



DSpace e Dataverse como repositórios institucionais de dados de pesquisa: um estudo comparativo

DSpace and Dataverse as institutional repositories of research data: a comparative study

Joel G. Perozo-Vasquez

Doutorando em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC).
E-mail: ufscjoel@gmail.com

Diogo Vieira Cardoso

Mestrando em Gestão de Unidades de Informação pela Universidade do Estado de Santa Catarina (UDESC).
E-mail: diogoinweb@gmail.com

Darli Damian da Silva

Especialista em Educação pela Universidade Federal de Santa Catarina (UFSC).
E-mail: darli1301@gmail.com

Thiago Magela Rodrigues Dias

Doutor em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG).
E-mail: thiagomagela@gmail.com

Divino Ignácio Ribeiro Junior

Doutor em Engenharia e Gestão do Conhecimento pela Universidade Federal de Santa Catarina (UFSC).
E-mail: divinoirj@gmail.com

RESUMO

Contexto/Problemática: A escolha de um sistema de repositório de dados de pesquisa adequado é essencial para a visibilidade, acessibilidade e preservação dos dados científicos. Objetivo: Comparar as plataformas de repositórios institucionais de dados de pesquisa Dataverse e DSpace, fornecendo uma análise detalhada para auxiliar instituições na escolha da plataforma mais adequada. Metodologia: Abordagem qualitativa exploratória baseada em análise documental, utilizando 17 critérios adaptados de estudos anteriores. Resultados: Ambos os softwares oferecem soluções robustas, mas com diferenças significativas, pois enquanto o Dataverse se destaca pela especialização em dados de pesquisa e conformidade com os princípios FAIR; o DSpace é mais versátil, sendo amplamente utilizado para a gestão de diversos tipos de documentos acadêmicos. Discussão: Foram identificados quatro cenários possíveis para a implementação dos repositórios, cada um com vantagens e desafios específicos. Conclusões: A escolha entre Dataverse e DSpace deve considerar fatores como capacidade técnica e orçamentária, volume e tipo de dados a serem gerenciados, e a estratégia institucional de longo prazo; além de compreender as capacidades e limitações de cada software é essencial para garantir a gestão eficiente dos dados científicos.

Palavras-chave: repositório de dados de pesquisa. repositório institucional. acesso aberto a dados de pesquisa. Dataverse. DSpace.

ABSTRACT

Context/Problem: Choosing a suitable research data repository system is crucial for the visibility, accessibility and preservation of scientific data. Objective: To compare the institutional research data repository platforms Dataverse and DSpace, providing a detailed analysis to assist institutions in choosing the most suitable platform. Methodology: An exploratory qualitative approach based on documentary analysis, using 17 criteria adapted from previous studies. Results: Both software packages offer robust solutions, but with significant differences; whilst Dataverse stands out for its specialisation in research data and compliance with FAIR principles, DSpace is more versatile, being

widely used for the management of various types of academic documents. Discussion: Four possible scenarios for the implementation of the repositories were identified, each with specific advantages and challenges. Conclusions: The choice between Dataverse and DSpace should take into account factors such as technical and budgetary capacity, the volume and type of data to be managed, and the institution's long-term strategy; furthermore, understanding the capabilities and limitations of each software is essential to ensure the efficient management of scientific data.

Keywords: research data repository. institutional repository. open access to research data. Dataverse. DSpace.

1 INTRODUÇÃO

A consolidação do movimento de Acesso Aberto (Open Access - OA) nos últimos 30 anos tem promovido a democratização do conhecimento e incentivado o compartilhamento de dados de pesquisa. A disponibilização pública desses dados é fundamental para a Ciência Aberta (Open Science - OS), pois permite que pesquisadores, instituições e o público em geral acessem, reutilizem e reproduzam dados utilizados em investigações anteriores de forma mais acessível. Nesse cenário, os repositórios institucionais de acesso aberto emergem como softwares essenciais para o armazenamento e divulgação desses dados, desempenhando um papel central na infraestrutura da ciência contemporânea (Dias; Dias, 2019; Lucas; Picalho; Caitano, 2021; Sayão; Sales, 2022).

Nesse contexto, emergem iniciativas de estruturação de repositórios institucionais, como sistemas digitais geridos por Instituições de Ensino Superior, Centros de Pesquisa e Desenvolvimento e outras instituições científicas e/ou acadêmicas, para preservar e disseminar digitalmente a produção intelectual de seus membros. Esses sistemas podem se apresentar como uma solução sustentável para a gestão de dados de pesquisa e também refletem o compromisso da instituição com a transparência e o acesso universal ao conhecimento científico (Campêlo; Barreto Neto, 2019).

O avanço desse movimento impulsionou o surgimento de múltiplos repositórios de dados de pesquisa, dotados de características, funcionalidades e diretrizes de acesso singulares. Consequentemente, tal heterogeneidade suscita a demanda por uma avaliação comparativa, visando orientar instituições acadêmicas e pesquisadores na seleção da plataforma mais pertinente.

A seleção institucional do sistema de repositório ideal constitui um complexo desafio, visto que incide diretamente sobre a visibilidade e a acessibilidade dos dados. Ademais, tal decisão afeta a preservação digital e o atendimento aos requisitos das agências de fomento, as quais demandam o compartilhamento em ambientes de acesso

aberto. Sob essa ótica, cada plataforma disponibiliza um arranjo exclusivo de recursos e funcionalidades, variando do suporte a múltiplos formatos de arquivo à implementação dos Princípios FAIR (acrônimo dos termos “Localizável, Acessível, Interoperável e Reutilizável”; *Findable, Accessible, Interoperable, Reusable*, em língua inglesa), que deverão se tornar o padrão internacional para disponibilização digital de dados de pesquisas (Shintaku; Appel; Oliveira, 2021).

Desse modo, repositórios de produtos e dados de pesquisa, tais como DSpace, Dataverse, Zenodo, OpenAIRE, OSF, entre outros, exibem particularidades, com vantagens e limitações inerentes à usabilidade, interoperabilidade, conformidade com padrões e normativas internacionais e suporte a diferentes áreas do conhecimento (Campêlo; Barreto Neto, 2019; Estevão; Arns; Strauhs, 2019; Lucas; Picalho; Caitano, 2021).

A título de ilustração, destaca-se o Re3data, um registro global criado em 2012 que cataloga uma ampla relação de repositórios alinhados a critérios específicos. Essa plataforma garante o armazenamento e o acesso perene a conjuntos de dados para agências de fomento, editoras e Instituições de Ensino Superior, consolidando-se como um recurso fundamental para facilitar a localização do ambiente de armazenamento mais pertinente. (Pavão *et al.*, 2018).

Em 2024, de 2524 softwares registrados nesse diretório em ordem decrescente de frequência, identificaram-se: Dataverse (191); DSpace (146); CKAN (103); MySQL (94); Fedora (48); EPrints (35); Nesstar (16); DigitalCommons (5); eSciDoc (5); dLibra (4); Opus (3); outros (692); e desconhecidos (1182).

O Dataverse (<https://dataverse.org>) é um software de repositório dedicado exclusivamente a dados de pesquisa, concebido pelo Instituto de Ciências Sociais Quantitativas (*The Institute for Quantitative Social Science - IQSS*) da Universidade de Harvard. Em contrapartida, o DSpace (<https://dspace.org>), desenvolvido pela Organização DuraSpace para o Massachusetts Institute of Technology (MIT) e a Hewlett-Packard (HP), atua como uma plataforma de repositório institucional genérico. No entanto, por apresentar atributos análogos aos repositórios de dados científicos, é frequentemente empregado com essa finalidade. Por sua vez, o CKAN (<https://ckan.org>) caracteriza-se apenas como um sistema de indexação de dados de pesquisa, não exercendo a função de repositório.

Diante do atual panorama de gestão de dados de pesquisa, este estudo objetiva apresentar um comparativo entre as principais plataformas de repositórios de acesso

aberto adotadas no Brasil: Dataverse e DSpace. Com isso, busca-se fundamentar o processo decisório relativo à implantação de repositórios institucionais, delineando uma visão geral desses sistemas para orientar gestores na escolha da solução mais adequada às suas demandas.

Para atingir tal objetivo, o estudo baseia-se na análise criteriosa de aspectos adaptados a partir dos estudos de Rodrigues (2004), Martins e Silva (2017), Pavão *et al.* (2018), e Campêlo e Barreto Neto (2019), tais como: Organização e Controle de Conteúdo, Descoberta de Conteúdo, Ferramentas de Relatórios, Recursos Sociais e Notificação, Interoperabilidade, Autenticação, Acessibilidade, Preservação e Curadoria Digital, Escalabilidade, Linguagem de Programação, Suporte a Banco de Dados, Suporte a Serviços e Manutenção, Formato de Arquivo de Suporte, Infraestrutura, Relatórios, Multimídia, e Treinamento e Suporte Técnico.

Como justificativa, a importância deste estudo reside, por um lado, na crescente demanda por repositórios eficientes e fáceis de usar, capazes de lidar com o volume e a diversidade dos dados de pesquisa atuais; e, por outro, com as exigências legais e institucionais de planejamento, julgamento objetivo, razoabilidade, eficiência, eficácia, e economicidade, entre outros princípios e diretrizes da Administração Pública (BRASIL, 1988; 2021).

Compreender, portanto, os atributos e limitações de cada software é essencial para garantir que os dados científicos sejam devidamente geridos, permaneçam acessíveis e sejam preservados a longo prazo. Nessa perspectiva, a análise comparativa proposta neste artigo contribuirá na discussão acadêmica do tema e fornecerá informações práticas que poderão ser aplicadas, com as suas devidas adaptações, por instituições de diferentes portes e contextos.

Além disso, espera-se que o artigo contribua para a melhoria das práticas de gestão de dados de investigação, promovendo a adoção de soluções mais adequadas e eficazes para a comunidade científica. Ao identificar os pontos fortes e fracos daqueles softwares, o estudo também pode contribuir para o desenvolvimento e fortalecimento desses softwares, estimulando, assim, o avanço da Ciência Aberta e um acesso mais democrático e universal ao conhecimento.

As próximas seções deste trabalho estão organizadas da seguinte forma: a Seção 2 trata sobre a fundamentação teórica, a Seção 3 apresenta a Metodologia, descrevendo a

coleta e tratamento de dados. Na seção 4, expõem-se os resultados e a respectiva discussão; e, na última seção 5, apresentam-se as conclusões como considerações finais.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, apresentam-se conceitos centrais referentes a: dados de pesquisa, movimento de acesso aberto, repositórios institucionais e critérios de avaliação utilizados para comparação dos softwares. A seguir, discutem-se cada um desses conceitos, abordando suas definições e relevância no cenário acadêmico e científico, além das implicações para a gestão e compartilhamento de dados de pesquisa.

2.1 DADOS DE PESQUISA E OPEN ACCESS

Dados de pesquisa se referem às informações coletadas, observadas ou geradas durante o processo de investigação científica. Esses dados podem assumir diversas formas, incluindo conjuntos de números, textos, gráficos, imagens, vídeos, algoritmos, entre outros formatos que variam de acordo com o domínio do conhecimento. A gestão eficaz dos dados de investigação inclui não apenas a coleta e o armazenamento, mas também a documentação, a preservação e a sua disponibilização para utilização futura por outros pesquisadores (Estevão; Arns; Strauhs, 2019).

A crescente exigência de transparência e reprodutibilidade na ciência deve levar à adoção de boas práticas na gestão de dados de investigação, incluindo a necessidade de os tornar acessíveis através de repositórios institucionais de acesso aberto. Ademais, o movimento Open Access (OA) surgiu, em meados dos anos 2000, como resposta à barreira imposta pelas assinaturas e licenças restritivas às publicações acadêmicas e científicas, propondo que os resultados da pesquisa, incluindo os dados, sejam livremente acessíveis ao público (Dias; Dias, 2019).

Nesse contexto, os princípios FAIR permitem a materialização, operacionalização e efetivação dos movimentos de Ciência Aberta e acesso aberto, conforme definido sinteticamente por Rocha *et al.* (2018, p. 9):

Para dados serem FAIR, estes devem ser atribuídos a identificadores únicos, persistentes e globais (F). Devem ser descritos por metadados indexáveis e ricos (F), representados em linguagens formais (I), aceitos pela comunidade (R), com atributos relevantes, precisos e úteis ao contexto (R), incluindo metadados de proveniência (R) e usando vocabulários controlados que seguem princípios FAIR (I). Estes dados devem ser recuperáveis pelo seu identificador através de um protocolo de comunicação padronizado, aberto e gratuito (A). Também devem

ser acompanhados de licenças claras e acessíveis (R), e referências qualificadas devem ligar (meta)dados (I).

Os dados de pesquisa, quando disponíveis sob os princípios do acesso aberto, reforçam o compromisso com a transparência científica, além de permitirem uma série de benefícios, tais como: aumento da visibilidade da investigação, promoção da colaboração interdisciplinar e aceleração potencial das descobertas científicas. No entanto, tornar os dados abertos e acessíveis requer plataformas adequadas para a sua gestão e distribuição. É neste contexto que os repositórios institucionais desempenham um papel central, fornecendo a infraestrutura necessária para o armazenamento seguro, curadoria e partilha de dados de investigação (Rocha *et al.*, 2018, 2021; Sayão; Sales, 2015, 2022; Vidotti; Torino; Coneglian, 2021).

2.2 REPOSITÓRIOS INSTITUCIONAIS DE DADOS DE PESQUISA

Repositórios institucionais são plataformas voltadas à gestão integral dos dados de pesquisa, abrangendo sua preservação e visibilidade. Mais do que simples depósitos, tais sistemas desempenham uma função estratégica ao adotar rigorosos critérios de arquivamento, transformando dados brutos em ativos valiosos, acessíveis e prontos para a reutilização pela comunidade científica. (Leite, 2012; Sayão; Sales, 2015, 2022).

Por sua vez, os repositórios de dados científicos visam democratizar o acesso à produção acadêmica e impulsionar o progresso da ciência. Sua evolução converge com o movimento de Acesso Aberto, proporcionando uma via de disseminação mais abrangente que o modelo tradicional. Este último, geralmente focado na publicação das conclusões finais, costuma omitir os dados gerados durante o processo investigativo, lacuna que esses sistemas buscam preencher. (Neves, 2018).

Nessa perspectiva, Silva e Lima (2021, p. 2) apresentam uma definição funcional de repositórios de dados como, a saber:

[...] sistemas de informação com funções gerenciais e técnicas que contribuem com o compartilhamento, o arquivamento, a representação, a disponibilização, a preservação, a descoberta, o acesso, a reutilização e a citação dos dados científicos, além de proporcionar a transparência e a validação dos resultados da pesquisa científica.

Tais repositórios são fundamentais para a preservação digital a longo prazo e para a adesão aos princípios da Ciência Aberta, sendo um facilitador entre a produção de conhecimento e a sua acessibilidade pública. Além disso, muitos repositórios seguem

padrões internacionais que facilitam a interoperabilidade entre plataformas digitais, softwares e aplicativos, permitindo que os dados possam ser descobertos, acessados e reutilizados em diferentes contextos. Destacam-se ainda, nesse contexto, os já citados princípios FAIR que estabelecem os parâmetros necessários para a governança de dados, assegurando que estes atendam aos requisitos de busca, acesso, interoperabilidade e reuso pela comunidade científica. (Lucas; Picalho; Caitano, 2021).

Nessa perspectiva, a diversidade de repositórios institucionais reflete distintas abordagens tecnológicas e operacionais. Sistemas amplamente adotados, como DSpace, Dataverse, Eprints, Zenodo e Dryad, oferecem recursos que variam em termos de usabilidade, curadoria, suporte a metadados e governança. A seleção da solução mais adequada depende das demandas específicas da instituição e de seus pesquisadores, além de condicionantes técnicos e financeiros (Sayão; Sales, 2015; 2022).

2.3 AVALIAÇÃO DE REPOSITÓRIOS INSTITUCIONAIS DE DADOS DE PESQUISA

A comparação entre diferentes repositórios de dados exige a adoção de métricas de avaliação amplamente aceitas. Esses critérios viabilizam uma análise objetiva das funcionalidades e limitações inerentes a cada ambiente digital, fundamentando a escolha da infraestrutura mais adequada.

Rocha *et al.* (2018) analisaram comparativamente o DSpace e o Dataverse segundo os princípios FAIR, os “Princípios de Citação de Dados” (*Data Citation Principles*; <https://force11.org/info/joint-declaration-of-data-citation-principles-final/>) e que atendessem a critérios estabelecidos para “repositórios digitais confiáveis” (RDC) que visam à obtenção de certificações, como *Data Seal Approval*, *Core Trust Seal*, Nestor e ISO 16363:2012.

Com base no Modelo de Referência para Repositórios “*Open Archival Information System*” (OAIS) (*Reference Model for an Open Archival Information System*; <https://public.ccsds.org/pubs/650x0m2.pdf>), determinado pela norma ISO 14721:2012, os autores propuseram 56 critérios para analisar as soluções tecnológicas, classificados de acordo com 7 categorias relacionadas “com critérios para certificação de repositório confiável de Core Trust Seal e com os princípios FAIR” Rocha et al. (2018, p.4).

Para fundamentar a comparação entre os diferentes ambientes de gestão de dados, Rocha et al. (2018, p. 12-13) elencam requisitos essenciais que abrangem desde a

infraestrutura técnica até as políticas institucionais. Os critérios sugeridos pelos autores detalham as capacidades esperadas em cada etapa do ciclo de vida dos dados:

Representação do Ambiente do Repositório: Recursos da solução tecnológica para representar o ambiente do repositório e seu modelo (políticas) de funcionamento;

Representação dos Conjuntos de Dados: Prover meios para representar um conjunto de dados (pacote);

Descrição e Documentação dos Conjuntos de Dados: Recursos para descrever e documentar os conjuntos de dados;

Produção dos Conjuntos de Dados: Recursos para a submissão dos dados ao repositório;

Armazenamento a Longo Prazo e Planejamento da Preservação: Recursos para o armazenamento seguro e a longo prazo da informação;

Acesso e Uso dos Conjuntos de Dados: Recursos para descoberta dos dados, restringir acesso a dados, entregar dados ao consumidor, prover acesso aos dados;

Uso, Desenvolvimento e Manutenção do Software: características do software quanto ao seu uso, desenvolvimento e manutenção.

Sob a influência da taxonomia proposta por Rocha *et al.* (2018), que oferece uma estrutura exaustiva de 56 requisitos, o presente estudo realizou uma síntese analítica para adequar tais parâmetros aos objetivos desta investigação. Assim, foram consolidados 17 critérios de análise, detalhados no Quadro 1, que priorizam as dimensões de usabilidade e governança mais críticas para o cenário atual. Esta seleção visa garantir a aplicabilidade do comparativo sem comprometer o rigor técnico. A seguir, detalham-se os procedimentos metodológicos que nortearam a execução da referida análise.

3 METODOLOGIA

Nesta seção, detalham-se o tipo de pesquisa, os critérios de seleção dos sistemas de repositórios, os métodos de coleta de dados e os critérios utilizados para a comparação. O presente estudo utiliza uma abordagem qualitativa de caráter exploratório, com base em uma análise documental.

A análise exploratória é apropriada para estudos que visam investigar um campo relativamente novo ou em constante evolução, como o dos repositórios institucionais Open Access (OA) e, especificamente, nos repositórios para armazenamento, acesso permanente e gestão de conjuntos de dados de pesquisa. A pesquisa qualitativa permite uma compreensão profunda das particularidades e funcionalidades de cada sistema, explorando suas características técnicas e operacionais. A análise documental envolve o exame de materiais como guias técnicos, manuais de usuário, publicações acadêmicas, relatórios institucionais e outras documentações relevantes que descrevem tais sistemas.

A seleção dos sistemas para o comparativo se baseou em critérios como: popularidade, relevância no cenário acadêmico brasileiro e sua adoção por universidades e centros de pesquisa (Rocha *et al.*, 2018). Assim, os sistemas de repositórios selecionados para este estudo incluem: DSpace, versão 8.0, um dos repositórios institucionais de produção acadêmica mais amplamente utilizados em todo o mundo, e o Dataverse, versão 6.4, com enfoque principal na gestão de dados de pesquisa.

A coleta de dados fundamentou-se na análise de documentos oficiais e registros disponíveis nos domínios digitais de cada repositório, abrangendo: a) documentação técnica e guias de usuário; b) relatórios de implementação publicados por instituições de pesquisa; e c) estudos de caso e artigos científicos sobre o impacto dessas ferramentas. Esse levantamento permitiu consolidar informações sobre as funcionalidades, políticas de preservação, licenciamento e suporte a padrões de metadados de cada sistema analisado.

3.1 CATEGORIAS DE AVALIAÇÃO E PROCEDIMENTOS DE COMPARAÇÃO

Como parte integrante da metodologia de pesquisa, descrevem-se os atributos e requisitos técnicos inseridos em cada uma das 17 categorias analíticas. Tais critérios foram sistematizados a partir da adaptação dos estudos de Rodrigues (2004), Martins e Silva (2017), Pavão *et al.* (2018), e Campêlo e Barreto Neto (2019). Essa estrutura fundamenta a avaliação técnica de cada sistema, conforme o detalhamento apresentado no Quadro 1.

Quadro 1 - Critérios de análise comparativa de plataformas de repositórios

No.	Critério de análise	Exemplificação de detalhamento de especificações que possam atender aos critérios
1	Organização e Controle de Conteúdo	Permitir a criação e gestão de coleções hierárquicas para organização lógica de datasets. Oferecer suporte à produção de conteúdo com controle granular de acesso por usuários e grupos. Facilitar a estruturação de arquivos dentro de datasets com metadados customizáveis.
2	Descoberta de Conteúdo	Implementar ferramentas avançadas de busca com suporte a filtros, como tipo de arquivo, nível de acesso e palavras-chave. Habilitar busca dentro de metadados e variáveis em arquivos tabulares. Oferecer funcionalidades para salvar e compartilhar resultados de buscas.
3	Ferramentas de Relatórios	Prover relatórios automatizados com métricas de uso, estatísticas de acesso e visualizações personalizáveis. Integrar ferramentas externas para análise e geração de relatórios com uso de APIs. Permitir fluxos de edição e publicação, incluindo revisão por pares.
4	Recursos Sociais e	Integrar funcionalidades para compartilhamento de datasets em redes sociais.

No.	Critério de análise	Exemplificação de detalhamento de especificações que possam atender aos critérios
	Notificação	Implementar notificações configuráveis para ações como atualizações de conteúdo e compartilhamento de arquivos.
5	Interoperabilidade	Garantir suporte a protocolos de interoperabilidade, como OAI-PMH, para colheita e compartilhamento de metadados. Disponibilizar APIs REST para integração com sistemas externos e scripts customizados. Oferecer compatibilidade com identificadores persistentes como DOI e Handle.
6	Autenticação	Suportar métodos de autenticação variados, como Shibboleth, OAuth2 e logins institucionais. Fornecer controle de autenticação e autorização para diferentes níveis de acesso.
7	Acessibilidade	Seguir diretrizes de acessibilidade da web (WCAG), garantindo compatibilidade com leitores de tela e navegação por teclado. Proporcionar interface responsiva e acessível para usuários com diferentes necessidades.
8	Preservação e Curadoria Digital	Implementar versionamento de arquivos e datasets para rastreabilidade histórica. Garantir integridade de dados com ferramentas de verificação e preservação a longo prazo. Suportar padrões de metadados que facilitem a curadoria digital.
9	Escalabilidade	Projetar o sistema para lidar com aumento de usuários e volume de dados sem perda de desempenho. Oferecer suporte a arquiteturas distribuídas para ambientes com alta demanda.
10	Linguagem de Programação	Definir a linguagem de desenvolvimento (ex.: Java) que suporte frameworks robustos e documentação ampla.
11	Suporte a Banco de Dados	Oferecer compatibilidade com sistemas de banco de dados confiáveis, como PostgreSQL.
12	Suporte a Serviços e Manutenção	Disponibilizar documentação técnica, FAQs e fóruns de discussão para suporte contínuo. Incluir sistemas de rastreamento de bugs e espaços para solicitação de novas funcionalidades. Prover suporte técnico via help desk ou canais dedicados.
13	Formato de Arquivo de Suporte	Permitir upload, armazenamento e pré-visualização de arquivos em diversos formatos, como texto, imagens, vídeos e dados tabulares.
14	Infraestrutura	Fornecer condições para instalação local ou hospedagem em nuvem. Garantir suporte técnico detalhado para configuração e manutenção do sistema.
15	Relatórios	Prover relatórios gráficos e indicadores automatizados para monitoramento e análise do uso do repositório.
16	Multimídia	Suportar diferentes tipos de mídias com visualizadores integrados para arquivos comuns, como vídeos e imagens.
17	Treinamento e Suporte Técnico	Disponibilizar recursos educacionais, como guias, tutoriais e webinars, para capacitação da equipe. Fornecer suporte técnico contínuo para usuários e administradores.

Fonte: Elaborado pelos autores (2025)

A análise comparativa fundamenta-se na avaliação detalhada de cada repositório à luz dos parâmetros estabelecidos. Para tanto, utilizam-se quadros comparativos que evidenciam as convergências e distinções entre os sistemas, proporcionando uma visão panorâmica de suas potencialidades e restrições. Tal abordagem viabiliza um exame

objetivo e estruturado, o que auxilia na identificação da solução tecnológica mais adequada às demandas de diferentes perfis de usuários e instituições.

4 RESULTADOS

Nesta seção, apresentam-se os resultados obtidos, os quais compreendem o quadro comparativo das variáveis investigadas, o exame pormenorizado de cada parâmetro e a respectiva discussão fundamentada nos achados da pesquisa.

4.1 COMPARATIVO DE CARACTERÍSTICAS TÉCNICAS

A partir do levantamento das informações disponibilizadas nos domínios oficiais de ambos os sistemas (Dataverse, 2024; DSpace, 2024), e fundamentando-se nos 17 parâmetros propostos, elaborou-se a síntese exposta no Quadro 2.

Quadro 2 - Comparativo técnico entre as soluções Dataverse e DSpace

No.	Critério de análise	Dataverse	DSpace
1	Organização e Controle de Conteúdo	Hierarquia de datasets e coleções; controle de acesso granular por usuário/grupo.	Comunidades e coleções hierárquicas; controle de acesso por permissões configuráveis.
2	Descoberta de Conteúdo	Apache-Solr	Apache-Solr
3	Ferramentas de Relatórios	Dashboard de métricas de uso embutidas; integração com APIs para relatórios customizados.	Estatísticas detalhadas de download e acesso por item e coleção; suporte a relatórios customizados.
4	Recursos Sociais e Notificação	Compartilhamento por link; notificações de atualização, compartilhamento de dados nas redes sociais.	Notificações configuráveis (NOAR); integração com redes sociais por meio de plugins.
5	Interoperabilidade	APIs RESTful; OAI-PMH; DOI; suporte nativo a padrões como DataCite e DDI.	APIs RESTful: DataCite; OAI-PMH; OAI-ORE; DOI/Handle. Compatibilidade com METS e MODS.
6	Autenticação	OAuth2, Shibboleth, OpenID Connect (OIDC), autenticação local.	Shibboleth, LDAP, SSO, Open ID Connect (OIDC), ORCID Authentication, autenticação baseada em IP e outros métodos.
7	Acessibilidade	WCAG 2.1 AA; interface responsiva.	WCAG 2.1 AA; suporte a leitores de tela e navegação por teclado.
8	Preservação e Curadoria Digital	Versionamento integrado; verificação de integridade com checksum; suporte a metadados para preservação (Dublin Core, DataCite).	Preservação via METS e PREMIS; suporte a versionamento; verificação de integridade com checksum.
9	Escalabilidade	Suporte a arquiteturas distribuídas e containerização.	Suporte a alta demanda e grandes repositórios; configurações para escalabilidade horizontal.
10	Linguagem de Programação	Java	Java
11	Suporte a Banco de Dados	PostgreSQL	PostgreSQL

No.	Critério de análise	Dataverse	DSpace
12	Suporte a Serviços e Manutenção	Documentação oficial, fóruns ativos; rastreamento de bugs via GitHub.	Documentação completa; suporte da comunidade LYRISIS; rastreamento de bugs no JIRA.
13	Formato de Arquivo de Suporte	Dados tabulares, texto, imagens, áudio e vídeo; pré-visualização de formatos estatísticos.	Formatos variados (texto, imagens, vídeos, áudio) - Bitstream; extensível com plugins para novos formatos.
14	Infraestrutura	Instalação local ou hospedagem em nuvem; suporte a clusters.	Instalação local ou hospedagem em nuvem; flexível para diferentes configurações.
15	Relatórios	Relatórios embutidos e exportáveis; integração com ferramentas externas via APIs Reporting Tools e Common Queries.	Estatísticas detalhadas com APIs IRUS, Google Analytics Statistics e SOLR Statistics. Extensibilidade via plugins
16	Multimídia	Pré-visualização integrada para arquivos: Multimídia, tabulares, textos, códigos (R, HTML) entre outros.	Visualizadores integrados para tipos comuns de mídia; extensível com plugins.
17	Treinamento e Suporte Técnico	Tutoriais, webinars, e fóruns online.	Manuais, tutoriais e suporte por meio de LYRISIS.

Fonte: Elaborado pelos autores (2024)

Na sequência, procede-se ao detalhamento de cada parâmetro que compõe a análise comparativa, fundamentando as distinções técnicas observadas.

4.1.1 Organização e Controle de Conteúdo

O Dataverse permite a criação de coleções organizadas de dados que podem ser configuradas com permissões específicas para diferentes papéis, como administradores, colaboradores e visualizadores. O controle de acesso é gerenciado por um sistema robusto de permissões, que pode ser configurado para usuários individuais ou grupos institucionais. Isso inclui a possibilidade de herdar permissões de coleções "parent" (configuração *Inherit Parent Role Assignments*) e definir requisitos de acesso detalhados para upload e gerenciamento de dados.

O DSpace 8.0, por sua vez, suporta a organização de conteúdos por coleções, comunidades e subcomunidades, permitindo controles avançados de acesso e gerenciamento de permissões. É possível configurar embargos e restrições específicas para dados sensíveis. A interface administrativa fornece ferramentas para criar e gerenciar essa hierarquia.

4.1.2 Descoberta de Conteúdo

O Dataverse oferece ferramentas avançadas de busca, baseadas em Apache-Solr, que incluem a capacidade de realizar buscas por metadados (tanto de citação quanto

específicos de domínio) e até mesmo por variáveis em arquivos tabulares. Ele também suporta busca geoespacial, com parâmetros para delimitação por latitude e longitude. Recursos de facetas permitem filtrar resultados por atributos como tipo de arquivo e nível de acesso.

O DSpace 8.0 também oferece funcionalidades avançadas de busca por meio do Apache-Solr, permitindo filtros por metadados, facetas e operadores booleanos. Isso facilita a descoberta de conteúdos baseados em critérios específicos definidos pelos usuários.

4.1.3 Ferramentas de Relatórios

As ferramentas de relatórios incluem a visualização de métricas agregadas e detalhadas das instalações do Dataverse, por meio de integrações com scripts em Python e R-Markdown. Por exemplo, o "dataverse-metrics" gera relatórios automáticos e gráficos personalizáveis sobre uso e desempenho. Consultas SQL pré-definidas estão disponíveis para extrair informações detalhadas dos bancos de dados. Disponibiliza integração com APIs externas e suporta um fluxo editorial que inclui revisão por pares antes da publicação final dos conjuntos de dados (*datasets*).

No DSpace 8.0, a revisão de metadados e os fluxos de publicação admitem logs detalhados de submissão e instâncias de aprovação, permitindo que os revisores monitorem alterações e acessem históricos em tempo real. O sistema disponibiliza, via painel administrativo, relatórios minuciosos sobre estatísticas de uso, como métricas de download e acessos, além de prever a integração com soluções externas de análise por meio de suas APIs.

4.1.4 Recursos Sociais e Notificação

No âmbito do Dataverse, os usuários dispõem de recursos para o compartilhamento de dados em mídias sociais e alertas sobre modificações nos registros. A infraestrutura suporta o uso de widgets para a incorporação de buscas e coleções em domínios digitais de terceiros, ampliando a visibilidade da produção em ambientes externos.

Por seu turno, o DSpace 8.0 pode ser configurado para enviar notificações automáticas sobre submissões, atualizações ou interações, com suporte a e-mails personalizados. Integração com plataformas sociais pode ser feita por meio de APIs

externas, com funcionalidades que permitem o compartilhamento fácil em redes sociais diretamente da página do item.

4.1.5 Interoperabilidade

O Dataverse suporta protocolos como OAI-PMH visando à coleta de metadados e integrações com ferramentas externas, como APIs RESTful para customização e automação. A integração com identificadores persistentes, como DOI e Handle, garante a interoperabilidade dos dados.

Nessa perspectiva, o DSpace 8.0 oferece suporte a protocolos como OAI-PMH, OAI-ORE, SWORDv2 e REST API, permitindo integração com outros sistemas, repositórios e serviços de indexação científica. A integração com Datacite é feita somente através de API. Adicionalmente, integra-se com sistemas que gerenciam identificadores persistentes como DOI ou Handle para garantir citações duradouras.

4.1.6 Autenticação

O Dataverse admite protocolos de autenticação via Shibboleth, OAuth2 e LDAP, o que viabiliza o acesso por meio de credenciais institucionais ou contas locais. Adicionalmente, a integração com SAML permite o uso de *Single Sign-On* (SSO), conferindo maior flexibilidade à gestão de identidades.

De forma análoga, o DSpace 8.0 suporta múltiplos métodos de autenticação, como LDAP, OAuth2, Shibboleth e CAS. Sua configuração modular permite a validação de usuários em ambientes acadêmicos, onde as permissões são vinculadas a papéis institucionais definidos pela administração do repositório.

4.1.7 Acessibilidade

O Dataverse segue as diretrizes WCAG (2.1 AA), garantindo acessibilidade para usuários com deficiências visuais ou motoras, incluindo suporte a leitores de tela. A interface é projetada para ser responsiva, com a finalidade de assegurar uma experiência adequada em dispositivos móveis e desktop.

A interface do DSpace 8.0 prioriza a inclusão ao alinhar sua arquitetura aos padrões WCAG 2.1 AA, proporcionando uma navegação compatível com tecnologias assistivas, como leitores de tela. A interface viabiliza a interação integral via teclado, o que amplia a

autonomia de uso para indivíduos com limitações motoras e assegura a conformidade com padrões globais de acessibilidade digital.

4.1.8 Preservação e Curadoria Digital

O Dataverse implementa versionamento automático para cada upload ou modificação em datasets, permitindo rastreamento histórico completo das alterações realizadas nos dados. Suporte para formatos abertos e metadados completos (Dublin Core) garante a longevidade dos dados, com backups regulares e funcionalidades para recuperação.

No DSpace 8.0, cada item pode ter múltiplas versões armazenadas, permitindo acesso histórico às versões anteriores dos dados publicados. A preservação é garantida pelo uso de bancos de dados robustos e armazenamento seguro no sistema de arquivos. Ferramentas para verificação de integridade de arquivos e gestão de formatos também estão disponíveis, além do suporte extensivo aos padrões de metadados Dublin Core e outros formatos que facilitam a curadoria digital.

4.1.9 Escalabilidade

Projetado para atender a múltiplos usuários e demandas crescentes, o Dataverse pode ser dimensionado horizontalmente com suporte a várias instâncias em nuvem (ex.: AWS) ou localmente, sem perda perceptível na performance do sistema devido à sua arquitetura modular baseada em microsserviços. Configurações específicas permitem ajustar o desempenho conforme o uso cresce.

Projetado para gerenciar grandes volumes de dados, o DSpace 8.0 suporta altas demandas mediante configurações avançadas de cache e balanceamento de carga. O sistema permite a escalabilidade horizontal por meio da adição de nós, mantendo a integridade do desempenho sob carga intensa. Adicionalmente, a infraestrutura admite o uso de volumes de armazenamento em nuvem (como Amazon S3), proporcionando elasticidade ao repositório. Essa arquitetura garante versatilidade operacional, suportando tanto implementações locais quanto em ambientes de nuvem (*cloud computing*).

4.1.10 Linguagem de Programação

No que tange à arquitetura de software, o Dataverse é edificado em linguagem Java, utilizando o *framework* GlassFish e bibliotecas associadas. Por sua vez, o *backend* do DSpace 8.0 fundamenta-se em Java (*Spring Boot*), enquanto o *frontend* adota Angular com TypeScript. Em ambos os sistemas, a escolha por essas tecnologias confere robustez à arquitetura e assegura a sustentabilidade da manutenção, beneficiando-se da maturidade e da vasta documentação técnica provida pela comunidade global de desenvolvedores.

4.1.11 Suporte a Banco de Dados

O Dataverse demonstra compatibilidade com PostgreSQL (recomendado) e MySQL, entre outros, utilizando-os para a gestão de metadados e ativos informacionais sob protocolos de transação ACID, o que assegura a integridade dos dados. Essa arquitetura permite a execução de consultas SQL customizadas para extrações e análises complexas.

De forma análoga, o DSpace 8.0 adota o PostgreSQL para a custódia de metadados e relações entre objetos. O sistema oferece alta flexibilidade de configuração via “*dspace.cfg*”, permitindo o ajuste fino de conexões, além de possibilitar a personalização de fluxos de submissão e validação de metadados por meio do arquivo “*input-forms.xml*”.

4.1.12 Suporte a Serviços e Manutenção

O Dataverse dispõe de um repositório documental abrangente, que abrange desde protocolos de instalação e diretrizes de APIs até a customização de metadados e resolução de problemas (*troubleshooting*). O suporte é complementado por comunidades ativas e listas de discussão técnica, utilizando o GitHub para o rastreamento de falhas e o gerenciamento de demandas evolutivas.

De modo similar, o DSpace 8.0 se sustenta em uma documentação extensa e em uma rede de colaboração global. O sistema utiliza o JIRA para a gestão rigorosa de solicitações, garantindo um acompanhamento eficaz na correção de erros e na implementação de novas funcionalidades sugeridas pela comunidade.

4.1.13 Formato de Arquivo de Suporte

Orientado pelos requisitos de interoperabilidade, o Dataverse suporta um extenso repertório de formatos de arquivos digitais, focando na sustentabilidade dos formatos tabulares para análise estatística (Stata, SPSS, SAS). A solução tecnológica permite o gerenciamento granular de permissões, possibilitando o bloqueio de extensões

indesejadas e a configuração de limites de armazenamento via arquivos JSON ou interface administrativa. Cada depósito no ambiente é validado por algoritmos de *checksum* e associado a metadados técnicos, como dimensões do arquivo e pré-visualizações automatizadas. Tal arquitetura garante que, independentemente da natureza do arquivo (documental, comprimido ou audiovisual), a integridade e a rastreabilidade do dado sejam preservadas.

Na mesma filosofia de interoperabilidade, o DSpace 8.0 também é projetado para suportar uma vasta gama de formatos de arquivo digitais, sem uma lista fixa ou exaustiva, pois utiliza um sistema de identificação *MIME types*, o que permite o reconhecimento automatizado de milhares de extensões. Entre as tipologias suportadas, destacam-se documentos de texto (PDF, DOC, DOCX, RTF, ODT, HTML, TXT); arquivos de imagens (JPEG, PNG, GIF, TIFF, BMP); recursos multimídia (MP3, WAV, MPEG, AVI, MP4); além de conjuntos de dados estruturados e comprimidos (JSON, XML, XLS, XLSX, CSV, ZIP, TAR.GZ). Esta flexibilidade se estende ainda a arquivos especializados, como códigos-fonte, arquivos CAD, e documentos em EPUB, LaTeX, PostScript, entre outros. Em cenários de exceção, a infraestrutura possibilita o registro manual de formatos personalizados. Adicionalmente, há visualizadores integrados disponíveis através do uso do ``dspace-api`` permitindo prévias rápidas.

4.1.14 Infraestrutura

O Dataverse admite implementações tanto em servidores locais quanto em infraestruturas de nuvem. Seus requisitos técnicos exigem o *Java Runtime Environment* (JRE), e o sistema oferece suporte à containerização via Docker, o que simplifica os processos de instalação e a manutenção evolutiva.

De igual modo, o DSpace 8.0 apresenta versatilidade para operação em ambientes locais ou remotos, sendo compatível com múltiplos sistemas operacionais. Para sua execução, recomenda-se a utilização do Apache Tomcat como servidor de aplicações (*Servlet Container*), responsável pelo processamento da lógica em Java e pela entrega dos serviços do repositório, com uma documentação técnica que provê diretrizes detalhadas para a governança da infraestrutura.

4.1.15 Relatórios

No Dataverse, relatórios automatizados incluem gráficos de uso e estatísticas detalhadas, acessíveis diretamente ou por meio de ferramentas externas, como scripts de visualização de dados.

No DSpace 8.0, relatórios incluem métricas de uso e exportação de dados sobre atividades no repositório. Fornece relatórios detalhados sobre uso através do painel administrativo, além da integração possível com Google Analytics. É possível a extração de relatórios utilizando filtros por metadados.

4.1.16 Multimídia

No Dataverse, pré-visualizadores podem ser configurados para suportar diferentes tipos de mídia, como imagens (formatos comuns como PNG/JPEG) e vídeos, diretamente na interface do usuário, podendo ser incorporados usando links externos.

O DSpace 8.0 oferece suporte a múltiplas tipologias de mídia, incluindo recursos de áudio, vídeo e objetos tridimensionais, viabilizando a visualização integrada via interface Angular sem a obrigatoriedade de download prévio. A infraestrutura utiliza reprodutores nativos em HTML5 para formatos sonoros (MP3, WAV, OGG) e audiovisuais (MP4, WebM), processando a reprodução via streaming de bitstreams. Documentos em PDF e arquivos de imagem contam com visualizadores embutidos, enquanto modelos 3D (GLTF, OBJ, STL) são renderizados por meio de tecnologias WebGL, como o Three.js. Essa funcionalidade de exibição dinâmica, controlada diretamente no navegador, amplia a interatividade e a acessibilidade dos ativos informacionais custodiados no sistema.

4.1.17 Treinamento e Suporte Técnico

No Dataverse, guias detalhados e workshops da comunidade ajudam na capacitação de equipes técnicas e usuários; webinars regulares são oferecidos pela equipe desenvolvedora. A documentação inclui exemplos práticos para usuários e administradores.

No DSpace 8.0, por sua vez, o treinamento da equipe é apoiado por manuais, guias online e fóruns interativos, com tutoriais online disponíveis junto à documentação técnica e webinars que são organizados periodicamente pela comunidade.

O suporte aos ambientes digitais Dataverse e DSpace é reforçado, no Brasil, pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), por meio da oferta de materiais instrucionais e técnicos. Para o DSpace, em particular, o Instituto mantém fórum

de discussão (<https://forum.ibict.br/c/dspace/16>) que integra a comunidade de usuários e desenvolvedores.

4.2 DISCUSSÃO

A análise comparativa entre o Dataverse e o DSpace revela que ambas as soluções constituem infraestruturas robustas para a implementação de repositórios institucionais de dados de pesquisa. Os resultados demonstram que ambos os sistemas atendem integralmente aos 17 requisitos parametrizados nesta investigação. Todavia, recomenda-se a realização de ensaios experimentais em ambos os ambientes antes da adoção definitiva, assegurando a aderência às especificidades operacionais de cada instituição.

No contexto dessa análise e para efeitos de discussão, identificam-se alguns cenários possíveis, orientativos para a tomada de decisão por parte dos gestores das instituições:

- 1) Para instituições que já possuem o DSpace como repositório acadêmico e pretendem implantar um repositório de dados, há duas opções:
 - a) adaptação do DSpace para servir como repositório de dados de pesquisa;
 - b) instalar, em paralelo ao DSpace (ou quaisquer outro sistema de repositório acadêmico), uma implantação do Dataverse exclusivamente para gestão de dados de pesquisa.
- 2) Para instituições que não possuem nem repositório acadêmico (como o DSpace), nem um repositório de dados (como o Dataverse), pode-se optar por uma das alternativas:
 - a) Implementar apenas o DSpace para ambos os propósitos;
 - b) Implementar o DSpace e o Dataverse simultaneamente.

No **cenário 1.a. (Adaptação do DSpace para servir como repositório de dados de pesquisa)**, as principais vantagens identificadas podem ser: a **integração de recursos**, pois a adaptação mantém todos os objetos digitais (artigos, dissertações, dados de pesquisa etc.) em um único sistema, simplificando o gerenciamento institucional; a **redução de custos**, já que não há necessidade de instalar e manter um software adicional, reduzindo esforços financeiros e técnicos; e o **suporte existente**, porque, em teoria, a equipe técnica já familiarizada com o DSpace pode aproveitar a sua experiência na personalização e operação do software.

Além das vantagens, nesse caso (1.a.) os principais desafios a enfrentar implicam em: **adaptações e customizações significativas necessárias** já que o DSpace, por padrão, não oferece suporte nativo tão robusto quanto o Dataverse para dados de pesquisa (por exemplo, metadados específicos como DDI e funcionalidades como visualização de dados tabulares); **disponibilidade de capacidade técnica**, pois a adaptação pode exigir expertise para implementar novas funcionalidades, especialmente relacionadas à curadoria e interoperabilidade exigidas por repositórios de dados de pesquisa; e **possíveis alterações no desempenho do sistema** devido ao aumento do volume de dados tabulares e multimídia que pode exigir otimizações adicionais do sistema após as alterações.

Sopesando esses fatores, as instituições podem optar por esta abordagem quando tiverem restrições orçamentárias, aquelas organizações que preferem centralizar todo o conteúdo acadêmico em um único sistema, e casos em que o repositório de dados não será usado em larga escala ou terá requisitos menos rigorosos.

No **cenário 1.b. (Instalar o Dataverse em paralelo ao DSpace)**, identificam-se como vantagens a destacar: a **especialização funcional**, pois o Dataverse é projetado especificamente para dados de pesquisa, oferecendo suporte nativo para metadados avançados (DataCite, DDI), integração com APIs RESTful e visualizações de arquivos tabulares, entre outras características singulares; o **cumprimento de diretrizes internacionais**, já que o Dataverse garante conformidade com padrões exigidos por iniciativas como FAIR e OpenAIRE; e a **interoperabilidade** entre o Dataverse e o DSpace, podendo ser integrados para sinergia no gerenciamento de dados e publicações.

Como desafios desta opção, ressalta-se que a necessidade de **gestão de múltiplos sistemas** pode aumentar a complexidade operacional e, frequentemente, os **custos adicionais** decorrentes da necessidade de recursos financeiros e técnicos para instalação, treinamento e manutenção de mais um sistema, neste caso o Dataverse.

Esta estratégia é recomendada a instituições dotadas de recursos financeiros e técnicos capazes de sustentar a operação de dois sistemas simultâneos. É indicada, sobretudo, a organizações que buscam repositórios altamente especializados para atender demandas de áreas como biociências, saúde e geociências, bem como a cenários que exijam conformidade estrita com padrões internacionais, requisito frequentemente mandatório para o acesso a fomento e financiamentos de agências de pesquisa nacionais e estrangeiras.

No caso das Instituições que não possuem nem repositório acadêmico (DSpace), nem repositório de dados (Dataverse), no **cenário 2.a (Implementar apenas o DSpace para ambos os propósitos)**, as vantagens apontadas são, dentre outras: uma **solução única**, quando o DSpace pode servir como um ambiente digital único para artigos, teses, dissertações e dados de pesquisa; e fácil acesso a uma **comunidade ativa**, já que a base de usuários do DSpace fornece suporte colaborativo e customizações baseadas em melhores práticas. Em contraponto, entre os desafios se listam: as **limitações no suporte a dados de pesquisa**, pois embora seja flexível, o DSpace não possui todas as funcionalidades nativas do Dataverse para gestão de datasets; e a **personalização obrigatória** para atender a requisitos rigorosos de interoperabilidade e curadoria de dados.

Essa abordagem é mais adequada para instituições com limitações financeiras que desejam começar com uma solução única e expandir futuramente; e para organizações que priorizam uma solução integrada para todo o conteúdo acadêmico.

No **cenário 2.b (Implementar o DSpace e o Dataverse simultaneamente)**, como vantagens destacadas, pontuam-se: a **abordagem especializada** em que cada sistema será usado para sua função ideal, garantindo máxima eficiência e conformidade com padrões internacionais; a **escalabilidade** que permite maior flexibilidade para expansão futura, com sistemas independentes que podem evoluir separadamente; e **atendimento pleno aos requisitos**, pois o Dataverse atenderá diretrizes específicas para dados de pesquisa, enquanto o DSpace gerenciará documentos acadêmicos.

No bojo dos desafios para esta abordagem, identificam-se: **maior complexidade operacional**, pois gerenciar dois sistemas simultaneamente pode demandar mais treinamento e recursos técnicos; e **investimento inicial elevado**, já que a instalação e manutenção de dois sistemas aumentam custos de curto prazo.

É sugerido optar por essa abordagem no caso de instituições que buscam um repositório robusto e aderente às melhores práticas internacionais para dados e publicações; e organizações com objetivos estratégicos claros de apoiar intensamente tanto a Ciência Aberta quanto a curadoria acadêmica.

Portanto, a escolha entre os cenários propostos depende principalmente de diversos fatores, tais como: a **capacidade técnica e orçamentária da instituição**, já que implementar ambos os sistemas pode ser a melhor opção para atender completamente às diretrizes e padrões, mas exige maior investimento; o **volume e tipo de dados a serem**

gerenciados – se o foco está em dados tabulares ou com requisitos de metadados específicos, o Dataverse oferece mais vantagens; e a **estratégia institucional de longo prazo**, já que as instituições que desejam crescer gradualmente podem começar com o DSpace adaptado e, futuramente, considerar a implantação do Dataverse.

5 CONCLUSÕES

A escolha de um repositório institucional de dados de pesquisa é crucial para a visibilidade, acessibilidade e preservação dos dados científicos. Este estudo comparativo entre Dataverse e DSpace permitiu fornecer uma análise detalhada visando auxiliar instituições na decisão de qual software adotar.

A análise criteriosa de aspectos como organização e controle de conteúdo, descoberta de conteúdo, ferramentas de relatórios, recursos sociais e notificação, interoperabilidade, autenticação, acessibilidade, preservação e curadoria digital, escalabilidade, linguagem de programação, suporte a banco de dados, suporte a serviços e manutenção, formato de arquivo de suporte, infraestrutura, relatórios, multimídia, e treinamento e suporte técnico forneceu uma visão abrangente e revelou insights valiosos sobre as capacidades e limitações de cada sistema.

Ambos os softwares oferecem soluções robustas para a gestão de dados de pesquisa, mas apresentam diferenças significativas em termos de funcionalidades, usabilidade e conformidade com padrões internacionais, a saber:

- O Dataverse se destaca por sua especialização em dados de pesquisa, oferecendo suporte nativo para metadados avançados, como DataCite e DDI, e funcionalidades específicas para visualização de dados tabulares. Sua conformidade com os princípios FAIR e a integração com APIs RESTful garantem uma interoperabilidade eficiente e um alto nível de conformidade com as diretrizes internacionais.
- Por outro lado, o DSpace é um software mais versátil, amplamente utilizado para a gestão de diversos tipos de documentos acadêmicos, incluindo artigos, teses e dissertações. Sua capacidade de adaptação para servir como repositório de dados de pesquisa, embora não tão especializada quanto o Dataverse, oferece uma solução integrada para instituições que buscam centralizar a gestão de todos os seus conteúdos acadêmicos em um único ambiente digital.

A escolha entre esses softwares deve considerar diversos fatores, como a capacidade técnica e orçamentária da instituição, o volume e tipo de dados a serem

gerenciados e a estratégia institucional de longo prazo. Instituições com recursos limitados podem optar pela adaptação do DSpace para gerenciar dados de pesquisa, enquanto aquelas com maior capacidade técnica e financeira podem implementar ambos os sistemas para maximizar a eficiência e conformidade com padrões internacionais.

Portanto, avaliar as capacidades técnicas de cada solução é fundamental para a governança e preservação dos dados de pesquisa. Ao sistematizar essas informações, este trabalho não apenas amplia o referencial teórico sobre o tema, mas orienta instituições na escolha de ferramentas estratégicas, promovendo práticas de gestão de dados mais resilientes e adequadas à realidade de cada entidade.

Como contribuições teórico-conceituais, enumeram-se: 1) a análise comparativa detalhada, entre os softwares Dataverse e DSpace, contribui para a literatura acadêmica ao fornecer uma base sólida para futuras pesquisas sobre repositórios de dados de pesquisa; 2) a ratificação da importância dos Princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) na gestão de dados de pesquisa, demonstrando como essas diretrizes podem ser aplicadas na prática para melhorar a acessibilidade e reutilização dos dados; e 3) a integração de metodologias de avaliação adaptadas de diversos autores (Rodrigues, 2004; Martins; Silva, 2017; Pavão *et al.*, 2018; Campêlo; Barreto Neto, 2019) proporciona uma abordagem robusta e abrangente para a análise de repositórios, contribuindo para o desenvolvimento de critérios de avaliação mais refinados e aplicáveis.

Como contribuições pragmáticas, o estudo provê diretrizes pragmáticas para a implementação de repositórios de dados de pesquisa, delineando cenários específicos junto aos seus respectivos benefícios e desafios, de modo a instrumentalizar gestores institucionais em processos decisórios fundamentados. Complementarmente, a prospecção e comparação de funcionalidades essenciais permitem que as instituições selecionem a arquitetura que melhor se ajuste às suas particularidades, ponderando critérios de organização, governança de conteúdo, interoperabilidade e segurança cibernética.

Entre as limitações desta pesquisa, identificam-se: a dependência de documentação disponível e estudos de caso existentes, que podem não refletir todas as nuances do uso real dos softwares e, além disso, a análise sendo baseada em versões específicas dos softwares, que podem evoluir e alterar as suas funcionalidades ao longo do tempo.

Futuras pesquisas podem explorar, entre outros assuntos: estudos de caso práticos com implementações reais dos sistemas em diferentes tipos de instituições; análises de custo-benefício detalhadas considerando diferentes cenários de uso; avaliação do impacto das atualizações de software nas funcionalidades e desempenho dessas soluções digitais; e investigação acerca da integração de novas tecnologias e padrões emergentes em repositórios de dados de pesquisa.

REFERÊNCIAS

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil de 1988**.

Brasília, DF: Presidência da República, 1988. Disponível em:

http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 3 abr. 2026.

BRASIL. **Lei nº 14.133, de 1º de abril de 2021**. Lei de Licitações e Contratos Administrativos.

Diário Oficial da União, Brasília, DF, 2 abr. 2021. Disponível em:

http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14133.htm. Acesso em: 3 abr. 2026.

CAMPÊLO, Leonard Richard Rodrigues Rufino; BARRETO NETO, Vanderlino Coelho.

Comparando Softwares gratuitos para criação de repositórios de dados abertos. **Ciência da Informação**, Brasília, v. 48, n. 3, 2019. DOI: <https://doi.org/10.18225/ci.inf.v48i3.5004>.

Disponível em: <https://revista.ibict.br/ciinf/article/view/5004>. Acesso em: 13 set. 2024.

DATAVERSE. **User Guides**. In: Dataverse.org. Disponível em:

<https://guides.dataverse.org/en/latest/#dataverse-documentation-v-version>. Acesso em: 20 out. 2024.

DIAS, Thiago Magela Rodrigues; DIAS, Patrícia Mascarenhas. Dados de pesquisa em acesso aberto: uma coleção de dados do conjunto de doutores cadastrados na Plataforma Lattes.

Ciência da Informação, Brasília, v. 48, n. 3, 2019. Disponível em:

<https://revista.ibict.br/ciinf/article/view/4997>. Acesso em: 13 set. 2024.

DSPACE. **Dspace - Documentation: Dspace 8.0**. In: Lyrasis.org. Disponível em:

<https://wiki.lyrasis.org/display/DSDOC8x>. Acesso em: 20 out. 2024.

ESTEVÃO, Janete Saldanha Bach; ARNS, Elaine Mandelli; STRAUHS, Faimara do Rocio. Gestão de dados de pesquisa: a practice to open the black-box of scientific research. **RDBCi: Revista**

Digital de Biblioteconomia e Ciência da Informação, Campinas, v. 17, 2019. Disponível em:

<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8656239>. Acesso em: 13 set. 2024.

LUCAS, Elaine Rosangela de Oliveira; PICALHO, Antonio Carlos; CAITANO, Vitória Maria Hartmann. Mapeamento e descrição de características de repositórios multidisciplinares de dados científicos abertos. **BiblioCanto**, Natal, v. 7, n. 1, p. 37-61, 2021. DOI:

<https://doi.org/10.21680/2447-7842.2021v7n1id25703>. Disponível em:

<https://periodicos.ufrn.br/bibliocanto/article/view/25703>. Acesso em: 13 set. 2024.

MARTINS, Dalton Lopes; SILVA, Marcel Ferrante. Critérios de avaliação para sistemas de bibliotecas digitais: uma proposta de novas dimensões analíticas. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 8, n. 1, p. 100-121, 2017. DOI: <https://doi.org/10.11606/issn.2178-2075.v8i1p100-121>. Disponível em: <https://www.revistas.usp.br/incid/article/view/125678>. Acesso em: 11 set. 2024.

NEVES, Bárbara Coelho. A Gestão de Dados Científicos: definições, conceitos e processo visando a internacionalização. In: VIRTUAL EDUCA ENCUEUNTROS, 2018, Salvador. **Ponencias** [...]. Salvador: Virtual Educa, 2018. Disponível em: <https://encuentros.virtualeduca.red/storage/ponencias/bahia2018/ya3Ss36UsaAC277pIRTjgsi1aFFIS6cTJMOXieIK.pdf>. Acesso em: 11 out. 2024.

PAVÃO, Caterina Groposo *et al.* **Acesso aberto a dados de pesquisa no Brasil**: repositórios brasileiros de dados de pesquisa: relatório 2018. Porto Alegre: UFRGS, 2018. Disponível em: <https://lume.ufrgs.br/handle/10183/185138>. Acesso em: 4 out. 2024.

RODRIGUES, Eloy *et al.* RepositóriUM: criação e desenvolvimento do Repositório Institucional da Universidade do Minho. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 8., 2004, Estoril. **Anais eletrônicos** [...]. Estoril: Associação Portuguesa de Bibliotecários, Arquivistas, Profissionais da Informação e Documentação, 2004. Tema: Nas encruzilhadas da informação e da cultura - (re)inventar a profissão. Disponível em: <https://publicacoes.bad.pt/revistas/index.php/congressosbad/article/view/621>. Acesso em: 11 nov. 2024.

ROCHA, Rafael Port da *et al.* **Acesso aberto a dados de pesquisa no Brasil**: soluções tecnológicas: relatório 2018. Porto Alegre: UFRGS, 2018. Disponível em: <http://hdl.handle.net/10183/185126>. Acesso em: 19 set. 2024.

ROCHA, Rafael Port da *et al.* Análise dos sistemas DSpace e Dataverse para repositórios de dados de pesquisa com acesso aberto. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 17, p. 1-25, 2021. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1572>. Acesso em: 19 set. 2024.

SAYÃO, Luís Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015. Disponível em: <https://www.aben.com.br/Arquivos/420/420.pdf>. Acesso em: 27 set. 2024.

SAYÃO, Luís Fernando; SALES, Luana Farias. Plataformas de gestão de dados de pesquisa. **Palavra Clave**, La Plata, v. 12, n. 1, 2022. Disponível em: <https://www.palabraclave.fahce.unlp.edu.ar/article/view/pce171>. Acesso em: 27 set. 2024.

SHINTAKU, Milton; APPEL, André; OLIVEIRA, Alexandre de. Tecnologias para gestão de dados de pesquisa segundo preceitos FAIR. In: SALES, Luana Farias *et al.* (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: IBICT, 2021. p. 129-146. Disponível em: <https://doi.org/10.22477/9786589167242.cap10>. Acesso em: 27 set. 2024.

SILVA, Lucas Henrique Alves da; LIMA, Izabel França de. Requisitos funcionais para repositórios de dados: subsídios teóricos. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 21., 2021, Rio de Janeiro. **Anais eletrônicos** [...]. Rio de Janeiro: ENANCIB, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/viewFile/460/461>. Acesso em: 7 nov. 2024.

VIDOTTI, Silvana Aparecida Borsetti Gregorio; TORINO, Emanuelle; CONEGLIAN, Caio Saraiva. #SejaJUSTOeCUIDADOSO: princípios FAIR e CARE na gestão de dados de pesquisa. In: SALES, Luana Farias *et al.* (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: IBICT, 2021. p. 201-214. Disponível em: <https://doi.org/10.22477/9786589167242.cap10>. Acesso em: 13 set. 2024.

Recebido em: 10 de janeiro de 2025

Aprovado em: 29 de agosto de 2026

Publicado em: 29 de agosto de 2026