



Tutorial para Instalação de Repositório de Dados Científicos

CKAN

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).

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. Comprehensive Knowledge Archive Network (CKAN)

1.1. Apresentação

O *Comprehensive Knowledge Archive Network* (CKAN) é um Software Livre de Gerenciamento de Dados desenvolvido com finalidade de facilitar a catalogação, preservação, publicação, disponibilização e utilização de dados por meio de uma aplicação web (portal de dados), no qual são armazenados os dados e seus metadados.

O CKAN foi concebido para ter grande interoperabilidade entre suas bases, de forma que uma base importe dados de outras bases CKAN ou de catálogos com tecnologias compatíveis, bastando para isso ter instalado o coletor de dados *harvester plugin*.

Os usuários do CKAN, por meio do mecanismo de busca, têm acesso rápido aos conjuntos de dados disponíveis, assim como, às informações sobre seu formato, se são de livre acesso ou privativas, etc, devido a seu aprimorado sistema de filtros por tags e à navegação entre dados relacionados.

1.2. Versão utilizada

Este tutorial foi desenvolvido com base na versão CKAN 2.5.2, utilizando o sistema operacional Ubuntu versão 14.04 e disponível em <http://ckan.uspdigital.usp.br>.

1.3. Passos para Instalação

Os passos para a instalação seguiram o manual disponibilizado pelo IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia), disponível em http://wiki.ibict.br/index.php/Manual_de_instalação. O manual desenvolvido pelo IBICT, por sua vez, baseou-se nas informações contidas no website do CKAN, disponível em <http://docs.ckan.org/en/ckan-2.5.2/maintaining/installing/index.html>

1.3.1. Passo complementar e adição e administrador

Seguindo os passos descritos no manual do IBICT torna-se necessária, ainda, uma configuração entre os passos **Configure o Jetty-Solr** e **Crie as tabelas no banco de dados PostgreSQL**. Neste ponto execute o comando:

```
pip install -U bleach
```

Após finalização da instalação do CKAN seguindo o manual do IBICT é preciso adicionar o administrador da plataforma. Para isto deve-se executar o comando seguinte, onde **admin1** é o nome do usuário administrador do repositório:

```
paster sysadmin add admin1 -c /etc/ckan/default/production.ini
```

1.3.2. Permissões de usuário no CKAN:

As configurações de *criação de usuário* e de *permissões* no repositório CKAN podem ser realizadas no arquivo `production.ini`:

```
pico /etc/ckan/default/production.ini
```

As permissões sugeridas são:

- Um usuário precisa estar registrado e ter entrado no CKAN para poder adicionar um novo conjunto de dados;
- Um usuário pode adicionar conjunto de dados mesmo que não esteja vinculado a uma organização;
- Um usuário pode criar e remover grupos;
- Um usuário não pode criar novas organizações;
- A criação de usuário deve ser feita somente pela interface web do repositório CKAN.

Com estas configurações é possível que um novo usuário seja cadastrado no repositório CKAN com as permissões para adicionar conjuntos de dados e criar grupos. Como resultado o arquivo `production.ini` deverá ficar com a seguinte configuração:

```
## Authorization Settings
ckan.auth.anon_create_dataset = false
ckan.auth.create_unowned_dataset = true
ckan.auth.create_dataset_if_not_in_organization = false
ckan.auth.user_create_groups = true
ckan.auth.user_create_organizations = false
ckan.auth.user_delete_groups = true
ckan.auth.user_delete_organizations = false
ckan.auth.create_user_via_api = false
ckan.auth.create_user_via_web = true
ckan.auth.roles_that_cascade_to_sub_groups = admin
```

O método que o Metabusador utiliza para obter os dados de cada repositório CKAN é o formato completo de exportação JSON. O material oficial pode ser obtido através do link: <http://docs.ckan.org/en/ckan-2.5.2/api/>.

A função de exportação dos dados pelo JSON é nativa no CKAN, não sendo necessária nenhuma configuração adicional. Para verificar se o JSON esta funcionando corretamente é preciso adicionar um conjunto de dados no seu repositório CKAN e visualizar o conjunto de dados já formatado a partir do seguinte endereço:

`http://<endereço-repositório-CKAN>/api/3/action/package_list`

Este método de extração permitirá a obtenção dos campos *Dublin Core* configurados atualmente, a saber:

- campos obrigatórios:
 - dc.title
 - dc.subject
 - dc.description
 - dc.contributor.author
 - dc.identifier.uri
 - dc.description.sponsorship

- campos opcionais:
 - dc.rights.holder
 - dc.date
 - dc.date.accessioned
 - dc.date.available
 - dc.date.issued
 - dc.format
 - dc.type

1.3.4. Habilitação do Protocolo OAI-PMH

O CKAN permite habilitar o protocolo OAI-PMH através da instalação do plugin **ckanext-oaipmh** obtido pelo endereço:

<https://github.com/kata-csc/ckanext-oaipmh>

É preciso instalar as seguintes dependências para que o protocolo funcione:

<https://github.com/kata-csc/ckanext-kata> (apenas a v3.0.0)

<https://github.com/kata-csc/ckanext-hierarchy>

<https://github.com/kata-csc/ckanext-dcat>

e executar o comando:

```
apt-get install libsasl2-dev python-dev libldap2-dev libssl-dev
```

Este plugin foi testado utilizando a versão 2.5.2 do CKAN, porém em testes com a versão 2.7.3 do CKAN o plugin apresentava inconsistências, não permitindo reiniciar o CKAN.

O protocolo OAI-PMH possibilita a extração de menos informações para o metabuscador, por isto foi adotada a extração completa via JSON.

1.3.5. Adição do termo de uso do repositório CKAN

O registro de novos usuários do CKAN 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 link para do termo de uso em pdf na tela de registro foi preciso alterar o arquivo:

```
new_user_form.html
```

Para isso é necessário a edição do arquivo que se encontra no seguinte endereço:

```
ckan/lib/default/src/ckan/ckan/templates/user/new_user_form.html
```

No final do arquivo, no `<div class="form-actions">` e antes do `<button class="btn` adicione o seguinte trecho:

```
<div class="form-actions">
  {% block form_actions %}
  <p>By proceeding to register you agree to the <a
href="http://ckan.uspdigital.usp.br/base/images/Termo-USP-
CKAN.pdf">terms of use</a>.</p>
  <button class="btn btn-primary" type="submit" name="save">{{
_("Create Account") }}</button>
  {% endblock %}
</div>
</form>
```

Onde o endereço <http://ckan.uspdigital.usp.br/base/images/Termo-USP-CKAN.pdf> aponta para o arquivo pdf do termo de uso.

A pasta onde o arquivo foi disponibilizado é:

```
ckan/lib/default/src/ckan/ckan/public/base/images/
```

O resultado final na tela de registro de usuário se encontra na figura 1:



Registrar uma nova Conta

Nome de usuário:

Nome Completo:

E-mail:

Senha:

Confirme:

By proceeding to register you agree to the [terms of use](#).

Registrar Conta

Figura 1: Tela de registro de usuários no CKAN com termo de uso.

1.3.6. Alteração de tradução de Mantenedor para Financiador

Na a versão em português pt_BR há um campo chamado “Mantenedor” que foi alterado para “Financiador”. Para alterar o campo é preciso seguir os seguintes passos:

1 – Abrir o arquivo ckan.po referente a língua portuguesa:

```
/ckan/lib/default/src/ckan/ckan/i18n/pt_BR/LC_MESSAGES/ckan.po
```

2 – Alterar os seguintes trechos do código substituindo “Mantenedor” por “Financiador” e substituindo “José da Silva” por “Nome do autor ou do financiador”:

```
#: ckan/templates/package/snippets/additional_info.html:37
#: ckan/templates/package/snippets/additional_info.html:42
#: ckan/templates/package/snippets/package_metadata_fields.html:20
msgid "Maintainer"
msgstr "Financiador"
```

e

```
#: ckan/templates/package/snippets/package_metadata_fields.html:14
#: ckan/templates/package/snippets/package_metadata_fields.html:20
#: ckan/templates/user/new_user_form.html:6
msgid "Joe Bloggs"
msgstr "Nome do autor ou do financiador"
```

3 – Ativar a virtualenv e ir no diretório do CKAN:

```
. /usr/lib/ckan/default/bin/activate
cd /usr/lib/ckan/default/src/ckan
```

4 – Executar o comando em python para gerar o ckan.mo:

```
python setup.py compile_catalog --locale pt_BR
```

5 – Reiniciar o apache:

```
service apache2 reload
```

6 – Após esta etapa a tela para inserção de novos conjuntos de dados deve estar conforme a figura 2:



Autor:

E-mail do autor:

Financiador:

E-mail do Mantenedor:

Campo customizado: Key: Value:

Figura 2: Tela de cadastro de conjunto de dados alterada.

7 – A tela com informações de dados já cadastrado modificada, ficando da

Informações Adicionais	
Campo	
Fonte	
Autor	
Financiador	
Versão	
Estado	
Última Atualização	
Criado	
chave-teste	

adicionar de um conjunto também deve ter sido seguinte forma (figura 3):

Figura 3: Tela de informações adicionais do conjunto de dados.

1.3.7. Cadastro de um novo conjunto de dados

Das informações obrigatórias definidas para o metabuscador, nem todas se encontram disponíveis no CKAN:

Identificador	Metadado	CKAN
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	

Quanto às informações opcionais, as seguintes se encontram no CKAN:

Identificador	Tipo	CKAN
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
H	dc.identifier.citation	
I	dc.format	Sim
J	dc.publisher	
K	dc.rights.holder	Sim
L	dc.type	Sim
M	dc.coverage.spatial	
N	dc.coverage.spatial.Lat	
O	dc.coverage.spatial.Long	
P	dc.coverage.temporal	
Q	dc.language	

Os campos que devem ser preenchidos no CKAN para atender aos itens solicitados pelo consórcio, considerando o identificador informado nas tabelas acima para indicar cada campo, são mostrados nas figuras 4, 5 e 6:

The image shows a screenshot of the CKAN dataset creation form. At the top, there are two tabs: '1 Criar um conjunto de dados »' (highlighted in green) and '2 Adicionar dados'. The form contains several fields, each with a red circled number or letter next to it:

- 1** **Título:** ex.: Um título descritivo. Below the field is a URL field: * URL: 200.144.254.177/dataset/<dataset> with an 'Editar' button.
- 3** **Descrição:** ex.: Algumas anotações úteis sobre os dados. Below the field is a note: 'Você pode usar formatação Markdown aqui'.
- 2** **Etiquetas:** ex.: economia, saúde mental, governo.
- K** **Licença:** Creative Commons Atribui... (dropdown menu). To the right, a note says: 'Definições de licenças e informações adicionais podem ser encontradas em opendefinition.org'.
- Organização:** UFABC (dropdown menu).
- Visibilidade:** Privada (dropdown menu).
- 5** **Fonte:** http://example.com/dataset.json
- Versão:** 1,0
- 4** **Autor:** José da Silva

Figura 4: Informações para cadastro de conjunto de dados

E-mail do autor:

6 Financiador:

E-mail do Mantenedor:

Campo customizado:

Campo customizado:

Campo customizado:

A licença de dados que você escolher acima aplica-se somente ao conteúdo de quaisquer recursos de arquivos que você adicionar a este conjunto de dados. Ao enviar este formulário, você concorda em lançar os valores de *metadados* que você incluir sob a licença [Open Database License](#).

Próximo: Adicionar Dados

* Campo obrigatório

Figura 5: Informações para cadastro de conjunto de dados

1 Criar um conjunto de dados » **2** Adicionar dados

URL:

Arquivo:

Nome:

Descrição:

Você pode usar [formatação Markdown](#) aqui

I Formato:

📘 Isto será estimado automaticamente. Deixe em branco se quiser

Anterior **Salvar & adicionar outro** **Finalizar**

Figura 6: Informações para cadastro de conjunto de dados

Os campos D, E, F, G e L são obtidos diretamente pelo formato de extração JSON.

1.3.8. Autorização dos usuários no CKAN

No CKAN, conforme a configuração de permissões, somente o administrador pode adicionar uma organização, fazendo com que todos os conjuntos de dados fiquem vinculados a esta organização. Se o repositório tiver somente uma organização, então todos os conjuntos de dados serão vinculados a ela.

Há três perfis de usuário no CKAN com as seguintes características:

Administrador da organização (coordenador) pode:

1. Ver os conjuntos de dados privados da organização;
2. Adicionar novos conjuntos de dados à organização;
3. Editar ou excluir qualquer um dos conjuntos de dados da organização;
4. Modificar o conjunto de dados quanto a visibilidade: público ou privado;
5. Adicionar usuários à organização e escolher o perfil do novo usuário: membro, editor ou administrador;
6. Alterar a função de qualquer usuário na organização, incluindo outros usuários administradores;
7. Remover membros, editores ou outros administradores da organização;
8. Editar os dados da organização (por exemplo: alterar o título, a descrição ou a imagem da organização);
9. Excluir a organização.

Editor da organização pode:

1. Ver os conjuntos de dados privados da organização;
2. Adicionar novos conjuntos de dados à organização;
3. Editar ou excluir qualquer um dos conjuntos de dados da organização.

Membro da organização pode:

1. Ver os conjuntos de dados privados da organização.

O coordenador (administrador) de uma organização pode adicionar usuários na organização e mudar o perfil deles. Para isto deve-se selecionar a organização e acionar o botão “Gerenciar”, conforme a figura 7:



Figura 7: Opção para gerenciar organizações.

Após clicar em gerenciar, o coordenador tem a opção de alterar o perfil dos membros para Administrador ou Editor e também pode adicionar novos membros na organização (figura 8).

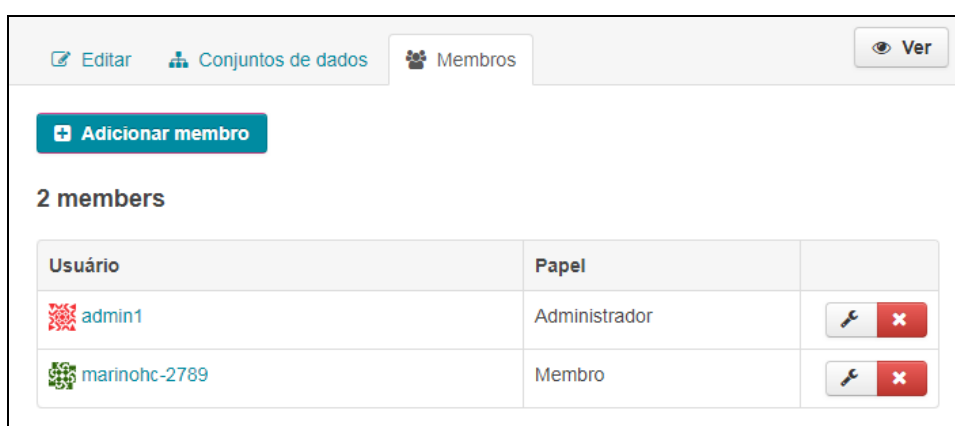


Figura 8: Tela para definir os perfis dos usuários.

Com relação ao conjunto de dados, o coordenador pode editar, remover ou tornar público/restrito um determinado conjunto que foi inserido na organização que coordena (figura 9).

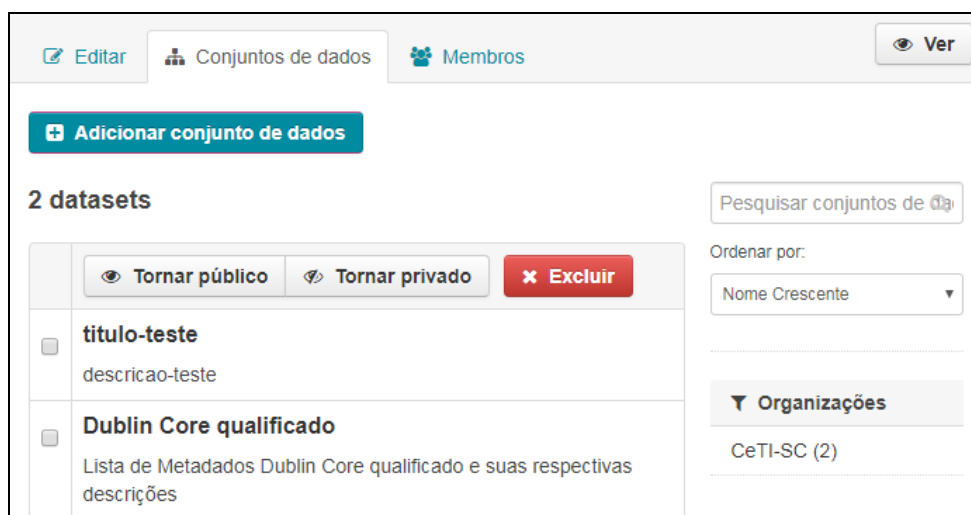


Figura 9: Tela para definir os perfis dos usuários.