

# ACESSO ABERTO A DADOS DE PESQUISA NO BRASIL

Planejamento para Implantação de Comunidade Produtora de  
Dados para o Repositório Rede de Dados de Pesquisa – RDP  
Brasil de Dados



Controle de Versionamento:

Versão	Autores/ Revisores	Data
1.0	<b>Universidade Federal do Rio Grande do Sul (UFRGS)</b> <b>Universidade Federal de Rio Grande (FURG)</b>	12/08/2019
1.1	<b>Rede Nacional de Ensino e Pesquisa (RNP)</b>	13/08/2019
1.2	<b>Universidade Federal do Rio Grande do Sul (UFRGS)</b>	13/08/2019
1.3	<b>Universidade Federal do Rio Grande do Sul (UFRGS)</b>	08/09/2019
1.3.1	<b>Universidade Federal do Rio Grande do Sul (UFRGS)</b>	31/03/2020

**Coordenador**

Rafael Port da Rocha (Universidade Federal do Rio Grande do Sul)

**Coordenadora adjunta**

Sônia Elisa Caregnato (Universidade Federal do Rio Grande do Sul)

**Pesquisadores (Universidade Federal do Rio Grande do Sul)**

Ana Maria Mielniczuk de Moura

Caterina Groposo Pavão

Paula Caroline Schifino Jardim Passos

Rene Faustino Gabriel Junior

Samile Andréa de Souza Vanz

**Pesquisadores (Universidade Federal do Rio Grande)**

Eduardo Nunes Borges

Luís Alberto Barbosa Azambuja

**Contato**

e-mail: [dadosdepesquisa@rnp.br](mailto:dadosdepesquisa@rnp.br) - site: <https://dadosdepesquisa.rnp.br>

Rua Ramiro Barcelos, 2777 - Campus Saúde Sala: 106 - Anexo 1 - Brasil - Porto Alegre - RS - CEP 90.035-007 - Telefone: +55(51)3308.5942



Como citar: ROCHA, Rafael Port da; BORGES, Eduardo Nunes; AZAMBUJA, Luís Alberto Barbosa; GABRIEL JUNIOR, Rene Faustino; MOURA, Ana Maria Mielniczuk; CAREGNATO, Sônia Elisa; PASSOS, Paula Caroline Schifino Jardim; PAVÃO, Caterina Groposo; VANZ, Samile Andrea de Souza. Acesso aberto a dados de pesquisa no Brasil: planejamento para implantação de comunidade produtora de dados para o repositório Rede de Dados de Pesquisa. Disponível em: <<http://hdl.handle.net/20.500.11959/1204>>

## **Acesso aberto a dados de pesquisa no Brasil: planejamento para implantação de comunidade produtora de dados para o repositório Rede de Dados de Pesquisa.**

### **Resumo**

Este documento apresenta o Planejamento para implantação de uma Comunidade Produtora de Dados para o Repositório Rede de Dados da Pesquisa - RDP Brasil. Este documento e o repositório piloto, Rede de Dados de Pesquisa são resultados da atuação do Grupo de Trabalho Rede de Dados de Pesquisa (GT-RDP), grupo selecionado pela Rede Nacional de Ensino e Pesquisa (RNP), em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), no atendimento à carta-convite RNP/IBICT de 01/2017, que busca a composição de um Grupo de Trabalho para atuar em parceria na identificação de práticas de Acesso Aberto a Dados de Pesquisa (AADP) no Brasil e no desenvolvimento de experimentos tecnológicos visando um futuro serviço de AADP.

### **Palavras-chave:**

Dados abertos, dados de pesquisa, comunidade produtora de dados, repositório, Dataverse.

## SUMÁRIO

<b>Acesso aberto a dados de pesquisa no Brasil: planejamento para implantação de comunidade produtora de dados para o repositório Rede de Dados de Pesquisa.</b>	<b>4</b>
<b>1 Introdução</b>	<b>6</b>
<b>2 Comunidade Produtora de Dados</b>	<b>8</b>
<b>3 Planejamento da Comunidade Produtora de Dados</b>	<b>12</b>
3.1 Os Produtores e os Consumidores dos dados	12
3.2 Os Gestores da Comunidade Produtora de Dados	13
3.3 O Conjunto de Dados	13
3.4 O Uso do Conjunto de Dados	14
3.5 A Representação do Conjunto de Dados	15
3.6 A Documentação do Conjunto de Dados	16
3.7 A Descrição do Conjunto de Dados	16
3.8 A Proveniência do Conjunto de Dados	23
<b>4 Criação da Comunidade no Dataverse</b>	<b>25</b>
4.1 Dataverse	25
4.2 Definição e Criação da Comunidade Produtora de Dados no Repositório	27
4.3 Definição e Criação dos Grupo de Usuários	28
4.4 Criação e Descrição de Conjunto de Dados	30
4.5 Definição das Termos de Uso e Condições de Acesso ao Conjunto de Dados	30
4.6 Criação de Arquivos de Conjunto de Dados	33
4.7 Publicação de Conjunto de Dados	34
<b>Referências/Bibliografia</b>	<b>35</b>
<b>Anexo</b>	<b>36</b>

# 1 Introdução

Este documento apresenta o **Planejamento para implantação de uma Comunidade Produtora de Dados para o Repositório Rede de Dados da Pesquisa - RDP Brasil**. Este documento e o repositório piloto, **Rede de Dados de Pesquisa**<sup>1</sup> são resultados da atuação do Grupo de Trabalho Rede de Dados de Pesquisa (GT-RDP), grupo selecionado pela Rede Nacional de Ensino e Pesquisa (RNP), em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), no atendimento à carta-convite RNP/IBICT de 01/2017, que busca a composição de um Grupo de Trabalho para atuar em parceria na identificação de práticas de Acesso Aberto a Dados de Pesquisa (AADP) no Brasil e no desenvolvimento de experimentos tecnológicos visando um futuro serviço de AADP.

O Planejamento para implantação de uma comunidade produtora de dados do Repositório foi elaborado com base no modelo de referência para Repositório, OAIIS (*Open Archival Information System*), e na metodologia ligada à OAIIS para identificar a interface entre o produtor e o repositório:

- CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. Producer-Archive Interface Methodology Abstract Standard. CCSDS 651.0-M-1. 2004 <https://public.ccsds.org/Pubs/651x0m1.pdf>
- CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. Reference model for an open archival information system (OAIIS). CSDS 650.0-M-2. 2012 <https://public.ccsds.org/pubs/650x0m2.pdf>

OAIIS é um modelo de referência que proporciona uma linguagem comum para facilitar a discussão entre comunidades interessadas no desenvolvimento de repositórios com responsabilidade em preservação digital. OAIIS descreve o **ambiente** em que um repositório reside, os seus **componentes funcionais** e a infraestrutura de **informação** que suporta o processo de arquivamento

O Planejamento para implantação de uma comunidade produtora de dados também considera os requisitos FAIR [v, viii] e critérios para obtenção de Certificação de Repositório Digital Confiável de *Core Trust Seal* [viii].

Repositório Digital Confiável é aquele cuja missão é fornecer acesso de longo prazo a recursos digitais gerenciados; que aceita a responsabilidade pela manutenção a longo prazo dos recursos digitais em nome de seus depositantes e em benefício dos usuários atuais e futuros; que projeta seu(s) sistema(s) de acordo com as convenções e os padrões comumente aceitos para garantir o gerenciamento, acesso e segurança contínuos dos materiais depositados nele; que estabelece metodologias para avaliação de sistemas que atendem às expectativas de confiabilidade da comunidade; que cumpre suas responsabilidades com depositantes e usuários de forma aberta e explícita; e cujas políticas, práticas e desempenho podem ser auditados e medidos [vi]

---

<sup>1</sup> Repositório Rede de de Dados de Pesquisa - <https://dadosabertos.rnp.br/>

Os princípios FAIR [v vii] indicam que dados devem ser localizáveis (*Findable*), acessíveis (*Accessible*), interoperáveis (*Interoperable*) e Reusáveis (*Reusable*). Para dados serem qualificados como 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, gratuito (A). Também devem ser acompanhados de licenças claras e acessíveis (R), e referências qualificadas devem ligar (meta)dados para enriquecer o conhecimento sobre os mesmos (I).

## 2 Comunidade Produtora de Dados

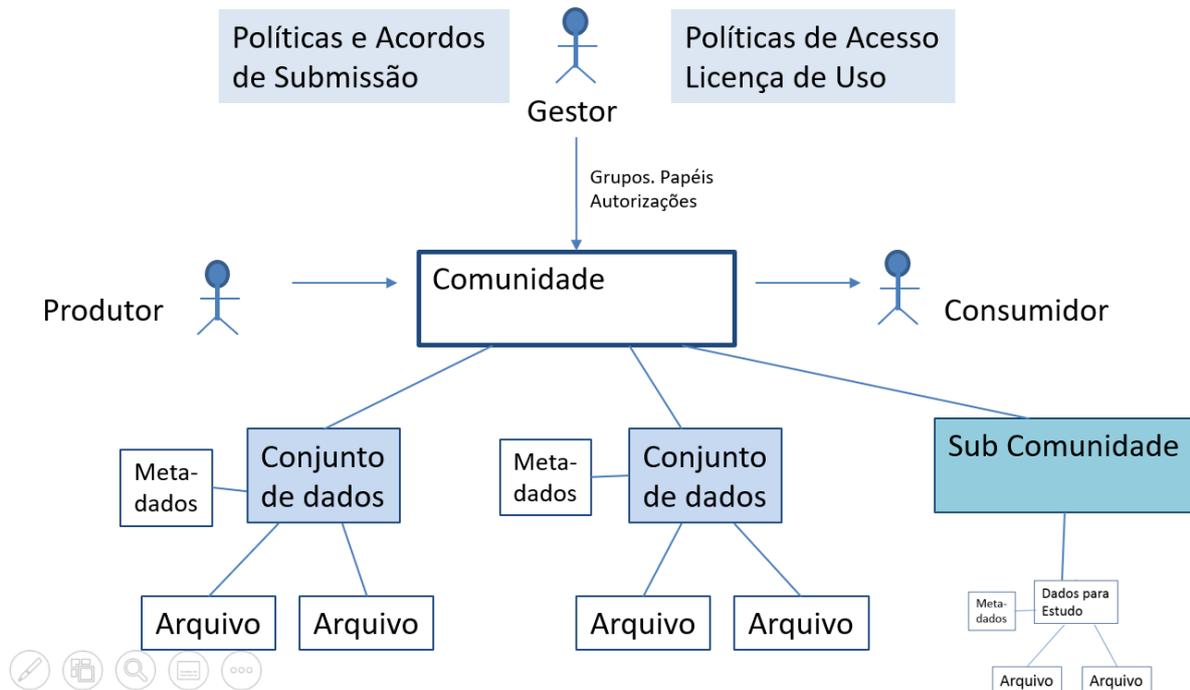
Considera-se uma **Comunidade Produtora de Dados** aquela que é capaz de:

- Atuar em uma temática de pesquisa identificada e caracterizada, cujos dados de pesquisa são de interesse para o compartilhamento e são compreendidos e capazes de serem utilizados por uma comunidade identificada de potenciais usuários.
- Produzir conjuntos de dados para fins de compartilhamento, nas seguintes condições:
  - Ter dados de interesse de compartilhamento, gerados pelas suas pesquisas e/ou dados extraídos de dispositivos, fornecidos por terceiros, reutilizados a partir de ações de compartilhamento, etc.
  - Ter domínio sobre a proveniência dos dados, entendendo como proveniência: “informações sobre entidades, atividades e pessoas envolvidas na produção de um dado ou coisa, que podem ser usadas para formar avaliações sobre sua qualidade, confiabilidade ou fidedignidade” [xi]
  - Ter direitos para conceder licenças e termos de uso dos dados.
  - Ser capaz de preparar os dados para o compartilhamento.
  - Dispor de unidade administrativa capaz de definir políticas e gerenciar a comunidade produtora de dados em repositório de dados.

Os principais componentes de uma **Comunidade Produtora de Dados** são (figura 1):

- **Produtor:** Fornece os dados a serem compartilhados. O produtor possui domínio sobre a proveniência dos dados, tem a responsabilidade pelos processos em que os dados foram trazidos à existência, possui direitos para conceder licenças e termos de uso sobre os dados, possui a responsabilidade em fornecer os dados já preparados para o compartilhamento e submete ao repositório os dados para compartilhamento.
- **Autor:** Cria os dados que serão compartilhados pela comunidade.
- **Consumidor:** Clientes que interagem com os serviços do repositório para descobrir e usar os dados produzidos pela comunidade, de acordo com as condições e termos de uso estabelecidos. São capazes de compreender e usar os dados.
- **Gestor:** Responsável pela gestão dos conjuntos de dados da comunidade, no repositório, isto é, gerir a produção e consumo dos dados da comunidade.
- **Conjunto de Dados:** Arquivos de dados descritos por uma única estrutura. Os conjuntos de dados podem ser decorrentes de **estudos**, ou de outras formas de produção de dados para fins de compartilhamento. Um **estudo** compreende na coleção de arquivos de dados resultantes da coleta intencional de dados por meio de solicitação, observação ou coleta de fontes secundárias para um propósito; utilizando instrumentos e metodologias de coleta descritos; e expressos como arquivos de dados com estruturas lógicas relacionadas ao instrumento de coleta de dados [xiv].
- **Metadados:** Dados estruturados que provêm informação sobre um ou mais aspectos de outros dados

Figura 1 – Comunidade Produtora de Dados



A implantação de uma **Comunidade Produtora de Dados** é composta pelas seguintes etapas:

- **Caracterização e Planejamento da Comunidade Produtora de Dados .**
- **Criação e Configuração da Comunidade Produtora de Dados em repositório.**
- **Operação da Comunidade Produtora de Dados.**

A **Caracterização e Planejamento da Comunidade Produtora de Dados** tem como objetivo levantar e identificar as características da comunidade, planejar e documentar a implantação dessa comunidade, a fim de servir como entrada para a criação e configuração da comunidade em repositório. Envolve identificar e caracterizar a comunidade, além de identificar e definir:

- Os produtores dos conjuntos de dados da comunidade
- Os consumidores dos conjuntos de dado da comunidade
- Os gestores da comunidade
- Um conjunto de dados da comunidade
- Uso de um conjunto de dados da comunidade
- A representação de um conjunto de dados da comunidade
- A descrição de um conjunto de dados da comunidade
- A documentação de um conjunto de dados da comunidade
- A proveniência de um conjunto de dados da comunidade

A **Criação e Configuração da Comunidade Produtora de Dados** desenvolve a comunidade no repositório de dados, através da criação e configuração de recursos e funcionalidades

para esse fim, disponibilizadas pela plataforma de software do repositório. O Quadro 1 apresenta as principais atividades a serem desenvolvidas nas etapas de planejamento e criação da comunidade, e suas relações.

Quadro 1 - Atividades de Planejamento e Criação da Comunidade Produtora de Dados

Recursos		Atividades de Planejamento	Atividades de Criação em Repositório
Comunidade Produtora de Dados		Identificar e caracterizar a comunidade.	Criação da entidade Dataverse
		Identificar e definir os produtores dos conjuntos de dados da comunidade.	Criação de Grupos responsáveis pela produção dos conjuntos de dados.
		Identificar e definir os consumidores dos conjuntos de dado da comunidade.	Criação de Grupos de Usuários (em caso de restrições de acesso)
		Identificar e definir os gestores dos conjuntos de dados.	Criação de Grupos responsáveis pela gestão da comunidade
Conjunto de Dados	Identificação	Identificar e definir o conjunto de dados.	Criar um Dataset, com o uso dos instrumentos de identificação e citação disponíveis pelo repositório
	Uso do conjunto de dados	Definir licenças, termos e condições de uso. Definir as restrições de acesso.	Configurar no módulo termos do Dataset as licenças, termos e condições de uso. Criar livros de visitas através do módulo Guestbook Definir restrições de acesso a arquivos.
	Representação do Conjuntos de Dados	Definir as estruturas do conjunto de dados. Definir as estruturas, os nomes e os formatos dos arquivos. Definir regras para o versionamento dos conjuntos de dados.	Inserir arquivos no <i>dataset</i> de acordo com as estruturas, nomes e formatos estabelecidos.

	Descrição dos conjuntos de dados	Definir os padrões para metadados descritivos. Definir regras preenchimento dos metadados .	Configurar os esquemas de metadados a serem usados no <i>dataset</i> . Produzir metadados descritivos a partir das regras estabelecidas.
	Documentação do Conjunto de Dados	Definir a Documentação dos Dados.	
	Proveniência de um conjunto de dados	Definir os depositantes dos dados. Verificar as autorizações legais necessárias para a publicação e uso dos dados. Definir o processo de submissão dos dados. Definir processo de preparação dos dados.	Configurar no <i>dataverse</i> o processo de submissão de arquivos do <i>dataset</i> através da configuração de grupos e papéis responsáveis pela carga, publicação dos arquivos.

# 3 Planejamento da Comunidade Produtora de Dados

## 3.1 Os Produtores e os Consumidores dos dados

Definir os produtores e os consumidores dos conjuntos de dados a serem compartilhados pela comunidade.

### **Produtores dos Conjuntos de Dados da Comunidade**

Pessoas que fornecem os dados a serem compartilhados. O produtor possui domínio sobre a proveniência dos dados, tem a responsabilidade pelos processos pelos quais os dados foram trazidos à existência, possui direitos para conceder licenças e termos de uso sobre os dados, possui a responsabilidade em fornecer os dados já preparados para o compartilhamento e submete ao repositório os dados para compartilhamento. Quem irá produzir os conjuntos de dados a serem compartilhados pela comunidade?

### **Consumidores dos Conjuntos de Dados da Comunidade**

Pessoas (ou sistemas) clientes que interagem com os serviços do repositório para descobrir e usar os dados produzidos pela comunidade de acordo com as condições e termos de uso estabelecidos. São capazes de compreender e usar os dados. Quais são os potenciais usuários dos conjuntos de dados da comunidade?

## 3.2 Os Gestores da Comunidade Produtora de Dados

Definir os responsáveis pela gestão da comunidade produtora de dados

### **Gestores da Comunidade de Compartilhamento de Dados**

Responsáveis pela gestão dos conjuntos de dados da comunidade no repositório.  
Quem será responsável por gerir a comunidade produtora de dados?

## 3.3 O Conjunto de Dados

Identificação de um conjunto de dados compartilhado pela comunidade.

### **Conjunto de Dados**

Podem ser decorrentes de **estudos**, ou de outras formas de produção de dados para fins de compartilhamento. Segundo DDI, um **estudo** compreende na coleção de arquivos de dados resultantes da coleta intencional de dados por meio de solicitação, observação ou coleta de fontes secundárias para um propósito; utilizando instrumentos e metodologias de coleta descritos; e expressos como arquivos de dados com estruturas lógicas relacionadas ao instrumento de coleta de dados (DDI).

Qual é o conjunto de dados a ser compartilhado pelo grup?

### 3.4 O Uso do Conjunto de Dados

Definir quem poderá utilizar o conjunto de dados (restrições de acesso), quais as licenças, termos e condições para usar o conjunto de dados.

#### **Restrições de Acesso ao Conjuntos de Dados**

Restrições de acesso, embargos, políticas e requisições para acesso e compartilhamento de dados

O conjunto de dados é de acesso restrito?

Quais são as restrições de acesso ao conjunto de dados?

#### **Licenças para o Conjunto de Dados**

Licença: instrumento legal através do qual o detentor dos direitos permite que uma segunda parte faça coisas que, de outra forma, iriam infringir os direitos por ele detidos

Quais são as licenças de uso do conjunto de dados ?

#### **Condições para Uso do Conjunto de Dados**

Termos ou condições que os usuários devem comprometer-se para poder usar os dados.

Existem condições especiais para o uso dos dados?

Quais são essas condições?

O usuário deverá comprometer-se com um termo de uso?

O usuário deverá responder a um questionário para poder usar os dados (*guestbook*)?

## 3.5 A Representação do Conjunto de Dados

Definir como serão as estruturas para representar o conjuntos de dados, as estruturas e os formatos dos arquivos que compõem o conjunto de dados e como será o versionamento deste conjunto.

### **Estrutura do Conjuntos de Dados**

Arquivos de dados e de documentação que compõem o conjunto de dados.

Quais são os arquivos de dados que compõem o conjunto de dados?

Quais são os arquivos de documentação que compõem o conjunto de dados?

Quais são as regras para os nomes de arquivos?

### **Formatos dos Arquivos do Conjunto de Dados**

Formatos dos arquivos que compõem o conjunto de dados.

Quais são os formatos para os arquivos de dados?

Quais são os formatos para os arquivos de documentação?

### **Versionamento do conjunto de dados**

Regras para o versionamento do conjunto de dados

O conjunto de dados é periodicamente acrescido de novos arquivos?

Quais são as regras que determinam uma nova versão do conjunto de dados?

## 3.6 A Documentação do Conjunto de Dados

Documentação referente a estrutura lógica dos arquivos de dados.

### **Estrutura do arquivo de dados**

*Codebook*: Livro que descreve o conteúdo, a estrutura e o layout de uma coleta de dados.

Um livro de códigos bem documentado deve conter informações destinadas a ser completas e autoexplicativas para cada variável em um arquivo de dados

Qual o conteúdo do arquivo de dados?

Qual a estrutura e o layout do arquivo de dados?

## 3.7 A Descrição do Conjunto de Dados

Compreende na definição dos metadados que serão usados para descrever o conjunto de dados. Esses metadados seguem esquemas de metadados que são padrões, com o intuito de facilitar a troca (interoperabilidade) de informações entre o repositório e outros serviços, como em serviço de busca federada. Um esquema de metadados também é chamado de vocabulário e compreende na especificação de um conjunto de elementos de metadados e das regras para o uso destes elementos.

Os repositórios normalmente permitem a criação e uso de vários esquemas de metadados. Para prover a interoperabilidade, esses repositórios desenvolvem *crosswalks*, que são especificações de mapeamentos que indicam como os elementos de metadados do repositório são mapeados para esquemas de metadados padrões. Dessa forma, os metadados do repositório podem ser exportados em representações que estão de acordo com esquemas de metadados padrões, favorecendo a interoperabilidade.

Dataverse é um software de repositório que permite a definição de esquemas de metadados quaisquer. Entretanto, disponibiliza esquemas de metadados pré-definidos, apropriados para diversos domínios da pesquisa. A especificação de cada um desses esquemas foi feita visando a compatibilidade com esquemas de metadados padrões, acompanhada de *crosswalks*, que apresentam mapeamentos para esses esquemas padrões. O quadro 2 apresenta, respectivamente, os esquemas de metadados disponibilizados por Dataverse e os esquemas de metadados padrões compatíveis, isto é, mapeáveis via *crosswalks*.

Quadro 2 – Esquemas de Metadados de Dataverse e esquemas padrões compatíveis

Esquema de Metadados de Dataverse	Compatibilidade dos esquemas de Dataverse com esquemas padrões
Citation Metadata	<ul style="list-style-type: none"> <li>• DDI Lite, DDI 2.5 Codebook</li> <li>• DataCite 3.1</li> <li>• Dublin Core's DCMI Metadata Terms</li> </ul>
Geospatial Metadata	<ul style="list-style-type: none"> <li>• DDI Lite, DDI 2.5 Codebook</li> <li>• DataCite</li> <li>• Dublin Core</li> </ul>
Social Science & Humanities Metadata	<ul style="list-style-type: none"> <li>• DDI Lite, DDI 2.5 Codebook</li> <li>• Dublin Core</li> </ul>
Astronomy and Astrophysics Metadata	<ul style="list-style-type: none"> <li>• International Virtual Observatory Alliance.VOResource</li> </ul>
Life Sciences Metadata	<ul style="list-style-type: none"> <li>• ISA-Tab Specification</li> </ul>

Fonte: Dataverse Users Guide [xvii]

O quadro 3 apresenta características dos principais esquemas de metadados que são compatíveis com os esquemas de Dataverse.

Quadro 3 – Esquemas de metadados padrões

DataCite	<ul style="list-style-type: none"> <li>• Visa a identificação de um recurso com o propósito de citação e recuperação.</li> <li>• É usado pelo serviço DataCite, que visa localizar, identificar e citar dados de pesquisa através do uso de DOI (<a href="https://datacite.org/">https://datacite.org/</a>)</li> <li>• Especificações: <a href="https://schema.datacite.org/">https://schema.datacite.org/</a></li> </ul>
DDI Codebook	<ul style="list-style-type: none"> <li>• Desenvolvido pela Data Documentation Initiative (DDI), organização voltada a desenvolver padrões internacionais para descrever dados produzidos por <i>surveys</i> ou por outros métodos observacionais, nas ciências sociais, comportamentais, econômicas e da saúde.</li> <li>• Compreende em elementos que estão presentes nos livros de códigos (codebooks), que servem para documentar conjuntos de dados</li> <li>• Especificação: <a href="https://ddialliance.org/Specification/DDI-Codebook/2.5/">https://ddialliance.org/Specification/DDI-Codebook/2.5/</a></li> </ul>
Dublin Core	<ul style="list-style-type: none"> <li>• Esquema de metadados simples para descrever recursos eletrônicos da web.</li> <li>• É esquema mínimo obrigatório para o protocolo OAI-PMH, usado para a construção de federações de repositórios</li> <li>• Especificação: <a href="https://www.dublincore.org/specifications/dublin-core/">https://www.dublincore.org/specifications/dublin-core/</a></li> </ul>
IVOA.VOResource	<ul style="list-style-type: none"> <li>• Dados astronômicos produzidos por comunidades ligadas a Observatório Virtual</li> <li>• Desenvolvido pela International Virtual Observatory Alliance</li> <li>• Esquema:</li> <li>• <a href="http://www.ivoa.net/documents/RM/20070302/index.html">http://www.ivoa.net/documents/RM/20070302/index.html</a></li> </ul>

ISA Framework	<ul style="list-style-type: none"> <li>• Investigation, Study, Assay</li> <li>• Ciências da Vida/Experimentos biomédicos e ambientais</li> <li>• Dados ligados saúde ambiental, genômica ambiental, metabolômica, (meta) genômica, proteômica, células-tronco, biologia de sistemas, transcriptômica e toxicogenômica, ácidos nucleicos, assinaturas celulares, etc.</li> <li>• Esquema: ISA-TAB</li> <li>• Especificação: <a href="https://isa-specs.readthedocs.io/en/latest/isatab.html">https://isa-specs.readthedocs.io/en/latest/isatab.html</a></li> </ul>
---------------	---

Dataverse permite a exportação dos metadados em esquemas como Dublin Core, DDI (Data Documentation Initiative Codebook 2.5) e DataCite 4, entre outros [xvii].

Definir os metadados e vocabulários para valores

### Seleção e definição de regras de uso dos metadados de “Citation Metadata”

O Dataverse disponibiliza um esquema básico de metadados, que é usado para descrever conjuntos de dados de qualquer domínio, chamado de “Citation Metadata”.

A aplicação de “Citation Metadata” para descrever conjunto de dados de um domínio envolve: a identificação de quais elementos serão usados, a definição de quais elementos serão de uso obrigatório e a especificação de regras de uso para esses elementos, caso necessárias.

Quais elementos de Citation Metadata serão usados para descrever o conjunto de dados?

Quais serão os elementos de preenchimento obrigatório?

Existem regras especiais para uso de alguns desses elementos? Quais?

Elemento do esquema Citation Metadata		Descrição / Uso do Elemento	Obrigatório Opcional Não usado
Título		Título completo no qual que o conjunto de dados é conhecido	Requerido por dataverse
Subtítulo		Subtítulo	
Título Alternativo		Como o conjunto de dados é conhecido ou abreviação	
URL Alternativa		Como projeto	
Outro identificador	Agência	Agência que gerou o identificador	
	Identificador	Identificador único que identifica o conjunto de dados	

Autor	Nome	Pessoa(s), órgão(s) corporativo(s) ou agência(s) responsável(is) pela criação da obra [xiv]	Requerido por dataverse
	Afiliação		
	Identificador	Identificador único	
	Esquema do identificador	Esquema do identificador (ORCID, ISNI)	
Contato	Nome	Contato para o conjunto de dados	Requerido por dataverse
	Afiliação		
	Email		Requerido por dataverse
Descrição	Texto	Um sumário descrevendo o propósito, a natureza e o escopo do conjunto de dados	Requerido por dataverse
	Data		
Assunto		Categorias de assuntos de domínio dos dados	Requerido por dataverse
Palavra chave	Termo	Palavra chave que descreve aspectos importantes do conjunto de dados	
	Vocabulário	Vocabulário controlado usado (LCSH,MESH)	
	URL do vocabulário		
Tópico	Termo	Tópicos e assuntos principais que os dados abrangem	
	Vocabulário	Vocabulário controlado usado (LCSH,MESH)	
	URL do vocabulário		
Publicação relacionada	Citação	Referência bibliográfica	
	Tipo de identificador	Tipo do identificador digital como DOI	
	Número do identificador	Identificador digital, como DOI	
	URL da publicação		
Notas		Informação adicional importante	
Idioma		Idioma do conjunto de dados	

Produtor	Nome	Pessoa ou organização com responsabilidade financeira ou administrativa sobre o conjunto de dados	
	Afiliação		
	Abreviação		
	URL		
	URL do logo		
Data de produção		Data em que a coleção de dados foi produzida	
Local de produção		Local em que a coleção de dados foi produzida	
Colaborador	Tipo	Organização ou pessoa responsável por coletar, gerenciar ou contribuir de alguma forma para o desenvolvimento do recurso.	
	Nome		
Grant	Agência	Agência financiadora	
	Número	Contrato ou número do projeto	
Distribuidor	Nome	Organização designada pelo autor ou produtor para gerar cópias da obra em particular, incluindo quaisquer edições ou revisões necessárias	
	Afiliação		
	Abreviação		
	URL		
	URL do logo		
Data de distribuição		Data em que se tornou disponível para distribuição	
Depositante		Pessoa ou o nome da organização que depositou o conjunto de dados no repositório	
Data de depósito		Data em que foi depositado no repositório	
Período de cobertura	Início	Período de tempo ao qual os dados se referem.	
	Fim		
Data da coleção	Início	Contém a(s) data(s) de quando os dados foram coletados	
	Fim		
Tipo de dado		Tipos dos dados incluídos nos arquivos	
Série	Nome	Nome e histórico da série a qual o conjunto de dados pertence	
	Informação		
Software	Nome	Software usado para gerar o conjunto de dados	
	Versão		

Material relacionado	Material relacionado	
Conjuntos de dados relacionados	Conjuntos de dados relacionados	
Outras referências	Referência que servem de base ou material de suporte	
Fontes dos dados	Lista de documentos ou arquivos de dados que serviram como fontes para a coleção de dados	
Origens das fontes	Informação sobre a origem das fontes e regras usadas para estabelecer as fontes	
Características das fontes	Avaliação de características e material de origem	
Documentação e acesso às fontes	Nível de documentação das fontes originais	

### Vocabulários controlados e Regras de Codificação

Vocabulários controlados e regras de codificação são usadas para controlar o preenchimento de valores de elementos de metadados.

Quais elementos do esquema de metadados possuem regras de codificação ou vocabulários controlados para seus valores?

Quais são as regras de codificação ou vocabulários?

Elementos com valores controlados por vocabulários ou regras	Vocabulário ou regra de codificação
Autor Produtor	Valores de Citation Metadata. <ul style="list-style-type: none"> <li>• ORCID, ISNI, LCNA, VIAF, GND</li> </ul>
Documentos e conjuntos de dados	Valores de Citation Metadata <ul style="list-style-type: none"> <li>• ark, arXiv, bibcode, doi, ean13, eissn, handle, isbn, issn, istc, lissn, lsid, pmid, purl, upc, url, urn</li> </ul>
Palavras-chave e Tópicos	Exemplos de vocabulários <ul style="list-style-type: none"> <li>• LCSH</li> <li>• MESH</li> </ul>
Assunto	Valores de Citation Metadata <ul style="list-style-type: none"> <li>• Agricultural Sciences</li> <li>• Arts and Humanities</li> <li>• Astronomy and Astrophysics</li> <li>• Business and Management</li> <li>• Chemistry</li> <li>• Computer and Information Science</li> <li>• Earth and Environmental Sciences</li> </ul>



## 3.8 A Proveniência do Conjunto de Dados

Definição de como os dados são carregados no repositório, com a identificação das entidades envolvidas pela produção dos dados e suas responsabilidades. Entende-se como proveniência: “informações sobre entidades, atividades e pessoas envolvidas na produção de um dado ou coisa, que podem ser usadas para formar avaliações sobre sua qualidade, confiabilidade ou fidedignidade” [xi]

### **O Produtor do Conjunto de Dados**

Depositante, pessoa ou instituição que forneceu os dados ao repositório. O produtor possui domínio sobre a proveniência dos dados, tem a responsabilidade pelos processos pelos quais os dados foram trazidos à existência, possui direitos para conceder licenças e termos de uso sobre os dados, possui a responsabilidade em fornecer os dados já preparados para o compartilhamento e submete ao repositório os dados para compartilhamento.  
Quem é o responsável pela produção dos dados?

### **Aspectos legais para publicação e uso dos dados**

Autorização, termo de autorização ou aspectos legais necessários para publicação e uso dos dados  
Quais são os aspectos legais e/ou termos que viabilizam a definição de licenças ?  
Quais são os aspectos legais e/ou termos que determinam condições para o uso dos dados?

### **Processo de preparação dos dados do conjunto de dados**

Necessário para preparar os dados para o compartilhamento.  
Há necessidade de um procedimento especial para preparar os dados para o compartilhamento? Qual ?

**Processo de submissão dos dados**

Realizado para carregar os dados já preparados no repositório, incluindo responsáveis pela carga e pela publicação do conjunto de dados ou de novas versões desse conjunto.

Quem serão os responsáveis pela carga dos arquivos de dados do conjunto de dados?

Quem serão os responsáveis pela publicação do conjunto de dados ou de uma nova versão deste conjunto?

# 4 Criação da Comunidade no Dataverse

## 4.1 Dataverse

Um repositório de dados que usa o software Dataverse é composto por entidades chamadas de **dataverses** e **datasets**. **Dataverse** é o recurso para representar uma instituição, uma unidade, um projeto ou um grupo que necessita compartilhar dados. Uma hierarquia organizacional também pode ser representada, visto que um **dataverse** pode conter outros **dataverses**. Uma comunidade produtora de dados é representada por um **dataverse**.

Um **dataverse** contém **datasets**. Um **Dataset** é o recurso que o ambiente disponibiliza para armazenar um conjunto de dados. Um **dataset** é composto por metadados, pelos termos de uso (como licenças) e por arquivos. Metadados descrevem o **dataset** e os arquivos.

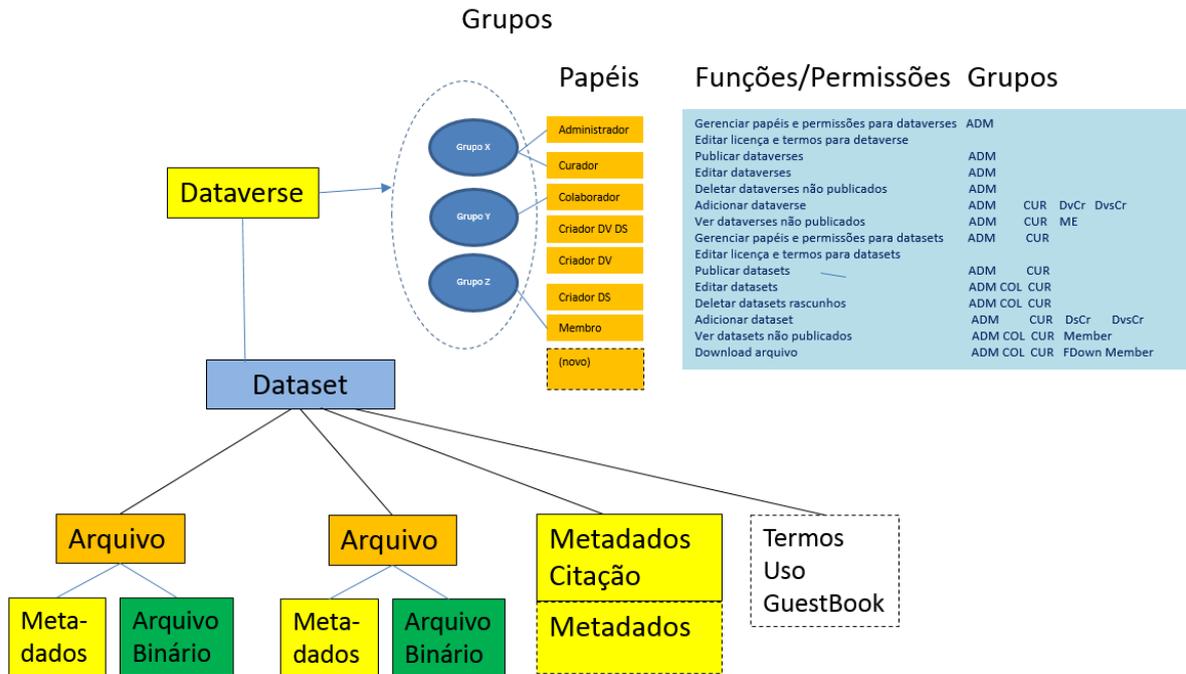
Um **dataverse** pode ser configurado como se fosse um repositório independente, com todas as autorizações para gerenciamento e operação (como definir papéis, permissões e grupos; criar e gerenciar **dataverses**, **datasets** e arquivos, etc). Para a definição de políticas, o software trabalha os seguintes conceitos:

- **Permissão:** representa a permissão para realização de uma determinada funcionalidade do ambiente, como publicar **dataverse**, editar **dataverse**, gerenciar papéis e permissões de **dataverse** ou **datasets**.
- **Papel:** corresponde a um conjunto de permissões, que caracteriza um perfil exercido por usuários na gestão ou operação do repositório. O ambiente já disponibiliza vários papéis (como Administrador, Publicador de Dataverse, Curador de Dataverse e Editor de Dataset, etc), e novos papéis podem ser criados.
- **Grupo:** representa um grupo de usuários, que são habilitados a atuar com determinados papéis.

Estudo sobre Dataverse realizado pelo GT-RDP:

- ROCHA, Rafael Port da; AZAMBUJA, Luís Alberto Barbosa; BORGES, Nunes Borges; GABRIEL JUNIOR, Rene Faustino; CAREGNATO, Sônia Elisa; PAVÃO, Caterina Groposo; VANZ, Samile Andrea de Souza; PASSOS, Paula Caroline Schifino Jardim. Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas: relatório 2018. Disponível em: <http://hdl.handle.net/10183/185126> >.

Figura 2 - Dataverse e Datasets



## 4.2 Definição e Criação da Comunidade Produtora de Dados no Repositório

O Grupo de Compartilhamento de Dados é representado no Repositório como uma Entidade Dataverse.

### Definir as Informações necessárias para a Criação do Dataverse

Qual o nome do Dataverse?

Como o Dataverse será identificado na web?

Qual a afiliação do Dataverse?

Como o dataverse poderia ser descrito (em um texto curto)?

Em que categoria o Dataverse pode ser classificado?

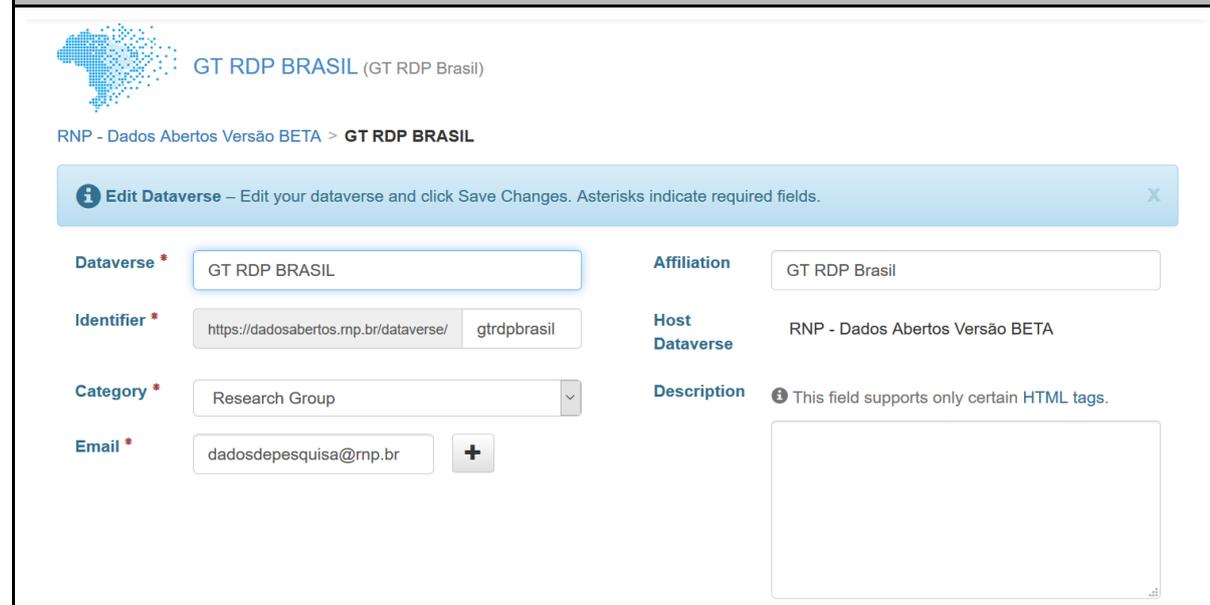
### Criar o Dataverse

▾

New Dataverse

(no Dataverse Raiz)

Exemplo:



 **GT RDP BRASIL** (GT RDP Brasil)

RNP - Dados Abertos Versão BETA > **GT RDP BRASIL**

**Edit Dataverse** – Edit your dataverse and click Save Changes. Asterisks indicate required fields. ✕

<b>Dataverse *</b>	<input type="text" value="GT RDP BRASIL"/>	<b>Affiliation</b>	<input type="text" value="GT RDP Brasil"/>
<b>Identifier *</b>	<input type="text" value="https://dadosabertos.mp.br/dataverse/"/> <input type="text" value="gtrdpbrasil"/>	<b>Host Dataverse</b>	RNP - Dados Abertos Versão BETA
<b>Category *</b>	<input type="text" value="Research Group"/> ▾	<b>Description</b>	<p><b>This field supports only certain HTML tags.</b></p> <div style="border: 1px solid #ccc; height: 60px; width: 100%;"></div>
<b>Email *</b>	<input type="text" value="dadosdepesquisa@mp.br"/> <input type="button" value="+"/>		

## 4.3 Definição e Criação dos Grupo de Usuários

Definir os grupos de usuários que serão responsáveis por gerenciar a Comunidade Produtora de Dados, criar e adicionar novos arquivos a um conjunto de dados e publicar conjuntos de dados ou novas versões de conjuntos de dados.

### Grupo dos Usuários Gestores

Responsável por gerir a Comunidade Produtora de Dados.

Esse grupo terá o papel de **Administrador**, cujos usuários poderão:

- Gerenciar papéis, permissões, licenças e termos para **dataverses** e **datasets**
- Adicionar, editar, publicar e remover **dataverses** e **datasets**
- Visualizar **dataverses** e **datasets** “drafts” (não publicados ou novas versões não publicadas)
- Baixar **arquivos**

Qual o nome do grupo de gestores?

### Definir o Grupo de Usuários Curadores de Datasets

Definir grupo de usuários responsável por gerir conjunto de dados do dataverse. Um usuário desse grupo terá autorização para criar, atualizar e publicar um conjunto de dados.

Esse grupo terá o papel de Curador, cujos usuários poderão:

- Gerenciar papéis, permissões, licenças e termos para **datasets**
- Adicionar, editar, publicar e remover **datasets**
- Visualizar **datasets** não publicados
- Baixar **arquivos**

Qual o nome do grupo de curadores?

### Definir o Grupo de Usuários Colaborador de Dataset

Definir grupo de usuários responsável por criar e atualizar datasets, que serão publicados pelos curadores. Usuários desse grupo não possuem autorização para publicar um conjunto de dados. Podem criar ou atualizar datasets ainda não publicados.

Esse grupo terá o papel de Colaborador, cujos usuários poderão:

- Editar **datasets**
- Deletar **datasets** não publicados (drafts)
- Visualizar **datasets** não publicados
- Baixar **arquivos**

## Criar os Grupos de Usuários no Dataverse

Contact Share Edit

 **RNP - Dados Abertos Versão BETA (RNP)**  
dvmp

General Information  
Theme + Widgets  
Permissions  
Groups

(no dataverse)

Exemplo:

+ Create Group

Group Name	Group ID	Membership	Action
Membro GT-RDP	&explicit/4-membrogrtdp	No members	 
Colaborador GT-RDP	&explicit/4-colaboradorgtrdp	No members	 
Curador GT-RDP	&explicit/4-curadorgtrdp	No members	 

## Atribuir papéis de Administrador, Curador e Colaborador aos grupos criados

Contact Share Edit

 **RNP - Dados Abertos Versão BETA (RNP)**  
dvmp

General Information  
Theme + Widgets  
Permissions

(no dataverse)

Exemplo:

Users/Groups ⌵ All the users and groups that have access to your dataverse.

Assign Roles to Users/Groups

4 Users/Groups

User/Group Name (Affiliation)	ID	Role	Action
Colaborador GT-RDP	&explicit/4-colaboradorgtrdp	Contributor	 Remove Assigned Role
Curador GT-RDP	&explicit/4-curadorgtrdp	Curator	 Remove Assigned Role
Dataverse Admin (UFRGS/FURG/RNP)	@dataverseAdmin	Admin	 Remove Assigned Role
Membro GT-RDP	&explicit/4-membrogrtdp	Member	 Remove Assigned Role

## 4.4 Criação e Descrição de Conjunto de Dados

Criar e descrever o conjunto de dados no repositório.

### Criar e descrever o dataset

Criar o dataset e preencher os metadados descritivos de do de acordo com definido na seção 3.7.

- New Dataverse
- New Dataset

(no dataverse)

Exemplo:

Citation Metadata ^

**Title \***

**Author \***

<b>Name *</b>	<input type="text" value="Admin, Dataverse"/>	<b>Affiliation</b>	<input type="text" value="UFRGS/FURG/RNP"/>	<input type="button" value="+"/>
<b>Identifier Scheme</b>	<input type="text" value="Select..."/>	<b>Identifier</b>	<input type="text"/>	

**Contact \***

<b>Name</b>	<input type="text" value="Admin, Dataverse"/>	<b>Affiliation</b>	<input type="text" value="UFRGS/FURG/RNP"/>	<input type="button" value="+"/>
<b>E-mail *</b>	<input type="text" value="dadosdepesquisa@mp.br"/>			

## 4.5 Definição das Termos de Uso e Condições de Acesso ao Conjunto de Dados

Definir os termos de uso dos dados e as condições de acesso aos dados.

### Definir como será o uso do Conjunto de Dados

Definir se os dados serão de domínio público (Licença CC0 - *Creative Commons Public Domain Dedication*) ou informar quais os termos para uso dos dados.

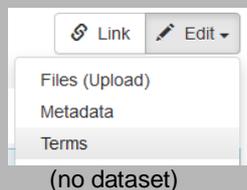
### Definir outras condições adicionais

Elaborar textualmente termos de confidencialidade, permissões especiais para acesso ao recurso, restrições para acesso ou uso, requisitos para citar os dados, informações que os usuários devem passar aos depositantes, informações adicionais para o uso dos dados, *disclaimer*

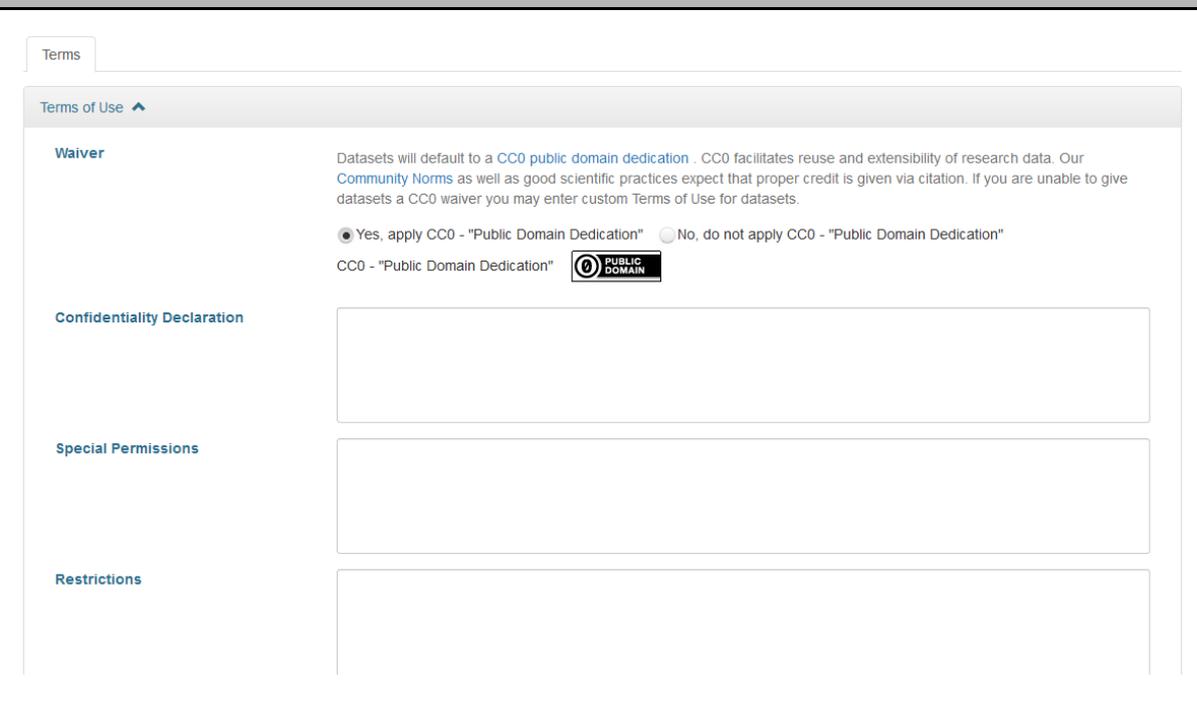
### Definir condições especiais para acesso ao Conjunto de Dados

Elaborar textos que informam como usuários podem obter acesso aos dados, local para acesso aos dados (caso não armazenados no repositório), arquivo original, estado de disponibilidade (embargo), contato para acesso, tamanho da coleção.

### Criar os Termos de Uso e Condições de Acesso no Dataset do Repositório



### Exemplo



Terms

Terms of Use

**Waiver**

Datasets will default to a [CC0 public domain dedication](#). CC0 facilitates reuse and extensibility of research data. Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. If you are unable to give datasets a CC0 waiver you may enter custom Terms of Use for datasets.

Yes, apply CC0 - "Public Domain Dedication"  No, do not apply CC0 - "Public Domain Dedication"

CC0 - "Public Domain Dedication" 

**Confidentiality Declaration**

**Special Permissions**

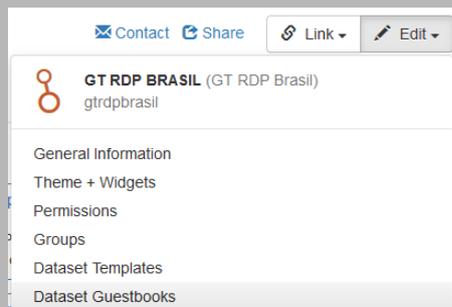
**Restrictions**

## Definir informações que o usuário deve fornecer ao baixar os dados

Definir questões que o usuário deve responder ao baixar os dados do dataset.

## Implementar as questões no Repositório

Criar as questões através da criação de um Gestbook ou usar um Gestbook já criado  
Associar o gestbook ao conjunto de dados no componente termos do dataverse



(no dataverse)

RNP - Dados Abertos Versão BETA > Create Dataset Guestbook

**Guestbook Name \***

**Data Collected** Dataverse account information that will be collected when a user downloads a file. Check the ones that will be required.

- Name
- Email
- Institution
- Position

**Custom Questions** Create your own questions to have users provide more than their account information when they download a file. Questions can be required or optional and answers can be text or multiple choice.

**Question Type**

Text

**Question Text**



Required field

## 4.6 Criação de Arquivos de Conjunto de Dados

### Definir as Informações necessárias para a Criação do Arquivo

Definir nome do arquivo, descrição do arquivo e etiquetas.

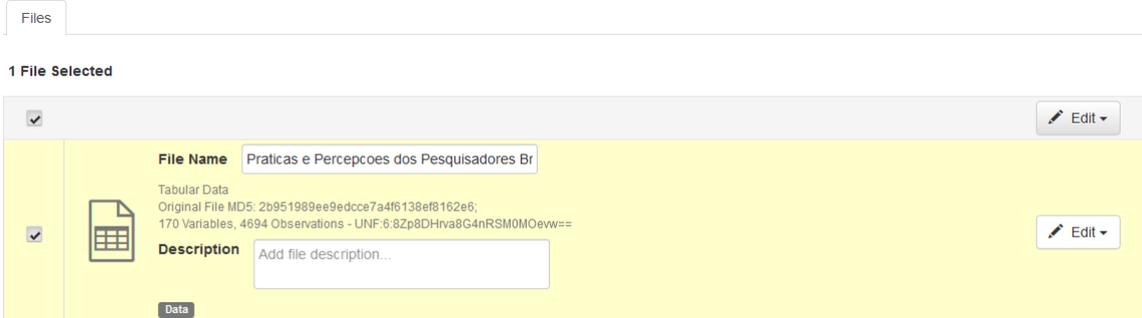
### Criar, descrever e carregar o arquivo no Repositório

Poderão criar arquivos usuários de grupos que possuem o papel de Administrador, Curador e Colaborador.

 + Upload Files

(no dataset)

Exemplo:



The screenshot shows a web interface for file management. At the top, there is a 'Files' tab. Below it, a message states '1 File Selected'. A table displays the details of the selected file:

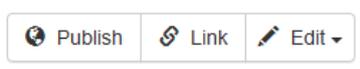
File Name	Description	Actions
Praticas e Percepcoes dos Pesquisadores Br	Tabular Data Original File MD5: 2b951989ee9edcce7a4f6138ef8162e6; 170 Variables, 4694 Observations - UNF:6:8Zp8DHva8G4nRSM0M0eww==	<input checked="" type="checkbox"/> Edit

Below the description, there is a text input field with the placeholder 'Add file description...'. A 'Data' label is visible at the bottom left of the file entry.

## 4.7 Publicação de Conjunto de Dados

### Publicar um Dataset ou uma versão não Publicada de um Dataset

Poderão publicar datasets os usuários de grupos que possuem o papel de Administrador e/ou de Curador.



(no dataset não publicado)

Exemplo:



The screenshot displays a dataset page with the following elements:

- Header:** Metrics (0 Downloads), Contact, Share, Publish, Link, Edit.
- Title:** Acesso aberto a dados de pesquisa no Brasil : repositórios brasileiros de dados de pesquisa (Draft).
- Description:** Conjunto de Dados produzidos pelo GT-RDP Brasil referente à investigação "Acesso aberto a dados de pesquisa no Brasil : repositórios brasileiros de dados de pesquisa".
- Subject:** Computer and Information Science; Social Sciences.
- Related Publication:** PAVÃO, C. et al. "Acesso aberto a dados de pesquisa no Brasil: repositórios brasileiros de dados de pesquisa", 2018.
- Navigation:** Files, Metadata, Terms, Versions.
- File Upload Section:** 1 File, Upload Files, Edit Files.
- File List:** Codebook Repositórios brasileiros de dados de pesquisa.pdf (Adobe PDF - 7.7 KB - Aug 12, 2019 - 0 Downloads, MD5: 7523e2c20846109ff6b48bc63a53e666), Download.

# Referências/Bibliografia

- [i] CONARQ. DIRETRIZES PARA A IMPLEMENTAÇÃO DE REPOSITÓRIOS DIGITAIS CONFIÁVEIS DE DOCUMENTOS ARQUIVÍSTICOS, 2014
- [ii] OAIS - CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. Reference model for an open archival information system (OAIS). CSDS 650.0-M-2. 2012 <https://public.ccsds.org/pubs/650x0m2.pdf>
- [iii] Wilkinson, M. D., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, vol. 3 (2016). <https://www.nature.com/articles/sdata201618>
- [iv] CTS - Core Trustworthy Data Repositories Requirements v01. [https://www.coretrustseal.org/wp-content/uploads/2017/01/Core\\_Trustworthy\\_Data\\_Repositories\\_Requirements\\_01\\_00.pdf](https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf)
- [v] Princípios Fair - <https://www.force11.org/group/fairgroup/fairprinciples>
- [vi] Jantz, R., Giarlo, M.: Digital Preservation - Architecture and Technology for Trusted Digital Repositories. Dlib Magazine, vol. 11, n. 6 (2006) <http://www.dlib.org/dlib/june05/jantz/06jantz.html>
- [vii] Wilkinson, M. D., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, vol. 3 (2016). <https://www.nature.com/articles/sdata201618>
- [viii] CTS - Core Trustworthy Data Repositories Requirements v01. [https://www.coretrustseal.org/wp-content/uploads/2017/01/Core\\_Trustworthy\\_Data\\_Repositories\\_Requirements\\_01\\_00.pdf](https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf)
- [ix] CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. Producer-Archive Interface Methodology Abstract Standard. CCSDS 651.0-M-1. 2004 <https://public.ccsds.org/Pubs/651x0m1.pdf>
- [x] UCPSR - What is a Codebook?  
<https://www.icpsr.umich.edu/icpsrweb/content/shared/ICPSR/faqs/what-is-a-codebook.html>
- [xi] Groth, P., Moreau, L.: PROV-Overview - An Overview of the PROV Family of Documents. W3C Working Group, Note 30 (2013). <https://www.w3.org/TR/prov-overview>
- [xii] Esquema DataCite - [https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel\\_v4.1.pdf](https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf)
- [xiii] Ball, A. (2014). 'How to License Research Data'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>
- [xiv] TDR-Metadata-Dictionary, 2016 <https://data.tdl.org/wp-content/uploads/2016/09/TDR-Metadata-Dictionary.pdf>
- [xv] DDI-Codebook v 2.5 - [www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field\\_level\\_documentation.html](http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation.html)
- [xvi] DDI.Structural Reform Group. Overview of the DDI Version 3.0 Conceptual Model. Working draft 4. [http://opendatafoundation.org/ddi/srg/Papers/DDIModel\\_v\\_4.pdf](http://opendatafoundation.org/ddi/srg/Papers/DDIModel_v_4.pdf)
- [xvii] Dataverse User Guide. atualização de agosto de 2019. <http://guides.dataverse.org/en/latest/user/index.html>

# Anexo

Conceitos:

<b>Repositório Digital</b>	Ambiente de armazenamento e gerenciamento de materiais digitais. Esse ambiente constitui-se de uma solução informatizada em que os materiais são capturados, armazenados, preservados e acessados. Um repositório digital é, então, um complexo que apoia o gerenciamento dos materiais digitais, pelo tempo que for necessário, e é formado por elementos de hardware, software e metadados, bem como por uma infraestrutura organizacional e procedimentos normativos e técnicos. <sup>i</sup>
<b>Repositório Digital Confiável</b>	Repositório cuja missão é fornecer acesso de longo prazo a recursos digitais gerenciados; que aceita a responsabilidade pela manutenção a longo prazo dos recursos digitais em nome de seus depositantes e em benefício dos usuários atuais e futuros; que projeta seu(s) sistema(s) de acordo com as convenções e os padrões comumente aceitos para garantir o gerenciamento, acesso e segurança contínuos dos materiais depositados nele; que estabelece metodologias para avaliação de sistemas que atendem às expectativas de confiabilidade da comunidade; que cumpre suas responsabilidades com depositantes e usuários de forma aberta e explícita; e cujas políticas, práticas e desempenho podem ser auditados e medidos. <sup>vi</sup>
<b>OAIS</b>	Modelo de referência que proporciona uma linguagem comum para facilitar a discussão entre comunidades interessadas na preservação. Descreve o ambiente em que um arquivo (repositório) reside, os componentes funcionais de um arquivo e a infraestrutura de informação que suporta o processo de arquivamento.
<b>Ambiente OAIS</b>	O ambiente de um repositório OAIS é composto pelo produtor, que fornece a informação a ser preservada, pelo consumidor, que interage com o ambiente para localizar e obter a informação preservada de seu interesse, e pela gestão, que determina as políticas, controlando as responsabilidades de gestão
<b>Produtor</b>	Papel desempenhado pelas pessoas ou sistemas clientes que fornecem as informações a serem preservadas, podendo incluir outros sistemas OAIS. <sup>ii</sup>

<b>Consumidor</b>	Papel desempenhado por pessoas ou sistemas clientes que interagem com os serviços OAIS para encontrar informações preservadas de interesse e para acessar essas informações em detalhes, podendo incluir outros sistemas OAIS <sup>ii</sup>
<b>Comunidade Designada</b>	Um grupo identificado de potenciais consumidores que devem ser capazes de entender um conjunto específico de informações. <sup>ii</sup>
<b>Gestor</b>	Papel desempenhado por aqueles que definem a política geral do OAIS como parte em um domínio de política mais amplo, por exemplo, como parte de uma organização maior. [ix]
<b>Metadados</b>	Dados estruturados que provém informação sobre um ou mais aspectos de outros dados. Podem ser classificados de diferentes formas, tais como descritivos, estruturais e administrativos (técnicos, direitos e de preservação digital).
<b>Estudo</b>	Coleção de arquivos de dados resultantes da coleta intencional de dados por meio de solicitação, observação ou coleta de fontes secundárias para um propósito; utilizando instrumentos e metodologias de coleta descritos; e expressos como arquivos de dados com estruturas lógicas relacionadas ao instrumento de coleta de dados
<b>Conjunto de Dados</b>	Arquivos de dados descritos por uma única estrutura
<b>Livro de Códigos Codebook</b>	Livro que descreve o conteúdo, a estrutura e o layout de uma coleta de dados. Um livro de códigos bem documentado deve conter informações destinadas a ser completas e autoexplicativas para cada variável em um arquivo de dados [x]
<b>Proveniência dos Dados</b>	Informações sobre entidades, atividades e pessoas envolvidas na produção de um dado ou coisa, que podem ser usadas para formar avaliações sobre sua qualidade, confiabilidade ou fidedignidade [xi]

Princípios FAIR:

**F Findable**

- F1 Meta(dados) com identificador persistente
- F2 Metadados ricos
- F3 (Meta)dados usados por recursos de busca
- F4 Metadados com identificador do conjunto de dados

**A Accessible**

- A1 Meta(dados) recuperável por identificador
  - A1.1 Por protocolo de recuperação aberto
  - A1.2 Autenticação e autorização
- A2 Metadados persistem quando dados são removidos

**I Interoperable**

- I1 Representação do conhecimento, vocabulários, ontologias
- I2 Vocabulários FAIR
- I3 Referências a outros meta(dados), contexto

**R Re-usable**

- R1 (Meta)dados descritos ricamente por atributos relevantes e precisos
  - R1.1 Licença
  - R1.2 Proveniência
  - R1.2 Padrões relevantes para a comunidade

Fonte: FAIR <sup>v</sup>