

METAOPENGOV: MODELO DE METADADOS PARA O PORTAL BRASILEIRO DE DADOS ABERTOS

MetaOpenGov: metadata model for the Brazilian Open Data Portal

Danielle Teixeira de Oliveira


Universidade Federal de Minas Gerais, Escola de Ciência da Informação
Belo Horizonte, MG, Brasil
biblio.danielleoliveira@gmail.com

<https://orcid.org/0000-0002-1958-9113> 

Patrícia Nascimento Silva

Universidade Federal de Minas Gerais, Escola de Ciência da Informação
Belo Horizonte, MG, Brasil
patricians@ufmg.br

<https://orcid.org/0000-0002-2405-8536> 

A lista completa com informações dos autores está no final do artigo 

RESUMO

Objetivo: propor um modelo de metadados para a representação e a organização da informação no Portal Brasileiro de Dados Abertos. Especificamente, buscou-se (i) identificar e analisar as tipologias e os formatos de dados abertos publicados, padrões e metadados utilizados no Portal e (ii) propor uma estrutura de representação para os dados abertos do governo federal brasileiro com a utilização de metadados.

Método: realizou-se uma pesquisa aplicada, exploratória e descritiva, com abordagem qualitativa e quantitativa, dividida em duas etapas, conforme os objetivos específicos. Foram utilizadas as técnicas de pesquisa bibliográfica e documental e a observação direta no Portal Brasileiro de Dados Abertos, entre julho e novembro de 2024.

Resultado: a partir da identificação do contexto brasileiro sobre dados abertos, padrões e elementos de organização e representação da informação, os resultados das etapas foram sistematizados para a elaboração do modelo de metadados MetaOpenGov. O modelo é baseado em três pilares: legislação e documentação; padrões e instrumentos de organização e representação da informação; e a estrutura existente no Portal, propondo recomendações para a estrutura de hierarquização, a divisão em categorias (organização, conjunto de dados, recurso e reúso) e 84 metadados em correspondência com o padrão *Data Catalog Vocabulary 3*, com sua respectiva validação.

Conclusões: concluiu-se que as recomendações dessa pesquisa podem apoiar mudanças significativas no contexto brasileiro, promovendo a interoperabilidade e a acessibilidade para a efetiva reutilização dos dados pela sociedade, destacando o papel essencial do profissional da informação no desenvolvimento das políticas de dados abertos.

PALAVRAS-CHAVE: Dados abertos. Portal Brasileiro de Dados Abertos. Representação da informação. Metadados. Interoperabilidade.

ABSTRACT

Objective: to propose a metadata model for the representation and organization of information on the Brazilian Open Data Portal. Specifically, we sought to (i) identify and analyze the types and formats of open data published, standards and metadata used on the Brazilian Open Data Portal and (ii) propose a descriptive representation structure for the Brazilian federal government's open data on the Brazilian Open Data Portal using metadata.

Method: this is an applied, exploratory and descriptive study with a qualitative and quantitative approach, divided into two stages according to the specific objectives. Bibliographic and documentary research techniques were used, as well as direct observation on the Brazilian Open Data Portal, between July and November 2024.

Results: after identifying the Brazilian context of open data, standards and elements for organizing and representing information, the results of the stages were systematized to create the MetaOpenGov metadata model. The model is based on three pillars: legislation and documentation; standards and instruments for organizing and representing information; and the existing structure of the Portal, proposing recommendations for the hierarchical structure, division into categories (organization, datasets, resource and reuse) and 84 metadata corresponding to the Data Catalog Vocabulary 3 standard and its respective validation.

Conclusions: it was concluded that the recommendations of this research can support significant changes in the Brazilian context, promoting interoperability and accessibility for the effective reuse of data by society, highlighting the essential role of information professionals in the development of open data policies.

KEYWORDS: Open data. Brazilian Open Data Portal. Information representation. Metadata. Interoperability.

1 INTRODUÇÃO

No contexto tecnológico vivenciado na segunda década do século XXI, os dados têm se tornado um recurso fundamental nos mais diversos setores da sociedade, motivando o desenvolvimento de aplicações e revolucionando o fluxo informacional. A produção e o consumo massivos de dados advindos de fontes e formatos heterogêneos caracterizam o fenômeno conhecido como *big data*, que impõe novos desafios aos profissionais da informação para realizar a gestão da informação na *web*.

Entre a variedade dos dados disponíveis, estão os dados abertos, que são aqueles que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa com, no máximo, uma exigência de atribuição de fonte e compartilhamento com as mesmas regras e possuem enorme potencial para promover transparência e engajamento social (Open Knowledge Brasil, [20–?]).

A discussão sobre a conceituação de dados e informação vem desde os primórdios da Ciência da Informação (CI), sendo que não existe um consenso absoluto entre os estudiosos da área. Davenport (2002) considera os dados como simples observações sobre o estado do mundo, e a informação como dados dotados de relevância e propósito. No cenário tecnológico contemporâneo, em que esses conceitos estão intrinsecamente relacionados, lidar com esse imensurável volume de dados tornou-se um desafio primário para diversos profissionais que trabalham com a informação, sobretudo em formato digital. Considerando que os dados estão diretamente relacionados à informação, eles também são propriedade de estudo da CI. Os dados abertos são subdivididos em vários tipos, podendo ser dados científicos, financeiros, culturais, ambientais e governamentais, sendo esta última tipologia abordada especificamente no presente trabalho.

O Portal Brasileiro de Dados Abertos é uma ferramenta oficial de disponibilização de dados e informações públicas do Governo Federal do Brasil. Ele se propõe a ser um serviço simplificado que organiza e padroniza o acesso aos dados públicos, priorizando o seu reuso e a utilização de tecnologias modernas. O portal funciona como um catálogo de dados federado, facilitando a busca e utilização de dados publicados pelos órgãos do governo, que os disponibilizam conforme os cronogramas estabelecidos em seus Planos de Dados abertos (PDAs) (Brasil, [20–?]). Tanto a elaboração do PDA quanto a publicação dos dados são responsabilidade de cada órgão.

Com a expansão dos Dados Governamentais Abertos (DGA), os problemas de representação e recuperação relacionados à grande quantidade e diversidade de dados e

informações são perceptíveis, sendo essencial conhecer as abordagens e os métodos utilizados para representar e recuperar esses dados (Nascimento Silva, 2023). Após diversas legislações, somadas aos padrões do governo, como a Infraestrutura Nacional de Dados Abertos (INDA), portarias, normativas e resoluções, observa-se um esforço em todas as esferas para o cumprimento das legislações que dispõem sobre o acesso à informação e a disponibilização de dados abertos. No entanto, a padronização existente ainda é limitada, refletida principalmente na Política de Dados Abertos, que preconiza recomendações e boas práticas, mas não detalha a organização dos dados e informações. Essa situação resulta em publicações de dados dispersos e fragmentados, dificultando sua recuperação e utilização e impossibilitando a sua reutilização em outros contextos (Nascimento Silva, 2018; Nascimento Silva, 2022).

Frente à importância de se representar os dados abertos do governo brasileiro de forma padronizada e interoperável, a questão de pesquisa que norteou este estudo foi: quais metadados podem ser utilizados para representar os dados abertos do governo federal brasileiro? O estudo tem como objetivo geral propor um modelo de metadados para a representação e a organização da informação no Portal Brasileiro de Dados Abertos. Especificamente, buscou-se (i) identificar e analisar as tipologias e os formatos de dados abertos publicados, os padrões e metadados utilizados no Portal e (ii) propor uma estrutura de representação da informação para os dados abertos do governo federal brasileiro com a utilização de metadados.

O estudo justifica-se por abordar uma lacuna da Representação do Conhecimento e da Informação na Web, que, conforme apontado por Lima (2020), envolve o grande quantitativo de dados não estruturados e a dificuldade de recuperação, sendo necessárias estruturas de representação e ferramentas de acesso que provenham o tratamento desse novo ambiente e tipo de recurso informacional, representado nesse estudo pelo conceito de *open government data*, dados abertos do governo, dados abertos governamentais, ou DGA.

Este artigo apresenta os resultados de uma pesquisa acadêmica defendida na Universidade Federal de Minas Gerais, em dezembro de 2024, por Oliveira (2025). Além disso, a pesquisa integra o grupo de pesquisa Observatório de Dados Abertos (Nascimento Silva, 2025), que monitora os dados abertos do governo brasileiro.

2 PROCEDIMENTOS METODOLÓGICOS



A pesquisa em questão caracteriza-se como aplicada, com a finalidade de resolver problemas reais; exploratória, objetivando desenvolver e esclarecer conceitos e proporcionar uma visão geral sobre um fato; e descritiva, visando descrever características de um determinado fenômeno e estabelecer relações entre variáveis (Gil, 2017). Como instrumentos, foram utilizados a pesquisa bibliográfica e documental e a observação direta com abordagem qualitativa e quantitativa.

A pesquisa foi dividida em duas etapas, uma para cada objetivo específico. Para o cumprimento do primeiro objetivo específico (Etapa 1): identificar e analisar as tipologias e os formatos de dados abertos publicados, os padrões e metadados utilizados no Portal Brasileiro de Dados Abertos, foi realizada uma observação direta em conjuntos de dados apresentados e a identificação de metadados nos manuais para publicadores de dados, disponíveis no Portal. A pesquisa bibliográfica e documental subsidiou o estudo de padrões e boas práticas existentes para a publicação de dados na *web*, um mapeamento das legislações vigentes no Brasil, bem como a identificação de trabalhos correlatos. Destaca-se que os resultados detalhados dessa etapa foram publicados em Oliveira e Nascimento Silva ([2025]).

Para o cumprimento do segundo objetivo específico (Etapa 2): propor uma estrutura de representação da informação para os dados abertos do governo federal brasileiro com a utilização de metadados, foi utilizada a observação direta, juntamente com a sistematização dos resultados da Etapa 1.

A observação direta consistiu na navegação contínua pelo Portal como usuário comum para identificar e compreender elementos da organização e representação da informação presentes na ferramenta. Durante os meses de julho a novembro de 2024, aproximadamente, o Portal Brasileiro de Dados Abertos foi acessado diariamente por meio do endereço eletrônico: <https://dados.gov.br/home> e foram examinadas as opções de navegação e consulta disponíveis. Foram acessados os menus disponíveis e utilizadas as ferramentas de busca para analisar a apresentação das informações para o usuário.

Após o mapeamento dos metadados do Portal, foram analisados os padrões de metadados compatíveis e selecionado o padrão Data Catalog Vocabulary 3 (DCAT-3) para realizar as proposições de representação dos dados abertos do governo federal. O padrão DCAT-3¹ foi selecionado por estar alinhado com as melhores práticas para dados na *web*, sendo, inclusive, uma das recomendações do World Wide Web Consortium (W3C), por

¹ <https://www.w3.org/TR/vocab-dcat-3/>.

promover a sustentabilidade ao utilizar-se de elementos advindos de outros vocabulários (Tomoyose, 2021) e por ser um padrão específico para a descrição de catálogos governamentais.

A escolha pelo padrão DCAT-3 justifica-se, também, pela ampla adoção desse modelo por comunidades internacionais, consolidando-o como uma referência para a descrição e catalogação de dados governamentais em diferentes contextos. Países membros da União Europeia e diversas iniciativas multilaterais já utilizam o DCAT como base para promover a interoperabilidade semântica, facilitar o intercâmbio de informações e garantir maior transparência na disponibilização de dados abertos. Ao adotar o DCAT-3 no contexto nacional, promove-se o alinhamento do Brasil às boas práticas internacionais, ampliando a sua compatibilidade com plataformas globais.

Com base na análise da Etapa 1 e as boas práticas para a publicação de dados na *web*, procedeu-se à organização hierárquica da estrutura de metadados observada no Portal, visando a uma sistematização preliminar dos elementos a serem modelados. Posteriormente, foi proposta uma estrutura de metadados, formulada a partir do conhecimento adquirido durante o desenvolvimento da pesquisa e com base nas necessidades específicas identificadas na observação direta para os dados abertos do governo brasileiro.

Uma preocupação surgida durante as etapas anteriores foi a adoção de boas práticas para a representação e a disponibilização de dados na *web* e, por isso, estudou-se com mais profundidade uma recomendação muito difundida e utilizada no cenário trabalhado para apresentar uma proposta consistente e pertinente à discussão apresentada: as Boas Práticas para Dados na Web² (*Data on the Web Best Practices*) traduzido por Lóscio, Burle e Calegari (2019).

Após a estruturação conceitual da proposta, foram estabelecidas correspondências entre os campos de metadados propostos e as propriedades do padrão DCAT-3 e, por extensão, dos padrões Dublin Core (DC), Friend of Friends (Foaf), vCard e Asset Description Metadata Schema (Adms), que são interoperáveis e complementares ao DCAT-3 e indicados para utilização na própria documentação do DCAT-3. Com a utilização do módulo gratuito da ferramenta Lucid Chart³, foi criado um diagrama de classes para apresentar os relacionamentos da estrutura proposta. Para a validação desta estrutura, ela

² <https://www.w3.org/TR/dwbp/>

³ <https://www.lucidchart.com/>

foi transcrita para a sintaxe *Turtle*, para expressar dados em *Resource Description Framework* (RDF), e testada na ferramenta gratuita IDLab Turtle Validator⁴. A proposta foi elaborada entre os meses de julho e novembro de 2024. Na próxima seção, serão apresentados os resultados obtidos na pesquisa.

3 RESULTADOS E DISCUSSÃO

Para a apresentação desta proposta, foram considerados três pilares, estabelecidos durante a execução das etapas anteriores da pesquisa:

- a) legislação e documentação:
 - legislações, normativas, portarias e diretrizes sobre acesso à informação;
 - política de Dados Abertos e legislações complementares;
- b) padrões e instrumentos de organização e representação da informação:
 - padrões de metadados (e-PMG, DC e DCAT-3);
 - vocabulários (Vocabulário Controlado de Governo Eletrônico - VCGE);
 - boas Práticas para Dados na Web (Lóscio; Burle; Calegari, 2019);
- c) estrutura existente no Portal:
 - aspectos, funcionalidades e metadados da estrutura existente.

Inicialmente, por meio da identificação da legislação e documentação, foi possível traçar um panorama das diretrizes existentes para conduzir a gestão de dados abertos no governo federal brasileiro e evidenciar as informações básicas a serem disponibilizadas obrigatoriamente, de acordo com a legislação. A seguir, foram estudados os instrumentos existentes que poderiam trazer contribuições para a otimização da representação da informação no Portal, e, por fim, analisou-se a estrutura vigente, na perspectiva do usuário comum, para identificar aspectos passíveis de melhorias. Nas subseções seguintes, serão detalhados os elementos considerados na proposta de metadados, que envolveram aspectos como: estrutura e organização do Portal, metadados sugeridos de acordo com os níveis de descrição propostos e sua correspondência com o padrão DCAT-3.

3.1 ESTRUTURA E ORGANIZAÇÃO

⁴ <https://www.lucidchart.com/>

A partir do mapeamento realizado no Portal na Etapa 1, foram identificados elementos da estrutura de organização para se levantar lacunas e realizar proposições de melhoria. As recomendações desta pesquisa visam ao aprimoramento da representação dos DGA do governo federal, de forma a impactar positivamente o **usuário consumidor**, ou seja, o cidadão, por isso, trabalhar-se-á com os itens visíveis para esse usuário.

Foi identificado que existe uma separação em blocos de preenchimento para as informações no Portal. Essa estrutura pode ser observada nos manuais para publicadores de dados. Dessa forma, a proposta sugere o **agrupamento por categorias**, em que os metadados relacionados possuem características comuns, tendo sido identificadas as seguintes categorias: Organização, Gestor de dados, Conjunto de dados, Recurso e Reúso.

Os blocos existentes aparentemente eram uma classificação interna, que não era mencionada de forma explícita nos documentos disponibilizados e nem refletida de forma tão evidente para o usuário consumidor na visualização dessas informações. Possivelmente, algumas delas não eram aparentes devido a questões de sigilo de informações privadas dos órgãos, como é o caso de dados pessoais e informações de segurança da informação.

Como apontado por Lira (2014), a qualidade e consistência da informação descritiva influencia diretamente no sucesso do consumo e da reutilização de dados abertos. Consequentemente, quanto mais ricamente um determinado dado for descrito, melhor será o seu reaproveitamento em diferentes contextos. Assim, para além das **sugestões dos metadados**, apresentou-se uma sugestão de **correlação destes com os padrões DCAT-3** e, por extensão, os padrões DC, Vcard, Foaf e Adms e as suas respectivas propriedades, classes e superclasses.

Destaca-se que a proposta apresentada considerou categorias e metadados existentes no Portal, tendo em vista o fato de que o instrumento já é utilizado pelos órgãos e tem apresentado bons resultados, propondo alterações e melhorias como recomendações para a otimização desse importante instrumento.

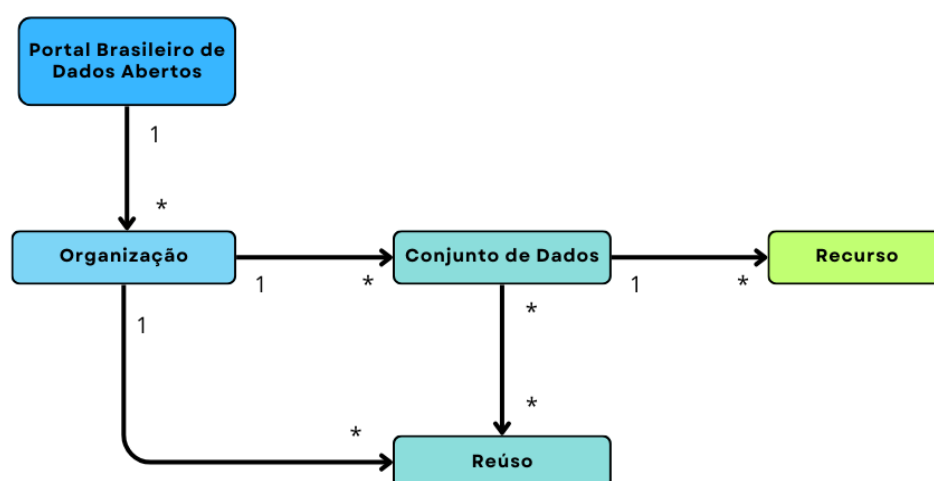
3.2 RELACIONAMENTOS

A primeira movimentação da proposta foi a explicitação da estrutura de hierarquização para a descrição com a divisão em categorias, considerando cada uma como uma classe que agrupa campos de descrição, sendo elas: Organização, Conjunto de dados, Recurso e Reúso. Foram mantidas as categorias identificadas no Portal que

continham dados que interessavam ao cidadão e, por isso, não se trabalhou com metadados da categoria Gestor de dados, pois entende-se que esta destina-se a dados sensíveis dos órgãos, que fogem ao escopo da pesquisa e não são visíveis para os consumidores.

Foi criada uma classe chamada Portal Brasileiro de Dados Abertos, adicionada para possibilitar a validação da proposta conceitual com a utilização do DCAT-3. Para isso, foram identificados atributos básicos para a representação do Portal, que abrangem todos os demais níveis de descrição, a saber: Organização, Conjunto de dados, Recurso e Reúso. A partir dessa definição, foi criado um diagrama para apresentar as classes utilizadas para representar a estrutura proposta para o Portal, conforme apresentado na Figura 1.

Figura 1 - Relacionamento entre as classes propostas



Fonte: elaborado pelas autoras (2025).

A classe Portal Brasileiro de Dados Abertos está relacionada à Organização, uma vez que o Portal terá muitas organizações cadastradas e apresentará um relacionamento de um para muitos. Nesse sentido, a classe Organização está relacionada às classes Conjunto de dados e Reúso, sendo os relacionamentos de um para muitos, nos quais uma organização pode publicar vários conjuntos de dados ou vários reúsos. Já a classe Recurso se relaciona com a classe Conjunto de dados em um relacionamento de um para muitos, na medida em que um conjunto de dados pode ter vários recursos associados a ele.

Destaca-se que o reúso está relacionado a uma única organização, considerando a estrutura existente em relação à responsabilização e à manutenção, mas limita consumidores e outros atores ao publicarem algum reúso. Contudo, tendo em vista o fato de que ele precisa ser validado e mantido por algum órgão, essa estrutura permaneceu na

recomendação. Por fim, a classe Reúso relaciona-se com a classe Conjunto de dados em um relacionamento de muitos para muitos, já que um reúso pode estar relacionado a vários conjuntos de dados e vários conjuntos de dados podem estar relacionados a muitos reúsos.

É importante destacar que todas as classes, quando representadas no banco de dados, terão um atributo identificador, id, do tipo número, criado de forma sequencial, que representará a chave primária e permitirá identificar unicamente as entidades para fazer os relacionamentos sugeridos.

3.3 METADADOS

A segunda movimentação da proposta foi a definição dos metadados, que considerou os requisitos da legislação, as Boas Práticas para Dados na Web, os padrões citados pelo governo e, também, o padrão DCAT-3, recomendado pelo W3C. Segundo Lóscio, Burle e Calegari (2019), os metadados fornecem informações complementares que ajudam os consumidores de dados a compreender melhor seu significado e sua estrutura, além de esclarecerem aspectos como direitos e termos de licença, detalhes da organização responsável, qualidade dos dados, métodos de acesso e cronograma de atualização dos conjuntos de dados. Na Etapa 1, foi mapeada a estrutura do Portal e identificados 34 metadados visíveis para o usuário consumidor, e, nessa proposta, sugere-se uma estrutura de metadados com 84 campos visíveis para o usuário consumidor, representando um aumento de 50 metadados em relação à composição atual, sendo:

- a) quatro referentes ao Portal Brasileiro de Dados Abertos;
- b) 23 referentes à classe Organização (manutenção de 14, adição de oito e exclusão de um);
- c) 27 referentes à classe Conjunto de dados (manutenção de 17, adição de 10 e exclusão de dois);
- d) 17 referentes à classe Recurso (manutenção de sete e adição de 10);
- e) 13 referentes à classe Reúso (manutenção de oito e adição de cinco).

Desses, sugere-se 61 metadados obrigatórios e 23 opcionais. O quadro com a estrutura completa foi apresentado no *site*: <https://observatoriodedadosabertos.eci.ufmg.br/metaopengov/> e para auxiliar na compreensão da proposta, utilizou-se um esquema de cores em que as linhas em fundo lilás representam os metadados identificados na estrutura atual do Portal e que foram

mantidos, as linhas em fundo **verde** caracterizam os metadados adicionados e as linhas em fundo **azul** representam os metadados que sofreram alguma modificação, como alteração de visibilidade ou de obrigatoriedade.

Metadados podem ser utilizados para facilitar tarefas como a descoberta e reutilização de conjuntos de dados e podem ser definidos com diferentes níveis de granularidade, desde a descrição de uma propriedade específica de um recurso até a de um conjunto de dados completo, ou mesmo de todos os conjuntos de dados pertencentes a uma organização específica (Lóscio; Burle; Calegari, 2019).

Consideram-se como **obrigatórios** todos aqueles campos que são imprescindíveis para a identificação e compreensão dos dados, como título e descrição, além daqueles que necessariamente irão existir e podem, inclusive, serem gerados automaticamente pelo sistema de informação, como data de publicação e última alteração. Os campos **opcionais** são aqueles que contêm informações consideradas complementares, como colaborador ou atributos, aqueles que nem sempre existem necessariamente como versão anterior ou vocabulário controlado associado.

Alguns metadados têm opções restritas de preenchimento e para eles foi sugerido um controle de vocabulário para conferir conformidade e padronização. São eles: 'Tipo de organização', 'Natureza jurídica', 'Esfera', 'UF', 'Município', 'Licença de uso', 'Temas aos quais o dado está associado' e 'Tipo'. Essa ação possibilita um agrupamento eficaz e consistente das informações associadas a esses metadados, pois já existem listas com valores controlados e evitam-se erros de preenchimento devido à utilização de linguagem natural, indicando tipologias de dados abertos, formatos, padrões e metadados utilizados.

Destaca-se que alguns desses campos já possuíam esse controle de vocabulário, como é o caso do campo 'Temas', que utiliza o VCGE, porém, acredita-se que a explicitação da adoção dessa prática e o fornecimento de documentação com os termos controlados favoreçam a compreensão e utilização desses dados.

Davanzo e Moreira (2014) conferem aos vocabulários controlados um papel de destaque na organização de documentos em ambientes informacionais, já que eles oferecem subsídios para a recuperação precisa das informações. Nessa perspectiva, sugere-se a adoção de vocabulários controlados específicos de cada área (por exemplo, Saúde, Educação, Transportes) para o preenchimento do campo 'Assuntos/Palavras-chave', quando possível, de forma a promover maior precisão e eficácia na comunicação entre os usuários e o sistema de informações (Kobashi, 2008) e possibilitar ao usuário a identificação de conceitos que o interessem a partir de uma ferramenta normatizada e

documentada. Nas subseções seguintes, serão explicitadas as categorias e os metadados propostos.

3.3.1 Organização

O primeiro nível de descrição visível para o usuário consumidor refere-se à categoria **Organização**, que abrange as informações sobre o órgão gerador/publicador dos DGA e visa atender principalmente a um requisito da lei de acesso à informação⁵ (LAI), que exige, em seu Art. 8º, parágrafo 1º, inciso I, que:

Art. 8º É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas.

§ 1º Na divulgação das informações a que se refere o caput, deverão constar, no mínimo:

I - registro das competências e estrutura organizacional, endereços e telefones das respectivas unidades e horários de atendimento ao público (Brasil, 2011).

São propostos 23 campos de metadados, sendo 14 originários da estrutura atual do Portal, oito acréscimos e uma exclusão. Para que os campos de metadados abranjam as informações requeridas pela LAI, foram mantidos **inalterados** os campos 'Esfera', 'Órgão SIORG', 'Código SIORG', 'Poder', 'Organização Pai', 'Imagem/Logo do órgão', 'Sigla', 'CNPJ', 'UF', 'Município', 'Site da organização', 'E-mail da organização', 'Status da organização' e 'Descrição'. Esses campos foram preservados, pois contêm as informações básicas que caracterizam uma organização e sua estrutura, além de fornecerem informações de contato e de localização.

Em relação às **alterações** propostas, considerando a arquitetura em vigor, os campos 'Tipo de organização', 'Órgão SIORG', 'Código SIORG', 'Site da organização' e 'E-mail da organização' não são visíveis para o usuário consumidor no Portal. Propõe-se que isso seja alterado, visto que são informações públicas, inclusive disponibilizadas em outras páginas do governo, como na que apresenta a estrutura do Sistema de Organização e Inovação Institucional do Governo Federal (SIORG).

Quanto ao campo 'Organização pai', este é considerado como de preenchimento obrigatório e, analisando-se o contexto, considera-se a possibilidade de a organização não

⁵ Lei nº 12.527, de 18 de novembro de 2011.

ter uma hierarquia superior. Portanto, sugere-se que seja de preenchimento opcional, **alterando-se** a obrigatoriedade.

Em relação à sugestão de **exclusão**, considerou-se que os campos ‘Órgão SIORG’ e ‘Nome do Órgão’ têm o mesmo objetivo de identificar o nome do órgão para o cadastro e, por isso, optou-se pela manutenção do primeiro, pois o segundo considera o “cadastro oficial dos órgãos e das entidades da administração direta, autárquica e fundacional do Poder Executivo Federal junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG” (Brasil, [20–?]a), sendo assim, ele possui certo controle de vocabulário, que tem o objetivo de prover consistência na representação e na recuperação, evitando a dispersão de recursos informacionais (Fujita; Tolare, 2019) e facilitando a padronização e o intercâmbio de dados.

Quanto à **adição** de metadados, sugere-se a incorporação de ‘Natureza jurídica’, ‘Logradouro’, ‘Número’, ‘Bairro’, ‘Complemento’, ‘CEP’, ‘Telefone de contato da organização’ e ‘Horário de atendimento ao público’. Esses metadados ajudam a satisfazer os requisitos da LAI, além de apresentarem uma estrutura semelhante à existente e disponível do SIORG, e esse nível de detalhamento do endereço também é encontrado no padrão Government Information Locator Service (GILS), mapeado no Padrão de Metadados de Governo Eletrônico (e-PMG) e adotado pelo governo federal dos Estados Unidos da América (EUA).

O principal objetivo dos metadados da categoria Organização é fornecer informações que identifiquem e caracterizem o órgão produtor e/ou mantenedor de dados públicos. Buscou-se descrever a estrutura organizacional, como requerido pela LAI, além de informações complementares que expressam as particularidades de cada órgão. Os metadados propostos também atendem às Boas Práticas para Dados na Web Nº 1 — Fornecer metadados e Nº 2 — Fornecer metadados descritivos (Lóscio; Burle; Calegari, 2019), possibilitando a compreensão dos dados e o conhecimento da estrutura da organização que os gerou.

Segundo Fujita (2004), a indexação que fornece a identificação de conceitos mais pertinentes ao conteúdo resulta em uma recuperação do documento mais condizente com a questão de busca e, por isso, acredita-se que a identificação detalhada e estruturada da Organização a partir desses metadados possibilita uma melhor indexação por mecanismos de busca e facilita a descoberta de dados para o usuário, na medida em que a utilização de vocabulários controlados em vários campos permite a homogeneidade no preenchimento

e deixa evidentes os conceitos-chave do conteúdo, além de possibilitar a busca a partir de uma diversidade de parâmetros.

3.3.2 Conjunto de dados

A categoria **Conjunto de dados** abrange uma coleção de dados que são publicados ou mantidos por uma organização e estão disponíveis para acesso ou *download* em um ou mais formatos (Albertoni *et al.*, 2024). Alguns exemplos de conjuntos de dados são uma coleção de tabelas relacionadas entre si, dados extraídos de um mesmo sistema de informações ou uma *Application Programming Interface* (API) de dados abertos (Brasil, [20-?]b).

Os Conjuntos de dados são considerados os elementos principais retornados a partir das buscas no Portal e cada um deles possui, pelo menos, um recurso associado. (Brasil, [20-?]a). Nesta categoria, são 27 metadados, sendo 17 da configuração vigente do Portal, 10 acréscimos e duas exclusões.

Conforme a Boa Prática para Dados na Web Web Nº 2 — Fornecer metadados descritivos — são sugeridos os seguintes metadados para a descrição dos Conjuntos de dados: ‘Título’, ‘Descrição’, ‘Palavras-chave’, ‘Data de publicação’, ‘Entidade responsável (publicadora de dados)’, ‘Ponto de contato’, ‘Cobertura Espacial’, ‘Período temporal’, ‘Data da última modificação’ e ‘Temas/categorias’. Destes, apenas o campo ‘Data da última modificação’ não existia e foi sugerida sua **adição** com a nomenclatura de ‘Última alteração’.

Além dos campos compatíveis com a Boa Prática para Dados na Web Nº 2, foram mantidos **inalterados** os campos ‘Item de inventário de origem’, que atendem à Boa Prática para Dados na Web Nº 3 — Fornecer metadados estruturais, ‘Licença de uso’ —, relacionado à Boa Prática para Dados na Web Nº 4 — Fornecer informações sobre a licença de dados —, e ‘Periodicidade de atualização’, correspondente à Boa Prática para Dados na Web Nº 21 — Fornecer dados atualizados.

No caso dos metadados ‘Área técnica responsável pelo dado’, ‘E-mail da área técnica responsável’, ‘Cobertura temporal início’, ‘Cobertura temporal fim’, ‘Descontinuado’, ‘Data de descontinuação’ e ‘Conjuntos de dados associados (interoperabilidade)’, a presença deles foi observada apenas no manual para publicadores de dados, mas, aparentemente, eles não são visíveis para o usuário consumidor, por isso, sugere-se uma **alteração** de visibilidade. Como não são campos que contêm informações sigilosas ou

dados sensíveis, considera-se que deve ser aplicado o princípio de transparência da publicidade como regra, e o sigilo, como exceção.

Em relação aos metadados ‘Área técnica responsável pelo dado’ e ‘E-mail da área técnica responsável pelo dado’, sugere-se a **alteração** de visibilidade para abarcar uma importante vertente apontada por Lira (2014), mas não contemplada nessa proposta: a disponibilização de informações que descrevem os metadados. Conforme o autor, devido às dificuldades para a criação e disponibilização de metadados para a descrição de dados abertos, deseja-se, também, a promoção do reúso de metadados para evitar redundância e colaborar com a padronização da informação e, portanto, é fundamental fornecer, além de metadados que descrevam os dados, informações que os descrevam.

Levando essa perspectiva em consideração, espera-se que esses metadados sirvam como um ponto de contato para dúvidas sobre os dados e os metadados que os descrevem caso as informações disponibilizadas sejam insuficientes para satisfazer as necessidades de informação do usuário. Relaciona-se esses campos também às Boas Prática para Dados na Web N° 1 — Fornecer metadados —, N° 2 — Fornecer metadados descritivos —, N° 5 — Fornecer informações de procedência de dados — e N° 29 — Coletar *feedback* de consumidores de dados.

Para os campos ‘Cobertura espacial’ e ‘Granularidade espacial’, sugere-se a **alteração** de visibilidade, pelo mesmo motivo relatado acima, e, também, a **alteração** da obrigatoriedade, originalmente campos opcionais, que sugere-se que sejam obrigatórios, pois entende-se que, sendo dados gerados por organizações vinculadas ao governo federal, a cobertura e granularidade espacial sempre existirão. Esses metadados aderem à Boa Prática para Dados na Web N° 13 — Usar representações de dados que sejam independentes de localidade — e, quando isso não for possível, fornecer metadados sobre os parâmetros regionais dos conteúdos de dados.

Já os metadados ‘Descontinuado’ e ‘Data de descontinuação’ podem associar-se à Boa Prática para Dados na Web N° 22 — Fornecer uma explicação para os dados que não estão disponíveis —, para que os consumidores saibam que os dados referidos, extraídos do conjunto de dados corrente, não estão disponíveis ou podem ser disponibilizados sob condições diferentes. Por fim, sugere-se a alteração de visibilidade para o metadado ‘Conjuntos de dados associados (interoperabilidade)’, que atende às Boas Práticas para Dados na Web N° 2, 3 e 5, citadas anteriormente.

Em relação às **exclusões**, sugeriu-se a exclusão desta categoria dos metadados ‘Versão’ e ‘Atualização de versão’, pois entende-se que é mais vantajoso relacioná-los a

um recurso, pois um conjunto de dados pode ter vários recursos e cada um deles pode ter diferentes versões.

Sugere-se a **adição** dos campos opcionais ‘Colaborador’, ‘API habilitada’, ‘Vocabulário controlado associado’, ‘Lista dos recursos cadastrados’, ‘Documentação’, ‘Granularidade temporal’ e ‘Metodologia’; do campo obrigatório ‘Identificador’; e dos campos de geração automática ‘Data de publicação’ e ‘Última alteração’.

Em relação ao campo ‘Colaborador’, acredita-se que ele seja adequado em um contexto para a promoção da interoperabilidade e do reúso de dados, nos quais pode haver intercâmbio de informações e, por isso, seria interessante registrar uma possível cooperação para manter um histórico. A Boa Prática para Dados na Web Nº 5 — Fornecer informações de procedência dos dados — alerta justamente sobre isso e justifica que a procedência é uma das formas que os consumidores têm para julgar a qualidade de um conjunto de dados e que a origem e a história ajudam a determinar a confiabilidade do dado e a fornecer contexto para interpretação. Além disso, é um campo presente no e-PMG e mapeado, também, nos padrões Australian Government Locator Service (AGLS), e-Government Metadata Standard (e-GMS) e Metainformação Para Interoperabilidade (MIP).

Já o campo ‘API habilitada’ atende à Boa Prática para Dados na Web Nº 23 — Disponibilizar dados por meio de uma API —, que garante aos consumidores maior flexibilidade e facilidade de processamento dos dados, e, também, à Boa Prática para Dados na Web Nº 35 — Citar a publicação original do conjunto de dados. Ainda que a utilização de uma API exija certo conhecimento técnico por parte do usuário, a disponibilização de dados por meio dela pode ser o incentivo para a criação de aplicações que atendam a usuários com todos os níveis de habilidade informacional.

O campo ‘Vocabulário controlado associado’ atende à Boa Prática para Dados na Web Nº 15 — Reutilizar vocabulários, dando preferência aos padronizados —, e pode, ainda, complementar os dicionários de dados que porventura estejam associados a algum vocabulário controlado. Segundo Davanzo e Moreira (2014, p. 4):

O vocabulário controlado agrega valor ao processo de organização e recuperação da informação porque sua utilização possibilitará o controle do repertório terminológico que será utilizado para descrever e/ou representar os conceitos que descrevem um determinado campo. Deve-se ressaltar, portanto, que a adoção de uma linguagem clara, padronizada e, na medida do possível, comum a todos diminuirá consideravelmente as dificuldades de comunicação que efetivamente existem entre usuários e sistema.

Em relação aos metadados ‘Documentação’, ‘Granularidade temporal’ e ‘Metodologia’, eles constam na Cartilha Técnica para Publicação de Dados Abertos, que é

uma publicação do próprio governo e está indicada como documento norteador para a publicação de dados no Portal. Sinaliza-se que o padrão de metadados constante nesta cartilha é o instrumento a ser seguido e, por isso, sugere-se a adição desses campos como opcionais. Entende-se que eles complementam a proposta, atendendo à Boa Prática para Dados na Web Nº 2 — Fornecer metadados descritivos.

O campo ‘Identificador’ atende à Boa Prática para Dados na Web Nº 9 — Usar URIs persistentes como identificadores de conjuntos de dados —, que permite a identificação básica dos dados e é pré-condição essencial para o gerenciamento e a reutilização dos dados de forma adequada. Dessa forma, os conjuntos de dados ou as informações sobre eles serão encontrados e citados com facilidade em qualquer momento, independentemente do *status*, da disponibilidade ou do formato desses dados.

Por fim, os metadados ‘Data de publicação’ e ‘Última alteração’ atendem à Boa Prática para Dados na Web Nº 21 — Fornecer dados atualizados — e auxiliam o consumidor de dados a conhecer um pouco do histórico do conjunto de dados acessado.

3.3.3 Recurso

Quanto à categoria **Recurso**, compreende-se que cada recurso representa uma fonte de dados, podendo ser um arquivo, como uma planilha, um método de *webservice*/API ou um documento. Segundo informações do Portal, cada recurso poderia ser catalogado como um conjunto de dados separado, entretanto, é aconselhável que recursos sejam agrupados quando forem oriundos da mesma base de dados ou possuírem metadados em comum, de forma a facilitar a busca e a compreensão do conteúdo (Brasil, [20–?]a). Assim, os recursos estão sempre associados a um conjunto de dados, ainda que eles mesmos possam ser conjuntos de dados. Eles correspondem a 17 metadados totais na proposta, sendo sete originários da estrutura em vigor do Portal e 10 adições.

Conforme a Boa Prática para Dados na Web Nº 2 — Fornecer metadados descritivos —, os metadados descritivos podem incluir as seguintes características genéricas de uma distribuição: ‘Título’, ‘Descrição’, ‘Data da publicação’ e ‘Formato de mídia’. Uma distribuição representa um determinado formato disponível de um conjunto de dados, como, por exemplo, um arquivo no formato CSV para *download* ou um API (Albertoni *et al.*, 2024). O único campo dentre esses que não estava contemplado originalmente era ‘Data da publicação’ e sugere-se a sua **adição**. Além desses, foram mantidos **inalterados** os campos ‘URL do recurso’, ‘Tipo’, ‘Data de criação’ e ‘Última alteração’.

Relacionam-se os campos ‘Formato de mídia’ e ‘Tipo’ também à Boa Prática para Dados na Web Nº 3 — Fornecer metadados estruturais —, os campos ‘Data da publicação’, ‘Data de criação’ e ‘Última alteração’ à Boa Prática para Dados na Web Nº 21 — Fornecer metadados atualizados — e o metadados ‘URL do recurso’ às Boas Práticas para Dados na Web Nº 5 — Fornecer informações de procedência de dados — e Nº 20 — Fornecer acesso em tempo real.

Sugere-se a **adição** dos campos ‘Versão’ e ‘Atualização de versão’, que eram originários da categoria Conjuntos de dados, pois entende-se que cada recurso de um conjunto de dados pode ter sua própria versão. Como complemento, sugere-se, também, a adição dos campos opcionais ‘Notas de versão’ e ‘Versão anterior’. Essas adições atendem à Boa Prática para Dados na Web Nº 7 — Fornecer indicador de versão — e à Boa Prática para Dados na Web Nº 8 — Fornecer o histórico de versão e o bom versionamento —, já que possibilitam aos consumidores compreender se uma nova versão está disponível e permitem comparações. Além desses, sugere-se a **adição** dos campos ‘Assuntos/Palavras-chave’, ‘Idioma’, ‘Tamanho do arquivo’, ‘Sigilo/Direitos de acesso’ e ‘Normativa/Observância legal’.

O campo ‘Assuntos/Palavras-chave’ é um descritor básico que visa realizar uma indexação do conteúdo do recurso de forma resumida, com a utilização de termos relevantes. Esse é um metadado que atende à Boa Prática para Dados na Web Nº 2 — Fornecer metadados descritivos — e, como sugerido para a categoria Conjuntos de dados, a utilização de um vocabulário controlado para preenchimento poderia aprimorar a representação da informação ao padronizar, também, o conteúdo.

Ainda que o Portal abranja dados produzidos por órgãos brasileiros, entende-se que não existe a obrigatoriedade de esses dados estarem no idioma português e, por isso, vê-se a necessidade do campo ‘Idioma’ para registrar essa especificação.

O campo ‘Sigilo/Direitos de acesso’ estabelece restrições ou permissões de acesso, delimitando quem pode visualizar ou acessar um recurso, e é complementado pelo campo ‘Normativa/Observância legal’, que busca esclarecer algum tipo de restrição que possa haver na utilização de um recurso devido, por exemplo, a alguma legislação, como a LGPD, ou como a previsibilidade na LAI de restrição temporária de acesso público em questões que envolvam a segurança da sociedade e do Estado.

3.3.4 Reúso

Por fim, tem-se a categoria **Reúso**, que se refere às aplicações criadas a partir de conjuntos de dados ou recursos já existentes, podendo ser aplicativos, *sítes*, entre outros. Este nível possui 13 metadados, sendo oito da estrutura vigente e cinco adições. Mantiveram-se **inalterados** os campos ‘Nome’, ‘URL’, ‘Descrição’, ‘Tipo’, ‘Imagem’ e ‘Organização a qual o recurso está associado’. Os metadados ‘Nome’, ‘Descrição’ e ‘Imagem’ associam-se à Boa Prática para Dados na Web Nº 2 — Fornecer metadados descritivos —. O campo ‘URL’ relaciona-se com as Boas Prática para Dados na Web Nº 5 — Fornecer informações sobre procedência dos dados — e Nº 20 — Fornecer acesso em tempo real —. O metadado ‘Tipo’ vincula-se à Boa Prática para Dados na Web Nº 3 — Fornecer metadados estruturais —, e o metadado ‘Organização a qual o recurso está associado’ atende às citadas Boas Práticas para Dados na Web Nº 3 e Nº 5.

Sugere-se a **alteração** de visibilidade para o campo ‘Conjuntos de dados de origem’, para que o usuário consumidor possa verificar a correlação entre os conjuntos de dados, caso assim desejar. Esse metadado atende às Boas Práticas para Dados na Web Nº 3 — Fornecer metadados estruturais —, Nº 5 — Fornecer informações de procedência dos dados —, Nº 31 — Enriquecer dados por meio da geração de novos dados — e Nº 35 — Citar a publicação original do conjunto de dados.

Propõe-se, ainda, a **alteração** do campo ‘Nome do responsável’ para ‘Área técnica responsável’, de forma a estabelecer um ponto de contato mais administrativo e menos pessoal, preservando informações sensíveis/sigilosas. Para complementar esse ponto de contato, sugere-se a **adição** do campo ‘E-mail da área técnica responsável’. Esses metadados atendem à Boa Prática para Dados na Web Nº 2 — Fornecer metadados descritivos — e podem, também, atender à Boa Prática para Dados na Web Nº 29 — Coletar *feedback* de consumidores de dados —, já que fornecem um ponto de contato.

Além disso, sugere-se a adição dos metadados ‘Data de criação do reúso’, ‘Data de publicação do reúso’ e ‘Última alteração’, sendo o primeiro de preenchimento obrigatório e os demais, de geração automática pelo sistema, com intenção de fornecer mais detalhamento sobre o histórico do reúso e atender à Boa Prática para Dados na Web Nº 21 — Fornecer dados atualizados.

Essa categoria é a que possui a menor quantidade de elementos cadastrados no Portal, contando com apenas 10 reúsos cadastros até 18 de novembro de 2024, segundo informações da página inicial. Uma hipótese para esse fato é a de que os usuários ainda

não se apropriam dos dados de forma a aproveitarem todo o seu potencial e, por isso, ocorre a baixa expressividade de reuso dos dados no Portal ou daqueles que o utilizam, porém não o cadastram como novo recurso no Portal. Soma-se a isso a circunstância de que o reuso se vale de conjuntos de dados ou recursos já cadastrados, quando parte da descrição já foi realizada, sendo necessários menos metadados para a identificação e especificação dessa categoria.

Dos 13 metadados listados, apenas 'Imagem do reuso' e 'E-mail da área técnica responsável' não têm correspondência no e-PMG, sendo que o campo 'URL do reuso', correlacionado com a propriedade 'Localização (*Location*)', é considerado um dos mais importantes, porque informações mais detalhadas e específicas podem estar disponíveis no local onde o reuso está hospedado.

Considerando-se a proposta geral, sempre que possível, foi sugerida a preservação da estrutura atual, de forma a gerar menor impacto e maior simplicidade em uma possível implementação. As sugestões de alterações e adições foram pautadas buscando atender às legislações pertinentes em sua totalidade e pensando no usuário consumidor, que se beneficiará quanto mais ricas e detalhadas forem as informações fornecidas.

3.4 CORRESPONDÊNCIA COM O PADRÃO DCAT-3

Segundo Lóscio, Burle e Calegari (2019), uma boa prática genérica para a publicação de dados na *web* é o uso de padrões. Já para Tomoyose (2021), os vocabulários podem ser utilizados para padronizar os termos de metadados empregados. Após a definição dos metadados propostos para a representação no Portal, cada um foi correlacionado com uma propriedade e uma classe utilizando o vocabulário DCAT-3, que possui diversas características que o adequam ao contexto, como a possibilidade de utilização de outros vocabulários, aumentando a interoperabilidade dos dados e a possibilidade de disponibilização de uma versão legível por máquina dos metadados descritivos.

Para a validação da estrutura proposta, após a correlação dos metadados com propriedades do DCAT, foi criada uma sintaxe em *Turtle*, com a descrição de classes e propriedades, cujo código foi validado em uma ferramenta gratuita chamada *Turtle*

*Validator*⁶. Na prática, isso significa que o código está corretamente descrito e, no contexto do trabalho, representa o primeiro passo para corroborar o esquema proposto.

Das sete classes principais do DCAT-3, foram utilizadas três para modelar a proposta de metadados de acordo com esse padrão, sendo elas: ***dcat:Catalog***, representando o Portal Brasileiro de Dados Abertos, o catálogo central de disponibilização de dados da administração pública federal; a classe genérica *dcat:Dataset*, que foi utilizada duas vezes, com ***dcat:Dataset1*** representando os Conjuntos de dados e ***dcat:Dataset2*** representando os Reúsos publicados no Portal; e ***dcat:Distribution*** para abranger os metadados que caracterizam cada versão de um Recurso associado a um conjunto de dados. Como complemento, utilizou-se a classe ***foaf:Organization***, indicada pela própria documentação do DCAT-3 para a descrição das Organizações publicadoras.

A partir dessa definição, foi criado um diagrama para apresentar as classes e propriedades utilizadas para representar a estrutura proposta para o Portal. O relacionamento entre as classes significa que o Portal Brasileiro de Dados Abertos (*dcat:Catalog*) pode ter inúmeras organizações publicadoras (*foaf:Organization*) cadastradas e essas instituições podem catalogar inúmeros conjuntos de dados (*dcat:Dataset1*), que podem ter inúmeros recursos (*dcat:Distribution*) relacionados, e as organizações publicadoras podem cadastrar, também, inúmeros reúsos (*dcat:Dataset2*), que podem estar associados a inúmeros conjuntos de dados (*dcatDataset1*).

4 MODELO METAOPENGOV

Percebeu-se que o governo federal tem atuado em diversas frentes para garantir que os princípios de dados abertos sejam cumpridos e que ferramentas para a disponibilização dos dados sejam implementadas e fornecidas para a utilização do público. Porém, é muito importante que, além do engajamento no desenvolvimento de novas ferramentas e aplicações, sejam avaliadas a usabilidade, a validade, a pertinência e a adequação do que já existe, inclusive para evitar retrabalho e incentivar o reúso. Por isso, apresenta-se uma proposta de metadados que leva em consideração todo o contexto existente, além de recomendações e boas práticas mapeadas na literatura científica.

O modelo proposto foi intitulado de 'MetaOpenGov', em uma alusão aos metadados para os DGA, e é apresentado, resumidamente, na Figura 2. Dos 84 metadados sugeridos

⁶ <http://ttl.summerofcode.be/>

na proposta, 66 existem no Portal (para o perfil de administrador) e indica-se a correlação de 59 com metadados do e-PMG, como demonstrado no *site*: <https://observatoriodedadosabertos.eci.ufmg.br/metadados/>, e, conseqüentemente, com os padrões mapeados por ele. Vê-se que o e-PMG possui informações pertinentes e válidas no cenário presente, porém as suas desobrigatoriedade e desatualização contribuem para a sua defasagem e o seu desuso. Quanto aos metadados da Cartilha Técnica para Publicação de Dados Abertos, contemplaram-se todos na proposta, pois entende-se que a indicação desses documentos como padrão para descrição deve ser atendida para que essa proposta seja consonante com as diretrizes governamentais.

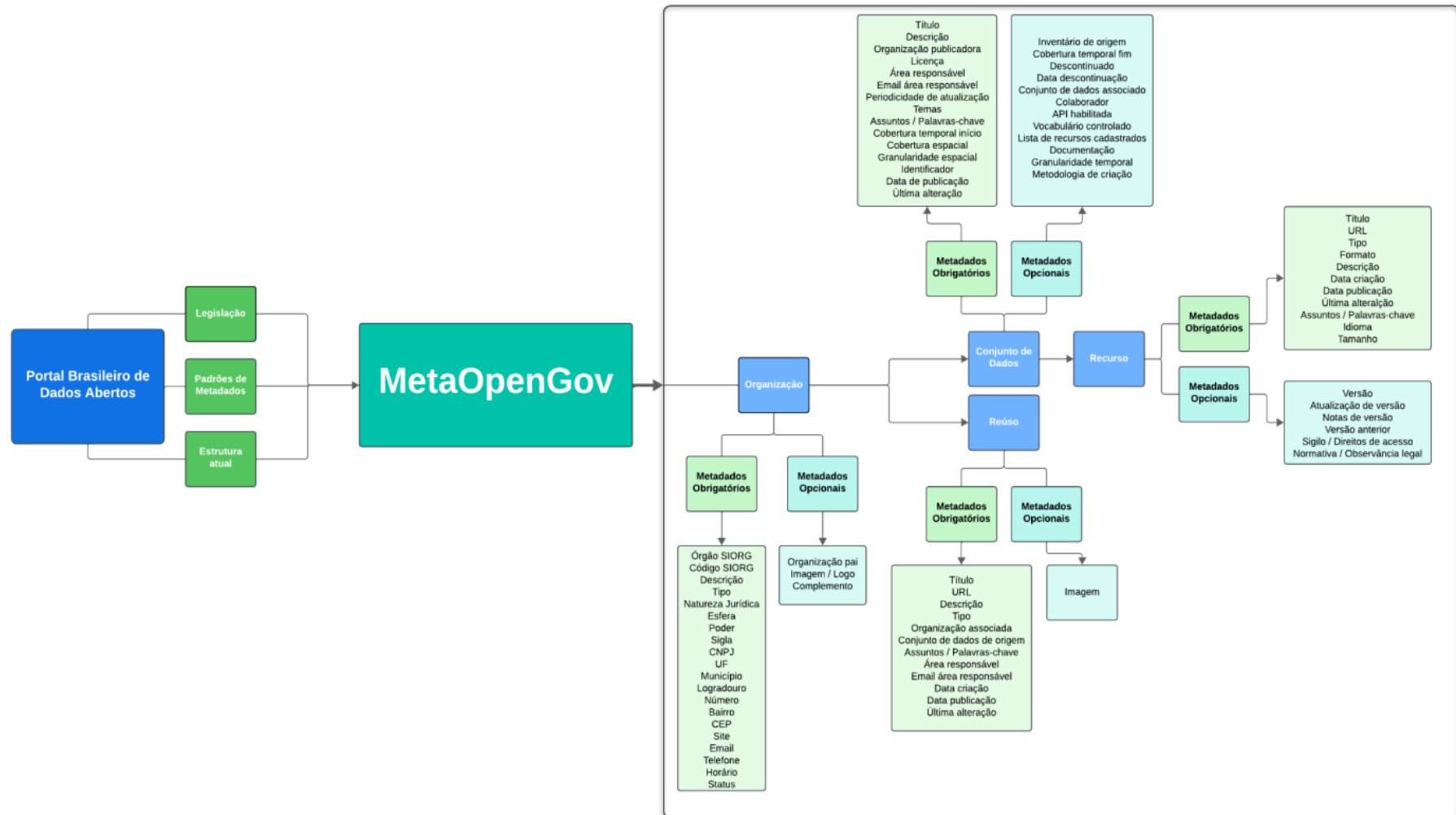
Em relação às Boas Práticas para Dados na Web (Lóscio; Burle; Calegari, 2019), todos os metadados sugeridos puderam ser correlacionados com alguma delas, o que demonstra a pertinência e a relevância da seleção desses metadados.

O desenvolvimento de sistemas e aplicações geralmente perpassa uma equipe de profissionais de áreas diversas, mas nem sempre se prevê um profissional da informação, que tem um olhar apurado e cuidadoso para questões de gestão e organização da informação que impactam todos os processos do ciclo de vida dos dados, incluindo sua representação e recuperação, que são refletidas diretamente na experiência do usuário. Pretendeu-se recomendar aprimoramentos na representação de dados abertos do Portal Brasileiro de Dados Abertos, uma importante ferramenta de disponibilização de informações públicas, sob a perspectiva da CI e seus instrumentos, e espera-se impactar positivamente o usuário consumidor, geralmente não especialista na manipulação de dados.

Essa proposição foi realizada com fundamentação nos conceitos da Organização da Informação (OI), com foco, principalmente, na representação da informação por meio do uso de metadados, com o objetivo final de proporcionar uma melhor compreensão dos dados apresentados e uma recuperação eficiente e satisfatória para o usuário.

Como evidenciado por Arakaki (2019), cada tipo de metadado desempenha um papel específico na representação de um recurso informacional. Nesse sentido, buscou-se incluir elementos diversificados, capazes de oferecer uma descrição rica e detalhada, apta a atender às diferentes necessidades informacionais. Espera-se contribuir para a mitigação da heterogeneidade dos dados com o oferecimento de opções padronizadas de descrição que favoreçam a interoperabilidade e a integração entre conjuntos de dados, por meio da observação da semelhança entre campos.

Figura 2 - Modelo de metadados para DGA – MetaOpenGov



Fonte: elaborado pelas autoras (2025).

5 CONCLUSÕES

Esta pesquisa foi desenvolvida com o objetivo geral de propor recomendações para otimizar a representação dos dados abertos do governo federal brasileiro com a utilização de metadados e padrões para a representação de dados abertos no Portal Brasileiro de Dados Abertos, no contexto da CI, por meio de uma pesquisa aplicada, que utilizou as técnicas de pesquisa bibliográfica e documental e observação direta.

Essa proposta foi composta por 84 metadados, sendo quatro para descrever informações básicas do Portal, 23 referentes à categoria Organização, 27 referentes à categoria Conjuntos de dados, 17 referentes à categoria Recurso e 13 referentes à categoria Reúso. Desses, foram sugeridos como obrigatórios 61 metadados considerados imprescindíveis para a identificação e compreensão dos dados, como, por exemplo, título e descrição, os que necessariamente irão existir, como data de publicação e última alteração, e, como opcionais, 23 metadados que contêm informações complementares ou que não existem em todos os contextos, como versão anterior ou vocabulário controlado associado.

O aumento expressivo na quantidade de metadados em relação à estrutura vigente reflete a necessidade de prover o máximo de informações possível ao usuário para que ele possa tomar decisões sobre a pertinência do conjunto de dados para satisfazer as suas necessidades e possibilitar o reúso dos dados em novas aplicações, que requerem o máximo de detalhamento possível.

Além da proposição conceitual, fundamentada na legislação e em boas práticas identificadas na literatura científica, propôs-se, também, a correlação dos campos com propriedades e classes do padrão DCAT-3, o que viabiliza a implementação da proposta. A partir desse panorama, fazem-se algumas considerações sobre os resultados obtidos, que são convergentes com a literatura, o que demonstra uma homogeneidade nas práticas e abordagens sobre os DGA, mas também acarreta preocupações sobre a recorrência de desafios ao decorrer do tempo.

Algumas recomendações práticas que podem beneficiar diretamente a representação da informação no Portal Brasileiro de Dados Abertos são a utilização de vocabulários controlados atualizados. Ainda que, devido à diversidade dos órgãos e das tipologias dos DGA, não seja viável a utilização de um único vocabulário geral para descrição, como o VCGE, a adoção de vocabulários dentro de cada área específica nortearia a compreensão dos dados e facilitaria o intercâmbio entre dados semelhantes.

Um desafio encontrado foi a fragmentação, dispersão e desatualização de publicações orientadoras, como normas, diretivas, legislações e manuais, além de *links* inexistentes ou inacessíveis e a dependência do poder público para uma implementação real da proposta, o que representa, também, um impedimento da validação com dados reais e, ainda que isso fosse concretizado, a falta de obrigatoriedade de utilização das soluções e ferramentas propostas levará, inevitavelmente, à obsolescência, restringindo sua relevância ao longo do tempo. Além disso, os dados obtidos por meio de observação direta podem ser imprecisos e estão sujeitos à subjetividade da pesquisadora, como qualquer estudo dessa natureza.

Com base nas contribuições e limitações apresentadas, sugerem-se algumas possibilidades para o desenvolvimento de trabalhos futuros, como realizar uma validação prática com um conjunto de dados real, possibilitando a avaliação da aplicabilidade e da eficácia das propostas em um contexto mais próximo da realidade para identificar eventuais ajustes necessários. Além disso, o desenvolvimento de um protótipo de *design* voltado para a organização e apresentação visual da estrutura seria uma complementação atrativa para a proposta apresentada. Esse protótipo possibilitará uma análise mais detalhada da usabilidade e da clareza na transmissão da informação, contribuindo para aprimorar a forma como os dados são apresentados e interpretados. Essas ações abrem caminho para uma análise mais prática e orientada ao usuário, reforçando o potencial de impacto social dos resultados na área. A realização dessa pesquisa com o usuário pode ser um adendo valioso para verificar a adequação da proposta às necessidades informacionais do cidadão e detectar pontos de melhoria, visando aumentar a participação cidadã.

Espera-se que esse trabalho represente uma base para discussões sobre organização e representação da informação no cenário de dados abertos nacional, fortalecendo a ligação entre os DGA e a CI, e inspire o desenvolvimento de novos modelos e proposições que abranjam o profissional da informação e a interação de diversos atores.

REFERÊNCIAS

ALBERTONI, R. *et al.* (ed.). **Data Catalog Vocabulary (DCAT)**. Version 3. 2024. Disponível em: <https://www.w3.org/TR/vocab-dcat-3/>. Acesso em: 26 ago. 2025.

ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados**: modelo baseado na família PROV. 2019. 139 f. Tese (Doutorado) - Curso de Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Marília, 2019. Disponível em: <http://hdl.handle.net/11449/180490>. Acesso em: 26 ago. 2025.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da República Federativa do Brasil**. Brasília, DF, 18 nov. 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 17 fev. 2025.

BRASIL. Governo Federal. Dados abertos. **Saiba como publicar um conjunto de dados**. [20-?]a. Disponível em: <https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-conjunto-de-dados>. Acesso em: 17 fev. 2025.

BRASIL. Governo Federal. Dados Abertos. **Sobre**. [20-?]b. Disponível em: <https://dados.gov.br/dados/conteudo/sobre>. Acesso em: 17 fev. 2025.

DAVANZO, L.; MOREIRA, W. O vocabulário controlado como ferramenta do processo de organização e recuperação da informação. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte. **Anais** [...] Belo Horizonte: ENANCIB, 2014. Disponível em: <https://brapci.inf.br/index.php/res/v/186303>. Acesso em: 26 ago. 2025.

DAVENPORT, T. H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 2002.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **RDBCi**: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, SP, v. 1, n. 1, p. 60–90, 2004. DOI: 10.20396/rdbci.v1i1.2089. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2089>. Acesso em: 26 ago. 2025.

FUJITA, M. S. L.; TOLARE, J. B. Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. **Informação & Informação**, Londrina, v. 24, n. 2, p. 93-125, 2019. DOI: 10.5433/1981-8920.2019v24n2p93. Disponível em: <https://doi.org/10.5433/1981-8920.2019v24n2p93>. Acesso em: 26 ago. 2025.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6 ed. São Paulo: Atlas, 2017.

KOBASHI, N. Y. **Vocabulário controlado**: estrutura e utilização. Brasília: Escola Nacional de Administração Pública, 2008. Disponível em: <https://repositorio.enap.gov.br/bitstream/1/1289/41/Vocabul%C3%A1rio%20controlado%20-%20estrutura%20e%20utiliza%C3%A7%C3%A3o.pdf>. Acesso em: 26 ago. 2025.

LIMA, G. Â. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em Ciência da Informação**, Belo Horizonte, p. 57–97, fev. 2020. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22283>. Acesso em: 26 ago. 2025.

LIRA, M. Â. B. **Uma abordagem para enriquecimento semântico de metadados para publicação de dados abertos**. 2014. 95 f. Dissertação (Mestrado em Ciência da Computação) - Centro de Informática, Universidade Federal de Pernambuco, Recife, 2014. Disponível em: <https://repositorio.ufpe.br/handle/123456789/11570>. Acesso em: 26 ago. 2025.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. (ed). **Boas práticas para dados na web**. Tradução: Núcleo de Informação e Coordenação do Ponto BR. [S. l.]: W3C, 2019. Disponível em: <https://www.w3.org/Translations/DWBP-pt-BR/>. Acesso em: 26 ago. 2025.

OLIVEIRA, D. T. **Metadados para representação de dados abertos**: proposta para o Portal Brasileiro de Dados Abertos. 2025. 221 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2025. Disponível em: <http://hdl.handle.net/1843/79947>. Acesso em: 26 ago. 2025.

OLIVEIRA, D. T.; NASCIMENTO SILVA, P. Modelos, frameworks e métricas para dados abertos do governo: uma análise da produção acadêmica brasileira. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, Brasil. [2025]. No prelo.

NASCIMENTO SILVA, P. **Dados governamentais abertos**: métricas e indicadores de reúso. 2018. 322 f. Tese (Doutorado) - Curso de Gestão e Organização do Conhecimento, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: <http://hdl.handle.net/1843/BUBD-AYNG4U>. Acesso em: 26 ago. 2025.

NASCIMENTO SILVA, P. Acesso à informação no Brasil: política de dados governamentais abertos. In: GERALDES, E. *et al* (org.). **Dez anos da lei de acesso à informação**: limites, perspectivas e desafios. São Paulo: INTERCOM, 2022. p. 748.

NASCIMENTO SILVA, P. Dados governamentais abertos sobre a COVID-19 no Brasil: avaliação do potencial de reúso. **Em Questão**, Porto Alegre, v. 29, e124820, 2023. DOI: 10.19132/1808-5245.29.124820. Disponível em: <https://doi.org/10.19132/1808-5245.29.124820>. Acesso em: 26 ago. 2025.

NASCIMENTO SILVA, P. **Observatório de Dados Abertos**: MetaOpenGov. 2025. Disponível em: <https://observatoriodedadosabertos.eci.ufmg.br/metaopengov/>. Acesso em: 26 ago. 2025.

OPEN KNOWLEDGE BRASIL. **Por que open?**. [20-?]. Disponível em: <https://ok.org.br/dados-abertos/>. Acesso em: 26 ago. 2025.

TOMOYOSE, K. **O Data Catalog Vocabulary (DCAT) para a publicação de dados de pesquisa nos princípios Linked Data**. 2021. 185 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de São Carlos, São Paulo, 2021. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/14116>. Acesso em: 26 ago. 2025.

NOTAS

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), projeto APQ-01609-24.
Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio à pesquisa.

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: D. T. Oliveira e P. Nascimento Silva

Coleta de dados: D. T. Oliveira

Análise de dados: D. T. Oliveira e P. Nascimento Silva

Discussão dos resultados: D. T. Oliveira e P. Nascimento Silva

Revisão e aprovação: P. Nascimento Silva

ORIGEM DA PESQUISA

Dissertação: OLIVEIRA, D. T. **Metadados para representação de dados abertos:** proposta para o Portal Brasileiro de Dados Abertos. 2025. 221 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2025. Disponível em:

<http://hdl.handle.net/1843/79947>. Acesso em: 26 ago. 2025.

Projeto: Observatório de Dados Abertos **Coordenadora:** Patrícia Nascimento Silva

PREPRINTS

O manuscrito não é um *preprint*.

USO DE INTELIGÊNCIA ARTIFICIAL

Não se aplica.

FINANCIAMENTO

Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Processo: 2070.01.0005296/2024- 8 - APQ-01609-24.

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo 303721/2025-1.

CONSENTIMENTO DE USO DE IMAGEM

Não se aplica.

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

CONFLITO DE INTERESSES

As pessoas autoras declaram não haver interesses conflitantes.

DISPONIBILIDADE DE DADOS DE PESQUISA E OUTROS MATERIAIS

Os dados foram publicados no próprio artigo.

LICENÇA DE USO

As autorias cedem à Encontros Bibli os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a Licença [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) (CC BY) 4.0 International. Essa licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. As autorias têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. As ideias expressadas neste artigo são de responsabilidade das pessoas autoras, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Edgar Bisset Alvarez, Patrícia Neubert, Genilson Geraldo, Camila de Azevedo Gibbon, Jônatas Edison da Silva, Luan Soares Silva, Marcela Reinhardt de Souza e Daniela Capri.

HISTÓRICO

17-02-2025 – Aprovado em: 08-08-2025 – Publicado em: 26-09-2025

