser descobertos, interpretados, gerenciados e efetivamente usados e reusados (Sayão; Sales, 2015, p. 20).

Nota-se, portanto, a importância da descrição adequada dos dados, garantindo sua contextualização, o que é feito por meio da representação por metadados em

infraestruturas tecnológicas como os repositórios. A gestão adequada dos dados de pesquisa, de acordo com Sayão e Sales (2016, p. 92), tem como componente central os repositórios de dados que "[...] rapidamente se tornam parte essencial da infraestrutura mundial de pesquisa", o que pode ser ligado a duas demandas essenciais no contexto atual: "[...] as políticas mandatárias das agências financiadoras de pesquisa e a incorporação pelas instituições de pesquisa e pela sociedade em geral dos valores e princípios preconizados pela Ciência Aberta".

O Aleia é um serviço voltado para o depósito de dados que foram coletados e produzidos pela comunidade científica do Ibict durante a pesquisa técnica ou científica, sendo comumente aceitos na comunidade para validar as descobertas e resultados de um estudo. Assim, seu principal objetivo é ser o repositório oficial do Ibict que custodia, publica, dissemina, preserva e compartilha conjuntos de dados da comunidade científica do Ibict constituída por seus servidores, pesquisadores visitantes, bolsistas e alunos do programa de pós-graduação (Ibict, 2024).

O repositório foi desenvolvido ao longo de oito etapas, sendo elas: 1) estudos para escolha do software em que seria desenvolvido (optou-se pelo Dataverse); 2) instalação do software escolhido em máquina de teste; 3) definição dos metadados; 4) definição das temáticas a serem definidas como comunidades; 5) criação e aprovação do design do repositório; 6) desenvolvimento dos conteúdos para página do repositório; 7) inclusão dos conteúdos no repositório; e 8) migração do repositório de ambiente de teste para ambiente de produção.

No que diz respeito à arquitetura informacional, o Aleia foi estruturado da seguinte forma: 1) Sobre; 2) Guia do usuário; 3) Política do repositório; 4) Termos; 5) Perguntas frequentes; 6) Contato; 7) Comunidades Dataverse, sendo elas: Tecnologias da Informação, Ensino e Pesquisa, Informação Científica, Informação para Sustentabilidade, Informação Tecnológica, Popularização da Ciência e Informação para a Sociedade; 8) Conjuntos de Dados; 9) Arquivos.

A arquitetura da informação foi definida pela equipe técnica do Aleia com o objetivo de agrupar assuntos gerais que mais são demandados pela comunidade científica do Ibict. Os formatos disponibilizados no Aleia são planilhas, imagens, vídeos, gráficos, arquivos tabulares, entre outros. Devem estar em formato natos digitais ou digitalizados, completos e prontos para disseminação.

Como mencionado anteriormente, o Aleia foi desenvolvido no Dataverse, software de código aberto direcionado para armazenamento e compartilhamento de dados de pesquisa. O Dataverse foi desenvolvido em 2006 pelo Instituto de Ciências Sociais Quantitativas da Universidade de Harvard em colaboração com profissionais de todo o mundo, por meio do *Dataverse Project* (Dataverse, 2023).

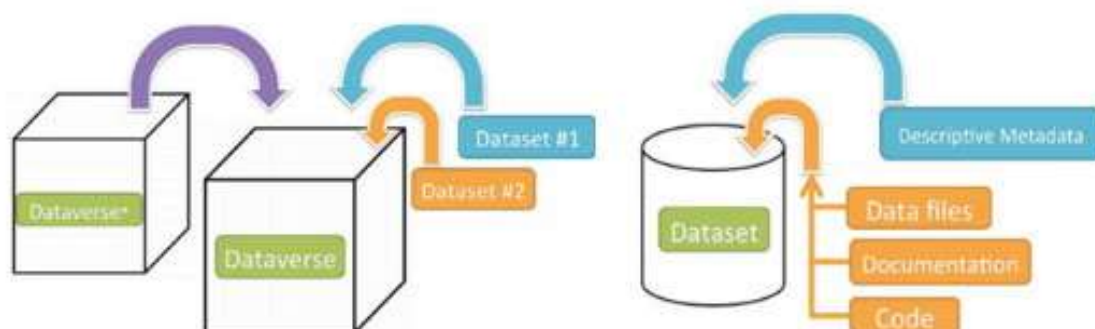
A essência do *Dataverse Project* é automatizar grande parte do trabalho do arquivista profissional, fornecer serviços e distribuir crédito ao criador de dados. Há a união da preservação dos dados e dos créditos ao autor, o que influenciou na escolha do Dataverse para ambos os repositórios de dados gerenciados pelo Ibict: Aleia e Deposita Dados. Além disso, já fala-se sobre como “[...]”

uma plataforma de repositório como o Dataverse pode facilitar muito a criação de dados científicos FAIR” (Silva; Rodrigues, 2021, p. 127). Corroborando com os autores, Boyd (2021, p. 5, tradução nossa) afirma que "A plataforma de software Dataverse oferece suporte a dados FAIR por meio de uma combinação de funcionalidade de infraestrutura padrão e recursos necessários e opcionais voltados para o usuário". Isso inclui, por exemplo, identificadores persistentes, metadados específicos do domínio e suporte para formatos de arquivo abertos.

Os princípios FAIR são um acrônimo para "*Findable*", "*Accessible*", "*Interoperable*" e "*Reusable*", que em português significa "Encontrável", "Acessível", "Interoperável" e "Reutilizável". São diretrizes que buscam enfatizar o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente, além de apoiar sua reutilização por usuários humanos. Com o volume cada vez maior de dados sendo produzidos e processados nas pesquisas, o que demanda o uso de computação intensiva, é essencial se pensar em formas de tornar os dados científicos mais FAIR, justamente pela otimização do processamento automático. De acordo com Sales *et al.* (2020), os princípios FAIR já são reconhecidos mundialmente como elementos-chave para boas práticas em todos os processos de gestão de dados. Logo, a adoção do Dataverse está consoante com essa tendência internacional.

O Aleia hospeda vários arquivos virtuais chamados Comunidades Dataverses. Cada comunidade contém conjuntos de dados chamados datasets, e cada *dataset* contém objetos digitais (arquivos) e metadados.

Figura 1 - Estrutura do Dataverse



Fonte: Crosas (2015).

Cabe destacar, ainda, que o Dataverse investe em uma série de boas práticas que podem ser consultadas, na íntegra, no site oficial (Ibict, 2024). Alguns exemplos de suma importância para os repositórios geridos pelo Ibict são:

- 1) crédito acadêmico: ao depositar, os pesquisadores tornam seus conjuntos de dados mais localizáveis na web. Um dos motivos é a oferta do recurso *widgets* pelo Dataverse, que permite que o pesquisador incorpore os conjuntos em qualquer site para ajudar outros pesquisadores a encontrá-los com mais facilidade;
- 2) citação dos dados: o Projeto Dataverse padroniza a citação de conjuntos de dados para tornar mais fácil para os pesquisadores publicarem seus dados e obterem crédito, bem como reconhecimento por seu trabalho. Quando se cria um conjunto de dados em um repositório Dataverse, a citação é gerada e apresentada automaticamente;
- 3) gestão dos dados: muitos periódicos, editores e agências de financiamento exigem que os

pesquisadores depositem conjuntos de dados em um repositório público. O Projeto Dataverse ajuda os pesquisadores a cumprir esse requisito, apoiando o depósito de conjuntos de dados e tornando-os facilmente detectáveis para outros pesquisadores reutilizarem e verificarem se um estudo pode ser replicado sem ter que, necessariamente, entrar em contato com os autores do estudo.

Nota-se, portanto, que a adoção do Dataverse traz uma série de benefícios para os depositantes, trazendo maior segurança para os pesquisadores que desejam compartilhar seus dados. Com isso em mente, o repositório Aleia permite o armazenamento, a gestão e a preservação dos dados de pesquisa de diferentes comunidades acadêmicas, apoiando e facilitando também a adoção do acesso aberto aos dados de pesquisa, fomentando novas pesquisas e promovendo uma Ciência Aberta, indo de acordo com o Manifesto de Acesso Aberto a Dados de Pesquisa Brasileira para Ciência Cidadã lançado em 2016.

Para acessar e depositar no Aleia, basta acessar o repositório em seu endereço eletrônico: aleia.ibict.br. Ao clicar na aba “Iniciar sessão”, clique no botão “ORCID”.

O usuário já estará logado e pronto para iniciar o depósito dos conjuntos de dados que deseja compartilhar. O depositante deve escolher a comunidade de acordo com o assunto principal do conjunto de dados, lembrando que o povoamento/alimentação do Aleia deverá ser realizado por autoarquivamento, ou seja, caberá ao autor realizar o depósito do seu próprio conjunto de dados, que passará pela revisão da equipe gestora do Aleia.

Figura 2 - Tela inicial



Fonte: (Ibict, 2024).

Os arquivos devem estar em formato natos digitais ou digitalizados, completos e prontos para disseminação. E, sempre que possível, devem ser escolhidos formatos que ajudem a preservar e garantir o acesso a longo prazo aos arquivos depositados no Aleia, com preferência por arquivos em formatos abertos e não-proprietários, o que evita barreiras de software.

CONSIDERAÇÕES FINAIS

Finalmente, almeja-se que o repositório possa atender à comunidade científica do Ibict, desempenhando o papel de uma fonte de informação confiável e segura. O Aleia tem como finalidade tornar os dados de pesquisa acessíveis, dar visibilidade aos dados de pesquisa

elaborados pela comunidade científica Ibict; garantir a reprodutibilidade e reuso dos dados de pesquisa da comunidade científica institucional do Ibict; funcionar como ferramenta para promover a transparência e por consequência a confiança e a solidez das pesquisas desenvolvidas, maximizando o retorno de aplicação de recurso público.

REFERÊNCIAS

BOYD, Ceilyn. Use of Optional Data Curation Features by Users of Harvard Dataverse Repository. **Journal of eScience Librarianship**, Worcester, v. 10, n. 2, p. 1-20, 2021. DOI: 10.7191/jeslib.2021.1191.

CROSAS, Merce. **Dataverse on the MOC**. 2015. Disponível em: <https://pt.slideshare.net/merceecrosas/dataverse-on-the-moc>. Acesso em: 12 jun. 2024. (Apresentação slides).

DATAVERSE. **The Dataverse Project**. 2023. Disponível em: <https://dataverse.org>. Acesso em: 15 jan. 2024.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Aleia**. Disponível em: <https://aleia.ibict.br>. Acesso em: 12 jun. 2024.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Repositório Aleia conquista certificação internacional CoreTrustSeal, que reconhece repositórios digitais confiáveis**. 2025. Disponível em: www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2025/junho/repositorio-aleia-conquista-certificacao-internacional-coretrustseal-que-reconhece-repositorios-digitais-confiaveis. Acesso em: 11 jun. 2025.

OPEN GOVERNMENT PARTNERSHIP. **4º Plano de Ação Nacional em Governo Aberto**. Brasília: Ministério da Transparência e Controladoria-Geral da União, 2018.

SALES, Luana *et al.* GO FAIR Brazil: A Challenge for Brazilian Data Science. **Data Intelligence**, Cambridge, v. 2, n. 1-2, p. 238-245, 2020. DOI: 10.1162/dint_a_00046.

SANCHEZ, Fernanda Alves; VIDOTTI, Silvana Aparecida Borsetti Gregório; VECHIATO, Fernando Luiz. A contribuição da curadoria digital em repositórios digitais. **Revista Informação na Sociedade Contemporânea**, Natal, v. 1, p. 1-17, 2017. DOI: 10.21680/2447-0198.2017v1n0ID12280.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, maio/ago., 2016. DOI: 10.5433/1981-8920.2016v21n2p90.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015.

SILVA, Fabiano Couto Corrêa da; RODRIGUES, Marcello Mundim. Implementação dos princípios FAIR em repositórios de dados científicos: uma análise comparativa das infraestruturas de software do DSpace e Dataverse. *In*: SALES, Luana Farias; VEIGA, Viviane dos Santos de Oliveira; HENNING, P Patrícia; SAYÃO, Luis Fernando (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: Ibict, 2021, p. 123-136.

Como citar este capítulo

SILVA, Tatyane Guedes Martins da; BONETTI, Leticia Guarany. Aleia: Repositório de dados de pesquisa do Ibict. *In*: AMARO, Bianca; CAMPOS, Phillipe de Freitas; BARCELOS, Janinne. (org.). **Infraestruturas de Ciência e de Acesso Aberto no Brasil:** iniciativas do Instituto Brasileiro de Informação em Ciência e Tecnologia. Brasília, DF: Editora Ibict, 2025. Cap. 6, p. 119-130. DOI: 10.22477/9788570132543.cap6