



# **Tutorial para Instalação de Repositório de Dados Científicos**

## **DataVerse**

Tutorial elaborado para orientar a instalação e configuração mínima de Repositórios de Dados Científicos e possibilitar sua integração com o metabuscador desenvolvido para atender à Política de Gestão de Dados da FAPESP ([www.fapesp.br/gestaodedados](http://www.fapesp.br/gestaodedados)).

### **Coordenadores:**

Profa. Dra. Fátima de Lourdes dos Santos Nunes  
Prof. Dr. João Eduardo Ferreira

### **Equipe Técnica**

Diego Lopes de Araujo  
Edmar Martineli  
Marino Hilário Catarino  
Mauro Cesar Bernardes  
Rodrigo Muller de Carvalho

**São Paulo, julho de 2018**

## 1 – Dataverse

### 1.1. Apresentação

O Dataverse é um grande repositório aberto de dados de todas as disciplinas, criado pelo Institute for Quantitative Ciências Sociais da Universidade de Harvard, disponível em <https://dataverse.org/>.

### 1.2. Versão utilizada

Este tutorial foi desenvolvido com base na versão Dataverse 4.8.3, utilizando o Ubuntu versão 14.04 e disponível em <http://dataverse.uspdigital.usp.br>.

### 1.3. Passos para instalação

Os passos para a instalação foram os apresentados pela página oficial do Dataverse: <http://guides.dataverse.org/en/latest/installation/>.

#### 1.3.1. Permissões de usuário no Dataverse:

As configurações de permissão que os usuários possuem no repositório Dataverse podem ser alteradas com a conta do administrador através da opção “Edit” (figura 1), logo após ter efetuado login no repositório.

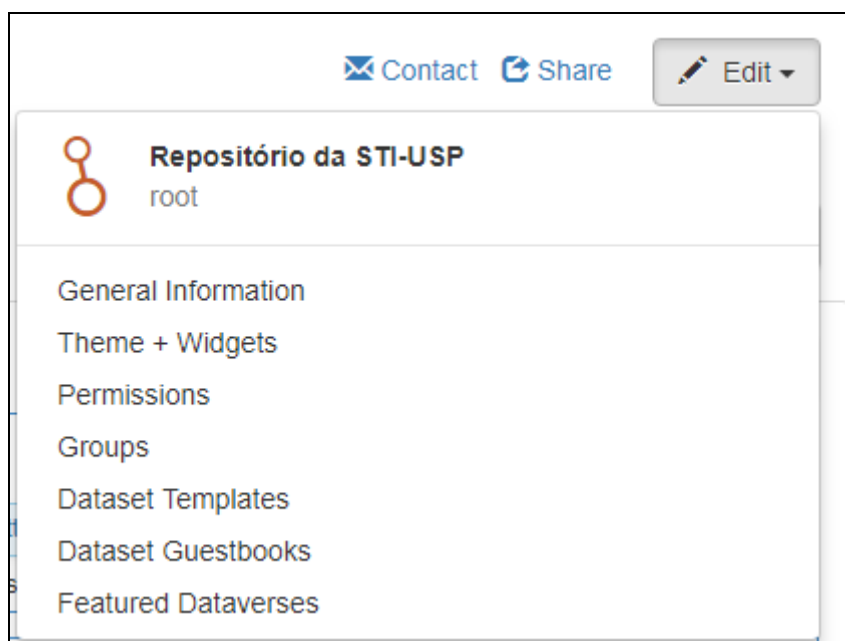


Figura 1: Tela de edição do Dataverse.

Ao seleccionar “Permissions” obtem-se a interface apresentada na figura 2:



Figura 2: Tela de permissões dos usuários no dataverse.

A partir da tela apresentada na figura 2 é possível gerenciar as regras para liberação dos usuários:

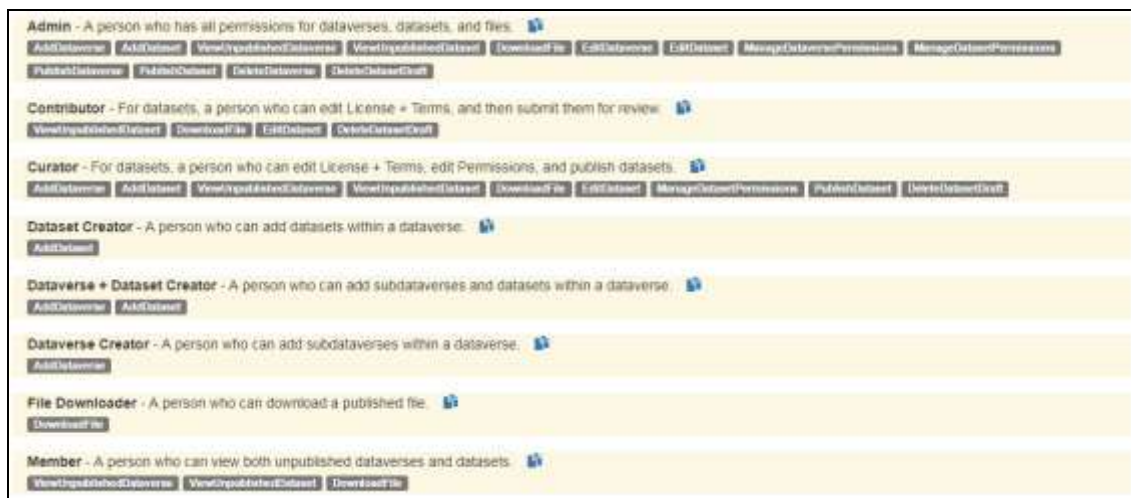


Figura 3: Regras de acesso para os usuários do Dataverse.

Pode-se definir regras para um usuário ou grupo específico através da opção “User/Groups”, conforme apresentado na figura 4:



Figura 4: tela para definir regras para um usuário ou grupo..

Para definir as opções de forma automática para todos os usuários que se cadastrarem no repositório Dataverse é preciso ir na opção “Permissions” e clicar em “Edit Access”. Nesta tela é possível determinar quais regras serão aplicadas para os usuários que se cadastrarem no repositório (figura 5).

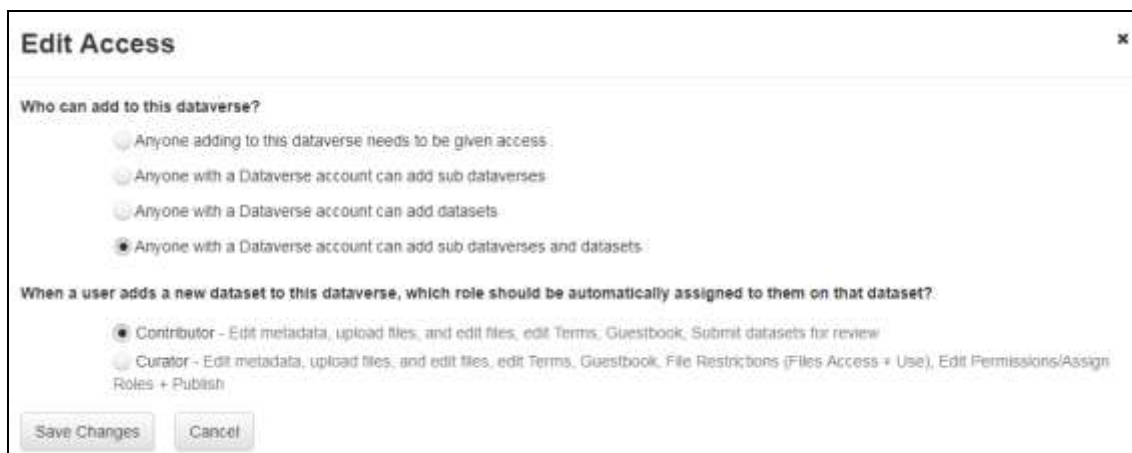


Figura 5: Tela para definir as permissões dos usuários no repositório.

No repositório configurado a regra definida para os usuários é a permissão de adicionar novos *dataverses* e *datasets*.

### 1.3.2. Habilitar os metadados solicitados pelo Consórcio:

Ao criar um novo Dataverse pode definir quais metadados deseja habilitar para que os usuários preencham quando forem inserir um novo dataset. Na tela de criação do novo Dataverse temos as opções em “Metadata Fields” (figura 6):

Dataverse Admin Dataverse (Dataverse.org) Unpublished

Repositório da STI-USP > **New Dataverse**

**Dataverse \***  **Affiliation**

**Identifier \***  **Host Dataverse**

**Category \***  **Description**

**Email \***

---

**Metadata Fields** Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse.

- Use metadata fields from Repositório da STI-USP
- Citation Metadata (Required) [\[+\] View fields](#)
- Geospatial Metadata [\[+\] View fields](#)
- Social Science and Humanities Metadata [\[+\] View fields](#)
- Astronomy and Astrophysics Metadata [\[+\] View fields](#)
- Life Sciences Metadata [\[+\] View fields](#)
- Journal Metadata [\[+\] View fields](#)

Figura 6: Tela para habilitar metadados no Dataverse.

Na tabela a seguir é apresentado os campos habilitados:

| Citation Metadata   |                   |
|---------------------|-------------------|
| Title               |                   |
| Subtitle            |                   |
| Alternative Title   |                   |
| Alternative URL     |                   |
| Author              | Name              |
|                     | Affiliation       |
|                     | Identifier Scheme |
|                     | Identifier        |
| Contact             | Name              |
|                     | Affiliation       |
|                     | E-mail            |
| Description         | Text              |
|                     | Date              |
| Subject             |                   |
| Keyword             | Term              |
|                     | Vocabulary        |
|                     | Vocabulary URL    |
| Related Publication | Citation          |
|                     | ID Type           |
|                     | ID Number         |
|                     | URL               |
| Language            |                   |
| Production Date     |                   |

|                            |                  |
|----------------------------|------------------|
| Grant Information          | Grant Agency     |
|                            | Grant Number     |
| Distribution Date          |                  |
| Deposit Date               |                  |
| Time Period Covered        | Start            |
|                            | End              |
| Date of Collection         | Start            |
|                            | End              |
| Kind of Data               |                  |
| <b>Geospatial Metadata</b> |                  |
| Geographic Coverage        | Country / Nation |
|                            | State / Province |
|                            | City             |
|                            | Other            |
| Geographic Bounding Box    | West Longitude   |
|                            | East Longitude   |
|                            | North Latitude   |
|                            | South Latitude   |

### 1.3.3. Extração de dados pelo Metabuscador USP

O método que o Metabuscador utiliza para obter os dados de cada repositório Dataverse é o formato completo de exportação JSON. O material oficial pode ser acessado através do link: <http://guides.dataverse.org/en/latest/api/native-api.html>.

A função de exportação dos dados pelo JSON é nativa no Dataverse, não é necessário fazer nenhuma configuração adicional. Para verificar se o JSON está funcionando corretamente é preciso adicionar um conjunto de dados no seu repositório e ir no endereço do navegador que irá listar todos os conjuntos de dados já formatados:

[http://<endereço-repositório-Dataverse>/oai?verb=ListRecords&set=<nome-dataverse>&metadataPrefix=dataverse\\_json](http://<endereço-repositório-Dataverse>/oai?verb=ListRecords&set=<nome-dataverse>&metadataPrefix=dataverse_json)

## 1.4 – Protocolo OAI-PMH

O Dataverse permite extrair os dados através do protocolo OAI-PMH porém esta opção possibilita a extração de menos informações para o metabuscador, por isto foi adotada a extração completa via JSON.

### 1.5 - Adição do termo de uso do repositório Dataverse:

O registro de novos usuários no Dataverse é feito de forma livre, e a tela de registro foi alterada de forma a incluir o termo de uso do repositório.

Para se inserir o termo de uso em pdf na tela de registro foi preciso:

O resultado final na tela de registro de usuário se encontra na figura 7:



The screenshot shows a registration form with the following fields: 'Affiliation', 'Position', and 'General Terms of Use'. The 'General Terms of Use' section contains the following text:

Este Repositório de Dados Científicos da Universidade de São Paulo é um experimento coordenado pela Superintendência de Tecnologia da Informação (STI) desta instituição, criado com o objetivo de avaliar a plataforma tecnológica em que se encontra. Pelo fato de estar em fase de experimentação, mesmo permitindo o livre acesso a todo o conteúdo, a STI não se responsabiliza pela preservação e integridade dos arquivos inseridos neste repositório.

Ao criar uma conta neste repositório e utilizá-la para adicionar arquivos, o usuário está ciente e aceita que:

1. concede direito não exclusivo à Universidade de São Paulo de armazenar (eventualmente em mais de uma cópia para fins de segurança, acesso e preservação) gratuitamente em seu Repositório de Dados Científicos e demais recursos que:

I have read and accept the Dataverse General Terms of Use as outlined above:

Figura 7: Tela de registro de usuários no Dataverse com termo de uso.

### 1.6 – Cadastro de um novo conjunto de dados:

Todas as informações definidas como obrigatórias encontram-se disponíveis para preenchimento no Dataverse:

| Identificador | Metadado                     | Dataverse |
|---------------|------------------------------|-----------|
| 1             | dc.title                     | Sim       |
| 2             | dc.Subject                   | Sim       |
| 3             | dc.description               | Sim       |
| 4             | dc.contributor.author        | Sim       |
| 5             | dc.identifier.uri            | Sim       |
| 6             | dc.description.sponsorship   | Sim       |
| 7             | dc.description.sponsorshipId | Sim       |

Quanto às informações opcionais, são as seguintes:

| Identificador | Tipo                | Dataverse |
|---------------|---------------------|-----------|
| A             | dc.rights           |           |
| B             | dc.coverage         |           |
| C             | dc.temporal         |           |
| D             | dc.date             | Sim       |
| E             | dc.date.accessioned | Sim       |
| F             | dc.date.available   | Sim       |
| G             | dc.date.issued      | Sim       |

|          |                          |     |
|----------|--------------------------|-----|
| <b>H</b> | dc.identifier.citation   | Sim |
| <b>I</b> | dc.format                | Sim |
| <b>J</b> | dc.publisher             | Sim |
| <b>K</b> | dc.rights.holder         |     |
| <b>L</b> | dc.type                  | Sim |
| <b>M</b> | dc.coverage.spatial      | Sim |
| <b>N</b> | dc.coverage.spatial.Lat  | Sim |
| <b>O</b> | dc.coverage.spatial.Long | Sim |
| <b>P</b> | dc.coverage.temporal     | Sim |
| <b>Q</b> | dc.language              | Sim |

Os campos que devem ser preenchidos no Dataverse para atender os itens solicitados pelo consórcio, considerando o identificador informado nas tabelas acima para indicar cada campo, são mostrados nas figuras 8, 9 e 10:

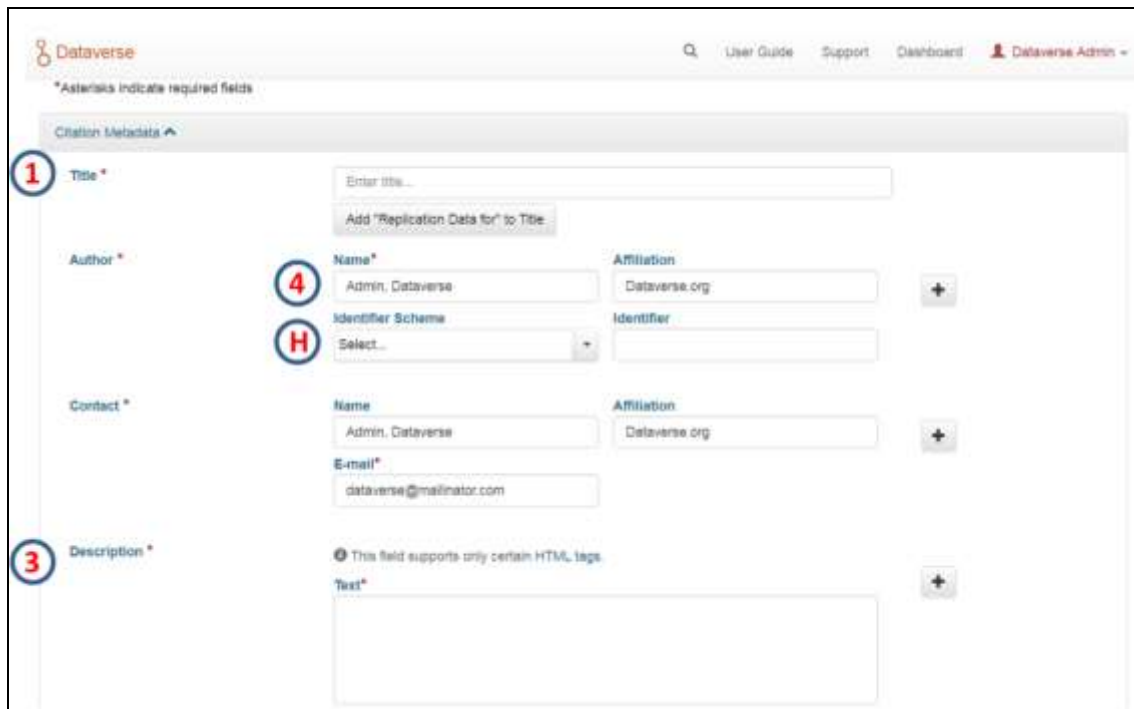


Figura 8: Informações para cadastro de conjunto de dados



The screenshot shows a registration form with several fields and sections:

- Date:** A text input field with a placeholder 'YYYY-MM-DD'. It is marked with a red circle 'D' and a blue circle 'G'.
- Subject:** A dropdown menu with a red circle '2' next to it. The options include Agricultural Sciences, Arts and Humanities, Astronomy and Astrophysics, Business and Management, Chemistry, and Computer and Information Science.
- Keyword:** A section containing a 'Term' input field, a 'Vocabulary' input field, and a '+' button. Below it is a 'Vocabulary URL' input field with the instruction 'Enter full URL, starting with http://'.
- Related Publication:** A section containing a 'Citation' input field and a '+' button. It is marked with a red circle 'H'.
- ID Type:** A dropdown menu with a red circle 'I' next to it, currently showing 'Select...'. It is followed by an 'ID Number' input field.
- URL:** An input field with a red circle '5' next to it and the instruction 'Enter full URL, starting with http://'.

Figura 9: Informações para cadastro de conjunto de dados

The screenshot shows a metadata form with the following sections:

- Notes:** A large text area for entering notes.
- Depositor:** An input field containing 'Admin, Dataverse'. It is marked with a red circle 'J'.
- Deposit Date:** An input field containing '2017-11-24'. It is marked with a red circle 'E'.
- Files:** A section with a 'Files' tab. It contains a link: 'For more information about supported file formats, please refer to the User Guide.' Below this is a button 'Select Files to Add' and a large area with the text 'Drag and drop files here.' At the bottom of this section, it says 'Files you upload will appear here'.

Figura 10: Campos para preenchimento dos metadados.

O preenchimento dos demais metadados é feito após finalizar o cadastro do dataset. Na tela apresentando o dado inserido temos a opção de menu “Metadata”, conforme mostrado na figura 11:

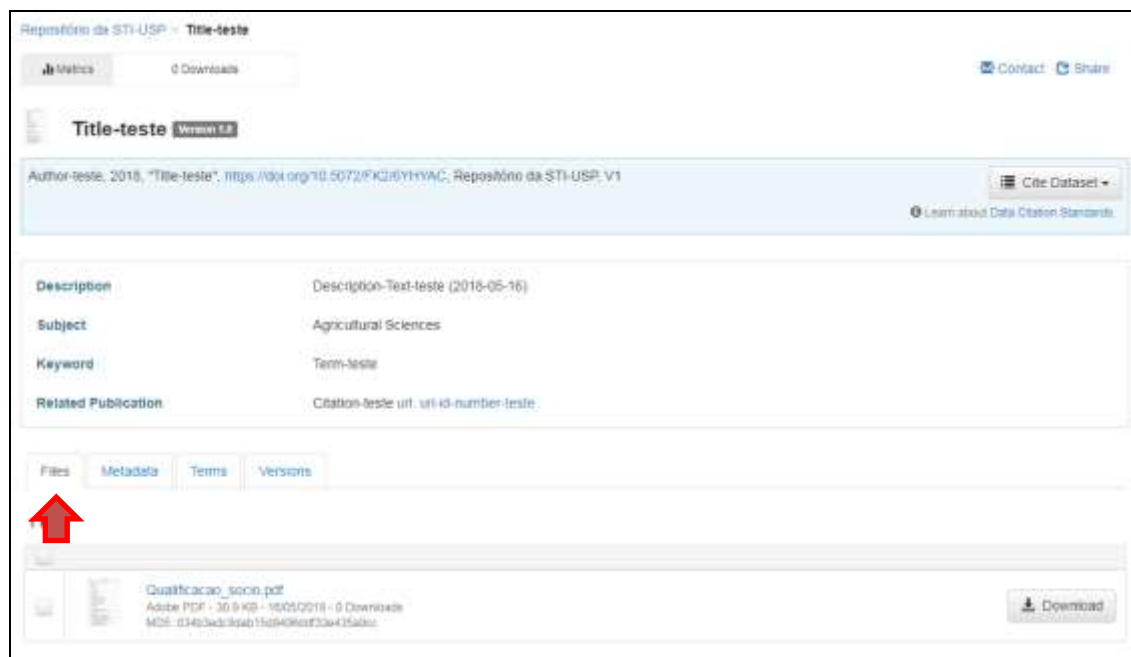


Figura 11: Campos para preenchimento dos metadados.

Ao clicar na opção “Metadata” deve habilitar a edição e o preenchimento das informações clicando em “Add + Edit Metadata” afim de preencher os seguintes campos para que todos os metadados sejam informados:




Figura 12: Opção para preencher complemento dos metadados.

O item dc.coverage.spatial é a agregação dos campos geográficos nesta ordem: City, State, Country; North Latitude e West Longitude, figura 13:

Figura 13: Opção para preencher complemento dos metadados.

### 1.7. Coordenador do Dataverse

No Dataverse, conforme a configuração de permissões, somente o administrador pode adicionar um Dataverse, fazendo com que todos os dataset fiquem vinculados a este dataverse. O administrador também pode tornar um membro administrador do dataverse, tornando-o o coordenador do repositório.

O coordenador também pode tornar público ou não um dataset adicionado por um usuário participante do repositório, desta forma atuando como mediador do repositório. Para que esta ação aconteça é preciso que o usuário envie o dataset para revisão do coordenador/administrador do dataverse, conforme a figura 14 apresenta:



Figura

14: Solicitando revisão do dataset para publicação.

O coordenador possui as seguintes funções:

- publicar o dataset enviado para revisão;
- retornar o dataset ao autor para modificação;
- tornar outro membro administrador/coordenador do dataverse.

Na figura 15 é apresentada a tela do coordenador com as opções referentes ao dataset:



Figura 14: Ações permitidas ao coordenador.

### 1.8. Configurar Harvesting

É necessário habilitar o harvesting na instalação do Dataverse. Após feito a habilitação é necessário configurar os conjuntos de datasets para os clientes. A documentação dessa etapa foi feita seguindo a documentação do Dataverse:

<http://guides.dataverse.org/en/latest/admin/harvestserver.html>.

Especificamente, para habilitar todos os datasets para harvesting, foi configurada a opção `dsPersistentId` com o endereço de DOI desse repositório.