Perfil DCAT BR para o CKAN

Foi feita uma versão inicial do perfil DCAT BR (govbr\_dcat\_ap) -Alterar para DCAT-BR- o plugin ckanext-dcat (<https://github.com/ckan/ckanext-dcat>) para uso no CKAN (<https://github.com/ckan/ckan>) da CGU, que serve como *backend* para o Portal de Dados Abertos do Governo Federal (<https://dados.gov.br>).

O perfil foi baseado nas orientações da tabela abaixo:



A seguir será apresentada a estrutura do perfil com algumas dúvidas que surgiram. Solicita-se gentilmente o esclarecimento dessas dúvidas bem como a validação do que foi feito.

# Conjunto de Dados

O conjunto de dados é representado por um recurso do tipo *dcat:Dataset*.

## Organização Associada

Os termos *foaf:Person* e *foaf:Organization* são na verdade recursos RDF e não propriedades. Portanto, mantive essa informação na propriedade *dct:publisher -utilizar a propriedade dct:creator, deixando a propriedade dct:publisher para a área técnica responsável-*, conforme o comportamento padrão do plugin *ckanext-dcat*.

A propriedade *dct:publisher* fica associada a um recurso *foaf:Organization* que se refere à URL da página do órgão no Portal de Dados Abertos. O recurso também possui a propriedade *foaf:name* com o nome da organização.

Quanto à possibilidade de usar um recurso do tipo *foaf:Person*, pelas regras do Portal de Dados Abertos atualmente não é possível cadastrar um conjunto sem organização associada. Portanto, este campo sempre terá um recurso do tipo *foaf:Organization*.

## Título do Conjunto

O título é representado literalmente pela propriedade *dct:title*.

## Descrição do Conjunto

A descrição é representada literalmente pela propriedade *dct:description*.

## Licença de Uso

A licença é representada literalmente pela propriedade *dcat:license*. Os valores possíveis habilitados nos Portal de Dados Abertos estão listados abaixo:

* cc-by
* cc-zero
* odc-odbl
* odc-pddl

O CKAN usa como referência para as licenças as definições do site <https://opendefinition.org/licenses/>.

**Dúvida:** a propriedade dcat:license deve se referir a uma URL desse site (Ex: <https://opendefinition.org/licenses/cc-by/>) ou deve ficar com o valor literal armazenado?

Ideal seria utilizar uma URL de algum site, por exemplo, o IANA que é de media types, mas apenas um exemplo (https://www.iana.org/assignments/media-types/media-types.xhtml).

## Periodicidade de atualização

A periodicidade é representada literalmente pela propriedade *dct:accrualPeriodicity*. Os valores possíveis habilitados no Portal de Dados Abertos estão listados abaixo:

* ANUAL
* DIARIA
* MENSAL
* OUTRAS
* QUINZENAL
* SEMANAL
* SOB\_DEMANDA
* TRIMESTRAL

**Dúvida:** a propriedade deve conter algum dos valores literais acima ou deve se referir a um esquema existente? Por exemplo, este é o esquema recomendado pelo Dublin Core: <https://www.dublincore.org/specifications/dublin-core/collection-description/frequency/>. Na Europa também existe essa especificação própria para este campo de periodicidade: <http://publications.europa.eu/resource/authority/frequency>.

Ideal seria utilizar uma URL de algum site, por exemplo, o IANA que é de media types, mas apenas um exemplo (<https://www.iana.org/assignments/media-types/media-types.xhtml>).

Será que o governo federal não tem algum esquema?

## Observância Legal

Este campo ainda não foi implementado no Portal de Dados Abertos, portanto ainda não foi incluído neste perfil.

Foi enviada em um segundo e-mail uma proposta do campo como nível de acesso, que atende o que precisamos nesse momento.

## Área técnica responsável pelo dado

Este campo é representado literalmente pela propriedade *dct:creator*.

Utilizar a propriedade dct:publisher

## E-mail da área técnica responsável

Este campo é representado literalmente pela propriedade *dcat:contactPoint*.

**Dúvida:** o padrão DCAT recomenda o uso da especificação VCARD (<https://www.w3.org/TR/vcard-rdf/>) para esta propriedade ([https://www.w3.org/TR/vocab-dcat-2/#Property:resource\_contact\_point](https://www.w3.org/TR/vocab-dcat-2/%23Property:resource_contact_point)). O VCARD, no entanto, já englobaria tanto um nome quanto e-mail. Devemos deixar esta propriedade apenas com o e-mail literal mesmo?

O ideal seria a utilização do VCARD, mas fica a critério da CGU, visto que não sabemos o trabalho que dará realizar essa atividade.

## Palavras-chave

As palavras-chave são representadas literalmente em propriedades *dct:keyword*.

## Versão

A versão é representada literalmente pela propriedade *adms:version*.

## Temas associados

Os temas associados são representados literalmente em propriedades *dcat:theme*.

Os temas disponíveis hoje no Portal de Dados Abertos são os seguintes:

* Agricultura, extrativismo e pesca (agricultura-extrativismo-e-pesca)
* Assistência e Desenvolvimento Social (assistencia-e-desenvolvimento-social)
* Ciência, Informação e Comunicação (ciencia-informacao-e-comunicacao)
* Comércio, Serviços e Turismo (comercio-servicos-e-turismo)
* Cultura, Lazer e Esporte (cultura-lazer-e-esporte)
* Dados Estratégicos (dados-estrategicos)
* Defesa e Segurança (defesa-e-seguranca)
* Economia e Finanças (economia-e-financas)
* Educação (educacao)
* Energia (energia)
* Equipamentos Públicos (equipamentos-publicos)
* Geografia (geografia)
* Governo e Política (governo-e-politica)
* Habitação, Saneamento e Urbanismo (habitacao-saneamento-e-urbanismo)
* Indústria (industria)
* Justiça e Legislação (justica-e-legislacao)
* Meio Ambiente (meio-ambiente)
* Pessoa, Familia e Sociedade (pessoa-familia-e-sociedade)
* Plano Plurianual (plano-plurianual)
* Relações Internacionais (relacoes-internacionais)
* Saúde (saude)
* Trabalho (trabalho)
* Transportes e Trânsito (transportes-e-transito)

**Dúvida:** o valor da propriedade deve ser o título do tema (Ex: Relações Internacionais) ou o id do tema (Ex: relacoes-internacionais) ?

Utilizar o título, visto que no futuro um RDF irá puxar essa informação.

Será que o governo federal não tem algum esquema?

## Recursos

Os recursos do conjunto de dados são representados nas propriedades *dcat:distribution*. Cada recurso é um elemento do tipo *dcat:Distribution* sem referência a URL (no Portal de Dados Abertos, não existe uma URL específica para o recurso, ele é exibido apenas na página do conjunto de dados).

### Título do Recurso

O título do recurso é representado literalmente pela propriedade *dct:title*.

### Descrição do Recurso

A descrição do recurso é representada literalmente pela propriedade *dct:description*.

### URL

A URL do recurso é representada pela propriedade *dcat:accessURL*.

### Tipo de Recurso

O tipo é representado literalmente pela propriedade *dct:type*. Os valores disponíveis no Portal de Dados Abertos são os seguintes:

* API
* DADOS
* DICIONARIO\_DE\_DADOS
* DOCUMENTACAO
* OUTRO

### Formato do Recurso

O formato é representado literalmente pela propriedade *dct:format*. O valor normalmente corresponde à extensão do arquivo (Ex: CSV, ZIP, JSON, etc).

**Dúvida:** a recomendação do padrão DCAT é usar o recurso *dct:MediaTypeOrExtent* ([https://www.w3.org/TR/vocab-dcat-2/#Property:distribution\_format](https://www.w3.org/TR/vocab-dcat-2/%23Property:distribution_format)) e no caso de formatos de arquivo se recomenda usar os tipos definidos pela IANA (<https://www.iana.org/assignments/media-types/>). Devemos apresentar este formato nos tipos definidos pela IANA?

Utilizar os tipos definidos pela IANA.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nome da propriedade | Propriedade enviada | Propriedade DCAT-BR v3 | Alteração | Tem no selo Bronze |
| Organização Associada | foaf:Person foaf:Organization | foaf:Person  foaf:Organization | dcterms:creator | OK |
| Título do Conjunto | *dct:title* | dcterms:title |  | OK |
| Descrição do Conjunto | *dct:description* | dcterms:description |  | OK |
| Licença de Uso | *dcat:license* | *dcterms:license* |  | OK |
| Periodicidade de atualização | dct:accrualPeriodicity | dcterms:accrualPeriodicity |  | OK |
| Observância Legal | dct:accessRights | dcterms:accessRights | seguir a proposta enviada por e-mail. Público, restrito e sigiloso igual ao SEI | OK |
| Área técnica responsável pelo dado | dct:creator | dcterms:creator | dcterms:publisher | OK |
| E-mail da área técnica responsável | dcat:contactPoint | dcat:contactPoint | Poderia utilizar VCARD, caso dê | OK |
| Palavras-chave | dct:keyword | dcat:keyword |  | OK |
| Temas associados | dcat:theme | dcat:theme |  | OK |
| Título do Recurso | dct:title | dcterms:title |  | OK |
| Descrição do Recurso | dct:description | dcterms:description |  | OK |
| URL | dcat:accessURL | dcat:accessURL |  | OK |
| Tipo de Recurso | dct:type | dcterms:type |  | OK |
| Formato do Recurso | dct:format | dcterms:format |  | OK |
| Versão | adms:version | dcat:version |  | OK |
| Área técnica responsável pelo dado |  | dcterms:publisher |  |  |

Teria que retirar do selo Bronze Possíveis alterações Inserir no selo bronze

Iremos enviar um arquivo com o novo selo bronze, ele está baseado no DCAT v3, aonde os nomes das propriedades alteram um pouco, favor verificar o arquivo para atualização das propriedades.