

# Preservação e Certificação de Repositórios Dataverse

*Prof<sup>a</sup> Laura Rezende*

# AGENDA

---

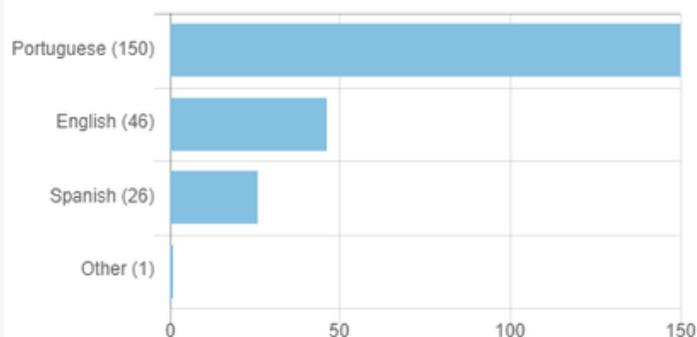
- **Panorama atual de depósitos de dados científicos;**
- **Qualidade nos repositórios de dados científicos;**
- **A certificação Core Trus Seal e os Repositórios Dataverse;**

# Repositórios de acesso aberto - Brasil

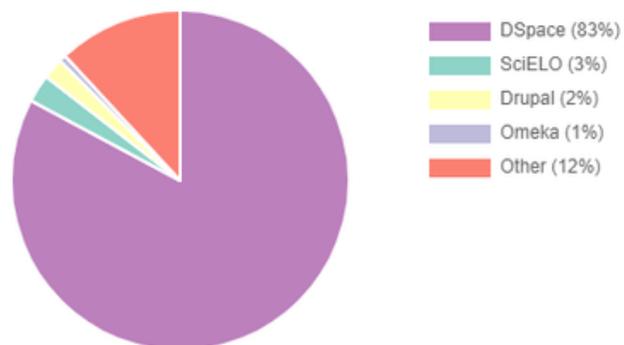
Number of items: 151.

An overview of this collection.

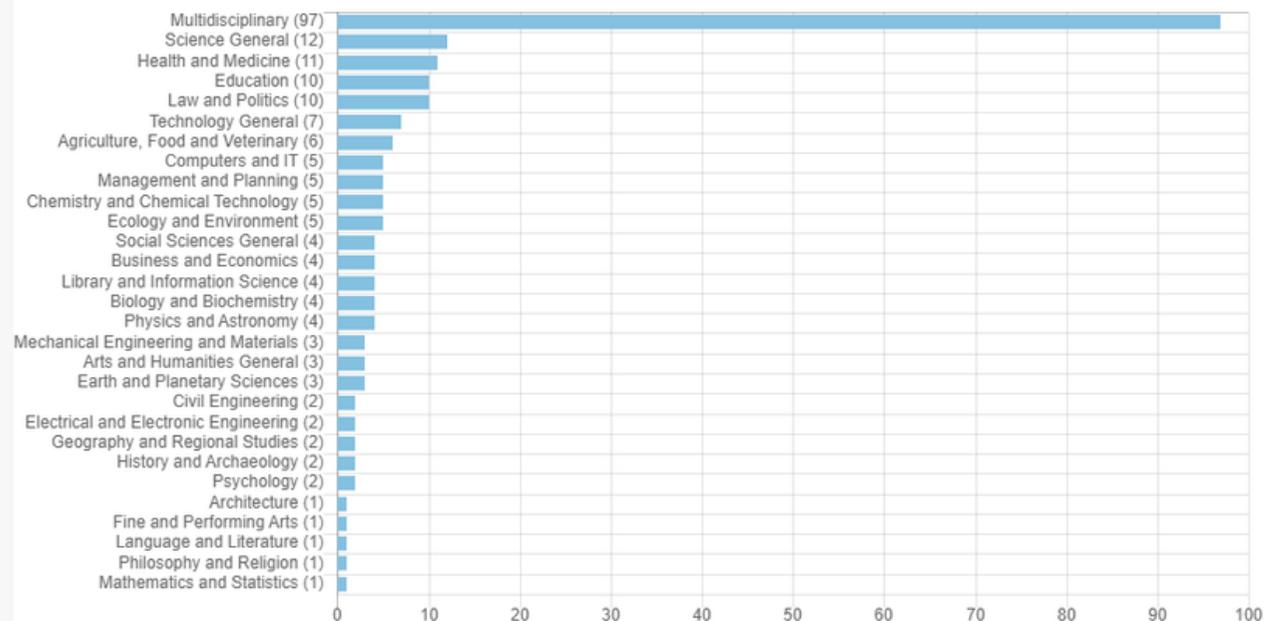
## Language of Content



## Software Platforms Overview



## Subject of Content



# Repositórios de acesso aberto - Brasil



## Registry of Open Access Repositories

[Home](#) [About](#) [Search](#) [Search Content](#) [Browse](#)

[Login](#) | [New Entry](#) | [Create Account](#)

Count matches any of "Brazil"

Brazil (185) ▼

Any Software ▼

Any Repository Type ▼

Sort by number of records ▼

[Search](#)

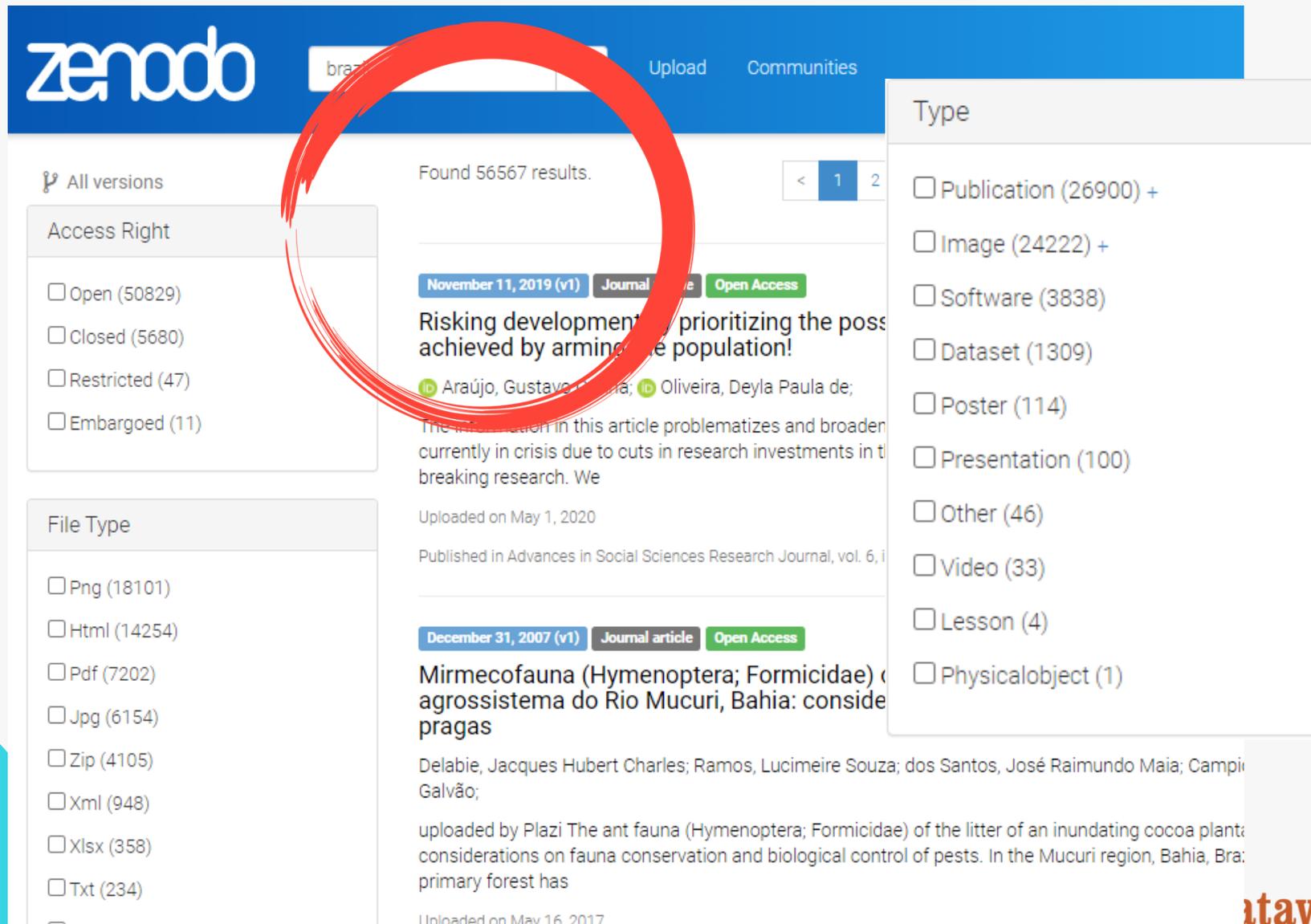
Displaying results 1 to 20 of 185.

[1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [6](#) | [7](#) | [8](#) | [9](#) | [10](#) | [Next](#)

# Repositórios de Dados Científicos - Brasil



# Pesquisadores Brasileiros e os repositórios internacionais

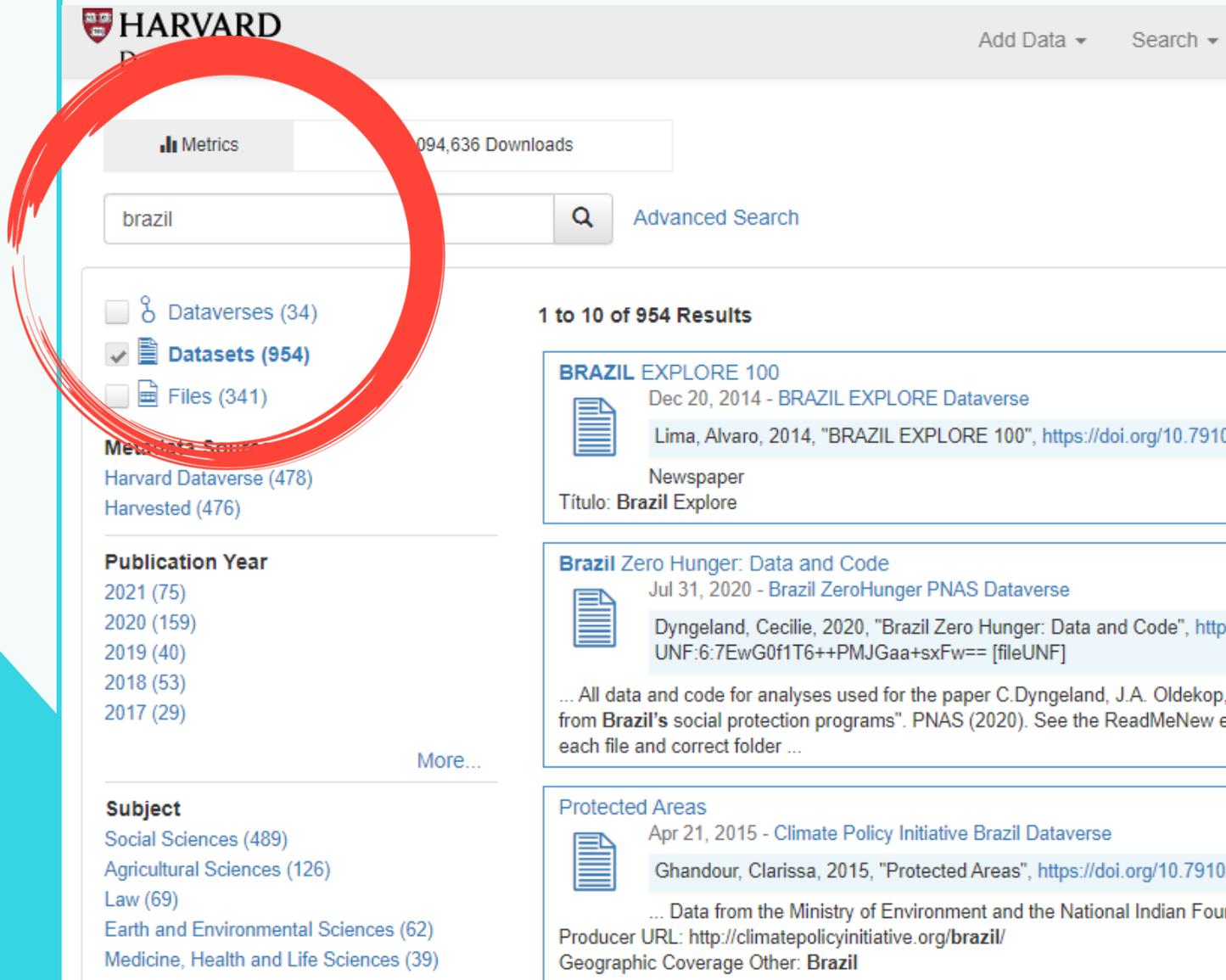


The screenshot shows the Zenodo website interface. At the top left is the Zenodo logo. A search bar contains the text 'bra'. To the right of the search bar are links for 'Upload' and 'Communities'. Below the search bar, the text 'Found 56567 results.' is displayed. A red circle highlights the search bar and the dropdown menu that appears below it. The dropdown menu is titled 'Type' and lists various file types with their respective counts:

- Publication (26900) +
- Image (24222) +
- Software (3838)
- Dataset (1309)
- Poster (114)
- Presentation (100)
- Other (46)
- Video (33)
- Lesson (4)
- Physicalobject (1)

Below the search results, two article entries are visible. The first entry is dated 'November 11, 2019 (v1)' and is a 'Journal article' with 'Open Access'. The title is 'Risking development by prioritizing the possible achievements by arming the population!'. The authors are 'Araújo, Gustavo' and 'Oliveira, Deyla Paula de'. The second entry is dated 'December 31, 2007 (v1)' and is a 'Journal article' with 'Open Access'. The title is 'Mirmecofauna (Hymenoptera; Formicidae) of the litter of an inundating cocoa plant: considerations on fauna conservation and biological control of pests. In the Mucuri region, Bahia, Brazil: primary forest has'. The authors are 'Delabie, Jacques Hubert Charles; Ramos, Lucimeire Souza; dos Santos, José Raimundo Maia; Campi Galvão;'. The text below the title reads: 'uploaded by Plazi The ant fauna (Hymenoptera; Formicidae) of the litter of an inundating cocoa plant: considerations on fauna conservation and biological control of pests. In the Mucuri region, Bahia, Brazil: primary forest has'. The upload date is 'Uploaded on May 16, 2017'.

# Pesquisadores Brasileiros e os repositórios internacionais



**HARVARD** Add Data ▾ Search ▾

Metrics 094,636 Downloads

brazil  Advanced Search

 Dataverses (34)

 **Datasets (954)**

 Files (341)

**Publication Year**

2021 (75)

2020 (159)

2019 (40)

2018 (53)

2017 (29)

[More...](#)

**Subject**

Social Sciences (489)

Agricultural Sciences (126)

Law (69)

Earth and Environmental Sciences (62)

Medicine, Health and Life Sciences (39)

**1 to 10 of 954 Results**

**BRAZIL EXPLORE 100**  
Dec 20, 2014 - BRAZIL EXPLORE Dataverse

 Lima, Alvaro, 2014, "BRAZIL EXPLORE 100", <https://doi.org/10.7910>

Newspaper  
Título: **Brazil Explore**

**Brazil Zero Hunger: Data and Code**  
Jul 31, 2020 - Brazil ZeroHunger PNAS Dataverse

 Dyngeland, Cecilie, 2020, "Brazil Zero Hunger: Data and Code", <http://UNF:6:7EwG0f1T6++PMJGaa+sxFw==> [fileUNF]

... All data and code for analyses used for the paper C.Dyngeland, J.A. Oldekop, from **Brazil's** social protection programs". PNAS (2020). See the ReadMeNew e each file and correct folder ...

**Protected Areas**  
Apr 21, 2015 - Climate Policy Initiative Brazil Dataverse

 Ghandour, Clarissa, 2015, "Protected Areas", <https://doi.org/10.7910/>

... Data from the Ministry of Environment and the National Indian Four  
Producer URL: <http://climatepolicyinitiative.org/brazil/>  
Geographic Coverage Other: **Brazil**

# **Armazenamento e Compartilhamento de Dados: Confiança é um fator essencial**

- **Agências de fomento: garantia de perenidade e retorno de investimento;**
- **Pesquisadores: garantia de que seus dados estão em local seguro, permanecendo acessíveis, utilizáveis e inteligíveis ao longo do tempo;**
- **Usuários: Buscam dados de alta qualidade que vem sendo preservados adequadamente;**

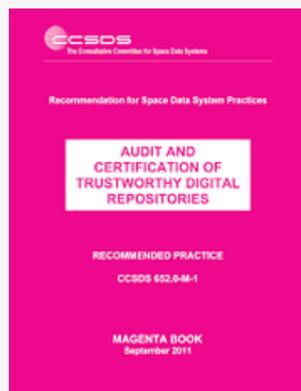
# Confiança em repositórios digitais



Levou à criação de certificações e metodologias de (auto)avaliação de repositórios:

- **Core Trust Seal**
- **Nestor Seal for Trustworthy Digital Archives** (Network of Expertise in long-term Storage of digital resources in Germany)
- **DRAMBORA (Digital Repository Audit Method Based on Risk Assessment)**: JISC (Joint Information Systems Committee), DCC - Digital Curation Centre;
- **DPCMM (Digital Preservation Capability Maturity Model)**: Dollar and Ashley;
- **Levels of Digital Preservation**: NDSA- National Digital Stewardship Alliance

# Visão Geral das Certificações



- **2010:** MoU (Memorandum of Understanding) estabelecendo a Estrutura Europeia de Auditora e Certificação:
  - the Data Seal of Approval, Audit and Certification of Trustworthy Digital Repositories Standard (ISO16363) and nestor Seal Criteria for Trustworthy Digital Archives (DIN31644).
1. **Certificação Básica / Essencial (Core):** concedida pelo antigo DSA e agora CTS;
  2. **Certificação Extendida:** exige ter o DSA ou CTS e uma auto auditoria revisada por membro externo baseada na TRAC (ISO 16363) ou Nestor Seal (DIN 31644);
  3. **Certificação Formal:** exige ter o DSA ou CTS e uma auditoria externa baseada na TRAC (ISO 16363) ou Nestor Seal (DIN 31644);



- **World Data System (WDS) - International Science Council**
- **Data Seal of Approval (DSA) - certificação anterior**

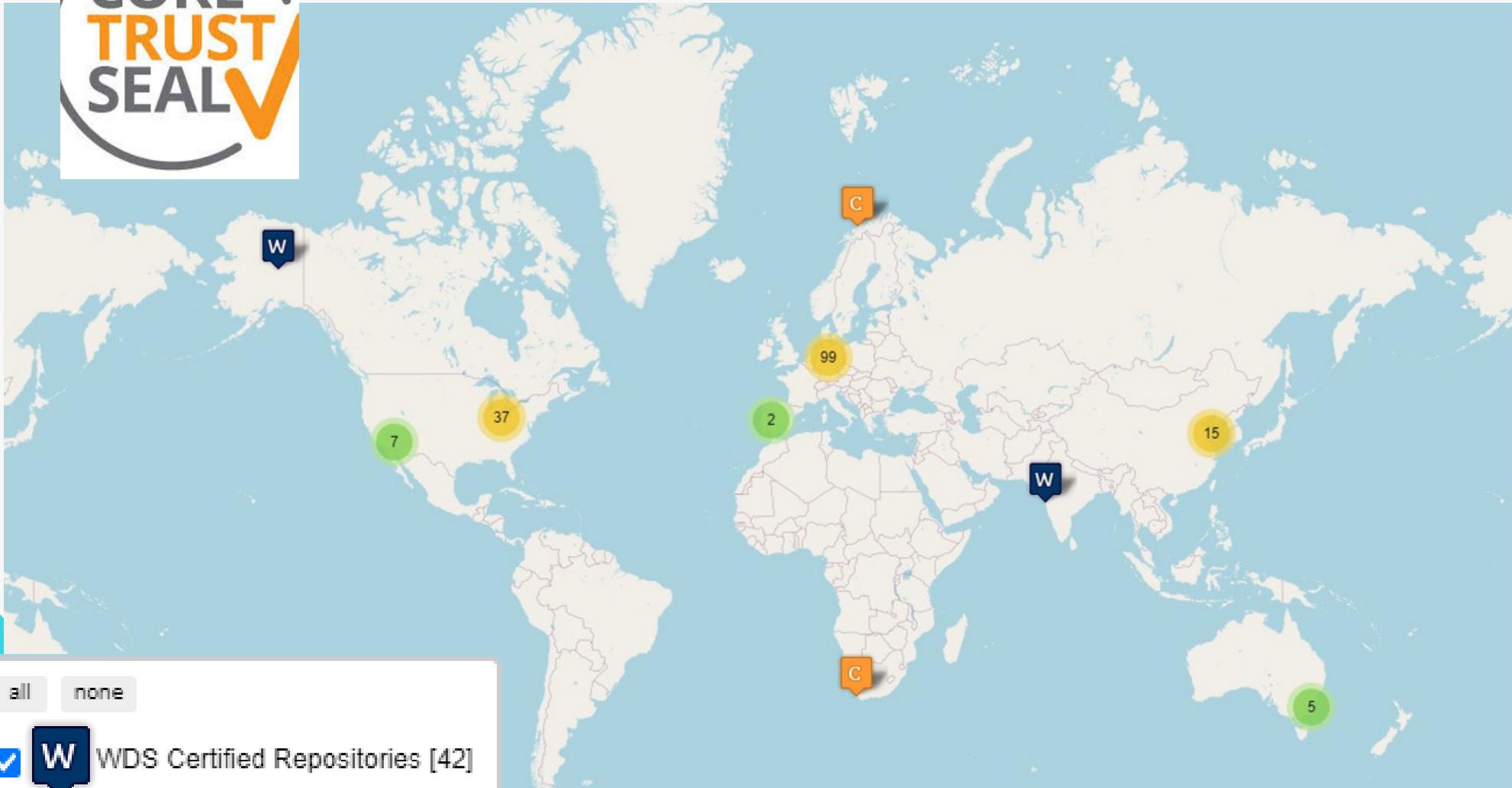
# CTS - Histórico

- (2008) DSA foi criado pela DANS (Data Archiving and Networked Services) atendendo à solicitação de 2 agências de fomento holandesas - Ciências Sociais e Humanidades; segue padrões de requisitos já existentes (TRAC - ISO 16363, Nestor, DRAMBORA);

Ideia inicial: selo de aprovação que garantisse que os dados digitais armazenados em repositórios pudessem ser localizados, compreendidos e utilizados futuramente.

- (2009) Data Seal of Approval passa a ser conduzido por um comitê internacional;
- 3 versões do conjunto de requisitos:
  - 2010
  - 2014 - 2017
  - 2017 - 2019
- (2017) DSA passa a ser CoreTrustSeal (CTS).
- Período de transição (2017 a 2018) - Aplicações em andamento recebem selo CTS;
- Repositórios que receberam DSA de 2014 a 2017 deveriam renovar seguindo os requisitos CTS ;

# Repositórios Certificados



all none

- W** WDS Certified Repositories [42]
- D** DSA Certified Repositories [16]
- C** CTS Certified Repositories [111]



# Visão Geral

- Processo auto avaliativo com evidências comprovadas;
- Válida por 3 anos;
- Aplicação submetida em Inglês;
- 16 Requisitos que refletem as características desejáveis de um repositório confiável;
- Todos os requisitos são mandatórios e possuem pesos iguais no processo avaliativo;
- Algumas sobreposições e/ou informações duplicadas são inevitáveis;
- Terminologia utilizada baseada no Modelo OAIS;
- Princípios FAIR estão implícitos nos requisitos;

# Visão Geral

- A equipe gestora do repositório deve indicar um nível de conformidade para cada um dos **Requisitos**:

0 – Não se aplica;

1 – O repositório ainda não considerou este requisito;

2 – O repositório possui um conceito teórico sobre este requisito;

3 – O repositório está em fase de implementação deste requisito;

4 – Este requisito foi totalmente implementado no repositório

# Visão Geral

- Os níveis de conformidade são indicadores do processo auto avaliativo do repositório;
- A avaliação de conformidade declarada nas respostas utiliza evidências comprovadas (links de documentos formais, websites...)
- Exige-se justificativa detalhada para resposta "não se aplica" conferida a algum requisito;
- Níveis de conformidade na fase (1) ou (2) não são suficientes para a obtenção da certificação;
- A certificação pode ser concedida se alguns dos requisitos estiverem em fase de implementação (3) e o restante em fase (4).



# Requisitos CTS alinhados ao software Dataverse



## **Repositórios Dataverse com certificação CTS:**

- Tilburg University Dataverse (Holanda)
- Qualitative Data Repository (USA)
- DataverseNO (Noruega)
- University of North Carolina Repository (USA)

# Informações Básicas

## R-0: Contexto do Repositório

Clara definição e descrição do repositório:

- Tipo (institucional, consórcio, temático, periódico, governamental, projeto de pesquisa...);
- Comunidade(s) alvo;
- Política institucional para o repositório (divulgada) que envolva:
  - Missão, escopo, descrição de utilização;
  - Relação com outros repositórios da instituição;
  - Gestão e povoamento do repositório;
  - Licenças de acesso e uso;

## R-0: Contexto do Repositório

### Nível de curadoria praticado: (Escolher uma opção)

A) Conteúdo depositado é o mesmo distribuído;

- Equipe de suporte do repositório não realiza revisão do conteúdo depositado ;

B) Curadoria básica:

- Equipe de suporte do repositório revisa os datasets antes de serem publicados (inclusão de metadados, documentação...);

C) Curadoria ampliada:

- Conversão de formatos (arquivos de tabela no Dataverse);
- Aprimora a documentação;
- Instruções sobre depósitos alinhadas às melhores práticas da área de conhecimento;
- Templates com exigência de preenchimentos de campos;
- Revisão de datasets antes e depois de serem publicados;
- Implementam ações de melhorias na descrição dos dados;

D) Curadoria ao nível de dados:

- Além da curadoria ampliada, a equipe de suporte revisa os arquivos de dados;
- Além do download de dados para verificação, utilizam ferramentas externas ao Dataverse para revisão dos dados;



# Informações Básicas



## R-0: Contexto do Repositório

- **Parceiros/colaboradores internos e externos:**
  - Descrição do tipo de parceria:
    - armazenamento e segurança dos dados;
    - registro de identificadores persistentes (DataCite);
    - Harvard IQSS - desenvolvimento do Dataverse;
- Outras informações relevantes:
  - Nível de citações dos datasets do repositório;
  - Papel que exerce no contexto (local, regional, nacional);

## R-1: Missão/Esopo do Repositório

- **Missão do repositório; Deve ser divulgada e explicitar:**
  - Responsabilidade sobre os dados, garantia de ambiente digital apropriado, período de tempo, preservação e acesso continuado aos dados;
  - Nível hierárquico em que a missão do repositório foi aprovada;
  - Customização da webpage do repositório:
    - Dataverse (última versão): permite customizar termos de uso, cabeçalho, rodapé da página principal;

## R-2: Licenças do Repositório

- **Licenças de acesso e uso do repositório (divulgadas);**
  - Comunidade Dataverse: incentiva CC0;
  - Repositório pode ter termos de acesso e uso específicos (propriedade intelectual, proteção a dados sensíveis, intenções de uso...);
  - Depositantes: liberdade de escolha dentre as licenças disponíveis;
  - Dataverse: Gestbook e Access Request workflow;
  - Níveis de acessos diferenciados (arquivos, datasets);

## R-3: Continuidade de Acesso

- **Plano de continuidade de acesso (governança);**
  - Garantia de disponibilidade e acesso aos dados (presente e futuro);
  - Orçamento pré-definido;
  - Equipe gestora: níveis de responsabilidade;
  - Plano e Política de Preservação dos dados;
  - Gestão de riscos;
  - Resposta à mudanças bruscas (orçamento, mudanças de interesses institucionais...)

## R-4: Confidencialidade / Ética (alinhado ao R2)

- **Garantia de boas práticas quanto aos dados: criados, tratados, acessados e utilizados conforme condutas legais e éticas específicas;**
  - Depositantes podem restringir ou conceder acessos diferenciados (arquivos, datasets);
  - Privacidade dos sujeitos das pesquisas;
  - Manual sobre depósito e download responsável;
  - Dataverse:
    - recursos disponíveis:
      - Deaccessioning (versões de datasets);
      - Controle de Permissão de edição a datasets;
      - Controle de acesso e download a arquivos;

## R-5: Infraestrutura Organizacional

- **Instituição reconhecida (garantia de estabilidade e sustentabilidade do repositório);**
- **Orçamento adequado e pré-definido (recursos humanos, TI);**
- **Equipe qualificada que garanta o cumprimento da missão do repositório; (treinamento continuado)**
  - Dataverse: Comunidades de prática que incrementam a expertise da equipe;
  - Redes de repositórios, Consórcios;

## R-6: Orientações de especialistas (R0, R5)

- **Orçamento adequado e equipe qualificada que garanta o cumprimento da missão do repositório**
  - Dataverse: Comunidades de prática que incrementam a expertise da equipe;
  - Redes de Repositórios, Consórcios...
  - Recebe suporte ou consultoria externos;

## R-7: Integridade e Autenticidade dos Dados

- **Repositórios Dataverse:**

- Suporte à múltiplos locais de armazenamento (redundância de dados);
- Controle de versões (datasets e arquivos publicados);
- Checagem de integridade (MD5 e UNF p/ tabelas);
- Controle de permissões e workflow "submit for review";
- Login com contas autenticadas para controles de acesso, edição e publicação de datasets;

## R-8: Monitoramento e Avaliação (R11)

- **Repositório aceita dados e metadados baseados em critérios previamente definidos garantindo relevância e inteligibilidade aos usuários;**
- **Repositórios Dataverse:**
  - Workflows que contemplam versão de rascunho para revisão prévia; ("Submit for review" e URL privada);
  - Exigência de metadados bibliográficos mínimos garantindo citação correta;
  - Conjunto de metadados em conformidade com padrões internacionais e áreas específicas;
  - Equipe de suporte pode rejeitar ou remover dados que não estão de acordo com as políticas estabelecidas;

## R-9: Procedimentos de armazenamento documentados (R15, R16)

### **Armazenamento de arquivos em conformidade com modelo OAIS;**

- Gestão de processos de armazenamento documentada;
- Estratégias para múltiplos locais de armazenamento (redundância de dados);
- Checagens que garantam consistência das cópias redundantes;
- Gestão de riscos;
- Monitoramento de obsolescência e deterioração de mídia;

### Dataverse (últimas versões):

- Oferece estratégias de armazenamento redundante e recuperação de dados;
- Checagem de integridade de arquivos (APIs);
- Integração com Archivemática; BagIT (padronização de layout para armazenamento e transferência de arquivos); OAI-ORE (Open Archives Initiative Object Reuse and Exchange);

## R-10: Plano de Preservação (Ações implementadas) (R2)

- **O repositório deve definir e explicitar as ações de preservação e os níveis de responsabilidade pelos dados (repositório, depositantes e usuários);**
  - Prever migrações futuras;
  - Monitoramento de obsolescência e deterioração de mídia;
  - Estratégias claras de transferência de custódia (depositante e repositório);
  - Padrões de informações de submissão;
  - Padrões de informações de arquivamento;
  - Monitoramento e avaliação destas ações;

### **Dataverse (últimas versões):**

- Oferece estratégias de armazenamento redundante e recuperação de dados;
- Checagem de integridade de arquivos (APIs);
- Integração com Archivemática; BagIT (padronização de layout para armazenamento e transferência de arquivos); OAI-ORE (Open Archives Initiative Object Reuse and Exchange);

## R-11: Qualidade dos dados (R8)

- Controle de qualidade dos dados documentado;
  - Garantia de integridade e inteligibilidade dos dados (padrões);
  - Workflows de curadoria documentados;
  - Níveis de segurança documentados;
- Dataverse:
  - Conversão de formatos (arquivos de tabela no Dataverse);
  - Aprimora a documentação;
  - Instruções sobre depósitos alinhadas às melhores práticas da área de conhecimento;
  - Templates com exigência de preenchimentos de campos;
  - Revisão de datasets antes e depois de serem publicados;
  - Implementam ações de melhorias na descrição dos dados;
  - Suporte a padrões de metadados (documentação e proveniência) : DDI, DataCite, ISA-Tab, PROV-O);

## R-12: Workflows

- Workflows das ações bem definidos e documentados;  
(Ingestão - publicação - preservação)
  - Depositantes tornam explícitas informações sobre tratamento dos dados;
  - Níveis de segurança e impacto nas ações;
  - Monitoramento e avaliações qualitativas e quantitativas do repositório;
- Dataverse:
  - descrição de acordo com OAIS model;

pré-ingest - **ingest** - archival storage - data management - access - administration



Draft Version (dataset) = Submission Information Package (SIP) (not preserved)

Draft Version (dataset) → Dissemination Information Package (DIP)  
(Dataverse não utiliza Archival Information Package)

## R-13: Identificação e descoberta dos dados

- Repositório permite ao usuário: Descoberta e identificação persistente dos dados;
- Repositórios devem ser indexados em diretórios;
- Dataverse:
  - Recursos para otimização das buscas:
    - Busca por metadados (catálogos);
    - Palavras-chave, tags, tesouros, multifacetadas;
    - Identificadores persistentes;
  - Permitir inclusão automática de metadados (importação);
  - Utilizar templates e padrões para metadados (bibliográficos, área do conhecimento...).
    - Dublin Core, DDI, DataCite, OAI-PMH...

## R-14: Reuso dos dados

- Repositório permite o reuso dos dados garantindo disponibilidade dos metadados dando suporte à inteligibilidade dos dados;
- Necessário conhecimento aprofundado da(s) comunidade(s) alvo em termos de: práticas, ambiente técnico e tecnológico, padrões.
- Dataverse:
  - Conversão de formatos de dados (tabelas);
  - Padrão: Datasets devem conter um conjunto mínimo de metadados visando garantir correta citação, localização do depositante...

## R-15: Infraestrutura Técnica (R3, R9 e R16)

- Repositório está hospedado em uma infraestrutura adequada de hardware e software;
  - Plano de desenvolvimento de infraestrutura;
  - Inventário e documentação de software;
  - Conectividade com boa disponibilidade de banda;
  - Plano de continuidade de negócio (procedimentos definidos para recuperação de hardware ou backup de serviços essenciais);
- Dataverse:
  - Suporte a tecnologias open source;



# Tecnologia



## R-16: Segurança (R9, R15)

- Repositório deve ter um sistema de segurança robusto;
  - Análise de riscos (DRAMBORA);
  - Especificação de níveis de segurança suportados;
  - Procedimentos de autenticação de acesso ao sistema;
- Dataverse:
  - Suporte a HTTPS, proxies;
  - protocolos de autenticação: OAuth, OpenID connect;
  - requisitos de senha forte;
-

# Considerações finais

Padronização, auditoria e certificação são evoluções naturais em direção à um processo de maturidade de produtos e serviços;

A certificação também fornece uma rotulagem clara daquilo que confere confiança ao ambiente onde o ciclo de vida dos dados (de pesquisa) faz parte;

# Considerações finais

O processo de certificação Core Trust Seal deve sempre apresentar uma abordagem flexível e responsiva voltada à comunidade e adaptável à mudanças;

Isto se faz necessário para que se adapte às necessidades da comunidade de dados de pesquisa e seus parceiros colaborativos em rápida evolução;

# Considerações finais

A certificação Core Trust Seal busca de maneira geral garantia de preservação de dados independente das fronteiras disciplinares;

A noção de confiança é crítica em todo o ciclo de vida dos dados;  
Repositórios ou arquivos historicamente se definiram como uma parte distinta do ciclo de vida dos dados;

Cada vez mais, padrões e melhores práticas de repositórios têm sido adotados em orientações de gestão de dados de pesquisa.

# Considerações finais

- Brasil: Repositórios institucionais de produção científica já existentes;
- Fase inicial: Implantações de repositórios de dados científicos considerando:
  - aspectos de confiabilidade;
  - certificações internacionais - CTS evitando retrabalho para futuras certificações;

# Referências:

CoreTrustSeal Standards and Certification Board. (2019, November 20).  
CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance  
2020–2022 (Version v02.00-2020-2022). Zenodo.  
<http://doi.org/10.5281/zenodo.3632533>

Dataverse Project Curation Team. Dataverse Software Guide for Core Trust Seal.  
2021. <https://dataverse.org/cts-guide>



**OBRIGADA!!!**

**[lauravil.rr@gmail.com](mailto:lauravil.rr@gmail.com)**

**[laura\\_rezende@ufg.br](mailto:laura_rezende@ufg.br)**