

## Catálogo Nacional de Dados do Governo Federal: análise da utilização do vocabulário DCAT-BR

*National Data Catalogue of the Federal Government: analysis of the use of DCAT-BR vocabulary*

**Katyusha Madureira Loures de Souza<sup>1</sup>**



<http://lattes.cnpq.br/9649842485745030>



<https://orcid.org/0000-0002-0617-8518>

**Mariana Baptista Brandt<sup>2</sup>**



<http://lattes.cnpq.br/3761037263199030>



<https://orcid.org/0000-0001-8119-7527>

**Silvana Aparecida Borsetti Gregorio Vidotti<sup>3</sup>**



<http://lattes.cnpq.br/7390573927636069>



<https://orcid.org/0000-0002-4216-0374>

### Resumo

O Catálogo Nacional de Dados (CND) foi criado pelo governo brasileiro com o objetivo de facilitar a descoberta de dados e promover a interoperabilidade de dados abertos entre órgãos públicos. Baseado no *Data Catalog Vocabulary* (DCAT) da W3C, o CND estrutura a descrição de conjuntos de dados na web com a especialização DCAT-BR, criada para este fim. Este artigo analisa a implementação do DCAT-BR no CND e sua eficácia para alcançar os objetivos de descoberta de dados e interoperabilidade. A pesquisa documental e bibliográfica revelou que a descoberta de dados é apenas parcialmente atendida, devido à falta de descrições detalhadas das variáveis dos conjuntos de dados. Em relação à interoperabilidade, identificou-se que o DCAT-BR não cobre todos os metadados necessários e apresenta redundâncias na relação entre alguns termos e metadados. Conclui-se, portanto, que são necessárias melhorias para que o CND cumpra seus objetivos de maneira mais abrangente.

**Palavras-chave:** dados abertos governamentais; catálogo de dados; vocabulário semântico; interoperabilidade; descoberta de dados.

### Abstract

The National Data Catalog (CND) was created by the Brazilian government to facilitate data discovery and promote the interoperability of open data among public agencies. Based on the Data Catalog Vocabulary (DCAT)

<sup>1</sup> Mestre em Ciência da Informação, Universidade de Brasília (UnB), Brasília, DF, Brasil. Analista em Ciência e Tecnologia, Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), DF, Brasil. [katyusha.souza@unesp.br](mailto:katyusha.souza@unesp.br).

<sup>2</sup> Doutora em Ciência da Informação, Universidade Estadual Paulista (Unesp), Marília, SP, Brasil. Analista legislativa, Câmara dos Deputados, Brasília, DF, Brasil. [mariana.brandt@unesp.br](mailto:mariana.brandt@unesp.br).

<sup>3</sup> Doutora em Educação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil. Professora do Departamento e do Programa de Pós-graduação em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Marília, São Paulo, Brasil. Coordenadora de Tecnologias Aplicadas (COTEA), Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasília, Distrito Federal, Brasil. [silvana.vidotti@unesp.br](mailto:silvana.vidotti@unesp.br).

from W3C, the CND structures the description of datasets on the web using DCAT-BR, a specialization created for this purpose. This article analyzes the implementation of DCAT-BR in the CND and its effectiveness in achieving the goals of data discovery and interoperability. Documentary and bibliographic research revealed that data discovery is only partially met due to the lack of detailed descriptions of dataset variables. Regarding interoperability, it was identified that DCAT-BR does not cover all necessary metadata and presents redundancies in the relationship between some terms and metadata. Therefore, it is concluded that improvements are needed for the CND to fully achieve its objectives in a more comprehensive manner.

**Keywords:** open government data; data catalog; semantic vocabulary; interoperability; data discovery.

## 1 INTRODUÇÃO

A disponibilização de dados abertos governamentais tem crescido nos últimos anos, especialmente após a promulgação da Lei de Acesso à Informação (LAI), que determina que os órgãos públicos devem divulgar as informações em seus sites de modo a “[...] possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina (Brasil, 2011)”. O governo brasileiro conta com um portal de dados abertos ([dados.gov.br](http://dados.gov.br)) que armazena 14.209<sup>4</sup> conjuntos de dados de 269 órgãos da administração pública. Além disso, vários órgãos publicam conjuntos de dados abertos de sua responsabilidade na própria página web do órgão.

Em dezembro de 2024 foi lançada uma nova ferramenta para disponibilização dos dados, o Catálogo Nacional de Dados (CND), com objetivos de descoberta de dados e interoperabilidade entre os órgãos (Brasil, 2024b). O CND utiliza o vocabulário semântico *Data Catalog Vocabulary* (DCAT), do *World Wide Web Consortium* (W3C), para descrição de conjuntos de dados, com uma especialização para o governo brasileiro denominada DCAT-BR.

A descoberta de dados é um objetivo alinhado aos objetivos bibliográficos, ou seja, os objetivos de um catálogo, que são: encontrar, identificar, selecionar, obter e navegar (IFLA, 2018). O objetivo deste artigo é analisar a implementação do DCAT no CND e sua potencialidade para atingir os objetivos de descoberta e de interoperabilidade dos dados propostos pela ferramenta. Para isso, foi realizada uma pesquisa documental nas fontes governamentais disponíveis sobre a elaboração do CND, nos portais de dados abertos dos órgãos e na documentação técnica do DCAT, além de pesquisa bibliográfica para referencial teórico.

---

<sup>4</sup> Consulta em 18 de janeiro de 2025.

## 2 O CATÁLOGO E SUAS FUNÇÕES

Por definição, um catálogo deve ser um instrumento eficaz e efetivo que permita ao usuário **encontrar** recursos bibliográficos numa coleção como resultado de uma pesquisa, utilizando atributos e relações entre recursos; **identificar** um recurso bibliográfico ou agente; **selecionar** um recurso bibliográfico que seja apropriado às necessidades do usuário; adquirir ou **obter** acesso a um item descrito e **navegar e explorar** dentro de um catálogo ou para além dele (IFLA, 2018). A *International Federation of Library Associations and Institutions* (IFLA) utiliza sempre o termo “recurso bibliográfico” devido a estar inserida no contexto bibliográfico, por ser uma Associação de Bibliotecas. Porém, os recursos descritos em um catálogo podem ser de outras naturezas sem alteração de suas funções e objetivos.

Em seu trabalho de 2008, Cunha e Cavalcanti, reuniram diversas definições sobre o termo ‘catálogo’ que abordam desde conceituações envolvendo a tipologia do documento - documento (secundário, lista, relação), função (analítico, entrada catalográfica), organização (alfabético, classificado por assunto, por autor, sistemático, analítico, assindético, bibliográfico, automatizado), por acesso (aberto, central, coletivo) e por função (comercial, de aquisição, de autoridade, dentre outros). À época, os autores já mencionavam o catálogo de dados com remissiva para o dicionário de dados (Cunha; Cavalcanti, 2008, p.73). Para esse termo, os autores propuseram a seguinte definição:

Inf 1. Programa associado com o sistema de gerenciamento de banco de dados e que mostra as definições dos dados usados pelo sistema. Funciona como auxiliar para a pessoa responsável pela administração e manutenção do banco de dados. 2. Listagem organizada de todos os elementos pertinentes ao sistema (Cunha; Cavalcanti, 2008, p. 73).

Os termos “catálogo de dados” e “dicionários de dados”, tratados como sinônimos à época, assumem características diferentes a partir dos avanços na produção, armazenamento e acesso a dados (big data). Diferente do dicionário de dados, que descreve dados de tabelas de bancos de dados internos das organizações e possuem função de documentação de sistemas, os catálogos de dados são ferramentas abrangentes para descrição de conjuntos de dados diversos, que podem ser de diferentes instituições e que são tratados aqui como recursos informacionais. Os catálogos de dados devem, assim, assumir as mesmas funções de um catálogo bibliográfico, permitindo aos usuários encontrar, identificar, selecionar e obter

os conjuntos de dados que descrevem, além de permitir a navegação, explorando as possibilidades de descoberta de dados no catálogo.

### 3 DADOS ABERTOS GOVERNAMENTAIS

Os dados produzidos pelos órgãos do governo são de interesse público e podem ser usados para vários fins. A transparência das informações é necessária para a manutenção dos governos democráticos e a publicação dos dados governamentais em formato aberto é uma das formas de viabilizar essa transparência. Assim, os vários órgãos publicam seus dados em seus próprios repositórios ou em repositórios coletivos. Esses repositórios necessitam de um cadastro dos dados publicados, ou seja, de um catálogo que traga informações sobre os dados, para que esses possam ser representados e encontrados.

Assim, o catálogo de dados torna-se um instrumento necessário para viabilizar a transparência ativa por meio da publicação dos dados produzidos: “Catálogos de Dados são ferramentas de gestão de metadados que apoiam a curadoria de dados oferecendo a capacidade de inventariar e descobrir dados em uma plataforma integrada, conectando a oferta com a demanda de dados”. (Jahnke; Otto, 2023, p. 90, tradução nossa).

Ou seja, para os autores o propósito fundamental é o inventariado e a descoberta desses dados. Além disso, os autores também mencionam outras possibilidades identificadas durante sua pesquisa, como a avaliação de dados (uso, qualidade, perfil), governança de dados (fluxos, papéis e responsabilidades, regras e políticas), colaboração de dados (*tagging*, compartilhamento, comentários) e análise de metadados (estória de dados, repositórios de aplicação de dados). Ademais, suas capacidades funcionais facilitam o apoio à visualização, administração e automação, incluindo características da inteligência artificial.

Como relatado na seção anterior, os catálogos buscam cumprir objetivos relacionados ao seu uso, em especial o de encontrar itens (livros, documentos, dados) necessários para o usuário. Esse é também o objetivo do Catálogo Nacional de Dados, quando se refere ao objetivo de descoberta de dados. O CND, objeto deste estudo, será descrito a seguir.

#### 3.1 Catálogo Nacional de Dados

O Catálogo Nacional de Dados é um projeto realizado a partir de uma parceria entre o Ministério da Gestão e da Inovação em Serviços Públicos (MGI) e a Controladoria-Geral da

União (CGU). Segundo o Guia rápido do Catálogo Nacional de Dados (Brasil, 2024c), o CND “é um instrumento para apoiar os órgãos na organização e disponibilização das informações sobre os dados em posse do governo a fim de que se tornem úteis para todos” e tem por objetivo “facilitar a descoberta e promover o reuso dos dados produzidos pelo governo”. O instrumento foi criado para identificar com clareza os dados existentes no governo, tornar os dados mais acessíveis, incentivar o uso e reuso de dados, transformar os dados em ativos informacionais e desenvolver produtos e políticas públicas (Brasil, 2024b).

Neste modelo, são catalogados os conjuntos de dados e os recursos em 3 níveis de catalogação, a depender dos metadados utilizados. Esses níveis são denominados “selos”, nas categorias ouro, prata e bronze. “Os conjuntos de dados recebem selos de acordo com a quantidade de campos preenchidos e a validação negocial dos campos. Total de 44 campos.” (Brasil, 2024a). A ideia é implementar os selos de acordo com um cronograma, que começa com o selo bronze e prevê a implementação do selo prata até outubro de 2025 e do selo ouro até abril de 2026. São esperados diversos benefícios advindos do catálogo, como o uso de dados para desenvolvimento de novos produtos e serviços, políticas públicas e para pesquisa e desenvolvimento. Segundo a cartilha, o catálogo “é uma ferramenta poderosa que pode ajudar o governo federal a melhorar seus serviços e a tomar decisões mais informadas.” (Brasil, 2024a, p. 6).

No dia 20 de dezembro de 2024, a Controladoria Geral da União (CGU), segundo matéria publicada em sua página eletrônica<sup>5</sup>, disponibilizou uma nova versão do Portal Brasileiro de Dados abertos com modificações para aumentar o grau de transparência e participação social a partir da incorporação das funcionalidades do Catálogo Nacional de Dados. Segundo a notícia publicada, o objetivo seria facilitar a descoberta de dados e promover a interoperabilidade entre os órgãos públicos que alimentam o portal.

Essa facilidade seria proporcionada pela catalogação também de conjuntos de dados não abertos, que segundo a CGU “promoverá maior transparência pública aos dados de posse dos órgãos públicos federais” e “trará mais robustez ao processo de elaboração dos Planos de Dados Abertos, pois contribuirá com a sua continuidade, ao registrar digitalmente o inventário de dados da Administração Pública”.

---

<sup>5</sup> Disponível em: <https://www.gov.br/cgu/pt-br/assuntos/noticias/2024/12/cgu-lanca-novas-funcionalidades-no-portal-de-dados-abertos>.

A notícia destaca melhorias no portal de dados, incluindo a criação do selo Bronze, aplicado a conjuntos de dados que atendem aos requisitos mínimos do padrão DCAT-BR Bronze, por meio de um *endpoint* de *harvesting* que busca RDFs diretamente das URLs de órgãos públicos que não utilizam CKAN. Outras melhorias incluem novos campos no detalhamento dos conjuntos de dados, checagem periódica de atualizações, identificação de dados sensíveis sobre raça/etnia para apoiar políticas públicas de ações afirmativas e aprimoramento da catalogação, com sinalização sobre a abertura dos dados, questões legais, relação com os Objetivos de Desenvolvimento Sustentável (ODS) e direcionamento para dados relacionados.

A atualização também incluiu a possibilidade de indicação da área técnica responsável pela catalogação (com respectivo e-mail, bem como sua versão e a existência do atributo de gênero). O cadastro da organização para publicação no Portal de Dados Abertos inclui o contato com a CGU e o preenchimento de um formulário com 18 campos (13 obrigatórios), que inclui o cadastro no Sistema de Organização e Inovação Institucional do Governo Federal (SIORG)<sup>6</sup>. Após o cadastro da organização também é necessário cadastrar o gestor da organização que ficará responsável pelo cadastramento dos dados. A indicação também deverá ser realizada através do contato com a CGU.

O cadastro de dados da organização pode ser realizado em duas etapas: conjunto de dados e recursos (que estão vinculados ao conjunto cadastrado). Na primeira etapa, é preciso informar o título do conjunto de dados, observância legal, descrição, licença de uso, área técnica responsável, periodicidade de atualização, tema, relacionamento com os Objetivos de Desenvolvimento Sustentável (uma parceria do Brasil com a ONU, lembrando também a Agenda 2030), se o conjunto possui dados sensíveis, palavras-chave que descrevem o conjunto de dados, sua cobertura temporal, espacial, versão (inclusive se foi descontinuado), visibilidade e associações (interoperabilidade). Na segunda etapa, referente aos recursos, basta informar o título do recurso, URL de acesso, tipo do recurso, formato e uma descrição.

O dicionário do que significa cada campo para que o usuário possa esclarecer dúvidas quanto ao seu preenchimento está disponível em um material do tipo "saiba mais" na página do CDN, no entanto não se identificou *tooltips* na tela de cadastramento do material.

---

<sup>6</sup> Saiba como publicar um conjunto de dados. Disponível em: <https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-conjunto-de-dados>. Acesso em 24 fev. 2025.

### 3.2 Vocabulários Semânticos para Descrição de Dados: DCAT e DCAT-BR

Em sua página, a W3C (2024) informa que o *Data Catalog Vocabulary* (DCAT), que está em sua 3a. versão, é um vocabulário do RDF desenvolvido para facilitar a interoperabilidade entre catálogos de dados publicados na web, na qual define um modelo e fornece exemplos para seu uso. A interoperabilidade pode ser organizacional, técnica ou semântica, e se refere a capacidade de trabalho conjunto entre sistemas e organizações garantindo que pessoas, organizações e sistemas computacionais possam interagir para trocar informações de maneira eficaz e eficiente (Brasil, 2012). No caso da interoperabilidade entre catálogos, trata-se da interoperabilidade semântica, que se refere

[...] a capacidade de dois ou mais sistemas heterogêneos e distribuídos trabalharem em conjunto, compartilhando as informações entre eles com entendimento comum de seu significado (Buranarach, 2004). A interoperabilidade semântica garante que os dados trocados tenham seu efetivo significado corretamente interpretado dentro do contexto de uma dada transação ou busca de informação, dentro da cultura, convenções e terminologias adotadas por cada setor ou organização e, assim, compartilhados pelas partes envolvidas. (Brasil, 2012, p.8)

Assim, segundo o W3C, o DCAT permite a um editor descrever conjuntos de dados e serviços de dados em um catálogo usando modelos padronizados e vocabulário que facilite o consumo e unificação de metadados de diversos catálogos. A solução tem como proposta tornar possível a publicação de catálogos de dados de forma descentralizada. A versão 3 deste catálogo buscou acrescentar casos de uso, requisitos e a experiência da comunidade envolvida (*Data eXchange Working Group, s.i.c.*) apoiando abordagens relacionadas à descrição e intercâmbio de dados.

A solução foi originalmente desenvolvida e hospedada na *Digital Enterprise Research Institute* (DERI), um fundo mantido pela *Science Foundation Ireland* (SFI) que era responsável pelo plano nacional de desenvolvimento em estratégias para ciência tecnologia e inovação no período de 2007 a 2013. Em 2014, o *eGov Interest Group* passou a trabalhar no refino e depois padronizado pelo *Government Linked Data* (GLD) *Working Group*, ambos do W3C. Sua segunda versão foi desenvolvida pela *Dataset Exchange Working Group* e publicada em 2020 e sua terceira versão foi publicada em agosto de 2024.

Segundo o W3C (2024), o DCAT incorpora termos de vocabulários pré-existentes onde poderia-se identificar uma terminologia clara com significado apropriado, mas também inclui

definições informais resumidas externas para conveniência. As definições autorizadas são disponibilizadas nas referências normativas.

O DCAT foi utilizado no Catálogo Nacional de Dados como uma versão brasileira denominada DCAT-BR: “O DCAT-BR, especialização do DCAT, é um vocabulário do Brasil, um perfil usado para descrever conjuntos de dados do setor público e vinculados de forma padronizada” (Brasil, 2024a, p. 14). Segundo o órgão gestor, o DCAT-BR teve seu vocabulário e taxonomias adequados aos aspectos e legislações brasileiras, como a Lei Geral de Proteção de Dados Pessoais (LGPD). Anteriormente, o DCAT já havia ganhado especializações de outros governos, como o DCAT-AP (*Area of Policy*, para dados de interesse público), da Comissão Europeia e o DCAT-AP-NO, do governo da Noruega, para dados e serviços específicos do país (Brasil, 2024b).

Espera-se que, com o uso do vocabulário DCAT-BR, seja possível estabelecer uma consistência entre os agregadores de conteúdo e portais de conjuntos de dados por meio da padronização da descrição dos metadados dos conjuntos de dados. Acredita-se também que o vocabulário irá fornecer bases para a interoperabilidade necessária com outras interfaces. A especialização brasileira do vocabulário é mantida no âmbito da Política de Governança de Dados Nacional, pelo Grupo de Trabalho DCAT-BR, que é responsável pelas atualizações e alterações nas especificações do vocabulário (Brasil, 2024b).

#### 4 PROCEDIMENTOS METODOLÓGICOS

Como mencionado, o universo desta pesquisa é o Catálogo Nacional de Dados (CND) do Governo Federal do Brasil, mantido pelo Ministério da Gestão e da Inovação em Serviços Públicos, em parceria com a Controladoria-Geral da União. O CND está disponível no endereço eletrônico <https://dados.gov.br/home>.

O acesso ao CND foi realizado entre os meses de janeiro e fevereiro do ano de 2025, com o objetivo de analisar empiricamente os dados e metadados utilizados, bem como a implementação do vocabulário semântico DCAT. Para embasar a análise também foram pesquisadas fontes documentais disponibilizadas pelos gestores do CND como materiais de

apoio<sup>7</sup>: Cartilha do Catálogo Nacional de Dados; Guia rápido do Catálogo Nacional de Dados e Catálogo Nacional de Dados: padrão DCAT-BR

Além disso, procedeu-se à pesquisa documental em fontes governamentais relacionadas à elaboração do CND, nos portais de dados abertos dos órgãos governamentais, bem como na documentação técnica do DCAT. Complementarmente, realizou-se uma pesquisa bibliográfica para fundamentação do referencial teórico. A análise da ferramenta considerou critérios relacionados à interoperabilidade, em consonância com as diretrizes estabelecidas na literatura especializada e nas normas vigentes no Brasil.

## 5 RESULTADOS

A partir da análise da documentação disponível sobre o CND, foram identificados os metadados utilizados para a descrição dos conjuntos de dados, denominados “campos” na Cartilha do CND (Brasil, 2024a), e os termos semânticos do vocabulário DCAT-BR, obtidos no Catálogo Nacional de Dados: padrão DCAT-BR (Brasil, 2024b). Com isso, foi realizado um mapeamento dos metadados por meio da sistematização e consolidação dessas informações em um quadro, considerando os termos do vocabulário semântico correspondentes em cada nível de descrição proposto pelo catálogo: bronze, prata e ouro. O resultado é apresentado no Quadro 1.

**Quadro 1: Listas de Metadados e termos semântico do vocabulário DCAT-BR no Catálogo Nacional de Dados por nível de descrição**

Bronze	Prata	Ouro
Organização (dcterms:Creator)	Área Geográfica (dcterms:spatial)	Substituído Por (dcterms:replaces)
Título (dcterms:title)	Data de Publicação (dcterms:issued)	Conjuntos de Dados Template
Descrição (dcterms:description)	Data de Alteração (dcterms:modified)	É Referenciado Por (dcterms:isReferencedBy)
Licença de uso (dcterms:license)	Idioma (dcterms:language)	Atualização de Versão (dcat:hasVersion)
Periodicidade de Atualização (dcterms:accrualPeriodicity)	Situação (adms:status)	URL de Obtenção de Acesso (dcat:landingPage)
Observância Legal (dcterms:accessRights)	Observância Legal Complementar (dcterms:rights)	Cobertura Temporal Fim (dcterms:temporal)

<sup>7</sup> Disponível em: <https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/catalogo-nacional-de-dados/materiais-de-apoio>

Bronze	Prata	Ouro
Área Técnica Responsável (dcterms:Publisher)	Possui Relação com Objetivos de Desenvolvimento Sustentável (ODS)? (não possui termo semântico)	Cobertura Temporal Início (dcterms:temporal)
E-mail da Área Técnica Responsável (dcat:contactPoint)	Possui Dados de Gênero? (não possui termo semântico)	Cobertura Espacial
Palavras-Chave (dcat:keyword)	Possui Dados de Raça/Etnia? (não possui termo semântico)	Granularidade Espacial
Versão (dcat:version)	URL de Acesso (dcat:accessURL)*	Data da Descontinuação
Temas (dcat:theme)	Tipo de Mídia (dcat:mediaType)*	Conjuntos de Dados Associados (Interoperabilidade)
Dados abertos (não possui termo semântico)	Descrição do Recurso (dcterms:description)*	Tamanho em Bytes (dcat:byteSize)*
Visibilidade (não possui termo semântico)	Formato (dcterms:format)*	Observância Legal do Recurso (dcterms:accessRights)*
13	Título do Recurso (dcterms:title)*	Data de Publicação do Recurso (dcterms:issued)*
	14	Licença de Uso do Recursos (dcterms:license)*
		Data de Alteração do Recurso (dcterms:modified)*
		Verificação de Conteúdo (spdx:checksum)*
		17

Fonte: Baseado na Cartilha do CND e no Vocabulário Padrão DCAT-BR (os metadados de recursos foram indicados com asterisco).

Conforme descrito na Cartilha, há metadados para descrever recursos (em negrito no quadro 1) e metadados para descrever conjuntos de dados. A Cartilha não explica os termos, mas foram encontradas na página “Saiba como publicar um conjunto de dados”, do portal de Dados abertos do Gov.br, as seguintes definições: “Conjunto de Dados: Refere-se a uma coleção organizada de informações ou dados relacionados a um tema específico. [...] Recursos: São as representações individuais ou arquivos específicos que compõem um conjunto de dados” (Brasil, 2024a). Em termos simples, os recursos são itens específicos dentro de um conjunto de dados. Como continuidade da análise, os metadados foram avaliados quanto a sua utilidade para a descoberta de dados, ou seja, quais campos podem auxiliar o usuário a encontrar os dados de que necessita? O resultado da análise é apresentado no Quadro 2.

**Quadro 2 - Metadados para descrição dos dados**

<b>Metadados para DESCRIÇÃO dos dados</b>
1. Título (dcterms:title)
2. Descrição (dcterms:description)
3. Palavras-Chave (dcat:keyword)
4. Temas (dcat:theme)
5. Área Geográfica (dcterms:spatial)
6. Idioma (dcterms:language)
7. Possui Dados de Gênero?
8. Possui Dados de Raça/Etnia?
9. Descrição do Recurso (dcterms:description)
10. Título do Recurso (dcterms:title)
11. Cobertura Temporal Fim (dcterms:temporal)
12. Cobertura Temporal Início (dcterms:temporal)
13. Cobertura Espacial
14. Granularidade Espacial
15. Organização (dcterms:creator)
16. Periodicidade de Atualização (dcterms:accrualPeriodicity)
17. Data de Publicação (dcterms:issued)
18. Observância Legal Complementar (dcterms:rights)
19. URL de Obtenção de Acesso (dcat:landingPage)
20. Data de Alteração do Recurso (dcterms:modified)
21. Data de Alteração (dcterms:modified)
22. Conjuntos de Dados Associados (Interoperabilidade)

Fonte: Elaboração própria.

Dos 44 metadados propostos pelo CND, observou-se que 22 atendem ao objetivo de descoberta de dados. Os metadados que possuem maior potencial de contribuir para o objetivo de descoberta de dados são os metadados de descrição, pois tratam do conteúdo informacional dos dados descritos. Atributos como tema, palavras-chave, título, descrição, abrangência geográfica e temporal, entre outros, se referem ao conteúdo dos dados e fornecem bases para que o usuário encontre, identifique e selecione o item, ou seja, para a descoberta de dados.

Contudo, apesar da descoberta de dados ser um dos objetivos do CND e de que metade dos metadados propostos para catalogação dos dados sirvam a esse propósito, observa-se que o conteúdo mais relevante dos dados em si, que corresponde às variáveis contidas no conjunto de dados, não é obrigatoriamente registrado na catalogação do item. Por exemplo, se o usuário busca dados sobre "Taxa de participação na força de trabalho, por gênero, raça e faixa etária, de 2000 a 2024", é preciso que o conjunto de dados contenha as variáveis:

- a) pessoas na força de trabalho (isso pode não estar em um metadado);
- b) gênero;

- c) raça e
- d) faixa etária (que possa contemplar faixas pré-definidas em uma tabela).

As informações de gênero e raça estariam registradas nos metadados “Possui Dados de Gênero?” e “Possui Dados de Raça/Etnia?”, e o período dos dados nos metadados “Cobertura Temporal Início” e “Cobertura Temporal Fim”. Assim, os dados sobre as demais variáveis necessárias (força de trabalho, considerando pessoas em idade de trabalhar e faixa etária), não seriam descritas por nenhum metadado e mesmo que o catalogador utilizasse o `dcat:keywords`, seria necessário um mínimo de padronização nos registros para evitar níveis diferentes de especificidade da descrição.

Atualmente, a única forma de identificar quais dados um conjunto apresenta acontece apenas se o catalogador informar, durante a catalogação, no metadado “descrição” ou “descrição do recurso”, quais são essas variáveis. A outra forma de identificação só é possível após a obtenção dos dados. Porém, sendo o catálogo um instrumento destinado à identificação e seleção de itens, não deveria ser necessário que o usuário acessasse o conjunto de dados para confirmar se o conteúdo atende às suas necessidades.

Quanto à interoperabilidade, a documentação analisada indica que ela será proporcionada pelo uso do vocabulário semântico DCAT-BR, uma vez que este “fornece interoperabilidade semântica com outras interfaces de consumo de metadados com base na reutilização de vocabulários controlados estabelecidos e mapeamentos para vocabulários de metadados existentes.” (Brasil, 2024b, p. 6). Contudo, a partir do mapeamento dos metadados e termos do DCAT-BR apresentados no Quadro 1, verificou-se a existência de metadados sem termo semântico correspondente, bem como metadados diferentes que utilizam o mesmo termo semântico do DCAT-BR. Tal configuração pode comprometer a interoperabilidade semântica. Além disso, ressalta-se que essa interoperabilidade ocorre entre os metadados de representação dos dados do catálogo, e não entre os dados contidos nos conjuntos de dados e recursos catalogados.

## 6 CONSIDERAÇÕES FINAIS

O Catálogo Nacional de Dados atende ao que se propõe no que tange a servir como uma ferramenta para inventariar os dados produzidos pelo governo federal e incentivar a interoperabilidade semântica por meio da descrição dos dados dos diferentes órgãos de forma

padronizada, conforme disposto na literatura e normas disponíveis. No entanto, considerando-se os objetivos do CND como os de promover a descoberta de dados e a interoperabilidade semântica, identificou-se a possibilidade de aperfeiçoar a ferramenta.

Em relação ao objetivo da descoberta de dados, sugere-se a complementação da descrição dos dados com a inclusão de metadados que informem as variáveis que o recurso possui. Com isso, seria possível que o usuário soubesse se o conjunto de dados o atende antes mesmo de acessá-lo, cumprindo a função de um catálogo. O vocabulário DCAT não atenderia a essa necessidade, já que o mesmo não possui previsão de descrição de variáveis de dados em seus termos. No entanto, isso poderia ser incluído na especialização DCAT-BR.

Quanto ao objetivo de interoperabilidade, os termos do DCAT-BR não encontraram correspondência em todos os metadados para catalogação dos recursos e conjuntos de dados, o que pode prejudicar o processo. Observou-se também o uso de termos iguais para metadados diferentes, podendo ocasionar problemas na interoperabilidade.

A interoperabilidade semântica permite a colaboração e troca de dados e informações, independentemente de diferentes estruturas, processos de negócio e formas de operação em cada instituição, mas demanda que as organizações envolvidas conheçam seus processos de trabalho de forma a seguir padrões já utilizados. Caso contrário, será necessário trabalhar em outras soluções que contemplem a coleta e tratamento de dados e informações para o alinhamento destes, o que deverá ser realizado sempre que for necessário obter dados novos ou atualizados. Com a interoperabilidade semântica, esse trabalho à parte não é necessário, basta que seja feita uma única análise da compreensão necessária. Assim, padrões e processos já utilizados são mantidos, sendo necessário apenas definir os momentos para acionar as interações entre os sistemas.

## REFERÊNCIAS

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Brasília, Presidência da República, 2011. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/ato2011-2014/2011/lei/l12527.htm](https://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm). Acesso em: 13 jan. 2025.

BRASIL. Ministério da Gestão e Inovação em Serviços Públicos. Secretaria de Governo Digital. **Cartilha do catálogo nacional de dados**. Brasília: MGI, [2024a]. Disponível em: <https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/catalogo-nacional-de-dados/materiais-de-apoio/cartilha-do-catalogo-nacional-de-dados/@@download/file>. Acesso em: 13 jan. 2025.

BRASIL. Ministério da Gestão e Inovação em Serviços Públicos. Secretaria de Governo Digital. **Catálogo nacional de dados**: vocabulário padrão DCAT-BR. Brasília: MGI, [2024b]. Disponível em: <https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/catalogo-nacional-de-dados/materiais-de-apoio/vocabulario-padrao-dcat-br-v1.0/view>. Acesso em: 13 jan. 2025.

BRASIL. Ministério da Gestão e Inovação em Serviços Públicos. Secretaria de Governo Digital. **Guia rápido do Catálogo Nacional de Dados**. Brasília: MGI, [2024c]. Disponível em: <https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/catalogo-nacional-de-dados/materiais-de-apoio/guia-rapido-do-catalogo-nacional-de-dados>. Acesso em: 13 jan. 2025.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Guia de interoperabilidade**: manual do gestor. Brasília: MP, 2012. Disponível em: <https://bibliotecadigital.economia.gov.br/handle/777/604>. Acesso em: 13 jan. 2025.

BRASIL. Portal Brasileiro de Dados Abertos. **Saiba como publicar um conjunto de dados**. [S. l.], [2025?]. Disponível em: <https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-conjunto-de-dados>. Acesso em 22 fev. 2025.

CUNHA, M. B.; CAVALCANTI, C. R. de O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008. 451 p.

IFLA. **Declaração dos Princípios Internacionais de Catalogação (PIC)**. 2018. Disponível em: <https://repository.ifla.org/items/195ded68-f032-4314-ac6e-a7ceafeef650>. Acesso em: 22 fev. 2025.

JAHNKE, N.; OTTO, B. Data catalogs in the enterprise: applications and integration. **Datenbank Spektrum**, n. 23, p. 89-96, jul. 2023. Disponível em: <https://doi.org/10.1007/s13222-023-00445-2>. Acesso em 22 fev. 2025.

WORLD WIDE WEB CONSORTIUM. **Data Catalog Vocabulary (DCAT)**: version 3. Disponível em: <https://www.w3.org/TR/vocab-dcat-3/>. Acesso em 13 jan. 2025.



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

**Como citar este trabalho:**

SOUZA, Katyusha Madureira Loures de; BRANDT, Mariana Baptista; VIDOTTI, Silvana Aparecida Borsetti Gregorio. *Catálogo Nacional de Dados do Governo Federal: análise da utilização do vocabulário DCAT-BR*. In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 8., Marília, SP. **Anais [...]**. Marília, SP: Universidade de Marília. 2025. DOI: <https://doi.org/10.22477/viii.widat.253>.