25°enancib
Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação
Informação, Decolonialidade e Direitos Difusos para o Desenvolvimento Sustentável
3 a 7 de novembro de 2025 - Rio de Janeiro

XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - XXV ENANCIB

GT 2 – Organização e Representação do Conhecimento

METAOPENGOV: MODELO PARA REPRESENTAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS

METAOPENGOV: *MODEL FOR OPEN GOVERNMENT DATA REPRESENTATION*

Danielle Teixeira de Oliveira - Universidade Federal de Minas Gerais (UFMG)
Patrícia Nascimento Silva - Universidade Federal de Minas Gerais (UFMG)

Modalidade: Trabalho Completo

Resumo: a crescente demanda por dados abertos no Brasil requer uma estrutura eficaz para garantir a acessibilidade e reutilização da informação. O objetivo desse trabalho é propor um modelo de metadados para a representação e organização da informação no Portal Brasileiro de Dados Abertos, por meio da utilização de metadados e padrões voltados à publicação de dados abertos. A pesquisa é caracterizada como aplicada, exploratória e descritiva e foi realizada em duas etapas: (i) análise bibliográfica, documental e observação direta do Portal; e (ii) sistematização dos dados coletados, realizada entre julho e novembro de 2024. A abordagem metodológica combinou análise qualitativa e quantitativa. Os resultados permitiram a elaboração do modelo de metadados MetaOpenGov, que se fundamenta em três pilares principais: a legislação e a documentação relacionada; os padrões e instrumentos de organização; e a estrutura existente no Portal. O modelo propõe 84 metadados, alinhados ao padrão *Data Catalog Vocabulary* 3, com recomendações para a hierarquização e divisão dos dados em categorias como Organização, Conjunto de dados, Recurso e Reúso. Conclui-se que as recomendações podem promover melhorias significativas na organização e acessibilidade dos dados, aumentando a interoperabilidade e facilitando sua reutilização pela sociedade. Além disso, o estudo destaca a importância do profissional da informação na implementação das políticas de dados abertos no Brasil.

Palavras-chave: metadados; Portal brasileiro de dados abertos; dados abertos; representação da informação; interoperabilidade.

Abstract: the growing demand for open data in Brazil requires an effective structure to guarantee the accessibility and reuse of information. The aim of this work is to propose a metadata model for the representation and organization of information on the Brazilian Open Data Portal. The research is characterized as applied, exploratory and descriptive and was carried out in two stages: (i) bibliographic and documentary analysis and direct observation of the Portal; and (ii) systematization of the data collected, carried out between July and November 2024. The methodological approach combined qualitative and quantitative analysis. The results enabled the development of the MetaOpenGov metadata model, which is based on three main pillars: legislation and related documentation; organizational standards and instruments; and the existing structure of the Portal. The model proposes 84 metadata, aligned with the Data Catalog Vocabulary 3 standard, with recommendations for hierarchizing and dividing data into categories such as Organization, Dataset, Resource and Reuse. It is concluded that the recommendations can promote significant improvements in the organization and accessibility of data, increasing interoperability and facilitating its reuse by society. The study also highlight the importance of information professionals in implementing open data policies in Brazil.

Keywords: metadata; Brazilian Open Data Portal; open data; information representation; interoperability.

1 INTRODUÇÃO

Na sociedade contemporânea, os dados assumem um papel estratégico na geração de conhecimento e no suporte à tomada de decisões. O fenômeno do *big data*, marcado pela produção massiva de dados em formatos e fontes heterogêneos, desafia os profissionais da informação a desenvolverem métodos eficientes de organização e gestão de dados e informações. Dentre os tipos de dados existentes, tem-se os dados abertos, que podem ser livremente utilizados, reutilizados e redistribuídos, com atribuição da fonte e destacando-se pelo potencial de promover transparência e participação social (Open Knowledge Brasil, [20–?]).

No campo da Ciência da Informação (CI), dados e informação são conceitos interligados, embora existam discussões teóricas sobre sua distinção. Para Davenport (2002), dados são observações sobre o mundo, enquanto a informação é resultado da contextualização desses dados. A crescente disponibilização de dados em formato digital reforça a necessidade de estruturas que possibilitem sua representação e recuperação de forma eficaz.

O Portal Brasileiro de Dados Abertos é o repositório oficial do Governo Federal para a publicação de dados públicos, organizado como um catálogo federado que reúne informações disponibilizadas por diferentes órgãos (Brasil, [20–?]). Ele utiliza, para fins de representação dos dados e da informação, o Padrão de Metadados do Governo Eletrônico (e-PMG), concebido com o objetivo de definir um conjunto mínimo de elementos para a recuperação e o gerenciamento de informações (Brasil, 2014). No entanto, esse padrão foi originalmente desenvolvido para atender ao contexto do governo eletrônico de forma ampla e, além disso, a última atualização do e-PMG ocorreu em julho de 2014, o que evidencia sua defasagem diante das evoluções tecnológicas.

Apesar dos avanços normativos, como o Plano de Dados Abertos (PDA) e a Infraestrutura Nacional de Dados Abertos (INDA), ainda há lacunas significativas na padronização e estruturação dos dados, o que dificulta sua recuperação e reutilização, sendo esta a problemática de pesquisa. Explora-se, nesse trabalho, um hiato apontado por Lima

(2020), que é a grande quantidade de dados não estruturados cuja recuperação e acesso tornam-se cada vez mais difíceis.

Diante desse cenário, este estudo busca responder: quais metadados podem ser utilizados para representar os dados abertos do governo federal brasileiro? O objetivo geral é propor um modelo de metadados para a representação e organização da informação no Portal Brasileiro de Dados Abertos, por meio da utilização de metadados e padrões voltados à publicação de dados abertos.

O trabalho contribui para a área de Representação da Informação na Web ao propor soluções para a estruturação e o acesso a Dados Governamentais Abertos (DGA), ainda marcados por dispersão e falta de uniformidade, demonstrando o seu potencial impacto tecnológico ao propor melhorias para sistemas de dados públicos; científico e educacional, por avançar o conhecimento sobre organização da informação; social e econômico, ao fomentar a transparência, o controle social e a inovação com base em dados; e cultural, valorizando o acesso à informação pública.

O trabalho é resultado de uma pesquisa acadêmica, defendida no Programa de Pós-graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais em dezembro de 2024.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta os conceitos de dados, informação, metadados e organização da informação, fundamentais para compreender os DGA, que, apesar de associados ao termo dados, envolvem diretamente os demais conceitos.

Setzer (1999) define dados como sequências de símbolos com caráter sintático e objetivo, processáveis por computadores, enquanto a informação é subjetiva, pois envolve significado e interpretação. Miranda (1999) aponta que dados, quando organizados, transformam-se em informação útil à tomada de decisão e, no caminho reverso, Setzer (1999) considera que a informação é uma abstração que precisa ser convertida em dados para ser processada, sendo essa distinção entre sintaxe (dados) e semântica (informação) central em sua abordagem.

Buckland (1991) complementa essas perspectivas ao definir a informação sob três perspectivas: como processo, conhecimento e coisa, esta última sendo a adotada neste trabalho, permitindo considerar os dados como representações informacionais reutilizáveis.

Paulo (2018) aponta que os DGA são elementos de base da informação, primários e estruturantes, permitindo a construção de informações significativas e a geração de novos conhecimentos e inovação para a sociedade. Dessa forma, entende-se que a fronteira entre dado e informação é sutil e, nesta pesquisa, os DGA são compreendidos como elementos que transitam entre esses dois conceitos, pois, ao serem disponibilizados e reutilizados, adquirem significado e utilidade. Assim, pressupõe-se que a interação dos usuários com os DGA os caracteriza como informações, e o reúso dessas informações potencializa a geração de conhecimento.

A organização e representação da informação são essenciais nesse processo. Para Bräscher e Café (2008), organizar a informação visa facilitar o acesso ao conhecimento. Ortega (2013) define a Organização da Informação (OI) como um conjunto de procedimentos que facilita o uso da informação conforme as necessidades dos usuários. Já a representação diz respeito aos meios descritivos utilizados para identificar e recuperar objetos informacionais, abrangendo tanto seus aspectos físicos quanto de conteúdo (Bräscher; Café, 2008).

Considerando que os DGA visam garantir o acesso público aos dados do governo, os processos de OI são indispensáveis para sua estruturação e reutilização. Lima (2020) destaca que os metadados atuam como intermediários entre o objeto e o usuário, sendo, assim, ferramentas de representação. Alves (2010) aponta que os metadados impulsionaram o desenvolvimento de padrões voltados à representação e ao acesso à informação digital e vincula seu desenvolvimento aos métodos de organização, tratamento e representação informacional.

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa caracteriza-se como aplicada, por ter como finalidade a resolução de problemas reais relacionados à representação dos DGA; exploratória, ao buscar desenvolver e esclarecer conceitos, além de proporcionar uma visão geral sobre práticas e padrões adotados; e descritiva, ao se dedicar à observação e análise de características do fenômeno estudado, estabelecendo possíveis relações entre variáveis (Gil, 2017). Como instrumentos metodológicos, foram utilizados a pesquisa bibliográfica e documental e a observação direta, com uma abordagem que integrou métodos qualitativos e quantitativos.

O desenvolvimento deste trabalho ocorreu em duas etapas, vinculadas aos objetivos específicos da pesquisa. A primeira etapa teve como foco a identificação e análise das

tipologias e dos formatos de dados abertos publicados, bem como os padrões e elementos de metadados utilizados no Portal Brasileiro de Dados Abertos, as legislações, os padrões internacionais e as boas práticas na *web*. A segunda etapa consistiu na proposição de uma estrutura de representação para os dados abertos do governo federal brasileiro, com base nos resultados da etapa anterior. Foi realizada uma observação direta, com navegação contínua no Portal Brasileiro de Dados Abertos entre os meses de julho e outubro de 2024. Os resultados obtidos nessas etapas estão detalhados em Oliveira (2025).

Com base no mapeamento realizado e nos conhecimentos obtidos sobre padrões de metadados, foi selecionado o padrão *Data Catalog Vocabulary 3 (DCAT-3)* para embasar a proposta de estrutura na terceira etapa. O DCAT-3 foi escolhido por ser uma recomendação do *World Wide Web Consortium* (W3C) por sua aderência às boas práticas para dados na *web*, por promover a sustentabilidade por meio da reutilização de elementos de outros vocabulários e por ser voltado especificamente à descrição de catálogos de dados governamentais (Tomoyose, 2021).

A estrutura proposta foi desenvolvida considerando-se, também, as Boas Práticas para Dados na Web (*Data on the Web Best Practices*), documento amplamente utilizado no contexto estudado (Lóscio; Burle; Calegari, 2019). Essa etapa da pesquisa foi desenvolvida entre os meses de agosto e novembro de 2024. A próxima seção apresenta os resultados obtidos.

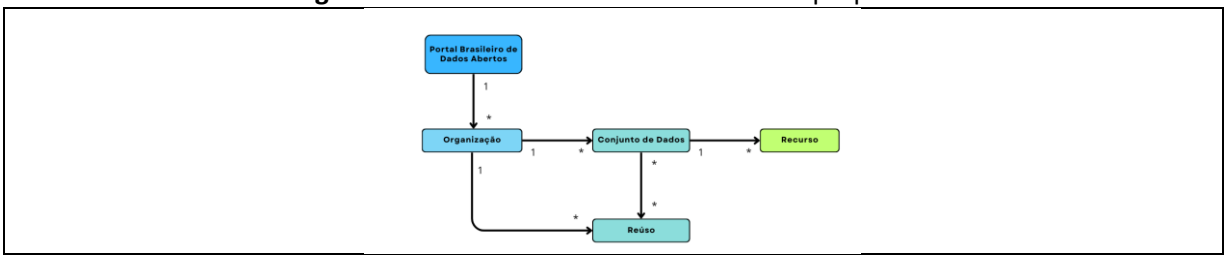
4 RESULTADOS E DISCUSSÃO

A proposta apresentada foi construída a partir da articulação de três pilares fundamentais, definidos ao longo da execução das etapas anteriores do estudo e essenciais para sustentar conceitual e tecnicamente o modelo de metadados para os DGA, considerando os marcos legais, os instrumentos da área da CI e a estrutura tecnológica já existente. São eles: **Legislação e documentação**: legislações, normativas, portarias e diretrizes sobre acesso à informação e Política de Dados Abertos e legislações complementares; **Padrões e instrumentos de organização e representação da informação**: padrões de metadados (e-PMG, *Dublin Core* - DC e DCAT-3), vocabulários (Vocabulário Controlado de Governo Eletrônico - VCGE) e Boas Práticas para Dados na Web (Lóscio; Burle; Calegari, 2019); e **Estrutura existente no Portal**: aspectos, funcionalidades e metadados da estrutura existente.

A análise da legislação e da documentação vigente permitiu identificar as diretrizes e as informações mínimas exigidas para a gestão de dados abertos no governo federal. Em seguida, foram estudados instrumentos que pudessem contribuir para a otimização da representação da informação no Portal. Por fim, analisou-se a estrutura atual da plataforma sob a ótica do usuário comum, visando identificar pontos de melhoria.

A partir do mapeamento realizado na primeira etapa dessa proposta, observou-se que o Portal já organiza os dados em blocos de preenchimento, ainda que essa estrutura não seja claramente apresentada. Com base nessa lógica, propôs-se o reagrupamento dos metadados em cinco categorias: Portal Brasileiro de Dados Abertos, Organização, Conjunto de Dados e Recurso e Reúso, considerando-se características comuns e compatibilidade com o que já está implementado. A categoria Gestor de Dados foi descartada nesta proposta por conter informações sensíveis, restritas ao uso interno dos órgãos. A estrutura geral da proposta é apresentada na Figura 1, por meio de um diagrama que ilustra as classes utilizadas e os relacionamentos, na representação da arquitetura do Portal.

Figura 1 – Relacionamento entre as classes propostas



Fonte: elaborada pelas autoras (2025).

A classe Portal Brasileiro de Dados Abertos associa-se com a classe Organização em uma relação de um para muitos, refletindo as diversas instituições públicas cadastradas. Cada Organização, por sua vez, pode publicar diversos Conjuntos de Dados e múltiplos Reúsos, também em relações de um para muitos. Já a classe Recurso está vinculada ao Conjunto de Dados em uma relação de um para muitos, considerando que um conjunto pode conter vários arquivos ou formatos. A classe Reúso conecta-se à classe Conjunto de Dados por meio de uma relação de muitos para muitos, uma vez que um reúso pode utilizar diferentes conjuntos.

Na segunda etapa da proposta, foram definidos os metadados associados a cada uma dessas classes, com base na legislação vigente, nas Boas Práticas para Dados na Web, nos padrões mencionados pelo governo brasileiro e no DCAT-3, recomendado pelo W3C. Segundo Lóscio, Burle e Calegari (2019), os metadados são essenciais para contextualizar os dados,

oferecendo informações sobre licenças, responsáveis, qualidade, formas de acesso e atualização.

A proposta amplia o número de metadados visíveis ao usuário de 34 para 84, todos voltados ao consumidor final, distribuídos da seguinte forma: 4 para o Portal Brasileiro de Dados Abertos, 23 para Organização (14 mantidos, 8 adicionados e 1 excluído), 27 para Conjunto de Dados (17 mantidos, 10 adicionados e 2 excluídos), 17 para Recurso (7 mantidos e 10 adicionados) e 13 para Reúso (8 mantidos e 5 adicionados).

Dos 84 metadados, 61 foram definidos como obrigatórios por serem essenciais à identificação e compreensão dos dados, incluindo campos que podem ser gerados automaticamente, como data de publicação ou última modificação. Os demais 23 são opcionais e contemplam informações complementares, como colaborador, vocabulários associados ou versões anteriores.

Para campos com opções de preenchimento restritas, propõe-se o uso de vocabulários controlados, assegurando padronização e consistência. Isso se aplica, por exemplo, aos campos ‘Tipo de organização’, ‘Natureza jurídica’, ‘Esfera’, ‘UF’, ‘Município’, ‘Licença de uso’, ‘Temas’ e ‘Tipo’. A adoção desses vocabulários evita inconsistências provocadas por linguagem natural e melhora o agrupamento e a recuperação das informações.

Além disso, recomenda-se o uso de vocabulários específicos por área, como Saúde, Educação e Transportes, no campo ‘Assuntos/Palavras-chave’, para favorecer a comunicação entre usuário e sistema e facilitar a identificação de conceitos relevantes, como sugerido por Kobashi (2008). Essa abordagem busca estruturar os DGA de forma mais clara e acessível, promovendo o seu efetivo aproveitamento por diferentes comunidades de usuários.

4.1 Organização

O primeiro nível de descrição visível para o usuário consumidor refere-se à categoria Organização, que reúne informações sobre o órgão responsável pela geração e publicação dos DGA, atendendo à exigência da LAI (Brasil, 2011).

Para essa categoria, foram propostos 23 campos de metadados: 14 provenientes da estrutura em vigor do Portal, oito novos campos adicionados e uma exclusão. Os campos mantidos incluem: ‘Esfera’, ‘Órgão SIORG’, ‘Código SIORG’, ‘Poder’, ‘Organização Pai’, ‘Imagem/Logo do órgão’, ‘Sigla’, ‘CNPJ’, ‘UF’, ‘Município’, ‘Site da organização’, ‘E-mail’, ‘Status’ e ‘Descrição’. Esses campos fornecem informações básicas e essenciais de

identificação, estrutura e contato da organização. Alguns desses campos, como 'Tipo de organização', 'Órgão SIORG', 'Código SIORG', 'Site' e 'E-mail', não são atualmente visíveis no Portal. Propõe-se que passem a ser exibidos, uma vez que são informações públicas, disponíveis, inclusive, em outras plataformas governamentais, como o Sistema de Informações Organizacionais do Governo Federal (SIORG).

O campo 'Organização Pai', embora atualmente obrigatório, deve ser tornado opcional, considerando-se que nem todas as organizações possuem hierarquia superior claramente definida. Já a exclusão sugerida refere-se ao campo 'Nome do Órgão' por ser redundante em relação ao campo 'Órgão SIORG', que adota vocabulário controlado a partir de cadastro oficial, favorecendo padronização, consistência e recuperação eficaz da informação (Fujita; Tolare, 2019).

Quanto aos campos adicionados, propõe-se a inclusão de: 'Natureza jurídica', 'Logradouro', 'Número', 'Bairro', 'Complemento', 'CEP', 'Telefone de contato' e 'Horário de atendimento'. Esses metadados ampliam o detalhamento necessário para satisfazer os requisitos da LAI e apresentam sintonia com a estrutura do SIORG.

As propostas se alinham às Boas Práticas para Dados na Web nº 1 (Fornecer metadados) e nº 2 (Fornecer metadados descritivos), conforme Lóscio, Burle e Calegari (2019), promovendo a compreensão contextual dos dados publicados.

4.2 Conjunto de dados

A categoria Conjunto de dados refere-se à coleção de dados publicada ou mantida por uma organização, acessível em um ou mais formatos. Trata-se do principal elemento retornado nas buscas realizadas no Portal, e cada conjunto de dados deve conter, no mínimo, um recurso associado.

Entre os campos existentes, sugere-se a manutenção de metadados como 'Título', 'Descrição', 'Palavras-chave', 'Item de inventário de origem', 'Licença de uso' e 'Periodicidade de atualização', por estarem em conformidade com as Boas Práticas para Dados na Web nº 2 (Fornecer metadados descritivos); nº 3 (Fornecer metadados estruturais); nº 4 (Fornecer informações sobre a licença dos dados); e nº 21 (Fornecer dados atualizados).

Em relação à visibilidade, recomenda-se a alteração de campos que atualmente estão disponíveis apenas para os publicadores, de modo que passem a ser exibidos, também, para os consumidores de dados, sendo eles: 'Área técnica responsável', 'E-mail da área técnica',

‘Cobertura temporal (início e fim)’, ‘Descontinuado’, ‘Data de descontinuação’, ‘Conjuntos de dados associados (interoperabilidade)’, ‘Cobertura espacial’ e ‘Granularidade espacial’. Essa sugestão fundamenta-se na premissa de que a publicidade é a regra e o sigilo a exceção no serviço público, e metadados que não envolvem riscos à segurança ou privacidade devem estar acessíveis ao cidadão.

Tais práticas estão em conformidade com as Boas Práticas para Dados na Web nº 1 (Tornar os dados disponíveis na web); nº 2 (Fornecer metadados descritivos); nº 5 (Fornecer informações de procedência dos dados); nº 13 (Usar representações de dados independentes de localidade); nº 22 (Fornecer explicação para dados indisponíveis); e nº 29 (Descrever a cobertura espacial).

Ainda no tocante à estrutura existente, propõe-se que os campos ‘Cobertura espacial’ e ‘Granularidade espacial’, opcionais, passem a ser de preenchimento obrigatório, baseando-se na premissa de que, por serem dados produzidos por organizações vinculadas ao Governo Federal, essas informações estão sempre presentes. Ambos os metadados se alinham à Boa Prática nº 13 (Usar representações de dados que sejam independentes de localidade).

No que diz respeito à exclusão de metadados, sugere-se a retirada dos campos ‘Versão’ e ‘Atualização de versão’ da categoria Conjunto de dados, com base no entendimento de que é mais apropriado relacionar esses metadados à categoria Recurso, uma vez que um mesmo conjunto pode conter diversos recursos, cada qual com diferentes versões.

Paralelamente, propõe-se a adoção de 10 novos metadados, sendo sete campos opcionais, um campo obrigatório e dois de geração automática. Entre os opcionais, estão: ‘Colaborador’, ‘API habilitada’, ‘Vocabulário controlado associado’, ‘Lista dos recursos cadastrados’, ‘Documentação’, ‘Granularidade temporal’ e ‘Metodologia’. O campo ‘Identificador’ é proposto como obrigatório, enquanto ‘Data de publicação’ e ‘Última alteração’ são sugeridos como campos de geração automática pelo sistema.

O campo ‘Colaborador’ é indicado como relevante em contextos de promoção da interoperabilidade e do reúso de dados. Essa proposta encontra respaldo na Boa Prática nº 5 (Fornecer informações de procedência dos dados).

O metadado ‘API habilitada’ está associado à Boa Prática nº 23 (Disponibilizar dados por meio de uma API), pois esse meio de acesso garante maior flexibilidade e facilidade de processamento por parte dos consumidores, e à Boa Prática nº 35 (Citar a publicação original do conjunto de dados), facilitando a referência e o consumo automatizado das informações.

A inclusão do campo ‘Vocabulário controlado associado’ está em consonância com a Boa Prática nº 15 (Reutilizar vocabulários), dando preferência aos padronizados.

A recomendação de inclusão dos metadados ‘Documentação’, ‘Granularidade temporal’ e ‘Metodologia’ se justifica pelo fato de constarem na Cartilha Técnica para Publicação de Dados Abertos, publicação oficial do governo federal que serve como documento norteador para os órgãos publicadores. Esses metadados complementam a descrição e estão alinhados à Boa Prática nº 2 (Fornecer metadados descritivos).

O campo ‘Identificador’, sugerido como obrigatório, responde à Boa Prática nº 9 (Usar URIs persistentes como identificadores de conjuntos de dados). Esse metadado é essencial para a identificação, o gerenciamento e a reutilização dos dados, garantindo que os conjuntos ou suas informações possam ser localizados e citados com facilidade.

Por fim, os campos ‘Data de publicação’ e ‘Última alteração’, de geração automática, alinham-se à Boa Prática nº 21 (Fornecer dados atualizados), pois fornecem ao consumidor um histórico de publicação e modificações do conjunto de dados.

4.3 Recursos

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, ainda, um documento (Brasil, [20–?]). No modelo proposto, essa categoria contempla 17 metadados no total, sendo sete oriundos da estrutura vigente do Portal e dez novos campos sugeridos.

Conforme a Boa Prática para Dados na Web nº 2 (Fornecer metadados descritivos), os metadados descritivos podem incluir características genéricas de uma distribuição, como ‘Título’, ‘Descrição’, ‘Data da publicação’ e ‘Formato’. Uma distribuição, por sua vez, representa um determinado formato disponível de um conjunto de dados, como um arquivo no formato CSV para *download* ou uma API (Albertoni *et al.*, 2024). O único desses campos que não estava contemplado originalmente no Portal era o campo ‘Data da publicação’, cuja adição é, aqui, recomendada. Além dele, os campos ‘URL’, ‘Tipo’, ‘Data de criação’ e ‘Última alteração’ foram mantidos inalterados, por estarem de acordo com boas práticas.

Adicionalmente, os campos ‘Formato’ e ‘Tipo’ relacionam-se, também, à Boa Prática nº 3 (Fornecer metadados estruturais). Os campos ‘Data da publicação’, ‘Data de criação’ e ‘Última alteração’ respondem à Boa Prática nº 21 (Fornecer metadados atualizados),

enquanto o campo 'URL' alinha-se às Boas Práticas nº 5 (Fornecer informações de procedência de dados) e nº 20 (Fornecer acesso em tempo real).

Propõe-se, ainda, a adição dos campos 'Versão' e 'Atualização de versão', anteriormente presentes na categoria Conjunto de dados. Essa realocação fundamenta-se no entendimento de que cada recurso pode possuir sua própria versão. Como complemento, recomendam-se os campos opcionais 'Notas de versão' e 'Versão anterior'. Essas adições atendem à Boa Prática nº 7 (Fornecer indicador de versão) e à Boa Prática nº 8 (Fornecer o histórico de versão).

Outros campos sugeridos incluem 'Assuntos/palavras-chave', 'Idioma', 'Tamanho do arquivo', 'Sigilo/direitos de acesso' e 'Normativa/observância legal'. O campo 'Assuntos/palavras-chave' constitui um descritor básico que permite a indexação do conteúdo do recurso de forma resumida, utilizando termos relevantes. Esse campo, além de atender à Boa Prática nº 2 (Fornecer metadados descritivos), pode ter sua eficácia aprimorada por meio do uso de vocabulário controlado, como já sugerido para a categoria Conjunto de Dados, o que contribuiria para a padronização da representação e a melhoria na recuperação da informação.

Embora o Portal abranja dados produzidos por órgãos brasileiros, compreende-se que não há obrigatoriedade de que esses dados estejam exclusivamente no idioma português. Por isso, recomenda-se a inclusão do campo 'Idioma'.

Já o campo 'Sigilo/direitos de acesso' tem como objetivo estabelecer restrições ou permissões de acesso, delimitando quem pode visualizar ou utilizar um determinado recurso. Esse metadado é complementado pelo campo 'Normativa/observância legal', o qual busca explicitar restrições decorrentes de legislações específicas, como a Lei Geral de Proteção de Dados (LGPD) ou a previsão, contida na Lei de Acesso à Informação (LAI), de restrição temporária ao acesso público de informações relacionadas à segurança da sociedade e do Estado.

Essas adições e esses ajustes visam garantir maior granularidade e transparência na descrição dos recursos, promovendo a conformidade com padrões reconhecidos e aprimorando a qualidade do acesso e da compreensão dos dados.

4.4 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 abranger desde aplicativos e *sites* até

outras soluções tecnológicas. Esta categoria contempla um total de 13 metadados, dos quais oito pertencem à estrutura atual do Portal e cinco são propostos como adições.

Mantiveram-se inalterados os campos 'Nome', 'URL', 'Descrição', 'Tipo', 'Imagem' e 'Organização à qual o recurso está associado'. Os metadados 'Nome', 'Descrição' e 'Imagem' estão alinhados à Boa Prática para Dados Na Web Nº 2 (Fornecer metadados descritivos); o campo 'URL' se relaciona com as Boas Práticas Nº 5 (Fornecer informações sobre a procedência dos dados) e Nº 20 (Fornecer acesso em tempo real); o campo 'Tipo' vincula-se à Boa Prática Nº 3 (Fornecer metadados estruturais); e o metadado 'Organização à qual o recurso está associado' atende tanto à Prática Nº 3 quanto à Nº 5.

Sugere-se tornar o campo 'Conjuntos de dados de origem' visível ao usuário, permitindo que o consumidor verifique a correlação entre os dados utilizados. Este metadado está em consonância com as Boas Práticas 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). Além disso, propõe-se a alteração do campo 'Nome do responsável' para 'Área técnica responsável', com o intuito de estabelecer um ponto de contato institucional em vez de pessoal, assegurando a proteção de dados sensíveis. Para fortalecer essa via de contato, também se sugere a adição do campo 'E-mail da área técnica'. Esses dois metadados contribuem para o cumprimento da Boa Prática Nº 2 (Fornecer metadados descritivos), além de atenderem, potencialmente, à Boa Prática Nº 29 (Coletar *feedback* de consumidores de dados), ao disponibilizarem um canal direto de comunicação.

Ainda como proposta de aprimoramento, incluem-se os metadados 'Data de criação do reúso', 'Data de publicação do reúso' e 'Última alteração'. O primeiro é de preenchimento obrigatório, enquanto os dois últimos podem ser gerados automaticamente pelo sistema. Esses campos visam oferecer um maior detalhamento sobre o histórico do reúso e atendem à Boa Prática Nº 21 (Fornecer dados atualizados).

Entre os metadados existentes, destaca-se o campo 'URL do reúso', que, ao estar correlacionado à propriedade 'Localização (*Location*)', é considerado um dos mais relevantes, uma vez que permite acesso a informações mais detalhadas e específicas disponíveis diretamente na plataforma onde o reúso está hospedado.

Considerando a proposta de forma geral, optou-se, sempre que possível, pela preservação da estrutura vigente, sendo que as sugestões de alterações e adições foram

elaboradas com base nas legislações pertinentes, priorizando a experiência do usuário consumidor, que se beneficiará à medida que as informações disponibilizadas forem mais completas, estruturadas e descritivas. A proposta completa pode ser visualizada no *site* https://observatoriodedadosabertos.eci.ufmg.br/metaopengov/ (Nascimento Silva, 2025).

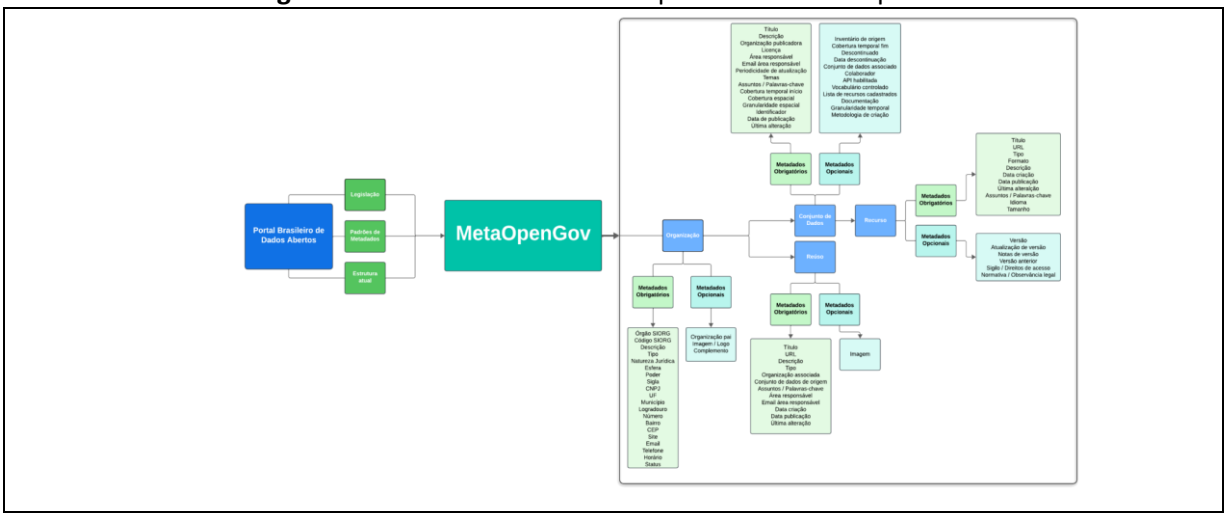
5 MODELO METAOPENGOV

Observa-se que o governo federal tem empreendido esforços em diversas frentes para assegurar a implementação dos princípios de dados abertos por meio do desenvolvimento e da oferta de ferramentas destinadas à disponibilização e uso público dos dados. No entanto, mais do que investir em novas soluções tecnológicas, é essencial avaliar criticamente a usabilidade, validade, pertinência e adequação dos recursos já existentes. Essa análise é fundamental não apenas para evitar retrabalho, mas também para promover o reúso de dados e estruturas consolidadas.

Com base nesse entendimento, apresenta-se uma proposta de metadados elaborada a partir do levantamento do cenário atual e das principais recomendações encontradas na literatura científica. O modelo desenvolvido foi denominado **MetaOpenGov**, em referência à sua aplicação à descrição dos DGA, e está resumido na Figura 2.

A proposta contempla 84 metadados e foi correlacionada com as Boas Práticas para Dados na Web (Lóscio; Burle; Calegari, 2019), tornando possível identificar vínculos entre todos os metadados sugeridos e, pelo menos, uma das práticas recomendadas, o que reforça a solidez, relevância e atualidade da estrutura desenvolvida.

Figura 2 – Modelo de metadados para DGA - MetaOpenGov



Fonte: elaborada pelas autoras (2025).

A proposta apresentada foi guiada pelos fundamentos da CI e pelo uso de seus instrumentos teóricos e técnicos com o objetivo de beneficiar, principalmente, o usuário final, que, na maioria das vezes, não possui familiaridade com a manipulação de dados.

O modelo foi fundamentado nos princípios da OI, com ênfase na representação da informação por meio do uso estruturado de metadados. O propósito é viabilizar uma compreensão mais clara dos dados e promover uma recuperação informacional mais eficiente e satisfatória. Como ressalta Arakaki (2019), cada tipo de metadado exerce uma função específica na representação dos recursos informacionais. Assim, buscou-se incorporar uma gama diversificada de elementos descritivos, capazes de oferecer uma representação rica, detalhada e funcional, adequada às distintas necessidades informacionais.

A proposta também visa contribuir para a redução da heterogeneidade dos dados ao propor campos padronizados de descrição que incentivem a interoperabilidade e facilitem a integração entre diferentes conjuntos de dados por meio da identificação de semelhanças estruturais.

6 CONSIDERAÇÕES FINAIS

Esta pesquisa foi desenvolvida com o objetivo de propor um modelo de metadados para a representação e organização da informação no Portal Brasileiro de Dados Abertos por meio da utilização de metadados e padrões voltados à publicação de dados abertos. A abordagem adotada foi a de uma pesquisa aplicada, com a utilização de técnicas de pesquisa bibliográfica, documental e observação direta, realizada ao longo do ano de 2024 no âmbito de um estudo acadêmico.

A proposta resultante compreende um total de 84 metadados, sendo quatro destinados à descrição de informações básicas do Portal, 23 voltados à categoria Organização, 27 à categoria Conjunto de dados, 17 relacionados à categoria Recurso e 13 destinados à categoria Reúso. Desse total (84), foram definidos como obrigatórios 61 metadados considerados essenciais para a identificação, compreensão e utilização dos dados, como título e descrição. Os 23 metadados restantes foram classificados como opcionais por tratarem de informações complementares ou que não estão presentes em todos os contextos, como vocabulários controlados associados ou versões anteriores.

O aumento expressivo na quantidade de metadados em relação à estrutura vigente reflete a necessidade de prover o máximo de informações ao usuário, permitindo uma

avaliação mais precisa sobre a pertinência de determinado conjunto de dados para atender às suas demandas, além de facilitar o reúso desses dados em novas aplicações, o que demanda um nível elevado de detalhamento.

Além da estrutura conceitual proposta, fundamentada em marcos legais e em boas práticas identificadas na literatura científica, foi realizada a correlação de todos os campos definidos com as propriedades e classes do padrão DCAT-3, o que viabiliza a aplicação técnica da proposta em ambientes que adotem os princípios da *web* semântica.

Esta pesquisa oferece uma contribuição teórica e técnica para o campo da CI e traz uma reflexão crítica sobre o papel dos dados abertos na promoção da transparência, do acesso à informação e da participação cidadã. Ao propor uma estrutura mais robusta, detalhada e alinhada a padrões internacionais, pretende-se fortalecer os alicerces dos DGA, para que eles possam ser utilizados, reutilizados e transformados em conhecimento socialmente relevante. Além disso, destaca a importância do profissional da informação na implementação das políticas de dados abertos no Brasil.

REFERÊNCIAS

ALBERTONI, R. *et al* (ed.). **Data Catalog Vocabulary (DCAT).** Version 3. W3C, [*s. l.*], 2024. Disponível em: https://www.w3.org/TR/vocab-dcat-3/. Acesso em: 15 abr. 2025.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2010. Disponível em: https://repositorio.unesp.br/server/api/core/bitstreams/e1fd7853-781a-46bd-beee-037375fb178f/content. 2010. Acesso em: 06 ago. 2025.

ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados**: modelo baseado na família prov. 2019. 139 f. Tese (Doutorado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Marília, 2019. Disponível em: https://repositorio.unesp.br/items/6b7dc0eb-1feb-4489-a81c-bc41ccbb4fb5. Acesso em: 30 abr. 2025.

BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais** […] São Paulo: ANCIB, 2008. Disponível em: https://brapci.inf.br/index.php/res/v/176535. Acesso em: 6 ago. 2025.

BRASIL. Governo Brasileiro. Comitê Executivo de Governo Eletrônico. **Padrão de Metadados do Governo Eletrônico (e-PMG)**.Brasília: MGI, 2014. 41 p. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/PMGVersao1_1.pdf. Acesso em: 06 ago. 2025.

BRASIL. Governo Federal. **Dados Abertos**: Sobre. [20–?]. Disponível em: https://dados.gov.br/dados/conteudo/sobre. Acesso em: 10 maio 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. **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: 12 maio 2025.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science (JASIS)**, New York, v. 45, n. 5, p. 351-360, 1991.

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.; TOLARE, J. B. Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. **Informação e Informação**, Londrina, v. 24, n. 2, p. 93-125, 2019. Disponível em: 10.5433/1981-8920.2019v24n2p93. Acesso em: 10 abr. 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: 30 abr. 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, 2020. Disponível em: https://periodicos.ufmg.br/index.php/pci/article/view/22283. Acesso em: 6 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. W3C, [*s. l.*], 2019. Disponível em: https://www.w3.org/Translations/DWBP-pt-BR/. Acesso em: 05 maio 2025.

MIRANDA, R. C. R. O uso da informação na formulação de ações estratégicas pelas empresas. **Ciência da informação**,Brasília, v. 28, p. 286-292, 1999. Disponível em: https://www.scielo.br/j/ci/a/r7L9msHr6FfrYpJ5PKk8fsS/. Acesso em: 06 ago. 2025.

NASCIMENTO SILVA, P. **MetaOpenGov.** Observatório de Dados Abertos, Belo Horizonte 2025. Disponível em: https://observatoriodedadosabertos.eci.ufmg.br/metaopengov/. Acesso em: 07 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: 06 ago. 2025.

OPEN KNOWLEDGE BRASIL. **Por que open**: definição de conhecimento aberto . [20–?]. Disponível em: https://ok.org.br/dados-abertos/. Acesso em: 17 abr. 2025.

ORTEGA, C. D. Contexto de desenvolvimento da organização da informação, com enfoque para a catalogação, na Escola de Ciência da Informação da UFMG. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 18, n. 2, p. 182-215, 2013. Disponível em: http://hdl.handle.net/20.500.11959/brapci/35682. Acesso em: 06 ago. 2025.

PAULO, J. S. **Dados abertos governamentais**: uma análise aplicada ao ministério público do trabalho. 2018. 83 f. Dissertação (Mestrado Profissional em Gestão Pública) - Universidade Federal do Espírito Santo, Vitória, 2018. Disponível em: https://repositorio.ufes.br/server/api/core/bitstreams/811af1b4-4fda-4d26-b660-1b45f3c9c436/content. Acesso em: 6 ago. 2025.

SETZER, V. W. Dado, informação, conhecimento e competência. **DataGramaZero Revista de Ciência da Informação**, v. 28, 1999. Disponível em: https://www.ime.usp.br/~vwsetzer/datagrama.html. Acesso em 11 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 em Ciência da Informação) - Universidade Federal de São Carlos, São Paulo, 2021. Disponível em: https://repositorio.ufscar.br/items/44c8251e-369a-4998-a11d-6c0205f9371a . Acesso em: 18 abr 2025.

AGRADECIMENTO

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, processo 303721/2025-1.