

---

# DIAGNÓSTICO DOS REPOSITÓRIOS DE DADOS NO BRASIL

---

*Diagnosis of Data Repositories in Brazil*

---

**Fabiano Couto Corrêa da Silva (1), Amanda Santos Witt (2), Larissa Weber Umpierre (3),  
Lucas George Wendt (4), Michele Krieger Bohnert (5), Wagner Silva Wessfl (6)**

(1) Universidade Federal do Rio Grande do Sul (UFRGS), Brasil, fabianocc@gmail.com

(2) amandawitt.asw@gmail.com

(3) larissaumpierreb@gmail.com

(4) lucas.george.wendt@gmail.com

(5) michele.bohnert@gmail.com

(6) wagnerwessfl@yahoo.com.br



## Resumo

Os dados digitais se tornaram essenciais para alavancar a pesquisa científica. Em parte, servem a um papel funcional e efêmero, mas sua importância se torna permanente e o seu valor como substrato da comunicação científica pode aumentar ou diminuir, dependendo das demandas e usos a que são destinados. A preservação dos dados passou a ser uma demanda tanto para pesquisadores quanto para universidades e centros de pesquisa. Em resposta, no Brasil têm surgido diferentes repositórios, que permitem armazenar dados em variadas áreas do conhecimento. Esta pesquisa objetiva localizar repositórios que reúnem dados primários de pesquisa científica no Brasil. Foram empreendidas buscas nos diretórios Re3Data, OpenDoar, ROAR e revisão dos repositórios das Universidades Federais cadastradas no Ministério da Educação. Isto possibilitou localizar 20 repositórios que abrangem dados e a análise das suas principais características em relação às áreas de cobertura temática e aos padrões de preservação. Os dados obtidos demonstram que, embora a maioria dos repositórios disponibilizem alinhamento com padrões internacionais, alguns não apresentam infraestruturas adequadas para os conjuntos de dados disponibilizados. Considera-se que pode haver falha na gestão de determinados repositórios, mas a maioria oferece adequação às necessidades para disponibilização e busca dos dados científicos em todos os indicadores.

**Palavras-chave:** Repositórios de Dados; Comunicação Científica; Dados Científicos; Gestão de dados; Preservação digital.

## Abstract

Digital data have become essential to leverage scientific research. In part, they serve a functional and ephemeral role, but their importance becomes permanent and their value as a substrate for scientific communication can increase or decrease depending on the demands and uses to which they are put. Data preservation has become a demand for researchers as well as for universities and research centers. In response, different repositories have emerged in Brazil, which allow storing data in various areas of knowledge. This research aims to locate repositories that gather primary scientific research data in Brazil. Searches were undertaken in the directories Re3Data, OpenDoar, ROAR and review of the repositories of the Federal Universities registered with the Ministry of Education. This made it possible to locate 20 repositories covering data and to analyze their main characteristics in relation to subject coverage areas and preservation standards. The data obtained show that, although most repositories provide alignment with international standards, some do not have adequate infrastructures for the datasets provided. It is considered that there may be a failure in the management of certain repositories, but most offer adequacy to the needs for availability and search of scientific data in all indicators.

**Keywords:** Data Repositories; Scientific Communication; Scientific data; Data management; Digital preservation.

## 1 Introdução

---

A tentativa de proporcionar uma definição precisa dos dados científicos é um desafio, pois implica adaptar-se ao contexto no qual se faz a pergunta. Eles compreendem uma área muito extensa (todas as disciplinas) e devido a isso a definição pode variar em relação às abordagens dos diferentes participantes (pesquisadores, instituições, financiadores, *etc.*) e com os diversos contextos nacionais (SILVA, 2019). Em quaisquer dos casos, uma definição em termos gerais pode ser: os dados científicos são definidos como registros de fatos - numéricos, textuais, imagens e sons - usados como fontes primárias para pesquisa e são cada vez mais aceitos na comunidade científica como necessários para validar os resultados de pesquisas (OECD, 2007).

O Gerenciamento de Dados Científicos está presente em todas as fases da pesquisa e abrange coleta, organização, documentação, armazenamento e preservação dos dados utilizados ou gerados durante um projeto de pesquisa. Desse modo, configuram-se, por exemplo, como dados estatísticos, medições, respostas de questionários, registros de entrevistas, fotografias e observações resultantes do trabalho de campo. Especificamente, Arano *et al.* (2011) distinguem como dados primários experimentos científicos; modelos e simulações; observações; dados derivados tanto do processo como da combinação de dados brutos; dados referenciais (por

exemplo, estruturas químicas ou sequências genômicas); e material suplementar como instruções de codificação e guias de entrevistadores, entre outros. Desse modo, os dados científicos são definidos como registros de fatos que são usados como fontes primárias de pesquisa e são cada vez mais aceitos pela comunidade científica como necessários para validar resultados (OCDE, 2007).

Diante do interesse pelos dados e a conseqüente necessidade de classificá-los com metadados que possibilitem a sua recuperação e eventual utilização por parte da comunidade científica, estamos passando por um momento de mudança na forma como a ciência é conduzida. Trata-se de um desafio seja para os pesquisadores, seja para as universidades e outras instituições afins. Conseqüentemente, estamos diante de um novo paradigma baseado no fato de que a ciência deve ser aberta, colaborativa e feita com e para a sociedade (Anglada e Abadal, 2018).

Parte-se da premissa que a maneira como a ciência atualmente é gerada não atende a valores básicos da comunicação científica, tais como: registro da descoberta, velocidade da publicação, solidez nos dados, replicabilidade e reprodutibilidade das pesquisas, bem como sua ampla distribuição. A expressão “ecossistema da Ciência Aberta” e seu conceito não estão ligados apenas às tecnologias da informação, visto que o seu modelo tem um alcance além da plataforma por onde é divulgada, abrangendo assim, uma simbiose entre as tecnologias, as pessoas e os processos do ciclo de vida da pesquisa.

A Ciência Aberta, entendida como uma das manifestações do movimento de livre acesso à informação, tem ganhado aceitação na comunidade científica. É uma consequência do desenvolvimento da Internet, sua massificação e sua atratividade de expandir o acesso a dados e informações resultantes de pesquisas acadêmicas (SILVEIRA, et. al, 2021). Por estarem disponíveis em formato digital, os usuários podem, portanto, acessar, recuperar, explorar, reproduzir e disseminar dados de pesquisa de acesso aberto gratuitamente. E aliás, por serem abertos, os dados também podem ser reutilizados por outros pesquisadores, transformados por novos métodos ou adicionados a outras fontes, evitando assim a duplicação de ensaios, garantindo a transparência de sua obtenção e permitindo sua validação (Melero, 2014).

Pesquisas anteriores mostram a viabilidade e importância de analisar repositórios institucionais a partir do estabelecimento de dimensões para tal, como por exemplo, softwares utilizados e áreas do conhecimento, no âmbito dos Estados Unidos (Lynch; Lippincott, 2005; Westrienen; Lynch, 2005; Markey et al.2007) e da América Latina (Costa; Leite, 2017). No Brasil, Lucas, Picalho e Caitano (2020) caracterizaram e mapearam repositórios de dados científicos abertos com fins de constituir uma alternativa para subsidiar a tomada de decisão institucional no âmbito da gestão de dados científicos. Os pesquisadores, para a descrição dos dados de cada repositório, empreenderam uma análise a partir das categorias, como tipos e formatos de arquivos aceitos, curadoria de conteúdo; identificadores persistentes de conteúdo, controle de versões, dentre outras. Concluíram que o conhecimento sobre esses repositórios é pertinente não apenas aos pesquisadores, mas especialmente às instituições. Estas, devem escolher entre as opções possíveis, qual se enquadra melhor a sua estrutura, de modo a elaborar políticas institucionais que incentivem a disponibilização de dados científicos segundo os parâmetros da Ciência Aberta (Open Science).

O portal *Re3data* e a Web foram utilizados por Santos e Freitas (2021) para identificar instituições que possuem repositórios de dados de pesquisa. Os dois repositórios da Embrapa chamaram a atenção de Santos e Freitas (2021) em função da proposta de compartilhamento de informação e pela possibilidade de interação social por eles viabilizada. Observaram em seu estudo, que muitos dos repositórios que integraram o *corpus* da pesquisa, não seguiam a padronização de acordo com os novos modelos de repositório de dados compatíveis com os princípios *Findable, Accessible, Interoperable, Reproducible* (FAIR), não tendo, outrossim, identificador digital persistente, apesar de estarem em acesso aberto. Saliendam que, apesar de este assunto ainda não ser familiar para muitos, ele tem se tornado mais comum em função das políticas fomentadas pelos movimentos favoráveis à temática e pelas demandas sociais por dados científicos. Os repositórios configuram assim, segundo a percepção das autoras, importantes instrumentos de organização e de visibilidade da informação e constituem uma oportunidade para o Brasil expandir a visibilidade dos seus pesquisadores, obter melhores resultados para as instituições de pesquisa em que estão inseridos e promover o diálogo com disciplinas variadas (Santos; Freitas, 2021).

Ademais, Santos e Freitas (2021) entendem que o surgimento de novos repositórios de pesquisa é uma realidade, uma vez que profissionais da informação, pesquisadores e instituições têm se conscientizado sobre os benefícios implicados, incluindo a contribuição para promover novos estudos científicos. Isto enseja um olhar mais atento para o caso brasileiro, que é objeto de estudo do presente artigo.

Peter Suber, um dos principais promotores do acesso aberto à ciência e um dos idealizadores da Declaração de Budapeste, define literatura de acesso aberto como aquela que é digital, online, gratuita e livre da maioria das restrições devido a direitos autorais (Suber, 2004 ). Desse modo, o acesso aberto pode ser alcançado através da publicação em revistas de acesso aberto, ou uma cópia do documento pode ser depositada em um repositório, seja ele institucional, temático ou entidade financiadora (Budapest Open Access Initiative, 2002). A Declaração da Bethesda sobre Publicação de Acesso Aberto (2003) define uma publicação de acesso aberto como aquela que atende a duas condições: o autor e o detentor dos direitos autorais garantem a todos os usuários o direito de acesso gratuito, irrevogável, mundial e perpétuo; e também ao depósito de uma versão e do material complementar em formato eletrônico em pelo menos um repositório online, pertencente a uma instituição acadêmica, ou a qualquer instituição que garanta acesso e preservação. A Declaração de Berlim (Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2002) reitera o apoio às duas declarações anteriores de acesso aberto e defende o papel fundamental da Internet como ferramenta de divulgação da pesquisa.

Conforme evidenciamos mediante o diagnóstico realizado nos repositórios que foram elencados, o compartilhamento de dados científicos atualmente é uma tendência crescente, principalmente nas ciências naturais, mas nem todos os campos do conhecimento a têm abraçado com a mesma intensidade. As idiossincrasias particulares dos pesquisadores nas ciências sociais e humanas resultam em poucos dados acessíveis. Como atenuante, a falta de compartilhamento pode ser devido à falta de conhecimento sobre onde e como depositar os dados científicos. A Ciência Aberta requer uma infraestrutura tecnológica adequada e um modelo de governança (OpenAire, 2017). Dados valiosos de pesquisa são perdidos ou destruídos devido à falta de estruturas para organizá-los, protegê-los e disponibilizá-los ao longo do tempo. Essas infraestruturas fazem parte de uma abordagem de compartilhamento e abertura de dados de acordo com os princípios FAIR

de "Findable, Accessible, Interoperable, Reusable" (Localizável, Acessível, Interoperável, Reutilizável).

O depósito de um conjunto de dados em um repositório é geralmente acompanhado pela captura ou coleta de informações (metadados) sobre os dados depositados, o que facilita a compreensão e interpretação dos dados (por exemplo, cobertura geográfica, temporal, etc.). Além dos metadados padrão como os metadados do Dublin Core que descrevem o autor, título, ano de criação, *etc.* de um conjunto de dados, um repositório geralmente oferece um conjunto de metadados específicos para o assunto, tema, disciplina (por exemplo, dados biológicos, astronômicos, ambientais, etc.) dos dados que hospeda.

O armazenamento e o depósito de dados são duas coisas distintas: enquanto o armazenamento diz respeito ao momento em que os dados são recolhidos e tratados, o depósito se constitui em um processo de preservação, partilha e divulgação que permitirá estabelecer uma ligação entre os dados e as publicações. O depósito dos dados é, portanto, garantia de transparência e reprodutibilidade da pesquisa. Os dados depositados em um repositório existem independentemente do artigo científico: eles devem ser descritos pelos metadados mais ricos possíveis, a fim de encontrá-los facilmente para incentivar o compartilhamento e a reutilização. A cada conjunto de dados é atribuído um identificador permanente ou número de acesso, que o torna visível, acessível e citável, da mesma forma que a publicação. Os Repositórios de Dados Científicos são, desse modo, bancos de dados projetados para hospedar, armazenar, tornar visíveis e acessíveis os dados oriundos de pesquisas científicas. Seu papel é permitir que os dados sejam depositados ou coletados, descritos, acessados e compartilhados para reutilização. Cada repositório geralmente tem uma política para o depósito, descrição e divulgação de dados. Essas infraestruturas fazem parte de uma abordagem de compartilhamento e abertura de dados de acordo com os princípios FAIR.

O objetivo geral deste trabalho é apresentar um diagnóstico dos repositórios de dados científicos no Brasil. Por sua vez, os objetivos específicos são:

- Estabelecer um conjunto de parâmetros e indicadores para avaliar a qualidade dos repositórios.

- Aplicar os atributos elencados baseados em infraestruturas globais de sistemas de arquivamento de dados na web.
- Determinar um panorama amplo sobre os mecanismos de preservação e os conteúdos contemplados nos repositórios analisados.

Um Repositório de Dados Científicos tem a função, portanto, de permitir o depósito, a sua descrição, o seu acesso e o seu compartilhamento com vista à sua reutilização. Um dos critérios para a escolha de um repositório pode ser a possibilidade de concessão de licença, exigindo a citação dos criadores dos dados quando estes forem reutilizados. Os Repositórios de Dados Científicos analisados no presente diagnóstico configuram-se como bancos de dados projetados para hospedar, armazenar, tornar visíveis e acessíveis os dados oriundos de pesquisas científicas. Seu papel é permitir que os dados sejam depositados ou coletados, descritos, acessados e compartilhados para reutilização. Desse modo, embora cada repositório geralmente tenha uma política para o depósito, descrição e divulgação de dados, o nosso foco reside na análise estrutural das infraestruturas disponíveis, no que concerne às adequações para tornar os dados devidamente registrados.

A fim de empreender um levantamento e análise dos Repositórios de Dados no Brasil, foi realizada uma pesquisa para localizar os Repositórios de Dados que estão atualmente em operação. Assim, representamos um diagnóstico sobre as principais áreas e recursos utilizados em âmbito nacional para oferecer infraestruturas que suportam a complexidade da preservação de dados e atendem a comunidade científica. De modo amplo, analisamos as principais características que envolvem os repositórios de dados científicos em âmbito nacional, em relação às infraestruturas e áreas predominantes.

## **2 Metodologia**

---

A pesquisa é de natureza descritiva-explicativa, pois teve como objetivo fornecer uma visão geral dos Repositórios de Dados no Brasil por intermédio de um levantamento de infraestruturas online que hospedam dados científicos. Existem diretórios ou registros deste tipo

de ferramentas que fornecem conhecimento global da existência de diferentes tipos de repositórios, incluindo os repositórios de dados. O problema identificado quando este trabalho começou foi a existência de repositórios de dados de instituições brasileiras que não estão incluídas nestes registros. Por este motivo, foi necessário realizar uma pesquisa sistemática para tentar dar conta do universo de portais e repositórios de dados de pesquisa no Brasil. O levantamento foi realizado no dia 19 de novembro de 2022.

Os passos para realizar esta pesquisa foram:

1. Revisão de cada um dos repositórios registrados nos seguintes diretórios: Registry of Research Data Repositories (Re3Data), Directory of Open Access Repositories (OpenDOAR) e Registry of Open Access Repositories (ROAR);
2. Identificação de registros de repositório com o mecanismo de busca Google, utilizando como filtro os domínios existentes no Brasil e as seguintes palavras-chave:
  1. portal dados abertos
  2. repositório dados pesquisa
  3. repositório dados científicos
  4. repositório institucional
  5. repositório;
3. Revisão de cada uma das 64 Universidades Federais registradas no portal do Ministério da Educação.

### **3 Resultados**

---

A seguir apresentaremos os resultados de acordo com os indicadores estabelecidos para analisar as infraestruturas de repositórios de dados científicos no contexto brasileiro, a saber: Identificação (instituição, estado, ano criação); Software, Identificador Persistente, Cobertura temática, Tipo de repositório, Tipologia de arquivos, Versionamento, Vínculo institucional, Data de criação, Região e Padrão de metadados.



## Identificação dos repositórios:

---

Foram identificadas 20 organizações responsáveis pelas infraestruturas que reúnem os conjuntos de dados por diferentes esquemas ou estruturas organizacionais, com o objetivo de proporcionar opções de acesso aos utilizadores e assegurar um crescimento ordenado dos repositórios analisados.

O levantamento realizado no diretório Re3Data apresentou 03 (três) consórcios em nível nacional: Lattes Data (CNPq), FAPESP COVID-19 DataSharing/BR e Scielo Data. Também pelo *Re3Data* foi possível identificar os repositórios que o Brasil participa em consórcios em nível internacional: WorldClim - Global Climate Data, International Ocean Discovery Program e o Global Collaboration Engine (GLOBE). Por meio do diretório *OpenDoar* identificamos os repositórios CarpeDIEN: Dados e Informações em Energia Nuclear e o Portal de Dados Abertos do TSE. No diretório *ROAR* identificamos o Portal Brasileiro de Dados Abertos, mas foi descartado, por ser um metabuscador e não um repositório de dados. Já no buscador Google, localizamos o Repositório de Dados da Universidade Federal do ABC; Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa), Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa) e a Sapiientia: Repositório do Instituto Butantan.

O primeiro repositório identificado é oriundo de uma parceria brasileira junto ao International Ocean Discovery Program, sob a supervisão da CAPES e de algumas instituições associadas, tendo iniciado em 2013. Em nível estritamente nacional, foi localizado, por intermédio do *Re3Data*, o Open Research Data @PUC-Rio, sendo o primeiro repositório, lançado em 2014. Dentre os repositórios observados, salienta-se a distinção entre os que o Brasil atua como colaborador em consórcios com outros países e os que são repositórios institucionais ou temáticos que reúnem conjuntos de dados de pesquisas das instituições em âmbito nacional. No caso da parceria com outros países, foram identificados os repositórios da *International Ocean Discovery Program* e o *WorldClim - Global Climate Data*.

Quadro 1- Identificação dos repositórios

Repositório	Instituição	Estado	Ano de criação
International Ocean Discovery Program	CAPES (e associadas internacionais)	---	2013
Open Research Data @PUC-Rio	PUC-RIO	RJ	2014
GLOBE	Instituto Nacional de Pesquisas Espaciais (e associados internacionais)	SP	2015
PPBio Data Repository: Repositório de Dados de Levantamentos Biológicos	Instituto Nacional de Pesquisas da Amazônia	AM	2016
WorldClim - Global Climate Data	Consórcio global	---	2016
IBICT Cariniana Dataverse Network	IBICT	DF	2016
Carpe Dien: Dados e Informações em Energia Nuclear	Instituto de Engenharia Nuclear - IEN	RJ	2016
Base de Dados Científicos da Universidade Federal do Paraná	UFPR	PR	2017
Centro de Documentação e Acervo Digital da Pesquisa (CEDAP)	UFRGS	RS	2018
Sapientia: Repositório do Instituto Butantan	Instituto Butantan	SP	2018
Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa)	Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)	DF	2018
Repositório de Dados da Universidade Federal do ABC	Universidade Federal do ABC (UFABC)	SP	2019
Repositório de Dados de Pesquisa da Unicamp	UNICAMP	SP	2019
Portal de Dados Abertos do TSE	TSE	DF	2019
FAPESP COVID-19 DataSharing/BR	FAPESP em cooperação com a Universidade de São Paulo, Fleury Institute, Hospital Israelita Albert Einstein, Hospital Sírio-Libanês	SP	2020
Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí	Instituto Federal de Educação Ciência e Tecnologia Goiano	GO	2020
Scielo Data	Scientific Electronic Library Online	SP	2021
ARCA DADOS	FIOCRUZ	RJ	2022
Repositório Institucional UNESP	UNESP	SP	2022
Lattes Data (CNPq)	CNPq	DF	2022

SILVA, Fabiano Couto Corrêa da; WITT, Amanda Santos; UMPIERRE, Larissa Weber; WENDT, Lucas George; BOHNERT, Michele Krieger; WESSFLL, Wagner Silva. Diagnóstico dos Repositórios de Dados no Brasil. *Brazilian Journal of Information Science: research trends*, vol. 17, publicação continua, 2023, e023031. DOI: 10.36311/1981-1640.2023.v17.e023031.

Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa)	Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)	DF	2022
---	---	----	------

Fonte: Dados da pesquisa

### Identificadores persistentes:

Neste quesito, buscamos analisar a univocidade e consistência na designação de conjuntos de dados, utilizando identificadores persistentes, observando os mecanismos adotados e a testagem de *URLs* ativas durante a identificação nos repositórios.

Em geral, estes sistemas de numeração, muitas vezes únicos dentro do seu próprio domínio, estão sendo incorporados em esquemas de nomeação mais universais nos repositórios analisados, adicionando um nome inequívoco para o domínio ou instituição. Conseqüentemente, este tipo de estrutura permite à organização a máxima flexibilidade na identificação local dos seus recursos, ao mesmo tempo que permite a incorporação de identificadores locais em um sistema global com a adição da componente de nome de autoridade apropriada. A maioria dos repositórios identificados apresentam uma identificação que permanece constante, adotando apenas um mecanismo para gerar *URLs*, mas o Repositório Institucional UNESP e a Base de Dados Científicos da Universidade Federal do Paraná adotaram mais de um modelo, especificamente o Handle e o DOI.

Quadro 2 - Identificadores persistentes

Repositório	Identificado Persistente
Base de Dados Científicos da Universidade Federal do Paraná	Handle, DOI
IBICT Cariniana Dataverse Network	Handle
ARCA DADOS	DOI
Centro de Documentação e Acervo Digital da Pesquisa (CEDAP)	Handle
Repositório Institucional UNESP	DOI e Handle
Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí	DOI
Repositório de Dados de Pesquisa da Unicamp	DOI
PPBio Data Repository: Repositório de Dados de Levantamentos Biológicos	DOI, URN

Open Research Data @PUC-Rio	DOI
Lattes Data (CNPq)	DOI
FAPESP COVID-19 DataSharing/BR	Handle
Scielo Data	DOI
WorldClim - Global Climate Data	Não utiliza
International Ocean Discovery Program	DOI
GLOBE	DOI
Carpe Dien: Dados e Informações em Energia Nuclear	Handle
Portal de Dados Abertos do TSE	Não utiliza
Repositório de Dados da Universidade Federal do ABC	Não utiliza
Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa)	Não utiliza
Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa)	DOI
Sapientia: Repositório do Instituto Butantan	Handle

Fonte: Dados da pesquisa

#### Software:

---

Os repositórios de dados científicos analisados têm características particulares, mas alinham-se como uma infraestrutura tecnológica que permite compartilhar, arquivar, acessar, explorar e citar dados científicos, de forma a dar-lhes visibilidade, potencializar o seu acesso, reutilização e garantir a sua preservação a longo prazo.

Existem mais de 4.700 repositórios de acesso aberto no mundo de acordo com o Registry of Open Access Repositories (ROAR). Mais da metade são multidisciplinares e a maioria utiliza o software DSpace. No entanto, a maioria são repositórios de literatura, que não foram originalmente concebidos para a publicação de conjuntos de dados de pesquisa. Por isso, requerem a implementação de adaptações técnicas para cumprir com requisitos, tais como, os princípios FAIR e outras diretrizes. Nos casos analisados no presente diagnóstico, desempenham muitas funções úteis para Ciência Aberta, sendo o principal o armazenamento de conjuntos de dados tornando-os disponíveis para uma possível reutilização. Além disso, entre outras opções, servem de apoio para a gestão dos dados que abarcam todas as áreas do conhecimento.

Entre os repositórios analisados, todos utilizam softwares que oferecem estruturas orientadas por dados, metadados de citação, metadados específicos. No entanto, os softwares mais utilizados, Dataverse e o Dspace (cada um presente em 7 instituições), oferecem mecanismos adicionais centrados em documentação adicional, gestão de versões, licenças e definição de unidades hierárquicas.

O caso específico da Embrapa destaca-se por ser a única instituição que adotou mais de um tipo de software. Contudo, essa decisão se mostra adequada ao considerar as diferentes finalidades dos conjuntos de dados contemplados. Enquanto os dados resultantes de pesquisa do campo das ciências agrárias estão agrupados no software Dataverse, os dados georreferenciais são depositados no software GeoNode. Trata-se de um aplicativo e plataforma baseado na web para o desenvolvimento de Sistemas de Informações Geoespaciais (GIS) e para a implantação de Infraestruturas de Dados Espaciais (SDI).

Há, ainda, soluções que não utilizam softwares desenvolvidos propriamente para o depósito de dados. São linguagens que permitem organizar os dados com adaptações específicas para as finalidades dos repositórios analisados. No caso do Repositório de Dados da Universidade Federal do ABC, foi desenvolvido com o Joomla, que é um sistema livre open source de gestão de conteúdo web desenvolvido em PHP e com base de dados MySQL. Algumas instituições optaram pelo desenvolvimento de portais em linguagem HTML e ferramentas próprias, possibilitando a construção de arquiteturas e a inserção de conjuntos de dados de forma básica, isto é, apenas os datasets, deixando à margem aspectos de preservação digital e recuperação dos dados que são contemplados por softwares específicos para esta finalidade.

Quadro 3- Software

Repositório	Software
Base de Dados Científicos da Universidade Federal do Paraná	DSpace
IBICT Cariniana Dataverse Network	Dataverse
ARCA DADOS	Dataverse
Centro de Documentação e Acervo Digital da Pesquisa (CEDAP)	DSpace
Repositório Institucional UNESP	DSpace
Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí	Dataverse
Repositório de Dados de Pesquisa da Unicamp	Dataverse
PPBio Data Repository: Repositório de Dados de Levantamentos Biológicos	Drupal
Open Research Data @PUC-Rio	Sistema Maxwell
Lattes Data (CNPq)	Dataverse
FAPESP COVID-19 DataSharing/BR	DSpace
Scielo Data	Dataverse
WorldClim - Global Climate Data	Portal desenvolvido em linguagem HTML
International Ocean Discovery Program	Portal desenvolvido em linguagem HTML
GLOBE	Portal desenvolvido em linguagem HTML
Carpe Dien: Dados e Informações em Energia Nuclear	DSpace
Portal de Dados Abertos do TSE	Dspace
Repositório de Dados da Universidade Federal do ABC	Joomla (CMS)
Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa)	GeoNode
Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa)	Dataverse
Sapientia: Repositório do Instituto Butantan	DSpace

Fonte: Dados da pesquisa

#### Cobertura temática:

Os repositórios analisados abarcam conjuntos de dados de acordo com a pertinência dos temas que são direcionados, como é o caso do Portal de Dados Abertos do TSE, que reúne dados dos campos jurídico e legislativo. Por outro lado, alguns repositórios multidisciplinares ainda apresentam conjuntos de dados limitando-se a poucas áreas ou cobertura temática mais restrita,

por exemplo, a Repositório de Dados da Universidade Federal do ABC, com dados concentrados no campo das Ciências Sociais Aplicadas.

Quadro 4 - Cobertura temática

Repositório	Cobertura temática
Base de Dados Científicos da Universidade Federal do Paraná	Ciências Humanas e Sociais, Ciências da Vida, Ciências Naturais, Ciências da Engenharia
IBICT Cariniana Dataverse Network	Ciências Humanas, Ciências Sociais, Ciências Naturais, Ciências da Engenharia, Ciências da Vida, Ciências Comportamentais, História da Ciência, História, Humanidade
ARCA DADOS	Ciências Humanas, Ciências Sociais, Ciências da Vida, Ciências Naturais, Ciências da Engenharia
Centro de Documentação e Acervo Digital da Pesquisa (CEDAP)	Ciências Humanas, Ciências Sociais, Ciências da Vida, Ciências Naturais, Ciências da Engenharia
Repositório Institucional UNESP	Ciências Humanas, Ciências Sociais, Ciências da Vida, Ciências Naturais, Ciências da Engenharia
Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí	Biologia, Agricultura, Silvicultura, Horticultura, Medicina Veterinária, Ciências da Vida, Química, Ciência da Computação, Engenharia Elétrica, Engenharia de Sistemas Ciências Naturais, Ciências da Engenharia
Repositório de Dados de Pesquisa da Unicamp	Ciências Humanas, Ciências Sociais, Ciências da Vida, Ciências Naturais, Ciências da Engenharia
PPBio Data Repository: Repositório de Dados de Levantamentos Biológicos	Zoologia, Ciências de Plantas, Ecologia Animal, Biodiversidade, Pesquisa de Ecossistemas, Ecologia de Plantas, Análise de Ecossistemas, Biologia, Ciências da Vida
Open Research Data @PUC-Rio	Engenharia de Sistemas, Ciência da Computação, Engenharia Elétrica, Engenharia de Sistemas, Ciências da Engenharia
Lattes Data (CNPq)	Multidisciplinar (sem restrições)
FAPESP COVID-19 DataSharing/BR	Ciências da Vida, Virologia, Epidemiologia, Biometria Médica, Informática Médica Microbiologia, Biologia, Medicamento
Scielo Data	Ciências Humanas, Ciências Sociais, Ciências da Vida, Ciências Naturais, Ciências da Engenharia
WorldClim - Global Climate Data	Ciências Atmosféricas e Oceanografia Geociências (incluindo Geografia) Ciências Naturais

International Ocean Discovery Program	Geociências (incluindo Geografia), Oceanografia Geofísica, Oceanografia Geodésia, Geologia, Paleontologia, Geoquímica, Mineralogia, Cristalografia, Ciências Naturais, Ciências Atmosféricas, Oceanografia
GLOBE	Biologia, Ciências Vegetais, Geociências (incluindo Geografia), Ciências Atmosféricas, Oceanografia, Ciências da Vida, Ciências Naturais
Carpe Dien: Dados e Informações em Energia Nuclear	Engenharia Nuclear, Física, Combustível Nuclear, dentre outros.
Portal de Dados Abertos do TSE	Jurídico, legislativo
Repositório de Dados da Universidade Federal do ABC	Ciências Sociais Aplicadas
Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa)	Ciências Exatas e da Terra; Ciências Biológicas; Engenharias
Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa)	Ciências Exatas e da Terra; Ciências Biológicas; Ciências da Saúde; Ciências Sociais Aplicadas; Ciências Humanas; Engenharias
Sapientia: Repositório do Instituto Butantan	Ciências Exatas e da Terra; Ciências Biológicas; Ciências da Saúde; Ciências Sociais Aplicadas; Ciências Humanas

Fonte: Dados da pesquisa

### Tipos de repositórios:

Considerando a finalidade do repositório, institucional ou temática, foi possível avaliar que tanto os Repositórios Temáticos quanto os Institucionais cumprem o papel de oferecer um conjunto de serviços que envolvem a difusão sem restrições dos materiais digitais criados pelas próprias instituições. Identificamos 15 repositórios institucionais, sendo que 7 deste total pertencem a universidades federais de ensino superior e 8 estão vinculados a centros de pesquisa próprios, mas igualmente servindo como órgãos federais.

Também foram identificados 5 repositórios temáticos reunindo conjuntos de dados das seguintes áreas: Biológicos, Covid-19, Atmosféricos, Oceanográficos, Terrestres e Espaciais.



Quadro 5- Tipos de repositórios

Repositório	Tipo de repositório
Base de Dados Científicos da Universidade Federal do Paraná	Institucional
IBICT Cariniana Dataverse Network	Institucional
ARCA DADOS	Institucional
Centro de Documentação e Acervo Digital da Pesquisa (CEDAP)	Institucional
Repositório Institucional UNESP	Institucional
Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí	Institucional
Repositório de Dados de Pesquisa da Unicamp	Institucional
PPBio Data Repository: Repositório de Dados de Levantamentos Biológicos	Temático
Open Research Data @PUC-Rio	Institucional
Lattes Data (CNPq)	Institucional
FAPESP COVID-19 DataSharing/BR	Temático
Scielo Data	Institucional
WorldClim - Global Climate Data	Temático
International Ocean Discovery Program	Temático
GLOBE	Temático
Carpe Dien: Dados e Informações em Energia Nuclear	Institucional
Portal de Dados Abertos do TSE	Institucional
Repositório de Dados da Universidade Federal do ABC	Institucional
Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa)	Temático
Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa)	Institucional
Sapientia: Repositório do Instituto Butantan	Institucional

Fonte: Dados da pesquisa

#### Tipos de arquivos/conteúdos:

Os dados oriundos dos repositórios analisados, organizados e representados com uma estrutura rígida, a qual foi previamente planejada para armazená-los, por exemplo, um banco de dados. São estruturados conforme a definição de um esquema, que estabelece como as tabelas e suas respectivas linhas e colunas serão armazenadas. Estão distribuídos por meio de bancos de dados que oferecem o acesso de forma organizada em diversos formatos.

Quadro 6 - Tipos de arquivos/conteúdos

Repositório	Tipos de arquivos/conteúdos
Base de Dados Científicos da Universidade Federal do Paraná	Bases de dados, Estatísticos, Código fonte, Gráficos estruturados, Aplicações de software, Imagens, Dados audiovisuais Texto simples
IBICT Cariniana Dataverse Network	Estatísticos, Texto simples, Dados brutos, Bancos de dados, outros
ARCA DADOS	Dados baseados em rede, Bancos de dados, Imagens, Gráficos estruturados, Audiovisuais, Estatísticos, Dados brutos, Texto simples, Texto estruturado, Aplicativos de software, Código fonte
Centro de Documentação e Acervo Digital da Pesquisa (CEDAP)	Bases de dados, Imagens, Dados audiovisuais, Estatísticos, Texto simples, Texto estruturado
Repositório Institucional UNESP	Bases de dados, Imagens, Dados audiovisuais, Estatísticos, Texto simples, Texto estruturado
Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí	Estatísticos
Repositório de Dados de Pesquisa da Unicamp	Dados baseados em rede, Bancos de dados, Imagens, Dados audiovisuais, Estatísticos, Código fonte, Texto simples, Texto estruturado
PPBio Data Repository: Repositório de Dados de Levantamentos Biológicos	Dados brutos, Estatísticos, Texto simples, Texto estruturado, Dados audiovisuais, Imagens
Open Research Data @PUC-Rio	Estatísticos
Lattes Data (CNPq)	Texto simples, Texto estruturado, Dados brutos, métricas, estatísticos.
FAPESP COVID-19 DataSharing/BR	Estatísticos
Scielo Data	Dados baseados em rede, Bancos de dados, Imagens, Gráficos estruturados, Dados audiovisuais, Estatísticos, Dados brutos, Texto simples, Texto estruturado, Aplicativos de software, Código fonte
WorldClim - Global Climate Data	Gráficos estruturados, Estatísticos
International Ocean Discovery Program	Imagens, Gráficos estruturados, Estatísticos, Dados brutos, Texto simples, Texto estruturado
GLOBE	Dados baseados em rede, Gráficos estruturados, Dados brutos
Carpe Dien: Dados e Informações em Energia Nuclear	Bases de dados, Estatísticos, Código fonte, Gráficos estruturados, Aplicações de software, Imagens, Dados audiovisuais e Texto simples.

Portal de Dados Abertos do TSE	Bases de dados, Estatísticos, Imagens, Dados audiovisuais e Texto simples.
Repositório de Dados da Universidade Federal do ABC	Administrativos, Estatísticos
Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa)	Estatísticos, Gráficos estruturados, Imagens, Texto simples
Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa)	Bases de dados, Estatísticos, Código fonte, Gráficos estruturados, Aplicações de software, Imagens, Dados audiovisuais, Texto simples
Sapientia: Repositório do Instituto Butantan	Amostras, Registros de coleta, Formulários, Modelos, Resultados experimentais, <i>Software</i> , Gráficos

Fonte: Dados da pesquisa

#### Versionamento:

---

Ao considerarmos que as pesquisas avançam gradativamente, tanto quanto os dados coletados sobre um mesmo domínio do conhecimento, é desejável que em cenários de edição monousuário ou multiusuário os conjuntos de dados depositados possam receber o acréscimo de novas versões, sem a necessidade de aplicar bloqueios ou gerar dados duplicados. O controle de versões facilita a continuidade de uma pesquisa, permitindo que os pesquisadores adicionem novos dados isoladamente em uma versão do conjunto de dados ou em várias sessões simultaneamente. Quando um pesquisador conclui uma coleção de edições, ele pode adicionar ou mesclar todas as alterações a partir da versão principal, ou seja, da qual a versão foi criada. A versão original de todas as versões de um conjunto de dados pode ser definida, portanto, como a “versão padrão”.

Neste critério de avaliação, buscamos analisar se os repositórios oferecem versionamentos de conjuntos de dados por meio de cópias separadas da Versão Padrão, ou seja, possibilitando rastrear as versões dos conjuntos de dados como extensão da primeira versão dos dados de uma pesquisa. Todos os repositórios apresentam conjuntos de dados científicos, isto é, relacionados aos dados brutos oriundos de pesquisas, exceto o Repositório de Dados da Universidade Federal do ABC, que tem como premissa armazenar informações quantitativas relacionadas à administração da Universidade.

Quadro 7- Versionamento

Repositório	Permite Versionamento
Base de Dados Científicos da Universidade Federal do Paraná	SIM
IBICT Cariniana Dataverse Network	SIM
ARCA DADOS	SIM
Centro de Documentação e Acervo Digital da Pesquisa (CEDAP)	SIM
Repositório Institucional UNESP	SIM
Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí	SIM
Repositório de Dados de Pesquisa da Unicamp	SIM
PPBio Data Repository: Repositório de Dados de Levantamentos Biológicos	SIM
Open Research Data @PUC-Rio	NÃO
Lattes Data (CNPq)	SIM
FAPESP COVID-19 DataSharing/BR	SIM
Scielo Data	SIM
WorldClim - Global Climate Data	SIM
International Ocean Discovery Program	SIM
GLOBE	NÃO
Carpe Dien: Dados e Informações em Energia Nuclear	SIM
Portal de Dados Abertos do TSE	SIM
Repositório de Dados da Universidade Federal do ABC	SIM
Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa)	SIM
Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa)	SIM
Sapientia: Repositório do Instituto Butantan	NÃO

Fonte: Dados da pesquisa

#### Padrão de metadados:

Embora o modelo de metadados predominante utilizado nos repositórios analisados tenha sido o Dublin Core (modelo de metadados desenvolvido pela DCMI - Dublin Core Metadata Initiative) na sua versão básica (15 Campos), os conjuntos de dados demonstraram uma necessidade de aumentar e normalizar o número de campos de metadados e modelos de metadados associados (esquemas), a fim de proporcionar uma maior qualidade semântica dos dados.

Do mesmo modo, é necessário diferenciar o ambiente de utilização dos metadados nos diferentes sistemas de informação documental uma vez que, em geral, não há clareza da diferença de utilização e aplicação dos metadados de entrada e de saída dos campos nestas duas instâncias. Para analisar os metadados, observamos o ambiente de utilização de entrada e saída, da seguinte maneira:

*Metadados de entrada:* Corresponde ao conjunto de campos adicionados ao recurso de informação no momento da sua criação ou registo em um sistema informático associado (Sistema de Informação). Estes campos são geralmente adicionados por humanos e máquinas (campos automáticos). O conjunto de campos associados ao modelo central com qualificadores de campo (por exemplo, o DC.relation.hasPart) é geralmente utilizado.

*Metadados de saída:* Corresponde ao conjunto de campos apresentados por meio de uma interface de saída do sistema de informação e que são transformados (diferentes modelos de metadados) de acordo com os requisitos do utilizador (Homem/Máquina).

O conjunto de diretrizes analisadas teve em conta os modelos de metadados existentes, a fim de promover arquiteturas mais abertas que envolvem novos âmbitos de campo, novas e diferentes tipologias de documentos e novos vocabulários especializados para a descrição normalizada dos recursos de registo dos conjuntos de dados.

Quadro 8 - Padrões de metadados

Repositório	Padrão metadados
Base de Dados Científicos da Universidade Federal do Paraná	Dublin Core
IBICT Cariniana Dataverse Network	DDI - Data Documentation Initiative
ARCA DADOS	Dublin Core
Centro de Documentação e Acervo Digital da Pesquisa (CEDAP)	Dublin Core
Repositório Institucional UNESP	Dublin Core
Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutaí	DataCite Metadata Schema; Dublin Core; DDI - Data Documentation Initiative; OAI-ORE - Open Archives Initiative Object Reuse and Exchange

Repositório de Dados de Pesquisa da Unicamp	Dublin Core
PPBio Data Repository: Repositório de Dados de Levantamentos Biológicos	EML - Ecological Metadata Language
Open Research Data @PUC-Rio	Dublin Core
Lattes Data (CNPq)	Dublin Core
FAPESP COVID-19 DataSharing/BR	Dublin Core
Scielo Data	Dublin Core
WorldClim - Global Climate Data	IPTC Core (International Press Telecommunications Council)
International Ocean Discovery Program	ISO 19115
GLOBE	Dublin Core
Carpe Dien: Dados e Informações em Energia Nuclear	Dublin Core
Portal de Dados Abertos do TSE	Não há formato específico. Os arquivos são disponibilizados com os próprios dados eleitorais em Comma-Separated Values (CSV) e em Portable Document Format (PDF).
Repositório de Dados da Universidade Federal do ABC	Dublin Core
Embrapa: Geoinfo (Infraestrutura de Dados Espaciais da Embrapa)	Dublin Core
Embrapa: Redape (Repositório de Dados de Pesquisa da Embrapa)	Metadata Citation
Sapientia: Repositório do Instituto Butantan	Dublin Core

Fonte: Dados da pesquisa

## 4 Conclusões

Os repositórios de dados analisados são bancos de dados centralizados que fornecem aos usuários acesso confiável não somente aos registros originais, como também aos processos e ações que devem ser considerados em relação à preservação digital. Estes repositórios ajudam a reduzir a duplicação de esforços, permitem o compartilhamento e a reutilização de dados entre a comunidade científica, auxiliam no treinamento e fornecem acesso aos dados. Tradicionalmente, têm se concentrado em dados submetidos por pesquisadores, mas os dados de projetos de software "*open source*" também estão sendo cada vez mais incluídos.

Os repositórios de dados no Brasil reúnem diversos tipos de análises, dados brutos, estruturados e não estruturados, em múltiplos formatos e em todas as áreas do conhecimento. Fomentar visibilidade a todo esse conteúdo tem sido uma preocupação constante tanto para as instituições quanto para os gestores dos repositórios elencados no presente artigo.

O diagnóstico demonstrou que os repositórios emergem como uma plataforma natural para o armazenamento de dados científicos e que estão se tornando cada vez mais importantes no contexto da Ciência Aberta. Esses sites especializados ou abrangentes também fornecem uma riqueza de recursos para publicação e suporte à reutilização de pesquisas, possibilitando desenvolver novos domínios para conjuntos de dados específicos. De maneira geral, os repositórios de dados no Brasil atendem aos requisitos mínimos para assegurar a preservação digital e o consequente acesso aos conjuntos de dados. Os aspectos contemplados nessa análise incluem Acesso aberto, uso de identificadores únicos e persistentes, interoperabilidade e conformidade com padrões, capacidade de embargo, capacidade de armazenamento, sistemas de metadados diferenciados por disciplina, métricas de uso e curadoria de dados que recebem mais menções.

Também foi possível verificar recursos de qualidade da Web voltados para a melhoria da experiência do usuário na interface. Considerando que não é necessário que um repositório apresente as funcionalidades de um site, esse recurso raramente é levado em consideração. Em sua maioria, também demonstraram classificação flexível, opções de exportação, rotulagem adequada, filtros de pesquisa, facilidade de uso, link para o artigo original e páginas separadas para cada conjunto de dados. Esses recursos são organizados em parâmetros e indicadores que podem auxiliar administradores dos repositórios a escolher um programa de gerenciamento adequado ou uma avaliação de repositório já implementada, servindo como fatores de tomada de decisão para implementar planos de gerenciamento de dados.

Por fim, cabe ressaltar que uma correta gestão de dados deve incluir aspectos de avaliação institucional e infraestrutura orientados para as demandas específicas de cada área ou organização, ou seja, considerando as necessidades de um Repositório Temático ou Institucional. Os repositórios que deixam claro a sua política de acesso aberto, devem traduzi-la em ações concretas

que garantam a gestão adequada e a sustentabilidade dos repositórios de dados, de acordo com a missão de criar e transmitir conhecimento.

## Referências

---

- Anglada, Luis; Abadal, Ernest. “¿Qué es la ciencia abierta?”. *Anuario ThinkEpi*, v.12, 2018, pp.292-298, <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/thinkepi.2018.43>. Acessado 22 dez 2022.
- Arano, Silvia et. al. “La comunidad Recursos y datos primarios de la Universitat Pompeu Fabra: los repositorios institucionales como infraestructuras científicas: estudio de caso”. *Revista Española de Documentación Científica*, v. 34, n. 3, 2011, pp. 385-407, doi: <http://doi.org/10.3989/redc.2011.3.834>. Acessado 25 set. 2022.
- Brasil e Ministério da Educação. *Cadastro nacional de cursos e instituições de educação superior: cadastro e-MEC*. <https://www.gov.br/mec/pt-br>. Acessado 19 nov. 2022.
- Budapest Open Access Initiative. *Budapest Open Access Initiative*. 2002, doi: <http://doi.org/10.4403/jlis.it-8629>. Acessado 19 nov. 2022.
- Costa, Michelli Pereira da e Lima, Fernando César Lima. “Repositórios institucionais da América Latina e o acesso aberto à informação científica”. IBICT, 2017, <https://repositorio.unb.br/handle/10482/23202>. Acessado 03 ago. 2023.
- Directory of Open Access Repositories (OpenDOAR). *Site institucional*. <https://v2.sherpa.ac.uk/opensoar/>. Acessado 18 nov. 2022.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Censo da Educação Superior*. <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior>. Acessado 27 nov. 2022.
- Lynch, C.; Lippincott, J. “Institutional repository deployment in the United States as of early 2005”. *D-lib Magazine*, v. 11, n. 9, 2005, [www.dlib.org/dlib/september05/lynch/09lynch.html](http://www.dlib.org/dlib/september05/lynch/09lynch.html). Acessado 03 ago. 2023.
- Lucas, Elaine Rosângela de Oliveira, Picalho, Antonio Carlos e Caitano, Vitória Maria Hartmann. “Mapeamento e descrição de características de repositórios multidisciplinares de dados científicos abertos”. *Bibliocanto*, v. 7, n. 1, p. 37 – 61, 2021, <https://dx.doi.org/10.21680/2447-7842.2021v7n1ID25703>. Acessado 03 ago. 2023.
- Markey, Karn et al. *Census of institutional repositories in the United States*. Council on Library and Information Resources, 2007.
- 
- SILVA, Fabiano Couto Corrêa da; WITT, Amanda Santos; UMPIERRE, Larissa Weber; WENDT, Lucas George; BOHNERT, Michele Krieger; WESSFLL, Wagner Silva. Diagnóstico dos Repositórios de Dados no Brasil. *Brazilian Journal of Information Science: research trends*, vol. 17, publicação continua, 2023, e023031. DOI: 10.36311/1981-1640.2023.v17.e023031.



- Melero, Remedios y Hernández-San-Miguel, Javier. Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. *Revista Española De Documentación Científica*, v. 37, n. 4, 2014, e066, doi: <https://doi.org/10.3989/redc.2014.4.1154>. Acessado 29 nov. 2022.
- Brasil e Ministério da Educação. *Cadastro nacional de cursos e instituições de educação superior: cadastro e-MEC*. <https://www.gov.br/mec/pt-br>. Acessado 19 nov. 2022.
- OpenAire. *Site institucional*. <https://www.openaire.eu/open-science-fair-2017>. Acessado 20 nov. 2022.
- Organisation for economic co-operation and development (OECD). *Principles and guidelines for access to research data from public funding*. Paris, 2007, <http://www.oecd.org/sti/sci-tech/38500813.pdf>. Acessado 20 nov. 2022.
- Registry of Research Data Repositories (Re3Data). *Site institucional*. [www.re3data.org/](http://www.re3data.org/). Acessado 18 nov. 2022.
- Registry of Open Access Repositories (ROAR). *Welcome to the Registry of Open Access Repositories*. [roar.eprints.org](http://roar.eprints.org). Acessado 17 nov. 2022.
- Sampaio, Rosana. “Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica”. *Rev. bras. fisioter.*, v. 11, n. 1, jan./fev. 2007, pp. 83-89, <https://www.scielo.br/j/rbfis/a/79nG9Vk3syHhnSgY7VsB6jG/?format=pdf&lang=pt>. Acessado 27 nov. 2022.
- Santos, Ana Cristina Gomes e Freitas, Judite Gonçalves. “Repositórios de dados científicos nas instituições de ensino e pesquisa no Brasil: um novo desafio”. *Biblionline*, v. 17, n. 4, 2021, pp. 3-22, <https://periodicos.ufpb.br/index.php/biblio/article/view/61618/34901>. Acessado 03 ago. 2023.
- Silva, Fabiano Couto Corrêa da. *Gestão de dados científicos*. Rio de Janeiro: Interciência. 2019.
- Silveira, Lúcia da et al. “Ciência Aberta na perspectiva de especialistas brasileiros: proposta de taxonomia”. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 26, 2021, pp. 1–27, doi: <http://doi.org/10.5007/1518-2924.2021.e79646>. <https://periodicos.ufsc.br/index.php/eb/article/view/79646>. Acessado 22 dez. 2022.
- Suber, Peter. *Open Access*. M.I.T. Press, 2012.
- Westrienen, Gerard van and Lynch, Clifford A. “Academic institutional repositories: deployment status in 13 nations as of mid 2005”. *D-Lib Magazine*, v. 11, n. 9. Set. 2005, <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>. Acessado 03 ago. 2023.

---

Copyright: © 2023 SILVA, Fabiano Couto Corrêa da; WITT, Amanda Santos; UMPIERRE, Larissa Weber; WENDT, Lucas George; BOHNERT, Michele Krieger; WESSFLL, Wagner Silva. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

---

Received: 03/01/2023

Accepted: 02/08/2023