

Correspondência dos autores

1 Universidade Federal de Minas Gerais
Belo Horizonte, MG - Brasil
oliveiracomercial88@gmail.com

2 Universidade Federal de Minas Gerais
Belo Horizonte, MG - Brasil
patricians.prof@gmail.com

Representação e recuperação de dados governamentais abertos: uma revisão de literatura

Danielle Teixeira de Oliveira¹ Patrícia Nascimento Silva²

RESUMO

Introdução: O desenvolvimento tecnológico da sociedade ao longo das últimas décadas permitiu uma maior interação e transparência entre o governo e a sociedade por meio da consolidação das iniciativas de governo aberto. Nesse contexto, surgiram os Dados Governamentais Abertos (DGA), que são informações do setor público disponibilizadas em formatos reutilizáveis para livre acesso e utilização. Contudo, para este novo cenário, os dados precisam ser representados, organizados, tratados e recuperados, o que atribui relação direta com os conceitos, as metodologias e os instrumentos da Ciência da Informação (CI). **Objetivo:** Investigar a produção acadêmica sobre a representação e recuperação dos DGA na CI. **Metodologia:** Pesquisa de natureza exploratória e descritiva por meio de uma revisão de literatura. O protocolo da revisão foi criado com base na literatura e executado entre julho e agosto de 2023. **Resultados:** Foram selecionados 40 documentos nos idiomas português, inglês e espanhol para análise de conteúdo. Foram identificadas sete categorias de análise a partir de documentos publicados entre 2011 e 2023, a maioria no idioma inglês e classificados como artigos de pesquisa. Dos 40 documentos analisados, 23 abordaram a representação dos DGA, sete abordaram a recuperação e 10 abordaram ambos os temas, que foram subdivididos em sete temáticas. **Conclusão:** Diante das abordagens variadas encontradas na literatura, ressalta-se a importância contínua dos DGA na esfera da CI e de áreas correlatas, proporcionando uma base para futuras pesquisas e aprimoramentos na representação, recuperação e utilização dos DGA, impulsionando, assim, avanços na transparência governamental, na inovação cívica e na ciência cidadã, aplicados a novos modelos de negócio.

PALAVRAS-CHAVE

Dados abertos. Dados governamentais abertos. Representação da informação. Ciência da Informação.

Open government data representation and retrieval: a literature review

ABSTRACT

Introduction: The technological development of society over the last few decades has enabled greater interaction and transparency between government and society through the consolidation of open government initiatives. In this context, Open Government Data (OGD) has emerged, which public sector information is made available in reusable formats for free access and use. However, for this new scenario, data needs to be

represented, organized, processed and retrieved, which is directly related to the concepts, methodologies and tools of Information Science (IS). **Objective:** To investigate academic production on the representation and retrieval of OGD in IS. **Methodology:** An exploratory and descriptive study using a literature review. The review protocol was created based on the literature and carried out between July and August 2023. **Results:** 40 documents in Portuguese, English and Spanish were selected for content analysis. Seven categories of analysis were identified from documents published between 2011 and 2023, the majority in English and classified as research articles. Of the 40 documents analyzed, 23 addressed the representation of OGD, seven addressed retrieval and 10 addressed both topics, which were subdivided into seven themes. **Conclusion:** Given the varied approaches found in the literature, the continued importance of OGD in the sphere of IS and related areas is highlighted, providing a basis for future research and improvements in the representation, retrieval and use of OGD, thus driving advances in government transparency, civic innovation and citizen science, applied to new business models.

KEYWORDS

Open data. Open government data. Information representation. Information retrieval. Information Science.

CRediT

- **Reconhecimentos:** Não aplicável.
- **Financiamento:** Não aplicável.
- **Conflitos de interesse:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Não aplicável.
- **Disponibilidade de dados e material:** Não aplicável.
- **Contribuições dos autores:** Conceitualização, Curadoria de dados, Análise formal, Investigação, Metodologia, Visualização, Escrita - rascunho original, Escrita - revisão & edição: OLIVEIRA, D. T.; NASCIMENTO SILVA, P.; Aquisição de financiamento, Administração do projeto, Supervisão, Validação: NASCIMENTO SILVA, P.
- **Imagem:** Foto extraída do Currículo Lattes.

JITA: IM. Open data.

ODS: 16. Paz, justiça e instituições eficazes.



Artigo submetido ao sistema de similaridade

Submetido em: 16/02/2024 – Aceito em: 30/10/2024 – Publicado em: 12/11/2024

Editor: Gildenir Carolino Santos

1 INTRODUÇÃO

O desenvolvimento tecnológico da sociedade, ao longo das últimas décadas, permitiu uma maior interação e transparência entre o governo e a sociedade por meio da consolidação das iniciativas de governo aberto. Essa visão inovadora da administração pública, que teve início a partir de 2009, impulsionou medidas destinadas à ampliação da transparência, ao combate à corrupção, ao estímulo à participação da sociedade civil e ao fomento do avanço de tecnologias emergentes (Bertin *et al.*, 2019).

Após as ações de transparência e o governo aberto, iniciaram-se, nos EUA, iniciativas sobre a disponibilização de dados abertos relacionados ao governo, *Open Government Data* (OGD), que fortaleceram as ações relacionadas ao tema e implicaram a disponibilização de bases de dados oficiais para a comunidade em geral (Ribeiro; Almeida, 2011).

Outro marco dessas ações foi a parceria para o Governo Aberto, que foi adotada por diversos países, inclusive no Brasil, e compreende princípios da transparência, da participação e da colaboração, relacionados com a prestação de contas, a responsabilização e a inovação tecnológica (Varela, 2017). Frente a isso, os governos têm acompanhado essas ações por meio de investimentos na adoção de legislações e políticas específicas, fomentando a transparência e o fortalecimento da democracia.

Os objetivos da publicação dos dados abertos governamentais, ou dados governamentais abertos (DGA), são ambiciosos e suas possibilidades de utilização são extensas. Contudo, mediante fenômenos como o *big data*, que envolvem um grande e heterogêneo volume de dados, surgem preocupações em relação à representação, ao acesso e à recuperação desses dados, a fim de permitir sua efetiva utilização, entre outros aspectos relacionados ao seu manuseio. Assim, com o aumento e a expansão dos DGA, os problemas de organização e recuperação relacionados às grandes quantidades de dados e informações são perceptíveis, sendo essencial conhecer as abordagens e os métodos utilizados para representá-los e recuperá-los (Nascimento Silva, 2023).

Segundo Davenport (2002), os dados são simples observações sobre o estado do mundo e a informação são dados dotados de relevância e propósito e exige necessariamente a mediação humana.

Os DGA estão relacionados ao conceito de dado, quando disponibilizados de forma primária, em conjuntos, e também à informação, quando esses conjuntos são contextualizados e apresentados em portais e painéis. Com isso, é na Ciência da Informação (CI) que os DGA estão fundamentados com as metodologias e os instrumentos de organização, representação e recuperação que **subsidiarão** a manipulação destes dados e informações para seu efetivo reúso. Contudo, a temática dos DGA ainda é recente e essa associação com a CI nem sempre é visualizada nas pesquisas que envolvem os dados abertos e os DGA.

Segundo Lima e Álvares (2012), na CI, a representação está relacionada às formas de simbolizar a informação e o conhecimento, e sua principal característica é a substituição do texto do documento por uma descrição abreviada. Para Mooers (1951), a recuperação da informação abrange aspectos intelectuais da descrição da informação e sua especificação para pesquisa, além de quaisquer sistemas, técnicas ou dispositivos utilizados para realizar a operação.

Diante deste cenário, a seguinte questão de pesquisa norteou este estudo: como a representação e a recuperação dos DGA têm sido investigadas no âmbito da CI? Assim, esta pesquisa objetivou mapear a produção acadêmica sobre a representação e a recuperação dos DGA na CI. Especificamente, buscou-se: (i) identificar a produção científica sobre a temática e (ii) mapear e delinear categorias sobre a representação e recuperação nos estudos sobre DGA. Para tanto, foi realizada uma revisão de literatura em bases científicas da área. Destaca-se que este estudo é um dos artefatos que integram uma pesquisa maior, intitulada: “Observatório de Dados Abertos”, que investiga a temática de dados abertos no contexto da CI.

2 REFERENCIAL TEÓRICO

Para a construção do referencial teórico e da fundamentação deste trabalho, foi realizado um levantamento bibliográfico para conceituar os termos dados, dados abertos, governo aberto, DGA, organização da informação, representação da informação e recuperação da informação no contexto da CI e áreas afins, já que a CI é considerada essencial para a efetivação da busca pelo acesso democrático de dados abertos (Pessoa; Sousa; Cardoso, 2020).

No que diz respeito à conceituação de dados, Souza (2006) os define como sequências de símbolos dotados de significado, os quais podem ser codificados, interpretados e processados por sistemas computacionais, além de serem transmitidos por meio de redes de comunicação. Com os movimentos de acesso aberto, iniciados nos anos 2000, é criado o conceito de dados abertos e seus princípios.

Os dados abertos fazem uma associação dos dados com a sua disponibilidade para reutilização em outras investigações, abrindo portas para variados tratamentos, aplicações e resultados (Sayão; Sales, 2013). Conforme o Guia de incentivo ao (re)uso de dados abertos (CEWEB, 2016), dados abertos são aqueles acessíveis a todos, devendo ser publicados, disponibilizados em forma legível por máquina e ter uma licença que permita o acesso, o uso e o compartilhamento para fins comerciais ou não.

O governo aberto surge de uma visão inovadora da administração pública, a partir de 2009 (Bertin *et al.*, 2019), com ações iniciadas nos EUA. O Brasil vem se destacando nesse cenário desde 2011, por meio da *Open Government Partnership*, quando, junto a outros sete países, assumiu compromissos nos campos da participação, da colaboração, do combate à corrupção e da transparência, contando com o suporte das novas Tecnologias de Informação e Comunicação (TICs) (Possamai; Souza, 2020).

Agune, Gregório Filho e Bolliger (2010) conceituam governo aberto como a disponibilização pela *Internet* de informações e dados governamentais de domínio público, permitindo seu uso livre pela sociedade, propiciando o acesso a dados primários para combinação, cruzamento e geração de novas informações. Marcos-Martín e Soriano-Maldonado (2011) complementam essa visão ao destacarem os pilares de transparência, participação e colaboração nas políticas de governo aberto.

Os DGA compreendem a interseção dos dados abertos com o governo aberto e, conforme seus oito princípios compreendem: dados completos, primários, atuais, acessíveis, processáveis por máquina, não discriminatórios, não proprietários e livres de licença, disponibilizados pelos governos por meio da *internet* para livre utilização, reutilização, informando autoria e compartilhados pela mesma licença (Agune; Gregório Filho; Bolliger, 2010; OGD, 2007).

Vaz, Ribeiro e Matheus (2011) enxergam os DGA como veículos de transparência enquanto Sinif e Bounabat (2018) ressaltam a capacidade das iniciativas de DGA de fornecerem informações governamentais a partes interessadas para reutilização, fomentando o desenvolvimento econômico, a transparência e a modernização. Os DGA, por sua vez, são produzidos pelo governo e disponibilizados para leitura, acompanhamento e reutilização em novos projetos, sites e aplicativos (W3C, 2011).

A respeito da organização, representação e recuperação da informação, Ortega (2013) define a organização da informação como um conjunto de procedimentos que tornam os

documentos utilizáveis para públicos específicos, visando suprir necessidades de informação em diversos contextos. Essa organização culmina na representação da informação, entendida como um conjunto de elementos descritivos que retratam os atributos de um objeto informacional específico. Em resumo, a organização da informação é “um processo envolvendo a descrição física e de conteúdo dos objetos informacionais e o produto desse processo descritivo é a representação da informação” (Bräscher; Café, 2008).

Ainda em relação aos processos de organização, Sant’Ana (2013) identifica três fases de acesso a dados: coleta, armazenamento e recuperação. A última engloba consultas e visualizações, efetivando esse acesso. Para Fernalda (2003), o processo de recuperação de informação consiste em identificar, no conjunto de documentos (*corpus*) de um sistema, que atendem à necessidade de informação do usuário. Com isso, para Baeza-Yates e Ribeiro Neto (2013), a recuperação da informação trata da representação, do armazenamento, da organização e do acesso a itens de informação, ou seja, elementos essenciais para atender às demandas dos usuários.

Por fim, o contexto dos DGA, que são disponibilizados pela *internet* e processáveis por máquina, envolve a organização de grandes volumes de dados (*big data*), novas tecnologias, plataformas e ambientes informacionais. Zafalon (2012) argumenta que ambientes digitais necessitam de padrões para a representação de documentos eletrônicos, garantindo sua recuperação e seu acesso. A autora também defende que a representação de recursos informacionais é vital para facilitar a comunicação entre unidades de informação e usuários, envolvendo a transmissão de documentos reunidos por suas características de conteúdo e expressão de forma individualizada. É nesse cenário que investigar a representação e a recuperação dos DGA é basilar para o seu efetivo reuso.

3 METODOLOGIA

A pesquisa é de natureza descritiva, por introduzir as características de determinada população ou grupo, e exploratória, pois tem como objetivo apresentar maior clareza em relação ao problema de pesquisa, tornando-o mais evidente, por meio de uma revisão bibliográfica elaborada para apresentar fundamentação teórica e identificar o estágio do conhecimento do tema na atualidade (Gil, 2017).

A revisão da literatura utiliza-se de dados que já foram estudados por outros pesquisadores e estão registrados, trabalhando a partir das contribuições dos autores dos estudos analíticos constantes nas bibliografias (Severino, 2013). O artigo de revisão de literatura também tem o objetivo de identificar lacunas na literatura e apresentar sugestões para a resolução de problemas (Koller, Couto, Von Hohendorff, 2014).

O **objetivo geral** desta revisão foi explorar o tema, analisar a produção científica relacionada à representação e à recuperação dos DGA e resolver as seguintes questões: quais trabalhos abordam a temática dos DGA na área da CI e afins? Como os DGA estão sendo representados? Como os DGA são recuperados? Diante da especificidade da temática, foram analisados alguns protocolos de revisão, como: Arlene Fink (2005), ChituOkoli; Duarte e Mattar (2019) e Galvão e Pereira (2014), no entanto, optou-se pela construção de um protocolo específico e adaptado para este estudo.

A **seleção das bases** de dados considerou a sua relevância acadêmica e foi utilizado o acesso ao Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Portal CAPES), que fornece acesso a bases de dados pagas por meio de convênio com a instituição de origem dos pesquisadores. Dessa forma, a realização das pesquisas ocorreu na rede local da instituição de vinculação.

Para definição das **estratégias de busca**, foram selecionadas sete bases de dados para avaliação de adequação, a partir do conhecimento prévio de utilização dos autores e da sua relevância no meio acadêmico: Base de dados referencial de artigos de periódicos em Ciência da Informação (BRAPCI), *Library, Information Science and Technology Abstracts* (LISTA-EBSCO), *Science Direct*, *Scientific Electronic Library Online* (SciELO), *Scopus*, *Web Of Science*, *Wiley Online Library*.

A estratégia de busca foi testada nas bases e compreendeu os descritores “dados abertos governamentais”, “ciência da informação” e suas variações. Ambas as expressões foram traduzidas para as línguas inglesa e espanhola. Após ajustes, a *string* geral de busca foi: ("dados abertos governamentais" OR "dados governamentais abertos" OR "*open government data*" OR "*datos abiertos del gobierno*") AND ("ciência da informação" OR "*informationscience*" OR "*ciencias de la informacion*"). Optou-se por realizar a busca em todos os campos, visto que era um parâmetro comum em todas as bases e não restringiu a busca.

A base BRAPCI não permitiu o relacionamento entre expressões, apenas entre termos, o que inviabilizou a replicação da *string*, por isso, ela foi eliminada nesta etapa. Nas demais bases, a *string* construída foi replicada, mantendo os termos de busca e selecionando o campo de busca que realizava a procura em todos os campos disponíveis em cada base.

O próximo estágio consistiu em definir a **elegibilidade, os critérios de inclusão e os critérios de exclusão** para os documentos recuperados. Nas buscas, foram considerados documentos nos idiomas: inglês, espanhol e português, por serem as línguas de maior familiaridade dos autores, e não houve delimitação de data, uma vez que a temática ainda é recente. Foram selecionados artigos de periódicos e trabalhos publicados em eventos, considerando que esses tipos de documentos recebem avaliação por pares. Como critérios de inclusão, foram considerados documentos que possuíam resumo e que estivessem alocados na CI ou em áreas correlatas, como computação, administração ou gestão pública. Foram estabelecidos como critérios de exclusão duplicatas, documentos escritos em idiomas diferentes dos três selecionados e documentos que não estivessem disponíveis na íntegra. Nessa etapa, ocorreu a eliminação da base de dados LISTA-EBSCO do escopo, porque os pesquisadores tiveram acesso apenas aos resumos dos documentos recuperados na busca.

Como procedimento de seleção e análise, considerando que a busca incluiu todos os campos das bases, em uma primeira etapa, foram selecionados todos os documentos cujos títulos possuíam pelo menos o termo ‘dados governamentais abertos’ ou suas variações, além da avaliação geral da adequação do conteúdo ao objetivo geral da pesquisa. Para a segunda etapa, o mesmo procedimento foi replicado nos resumos dos documentos. A terceira e última etapa correspondeu à análise dos documentos selecionados para leitura completa utilizando a técnica de análise de conteúdo, seguindo simplificadaamente as fases de pré-análise, exploração do material e tratamento dos resultados, de Bardin (2016). Esta análise dos documentos para seleção e posterior discussão foi realizada com base nas temáticas inicialmente evidenciadas nos resumos e, posteriormente, nos objetivos dos estudos, no referencial teórico, nos resultados

obtidos e nas conclusões dos trabalhos. A partir disso, os estudos foram agrupados conforme as similaridades apresentadas e foi identificada a abordagem utilizada.

O processo partiu dos temas mais gerais, como “representação e recuperação da informação”, até chegar a temáticas mais específicas, que foram categorizadas, conforme indica a técnica de Bardin (2016). Foram excluídos, após a análise completa, os documentos que não possuíam abordagem conceitual, teórica ou metodológica sobre DGA. O resumo do protocolo construído foi apresentado no Quadro 1.

Quadro 1. Definição de protocolo de revisão de literatura

Protocolo de revisão de literatura	
Critérios	Descrição
Objetivo geral	Identificar artigos de periódicos científicos e trabalhos publicados em eventos sobre dados governamentais abertos, sua representação e recuperação.
Questões a serem resolvidas	Quais trabalhos abordam a temática de dados governamentais abertos na área da Ciência da Informação e afins?
	Como os dados governamentais abertos estão sendo representados?
	Como os dados governamentais abertos são recuperados?
Fontes de informação pesquisadas	Base de dados: LISTA, SciELO, Science Direct, Scopus, Web Of Science, Wiley
Critérios de elegibilidade	Idioma: inglês, espanhol, português
	Sem delimitação de data
	Tipologia documental: artigos de periódicos, trabalhos de eventos (revisão por pares)
Critérios de inclusão e de exclusão	Inclusão: documentos que possuem resumo;
	Documentos nos idiomas inglês, português e espanhol;
	Trabalhos na área de Ciência da Informação ou áreas afins (Computação, Administração, Gestão Pública);
	Exclusão de documentos duplicados;
	Exclusão de documentos que não foram escritos em português, inglês e espanhol;
	Documentos cujos títulos e resumos não possuam abordagem sobre dados governamentais abertos;
Documentos que não estejam disponíveis na íntegra via Portal Capes.	
Campos de busca	Todos os campos
Expressões de busca	dados governamentais abertos e suas variações E ciência da informação
	Obs.: as expressões foram utilizadas em português, inglês e espanhol
String	("dados abertos governamentais" OR "dados governamentais abertos" OR "open government data" OR "datosabiertosdelgobierno") AND ("ciência da informação" OR "informationscience" OR "ciencias de lainformacion")
Procedimentos de seleção dos documentos recuperados	Inicialmente, a leitura dos títulos dos documentos recuperados com o intuito de verificar a pertinência do conteúdo ao objetivo geral da pesquisa e, em seguida, o mesmo procedimento com os resumos.
Procedimentos de análise	Leitura completa dos documentos a fim de identificar elementos de representação e recuperação de informação em dados governamentais abertos.
Critério de exclusão após análise dos documentos	Trabalhos que não possuem abordagem conceitual, teórica ou metodológica sobre dados governamentais abertos.

Fonte: elaborado pelas autoras (2024)

Após a execução do protocolo em todas as bases selecionadas, todos os resultados foram exportados e agrupados em *softwares* de planilhas eletrônicas. O próximo passo foi realizar uma comparação dos resultados para identificar aqueles que estavam duplicados. A partir disso, foram aplicados os critérios de inclusão, exclusão e análise definidos no protocolo da revisão.

A análise de conteúdo permitiu agrupar os documentos em categorias de análise, relacionando técnicas e conteúdos sobre as temáticas de representação e recuperação utilizadas no contexto dos DGA.

A coleta de dados foi realizada entre os meses de julho e agosto de 2023, com acesso do Portal Capes, e a pesquisa, no segundo semestre de 2023.

Nas próximas seções, os trabalhos selecionados serão discutidos e contextualizados na temática de pesquisa.

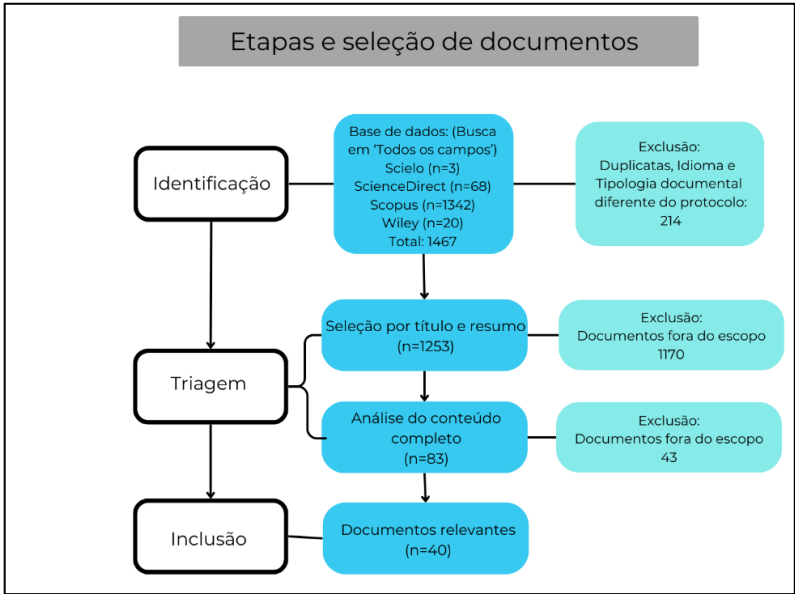
4 RESULTADOS

Conforme destacado no protocolo desta revisão de literatura, inicialmente, a busca considerou todos os campos das bases de dados e, desta forma, recuperou 1467 resultados. Após a aplicação dos critérios de idioma, tipologia documental e a exclusão de documentos duplicados, 1253 foram submetidos aos procedimentos de seleção, análise e categorização. A seleção analisando o título (primeira etapa) descartou 1.014 documentos e selecionou 239. A seleção analisando os resumos (segunda etapa) descartou 156 e selecionou 83 documentos para a análise completa do conteúdo na terceira etapa.

Da amostra de 83 documentos selecionados para leitura completa, foram eliminados 12 por estarem em outro idioma ou pela falta de acesso ao conteúdo na íntegra e 30 documentos foram descartados durante a leitura completa pela inadequação do conteúdo aos objetivos da revisão. Houve dois documentos que se tratava da mesma pesquisa, porém em tipologias documentais diferentes. Para este caso, optou-se por eliminar o artigo de evento e manter o artigo de pesquisa por ser mais completo e atual. Desse modo, foram selecionados 40 documentos que atenderam a todos os critérios e objetivos estabelecidos na revisão para compor os resultados. O resumo da aplicação dos critérios da revisão e as respectivas etapas também estão apresentados na Figura 1.

O Quadro 2 apresenta a identificação, o título, os autores, os anos de publicação, a fonte, o idioma e a tipologia dos 40 documentos selecionados na revisão.

Figura 1. Etapas e seleção de documentos na revisão de literatura



Fonte: elaborada pelas autoras (2024)

Quadro 2. Identificação dos documentos selecionados para análise

ID	Título	Autores	Ano	Fonte	Idioma	Tipologia
D1	<i>A Chatbot for Searching and Exploring Open Data: Implementation and Evaluation in E-Government</i>	Iván Cantador, Jesús Viejo-Tardío, María E. Cortés-Cediel, Manuel Pedro Rodríguez Bolívar	2021	Scopus	Inglês	Artigo de Evento
D2	<i>A Methodology for Retrieving Datasets from Open Government Data Portals Using Information Retrieval and Question and Answering Techniques</i>	Raissa Barcellos, Flavia Bernardini, Jose Viterbo	2022	Scopus	Inglês	Artigo de Evento
D3	<i>A systematic review of open government data initiatives</i>	Judie Attard, Fabrizio Orlandi, Simon Scerri, Soren Auer	2015	Scopus	Inglês	Artigo
D4	<i>Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica</i>	Jacquelin Teresa Camperos Reyes, Ricardo César Gonçalves Sant'Ana	2021	Scopus	Espanhol	Artigo
D5	<i>Allocation of Attention to Metadata and Retrieval Functions: Implications for Perceived Value and Open Data Discovery and Reuse</i>	Ping Wang, Yufei Xie, Yufei Xie, Qiao Li	2023	Web Of Science	Inglês	Artigo
D6	<i>An Assessment of the Quality of Open Government Data in Saudi Arabia</i>	Nada Faisal Alogaiel, Omer Abdulaziz Alrwais	2023	Scopus	Inglês	Artigo
D7	<i>An International Analysis of the Quality of Open Government Data Portals</i>	Alejandro Sáez Martín, Arturo Haro de Rosario, María Del Carmen Caba Pérez	2016	Scopus	Inglês	Artigo
D8	<i>Analysis of standard vocabulary use of the open government data: the case of the public data portal of Korea</i>	Haklae Kim	2019	Scopus	Inglês	Artigo
D9	<i>Base de Normas Jurídicas Brasileiras: uma iniciativa de Open Government Data</i>	Hudson De Martim, João Alberto De Oliveira Lima, Lauro César Araujo	2018	SciELO	Português	Artigo
D10	<i>Characteristic of Open Government Data (Ogd) Around the World: a Country-Based Comparative Meta-Analysis</i>	Rong Tang, Jie Jiang	2021	Science Direct	Inglês	Artigo
D11	<i>Cross-Portal Metadata Alignment – Connecting Open Data Portals Through Means of Formal Concept Analysis</i>	Milos Bogdanovic, Milena Frtunic, Gliglorijevic, Natasa Veljkovic, Darko Puflovic	2023	Science Direct	Inglês	Artigo
D12	<i>Documenting Context-Based Quality Assessment of Controlled Vocabularies</i>	Riccardo Albertoni, Monica de Martino, Alfonso Quarati	2021	Scopus	Inglês	Artigo
D13	<i>Electronic government actions: Searching for semantics in the publication of the Brazilian Social Security System's Open Data</i>	Cláudio José Silva Ribeiro, Reinaldo de Figueiredo Almeida	2012	Scopus	Inglês	Artigo de evento
D14	<i>Enabling Semantic Interoperability in Bhutan's E-Government: An Ontology-based Framework</i>	Youten Tshering, Chutiporn Anutariya	2022	Scopus	Inglês	Artigo de evento
D15	<i>Estudo Comparativo de Datasets Governamentais do Brasil e da Colômbia, com Dados de</i>	Jaqueline Teresa Camperos Reyes, Ricardo Cesar	2018	Web Of Science	Português	Artigo

	Agricultura e Desenvolvimento Rural	Gonçalves Sant'ana, José Eduardo Santarém Segundo				
D16	<i>Framework for open data mining in e-government</i>	Petar Milic, Natasa Veljkovic, Leonid Stoimenov	2012	Scopus	Inglês	Artigo de evento
D17	<i>From Open Data to Data-Intensive Science Through CERIF</i>	Keith G. Jeffery, Anne Asserson, Nikos Houssos, Valerie Brasse, Brigitte Jorg	2014	Science Direct	Inglês	Artigo
D18	<i>Improving the speed and ease of open data use through metadata, interaction mechanisms, and quality indicators</i>	Anneke Zuiderwijk, Marijin Janssen, Iryna Susa	2016	Scopus	Inglês	Artigo
D19	Inferência de Tópicos para Identificação de Subáreas Temáticas de Projetos Culturais	Nádia Felix Felipe Da Silva, Núbia Rosa Da Silva, Kátia Kelvis Cassiano, Douglas Farias Cordeiro	2021	Web Of Science	Português	Artigo
D20	Iniciativas Brasileiras para Dados Governamentais Abertos Conectados: uma Análise do Nível de Abertura dos Dados nas Plataformas Governamentais Brasileiras	Juliana Vasconcelos Braga, Feliz Alberto Ribeiro Gouveia	2022	Web Of Science	Português	Artigo
D21	<i>Investigating Metadata Adoptions for Open Government Data Portals in Us Cities</i>	Fanghui Xiao, Wei Jeng, Daqing He	2018	Wiley	Inglês	Artigo de evento
D22	<i>Linked government data hub, a nontology agnostic data harvester and API</i>	Victoria Kalogirou, Sander van Dooren, Ilias Dimopoulos, Yannis Charalabidis, Jean-Paul De-Baets, Georges Lobo	2020	Scopus	Inglês	Artigo de evento
D23	Metadados para descrição de dataset e recursos informacionais do "portal brasileiro de dados abertos"	Fábio Mosso Moreira, Zaira Regina Zafalon, Plácida Leopoldina Ventura Amorim da Costa Santos, Ricardo César Gonçalves Sant'Ana	2017	Scopus	Português	Artigo
D24	Métrica para Dados Governamentais Abertos	Patrícia Nascimento Silva, Marta Macedo Kerr Pinheiro	2018	Web Of Science	Português	Artigo
D25	Modelo Baseado em <i>Frictionless Data</i> Aplicado aos Dados Abertos Governamentais	Melissa Figueira Fagundes, Divino Ignácio Ribeiro Junior	2020	Web Of Science	Português	Artigo
D26	<i>OGD Metadata Country Portal Publishing Guidelines Compliance: a Multi-Case Study Search for Completeness and Consistency</i>	Barbara Slibar, Enrique Mu	2022	Science Direct	Inglês	Artigo
D27	<i>Open Geodata Reuse: towards Natural Language Interfaces to Web APIs</i>	Auriol Degbelo, Ang Sherpa	2020	Scopus	Inglês	Artigo de evento
D28	<i>Open Government Data: Usage Trends and Metadata Quality</i>	Alfonso Quarati	2017	Web Of Science	Inglês	Artigo

D29	<i>Planning and Designing Open Government Data Programs: an Ecosystem Approach</i>	Sharon S. Dawes, Lyudmila Vidiasova, Olga Parkhimovich	2016	Science Direct	Inglês	Artigo
D30	Qualidade dos Dados na Web: Revisão Integrativa Sobre Diretrizes para Publicação	Morgana Carneiro De Andrade, Maria José Baños Moreno, Juan-Antonio Pastos-Sánchez	2023	SciELO	Português	Artigo
D31	<i>Systematic and ontology-based approach to interoperable cross-domain open government data services</i>	Hadi Masoumi, Bahar Farahani, Fereidoon Shams Aliee	2022	Scopus	Inglês	Artigo
D32	<i>The Process of Open Data Publication and Reuse</i>	Alberto Abella, Marta Ortiz-De-Urbina-Criado, Carmen De-Pablos-Herederó	2019	Web Of Science	Inglês	Artigo de opinião
D33	<i>The U.S. Government manual in XML: A case study of a data.gov open data set</i>	Rebecca Kunkel	2016	Scopus	Inglês	Artigo
D34	<i>To Open or Not to Open? Determinants of Open Government Data</i>	Tung-Mou Yang, Jin Lo, Jing Shiang	2015	Web Of Science	Inglês	Artigo
D35	<i>Towards a tabular open data search engine for public sector information</i>	Alberto Berenguer, Jose-Norberto Mazón, David Tomás	2021	Scopus	Inglês	Artigo de evento
D36	<i>Transforming open data to linked open data using ontologies for information organization in big data environments of the Brazilian government: the Brazilian database government open linked data dbgoldbr</i>	Marcio Victorino, Maristela Terto de Holanda, Edison Ishikawa, Edgard Costa Oliveira, Sammohan Chhetri, George Ghinea***	2018	Scopus	Inglês	Artigo
D37	<i>TWC LOGD: A Portal For Linked Open Government Data Ecosystems</i>	Li Ding, Timothy Lebo, John S. Erickson, Dominic Difranzo, Gregoy Todd Williams, Xian Li, James Michaelis, Alvaro Graves, JunGuang Zheng, Zhenning Shangguan, Johana Flores, Deborah L. McGuinness, James A. Hendler	2011	Web Of Science	Inglês	Artigo
D38	<i>Two Procedures for Analyzing the Reliability of Open Government Data</i>	Davide Ceolin, Luc Moreau, Kieron O'hara, Wan Fokkink, Willem Robert Van Hage, Valetina Maccatrozzo, Alistair Sackley, Guus Schreiber, Nigel Shadbolt	2014	Web Of Science	Inglês	Artigo de evento
D39	Uma proposta de ecossistema de big data para a análise de dados abertos governamentais conectados	Marcio de Carvalho Vcitorino, Marcelo Shiessl, Edgard Costa Oliveira, Maristela Terto de Holanda,	2017	Scopus	Português	Artigo

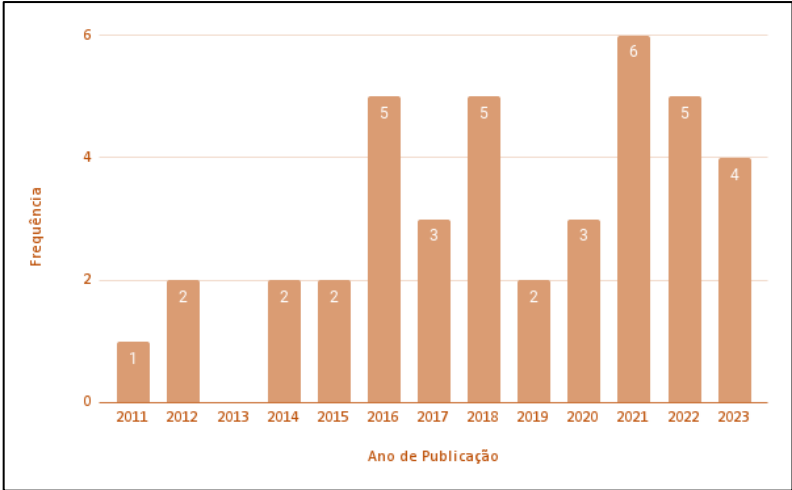
		Marçal de Lima Hokama				
D40	<i>Visual Query Interfaces for Semantic Datasets: an Evaluation Study</i>	Guillermo Vega- Gorgojo, Laura Slaughter, Martin Giese, Simen Heggestoyl, Ahmet Soylu, Arild Waaler	2016	Science Direct	Inglês	Artigo

Fonte: elaborado pelas autoras (2024)

Dos 40 documentos selecionados, um foi escrito em língua espanhola (2,5%), nove em língua portuguesa (22,5%) e os outros 30 (75%) em língua inglesa.

A maioria dos documentos foi classificada como artigo de pesquisa (29 documentos), seguida por trabalhos de evento (10 documentos) e artigo de opinião (um documento). Os resultados estão distribuídos em 31 locais de publicação (base de dados ou indexador). As publicações ocorreram entre os anos de 2011 e 2023, conforme apresentado no Gráfico 1.

Gráfico 1. Distribuição das publicações por ano



Fonte: elaborado pelas autoras (2024)

Na próxima seção, os dados obtidos serão descritos e analisados.

5 ANÁLISE DOS RESULTADOS

A fim de otimizar a análise dos documentos selecionados, após a leitura completa, eles foram categorizados quanto à área principal, relacionada ao objetivo da revisão (representação de informação, recuperação de informação, representação e recuperação) e a temática predominante. Foram encontrados 23 documentos que abordaram a representação (58%), sete que abordaram a recuperação (18%) e 10 que abordaram ambos os temas (25%), conforme apresentado no Quadro 3.

Destaca-se que a categorização foi realizada pelos autores, com base no conhecimento prévio, na experiência na área e com a identificação dos temas apontados pelos próprios autores dos documentos selecionados. Inicialmente, foram analisadas as áreas abordadas no conteúdo

de cada documento que interessam à CI, resultando em um outro agrupamento, mais específico, por temática, a saber: Formatos de Publicação / Tipos / Princípios / Padrões / Metadados; Vocabulários controlados / Ontologias / Dados ligados; Qualidade de Dados; *Frameworks*, Modelos e Métricas; Processamento de Linguagem Natural; Interoperabilidade semântica, similaridade semântica e Mineração de dados / texto, como indicado no Quadro 3.

Quadro 3. Agrupamento temático dos documentos

Temática		Área		
ID	Nome	Representação de Informação	Recuperação de Informação	Representação e Recuperação
T1	Formatos de Publicação / Tipos / Princípios / Padrões / Metadados	D3, D4, D6, D10, D12, D15, D17, D20, D23, D26, D28, D29, D30, D32	D34	D5, D11, D18, D21, D39
T2	Vocabulários controlados / Ontologias / Dados ligados	D8, D12, D13, D14, D20, D31, D33, D37	D40	D5, D22, D36, D39
T3	Qualidade de dados	D6, D8, D12, D18, D24, D26, D28, D30		D7, D21
T4	<i>Frameworks</i> , Modelos e Métricas	D6, D13, D14, D25		D24, D39
T5	Processamento de Linguagem Natural	D9	D1, D2, D27	
T6	Interoperabilidade semântica, similaridade semântica	D14	D35	
T7	Mineração de dados / texto		D16	D19

Fonte: elaborado pelas autoras (2024)

O Quadro 3 foi ordenado conforme a frequência das temáticas, da maior para a menor quantidade, numeradas de T1 a T7. Nos parágrafos a seguir, os documentos e as temáticas presentes são apresentados e discutidos com base na literatura da área. Alguns documentos repetem-se no agrupamento por apresentarem mais de uma temática, no entanto, para essa discussão, cada documento foi apresentado somente uma vez, acrescentando-se, quando necessário, comentários pertinentes a outras temáticas das quais fazem parte. Destaca-se que foi utilizado o identificador do trabalho, conforme definido no Quadro 2 para maior fluidez e otimização da identificação dos trabalhos no texto.

5.1 Formatos de Publicação / Tipos / Princípios / Padrões / Metadados

Esta categoria incluiu documentos que tratam da representação descritiva de dados e da utilização de metadados, considerados por Alves (2010) como elementos fundamentais no processo de tratamento descritivo da informação. As temáticas mais abordadas nos documentos recuperados se relacionam com “Formatos de Publicação / Tipos / Princípios / Padrões / Metadados”, identificadas no Quadro 4 como T1, presentes em 20 trabalhos, sendo eles: **D3, D4, D5, D6, D10, D11, D12, D15, D17, D18, D20, D21, D23, D26, D28, D29, D30, D32, D34 e D39.**

O estudo **D3** trouxe uma revisão sistemática de literatura sobre iniciativas de DGA e aborda brevemente a temática da representação dos DGA, formatos de publicação e recuperação de informações relevantes. Explica-se sobre como uma representação ideal de um conjunto de dados acompanharia a sua evolução ao longo do tempo e fala-se também sobre como a heterogeneidade dos formatos de dados e a heterogeneidade de sua representação são um revés para as iniciativas de dados abertos.

D4 descreveu a adesão ao uso de metadados em conjuntos dos DGA no contexto Ibero-americano, por meio de uma observação crítica dos documentos que abordaram o tratamento descritivo de conjuntos dos DGA. Os autores destacaram a importância de uma descrição ampla e correta dos metadados para aproximar os usuários dos modelos de dados.

O artigo **D5** trouxe contribuições relacionadas a dados abertos que podem ser aplicadas a contextos mais específicos como DGA. Os autores consideraram que as funções de recuperação de dados, como a recuperação por palavra-chave, classificação, filtragem e análise de resultados de pesquisa, são essenciais para esses sistemas. Para os autores, os metadados desempenham um papel essencial na representação de dados.

No estudo **D6**, foi realizada uma análise sobre a qualidade dos DGA da Arábia Saudita e, nos pontos analisados, exploraram-se os metadados e a representação desses dados, mencionando a importância de apresentá-los de forma clara e compreensível para um público amplo, envolvendo a escolha de uma representação adequada.

Em **D10**, comparou-se, a partir de uma meta-análise de 50 publicações, os tipos de formatos de dados, o número de conjuntos de dados e os tipos de categorias de dados de 67 países. O trabalho trouxe contribuições ao apresentar um panorama mundial dos DGA e criar subsídios para a criação de uma diretriz, ou recomendação, para a construção de um site ou portal dos DGA em que os formatos e as categorias dos dados mapeados seriam considerados um requisito mínimo.

O documento **D11** tratou da problemática de conexão de portais de dados abertos, por meio do alinhamento de metadados. O objetivo não se relaciona especificamente com DGA, mas os princípios poderiam ser aplicados a eles com a melhora da pesquisa entre portais e o enriquecimento de metadados por meio da categorização de dados abertos.

Em **D12**, os autores definiram diretrizes que promovem e implementam a adoção das recomendações do consórcio W3C para fornecer metadados de qualidade legíveis por máquina, aumentando a confiabilidade e a reutilização.

No estudo **D15**, foi apresentada a análise de *datasets* governamentais do Brasil e da Colômbia, especificamente na área da agricultura, levando em consideração as recomendações dos Oito Princípios de Sebastopol (*Open Government Data*, 2007) para realizar um estudo comparativo. Neste estudo, a representação de dados estava relacionada ao formato de publicação, em formatos estruturados processáveis por máquina e não proprietários.

O estudo **D17** considerou que os metadados são a chave para gerenciar dados abertos, sendo eles necessários para descoberta, para contexto (compreensão do conjunto de dados e avaliação de qualidade e relevância) e para utilização. Os autores propõem a utilização do *Common European Research Information Format* (CERIF) para integrar DGA e conjuntos de dados de pesquisas com financiamento público.

No trabalho **D18**, foi utilizada uma abordagem de *design science research* para avaliar se metadados, mecanismos de interação e indicadores de qualidade podem melhorar o uso dos

DGA. Foi identificado que a manipulação de dados pode ser aprimorada por meio de um modelo de metadados abrangente e de funcionalidades relacionadas, como classificação e filtragem, pesquisa por palavras-chave, suporte a idiomas, visão geral de metadados, visualização de conjuntos de dados *online* e ferramentas de visualização (em tabelas, gráficos e mapas), ajudando a atingir esse objetivo e tornando a tarefa de localizar, entender, analisar, visualizar e avaliar dados menos trabalhosa e demorada.

Em **D20**, portais de transparência brasileiros foram avaliados quanto à abertura dos seus dados e foi identificado o grau de maturidade em que se encontram, considerando o esquema de classificação cinco estrelas para dados abertos e os 14 princípios para os DGA de Tauberer. Os resultados apontam que a escolha do formato de dados tem amplas implicações sobre os aplicativos que podem ser construídos sobre os dados e questiona se os dados arquivados provavelmente serão utilizáveis no futuro.

No artigo **D21**, foi apresentada uma investigação sobre os portais de DGA em nível municipal nos EUA, com foco na adoção de plataformas, qualidade dos metadados e problemas relacionados à representação e à recuperação dos DGA. Este trabalho relacionou a representação à atribuição de metadados, destacando os atributos fornecidos em três plataformas comumente usadas: Socrata, *Comprehensive Knowledge Archive Network* (CKAN) e ArcGIS, como título, categorias, *tags* e informações geográficas. O estudo destacou a importância da apresentação e da qualidade dos metadados nos portais para a recuperação eficiente de informações, analisando a presença de elementos específicos nos metadados, como formato de arquivo, tamanho do arquivo, licença e propriedade dos dados.

No trabalho **D23**, acredita-se que a descrição eficiente dos recursos informacionais e a representação adequada dos conjuntos de dados podem amenizar problemas como o acesso parcial, superficial e de difícil compreensão dos DGA, contribuindo para a obtenção de resultados relevantes no processo de recuperação, transparência, *accountability* e interoperabilidade. Os autores questionaram quais metadados os serviços de disponibilização dos DGA estão utilizando para representar seus recursos informacionais. Foram estudados os *datasets* disponibilizados pelo Portal Brasileiro de Dados abertos, e os elementos de metadados utilizados na representação dos *datasets* foram relacionados aos elementos de metadados do Padrão Metadados do Governo Eletrônico (e-PMG) 1.1.

O trabalho **D26** centrou-se principalmente na qualidade dos metadados e na conformidade com as diretrizes de publicação de metadados para DGA, fornecidas pelos próprios portais, envolvendo quais campos publicar (completude) e como registrar seus valores (consistência), de três líderes nacionais em DGA (Canadá, EUA e Nova Zelândia), conforme denominado pelos autores. Em consonância com outros estudos, eles apontam que a qualidade dos metadados é fundamental para o uso e o reúso efetivos dos DGA e que metadados precisos e completos ajudam a garantir que os dados possam ser encontrados, compreendidos e usados corretamente.

No artigo **D28**, foi apresentado um estudo sobre DGA alinhado às temáticas da CI. O autor enfatiza que a qualidade dos metadados dos conjuntos dos DGA é um dos principais fatores que afetam seu uso. Ao fazer isso, o artigo chama a atenção para a necessidade de fornecer metadados precisos, completos e bem estruturados para melhorar a usabilidade e a descoberta dos DGA.

Utiliza-se também os princípios FAIR (*findable, accessible, interoperable and reusable*) para avaliação de portais, em consonância com outros estudos analisados. O estudo cita plataformas dedicadas, como CKAN e Socrata, e que a exploração delas nos portais de DGA ajuda a entender as suas características e funcionalidades e a sua influência na usabilidade dos dados.

O trabalho **D29** discutiu o planejamento e o *design* de programas de DGA, incluindo fatores que influenciam a publicação de dados, estabelecendo uma relação com a representação de informação. Eles analisam programas dos DGA em Nova York (EUA) e São Petersburgo (Rússia), identificando as características dos ecossistemas de dados abertos em cada cidade e as práticas que contribuíram para o sucesso desses programas. Os autores aplicam critérios para avaliação de usabilidade de conjuntos de dados como formato dos dados, metadados e meios de acesso, temáticas relacionadas à ciência da informação.

Em **D30**, a representação dos DGA foi relacionada à qualidade dos dados e à satisfação das necessidades dos usuários de sistemas de informação. Para os autores, a má qualidade dos dados está relacionada ao resultado de uma má representação do mundo real.

O trabalho **D32** discutiu o processo de publicação de dados abertos desde os sistemas internos das organizações até o reuso desses dados pelos usuários. Foi explicado que a etapa de publicação de dados abertos geralmente implica coleta de metadados e sua junção aos dados brutos, incluindo informações como autoria, frequência de atualização, tamanho, descrição do modelo de dados, estrutura, ajuda, contato e licença, entre outros.

O artigo **D34** mencionou a importância da recuperação dos DGA por meio de tecnologias e metodologias específicas, como o *framework Reveal, Refine, Reuse, Release e Run (R5)* e o uso de metadados como um habilitador-chave para facilitar o uso dos DGA. Os autores também discutiram os efeitos negativos da divulgação dos DGA e a importância da interoperabilidade entre políticas, gestão e tecnologia para o sucesso das iniciativas de dados abertos.

O trabalho **D39** apresentou a especificação de um ecossistema de *big data* para dar suporte à produção e ao consumo dos DGA de qualidade no âmbito do governo brasileiro.

Observa-se que muitos autores chegaram a conclusões similares sobre os desafios representados pela quantidade e heterogeneidade dos dados disponíveis e como a descrição e a representação adequadas representam uma solução para amenizar essas questões. No contexto dos DGA, os metadados foram considerados essenciais para a representação em ambientes digitais e foram relacionados à qualidade. Com isso, uma boa definição desses metadados implicará diretamente na melhor caracterização dos DGA.

5.2 Vocabulários controlados / Ontologias / Dados ligados

Esta categoria incluiu documentos que abordam linguagens documentárias verbais, que “visam o tratamento e a recuperação da informação pela representação do assunto” (Almeida, 2011), como os vocabulários controlados e as ontologias, assim como os dados ligados (*linked data*), que são dados semânticos conectados por meio de tecnologia.

Os trabalhos agrupados na temática Vocabulários controlados / Ontologias / Dados ligados (T2) foram: **D5, D8, D12, D13, D14, D22, D31, D33, D36, D37, D39 e D40**. O objetivo de um vocabulário controlado é “consistência na representação e na recuperação de modo a evitar a dispersão de recursos informacionais relacionados aos termos do vocabulário controlado” (Fujita; Tolare, 2019), e as ontologias são um tipo específico de vocabulário controlado e ferramentas amplamente utilizadas para tratar de dados ligados.

No artigo **D5**, os autores sugeriram tornar o resumo dos dados mais compreensível com a utilização de formatos abstratos estruturados, empregando um vocabulário controlado, que tem a função de padronizar os termos que irão compor o repertório terminológico utilizado para a representação e a recuperação de informação (Davano; Moreira, 2014) para representar o tópico dos metadados e ajudar os usuários a pesquisarem conteúdos específicos.

D8 trouxe uma pesquisa acerca da representação de informação, realizando uma análise do uso de um vocabulário padrão dos DGA, com foco específico em portais da Coreia. Destacou-se a importância da representação adequada dos dados para garantir sua qualidade e relevância e foram apresentadas sugestões para atingir esse objetivo, como o uso de um vocabulário padrão e a integração de termos dos padrões administrativos e de dados públicos. A utilização desse tipo de vocabulário foi considerada eficaz para conectar diferentes conjuntos de dados e para melhorar a sua qualidade.

Em consonância com **D8**, **D12** abordou uma temática que pode trazer grandes contribuições à representação de informação no contexto dos DGA, que é a análise de qualidade de vocabulários controlados dos DGA. A adoção de um vocabulário para o domínio dos DGA é importante para a padronização terminológica, possibilitando a ligação dos dados.

O trabalho **D13** trouxe preocupações sobre a semântica na publicação dos DGA e discorreu sobre ontologias e seu auxílio na interoperabilidade entre dados, propondo um *framework* para facilitar essa integração. Os autores utilizaram a abordagem da filosofia analítica para tratar da representação, afastando-se ligeiramente da temática da CI.

O documento **D14** considerou a representação de sistemas heterogêneos por meio de padrões e vocabulários comuns como uma solução para a interoperabilidade e propôs um *framework* baseado em ontologia para viabilizar a interoperabilidade semântica no governo eletrônico, com foco no Butão.

Em **D20**, foram estudadas iniciativas brasileiras para os DGA conectados a partir de uma análise da abertura de dados em plataformas governamentais. Os autores utilizaram o esquema de classificação de Tim Berners-Lee (2006), que é uma referência para avaliar dados ligados.

D22 apresentou um modelo de *hub* dos DGA vinculados com utilização de vocabulários controlados que agrega e categoriza dados de vários serviços para melhorar a transparência e a acessibilidade dos DGA. Para se tornar um *hub* de dados vinculados utilizando uma ontologia agnóstica que coleta, traduz e formata dados em uma ontologia interoperável, que pode ser apresentada por meio de um site ou de uma *Application Programming Interface* (API), foi proposto um site único e integrado para as administrações públicas, funcionando como um portal único. Segundo Farinelli, Almeida e Melo (2013), o problema da heterogeneidade semântica é elegível para ser tratado a partir de ontologias, como apresentado pelos autores de **D22**.

Em **D31**, foi mencionada a representação, que é uma das etapas do processo proposto para alcançar a interoperabilidade dos DGA. Essa representação foi feita por meio de ontologias, que permitiram a descrição semântica dos dados e a criação de um vocabulário comum para diferentes domínios, facilitando a recuperação e o uso dos dados por parte de diferentes usuários e sistemas. O objetivo dos autores foi propor uma abordagem sistemática para a criação de serviços dos DGA interoperáveis. Os resultados apresentados mostram que a abordagem proposta pode melhorar a qualidade e a relevância dos dados publicados, reduzir custos e aumentar a eficiência dos processos de publicação e, para isso, eles fornecem instruções, passo a passo, para que as organizações governamentais publiquem seus dados.

O trabalho **D33** abordou a representação de dados em XML e como isso permitiu a criação de uma ontologia para o Manual do Governo dos Estados Unidos; ontologia essa criada para representar entidades como agências governamentais, comitês, legisladores e seus cargos. Essa representação possibilitou que essas entidades fossem mapeadas de forma mais precisa na ontologia, facilitando a recuperação e o uso desses dados.

Em **D36**, os autores discutiram sobre a utilização de ontologias para organização de informações em ambientes de *big data* do governo brasileiro, com o objetivo de transformar dados abertos em dados abertos vinculados. A representação foi feita por meio da criação de ontologias que permitiram a integração de dados de diferentes fontes e a criação de uma base de conhecimento unificada, permitindo, assim, que as informações estivessem facilmente acessíveis e utilizáveis por diferentes usuários e sistemas, por meio da representação de dados em formato de dados abertos vinculados. Os autores apresentam um modelo conceitual, uma arquitetura técnica e um protótipo de implementação da ferramenta denominada DBgoldbr, projetada para classificar informações públicas governamentais.

No trabalho **D37**, há uma proposta de um portal para tratar DGA vinculados, assim como no artigo D36, que se preocupou com a integração de dados para facilitar o acesso e a utilização. No contexto deste último artigo, a representação foi utilizada para apoiar a produção e o consumo dos DGA vinculados. Segundo Tim Berners-Lee (2006), os dados vinculados são uma forma de se publicar e interligar dados estruturados na *Web*. Houve também o desenvolvimento de um modelo de organização de dados com ferramentas para permitir uma infraestrutura de produção dos DGA vinculados de forma rápida, persistente e extensível.

O trabalho **D39** abordou um assunto semelhante ao **D36** apresentando a especificação, em um alto nível de abstração, de um ecossistema de *big data* para dar suporte à produção e ao consumo dos DGA de qualidade no âmbito do governo brasileiro. O objetivo foi possibilitar o armazenamento dos dados oriundos de diversas origens para serem tratados e, posteriormente, utilizados como subsídio para a avaliação e o monitoramento de programas sociais.

Em **D40**, foi estudada a usabilidade de interfaces de consulta baseadas em ontologias e seu impacto na recuperação de dados em portais de DGA. Para isso, foram realizados testes de usuários com dois tipos de interfaces: uma baseada em gráficos e outra baseada em formulários.

Os autores demonstraram-se alinhados com práticas da CI, utilizando-se metadados, tesouros, taxonomias e principalmente as ontologias para auxiliar na contextualização e indexação das informações do ecossistema. Foram implementadas também camadas de apresentação com interfaces que seguiram princípios de usabilidade para facilitar a recuperação de informação.

5.3 Qualidade de Dados

Esta categoria incluiu documentos que se propõem a avaliar alguma dimensão de qualidade dos DGA. Os documentos relacionados à Qualidade de Dados, agrupados em T3, foram: **D6, D7, D8, D12, D17, D18, D21, D24, D26, D28, D30**.

No trabalho **D6**, os autores destacaram a qualidade dos DGA da Arábia Saudita, e em **D7**, propuseram-se a avaliar a qualidade de portais de DGA, a nível internacional, com uma abordagem similar à D6, ao analisar o conteúdo. Dois aspectos analisados que se relacionam com a representação de informação foram a semântica e o conteúdo.

Os autores de **D7** afirmaram que a qualidade dos portais dos DGA é um fator que pode afetar a capacidade dos usuários de localizar os dados de interesse, estabelecendo, assim, uma relação entre qualidade e recuperação de informação. Eles trouxeram também uma perspectiva interessante ao se preocuparem com a recuperação de informação para dois tipos de usuários: comum e avançado (programadores) e concluíram que para um portal dos DGA desenvolver seu potencial máximo, com maior nível de qualidade possível, os dados públicos devem ser fornecidos em um formulário padrão, possibilitando sua reutilização futuramente.

D8, em consonância com os documentos número **D6** e **D7**, destacou a importância da representação adequada dos dados para garantir sua qualidade e relevância e trouxe sugestões para atingir esse objetivo.

Os documentos **D12, D17, D18** e **D21** foram discutidos nas seções anteriores e destacaram a abordagem utilizada para tratar de qualidade de dados. **D24** compilou aspectos relevantes para mensurar a qualidade dos dados publicados e, conseqüentemente, seu reuso.

Nesta seção, os trabalhos apresentaram soluções que envolveram várias técnicas da CI para alcançar a qualidade dos DGA, mas é importante destacar que essas técnicas também foram combinadas com questões tecnológicas e requisitos de sistemas.

| 19

5.4 Frameworks, Modelos e Métricas

Esta categoria incluiu documentos que propõem aplicações práticas ou conceituais em forma de modelos, métricas e ou *frameworks* para aplicação em DGA. Os documentos que abordam *Frameworks*, Modelos e Métricas, agrupados em T4, foram: **D6, D13, D14, D24, D25** e **D39**.

Em **D6**, propôs-se uma estrutura para avaliar a qualidade dos DGA usando métricas compatíveis com o Portal de Dados Abertos da Arábia Saudita. Os trabalhos **D13** e **D14** propuseram *frameworks* baseados em ontologias para facilitar a integração entre dados.

D24 propôs uma métrica para avaliação do potencial de reuso dos DGA fundamentada em padrões e princípios da publicação dos DGA combinada com elementos da representação e recuperação dos dados publicados. A métrica incorporou modelos, princípios e boas práticas de publicação como: os oito princípios de Sebastopol e as cinco estrelas de Tim Berners-Lee, que definem critérios de avaliação que se relacionam com a representação e a recuperação de informação.

Em **D25**, foi divulgada uma pesquisa inédita com a apresentação de um modelo promissor para a publicação dos DGA: a iniciativa chamada *Frictionless Data*, criada pela *Open Knowledge Foundation*, que foi apresentada como uma alternativa para remover o “atrito”, ou seja, a dificuldade no trabalho com os dados, com foco na simplicidade e em uma estrutura básica que permita a qualquer pessoa publicar e usar conjuntos de dados com mais facilidade, sem passar por um intermediário.

O trabalho **D25** apresentou uma contribuição relevante com a aplicação inédita dos princípios FAIR, que, segundo Henning *et al* (2019), visam implementar um conjunto de metadados definidos, para uso por mecanismos computacionais automatizado ou por pessoas, aos DGA por meio de um modelo baseado em *Frictionless Data*. Estepode contribuir com o reuso e o compartilhamento dos dados, permitindo a construção de modelos semelhantes para o reconhecimento de padrões dos DGA, além de alertar para a necessidade de um *framework* ou modelo de disponibilização de dados dos governos, com o objetivo de trazer maior unicidade e respeito a determinados critérios internacionalmente aceitos, em consonância com as conclusões de Silva (2018).

Em **D39**, os autores apresentaram a especificação de um modelo de ecossistema de *big data* para dar suporte à produção e ao consumo dos DGA de qualidade no âmbito do governo brasileiro.

5.5 Processamento de Linguagem Natural

Esta categoria incluiu documentos que abordaram a recuperação de informação com tecnologias de processamento de linguagem natural, que é uma abordagem que realiza a “análise de textos em um ou mais níveis linguísticos, com o propósito de simular o processamento humano da língua” (Ferneda, 2012). O **Processamento de Linguagem Natural**, agrupado no Quadro 4 como **T5**, foi identificado nos trabalhos **D1**, **D2**, **D9** e **D27**.

Em **D1**, foi apresentado um *chatbot* inovador que explorou o conteúdo de coleções dos DGA, por meio de consultas realizadas por um usuário não especializado. Em **D2**, os autores também apresentaram técnicas de perguntas e respostas e o aprendizado de máquina para recuperação. Métricas de precisão e revocação foram utilizadas para mensurar a eficácia da metodologia proposta.

O estudo **D27** combinou o processamento de linguagem natural com interfaces de *APIsweb* para recuperar dados geoespaciais. Foi feito um recorte específico, no entanto, essa abordagem pode ser replicada em outros contextos, como a utilização da consulta em linguagem natural em APIs para recuperação dos DGA. As abordagens utilizadas nos estudos estão em consonância com temáticas estudadas na CI e Ciência da Computação, fazendo uma intersecção entre elas.

O trabalho **D9** também utilizou o processamento de linguagem natural, porém com foco na construção de *datasets* para representação da informação. Ainda que o trabalho tenha foco em uma área específica, o modelo de organização da informação para publicação de normas jurídicas pode ser replicado em ferramentas de áreas distintas, no intuito de contribuir com cientistas da informação e pesquisadores em geral, por meio de *datasets* que podem ser entendidos como uma plataforma de dados para trabalhos futuros.

5.6 Interoperabilidade semântica / Similaridade semântica

A interoperabilidade refere-se à capacidade de diversos sistemas e organizações trabalharem em conjunto (interoperar), garantindo que pessoas, organizações e sistemas computacionais interajam para trocar informações de maneira eficaz e eficiente (Brasil, 2023).

Esta categoria incluiu documentos que tratam da interoperabilidade semântica, definida por Rio Branco (2020) como a capacidade de sistemas de informação de trocar, usar e reusar informação legível por máquinas.

A interoperabilidade semântica, similaridade semântica (T6), foi identificada em dois trabalhos: **D14** e **D35**. Ainda que o artigo **D14** foque sua abordagem na representação e o **D35** na recuperação, ambos apresentam propostas baseadas em interoperabilidade semântica. Além disso, eles as relacionam a questões de elaboração e gerenciamento de linguagens de indexação, com o propósito de utilizá-las para indexação e recuperação de informação de maneira integrada em contextos diversos (Rio Branco, 2020).

O artigo **D35** tratou da recuperação de dados tabulares abertos com a incorporação de palavras para calcular a similaridade semântica entre os dados, considerando metadados no cenário de informações do setor público. O estudo apresentou também uma nova interface de pesquisa que recupera dados tabulares considerando o conteúdo da tabela e não apenas os metadados que a representam. Essa interface preencheu a lacuna da dependência da qualidade dos metadados para representar o conteúdo tabular de dados abertos para realizar buscas.

Importante destacar que a interoperabilidade semântica está intrinsecamente relacionada às temáticas agrupadas em T2, sobre controle de vocabulário, sendo um possível caminho a ser seguido.

5.7 Mineração de dados / texto

Esta categoria inclui documentos que tratam da técnica de utilização de algoritmos e sistemas para explorar, processar e analisar grandes quantidades de dados.

A temática Mineração de dados / texto (T7) foi identificada nos trabalhos **D16** e **D19**. O primeiro propôs um *framework* para mineração dos DGA e o segundo abordou uma técnica conhecida como *Knowledge Discovery from Textual Databases (KDT)*.

Em **D16**, foi justificada a aplicação da mineração de dados na obtenção de conhecimento e informações úteis que são acessíveis e compreensíveis e ajudam os cidadãos e outras partes interessadas a processarem dados abertos, de acordo com suas necessidades. Este princípio, segundo Amo (2004), está alinhado com o propósito da recuperação de informação, já que a mineração de dados é uma área de pesquisa multidisciplinar e abrangente.

Em **D19**, foi explicitado como a técnica KDT, que aborda como textos representados nos mais diversos formatos (e-mails, arquivos em formato pdf, conteúdo de páginas *web*, textos eletrônicos digitalizados) podem ter seu conteúdo extraído por meio de análise semântica. Os autores mencionaram que os avanços em ciência da computação e áreas afins desencadearam um aumento exponencial na capacidade de manipulação, recuperação e utilização dos dados sob diversas perspectivas, o que está diretamente relacionado à representação e à transformação de dados em informação.

Para Setzer (1999), dado é uma sequência de símbolos quantificados ou quantificáveis e informação é uma abstração informal, que está na mente de alguém, representando algo significativo para essa pessoa. Ainda segundo esse autor, os dados, desde que inteligíveis, são sempre incorporados por alguém como informação. A contribuição do estudo D19 é significativa por oferecer um exemplo prático para a representação de documentos a partir de descrições que mantêm relações estatísticas para classificação, sumarização e recuperação de informações, aplicadas aos DGA.

Após a apresentação das sete áreas e considerando uma análise geral, todos os trabalhos avaliados abordaram, em algum aspecto, a representação de DGA ou, em alguns casos, de dados abertos com abordagens que poderiam ser replicadas ao contexto dos DGA. Também foi possível identificar que a representação esteve atrelada a uma melhora na utilização e reutilização dos DGA.

A recuperação dos DGA apareceu nos documentos de forma mais sutil, sendo citada explicitamente em poucos estudos, sem um foco principal, com a apresentação de alguns métodos e técnicas de recuperação de informação no contexto dos DGA. Essa é uma área promissora, já que, como apontado pela *Open Knowledge Foundation* ([s.d]), os DGA são um tremendo recurso e amplamente inexplorados, tendo em vista que o governo é uma das organizações que mais coletam dados, que geralmente são públicos devido às suas propriedades.

Um exemplo de aplicação nessa área é a implementação das tecnologias de Inteligência Artificial (IA) para recuperação dos DGA, como apresentado por Oliveira, Nascimento Silva e Mafra Pereira (2024), que analisam técnicas de IA para aprimorar a experiência do usuário na recuperação dos DGA no Portal Brasileiro de Dados Abertos.

Princípios e práticas sobre dados abertos foram observados nos trabalhos, com destaque para as cinco estrelas de Tim Berners-Lee, os 8 princípios de Sebastopol e os princípios FAIR, fundamentos basilares sobre a temática que devem ser observados.

O agrupamento temático proposto para este estudo (T1 a T7) permitiu identificar tópicos que são objetos de estudo da CI e em interseção com a Ciência da Computação. No entanto, apesar da segmentação, alguns temas estão diretamente relacionados, como vocabulários controlados, dados ligados e interoperabilidade. Esta relação sobre as temáticas permitiu observar uma certa tendência de utilização dessas linguagens artificiais para estabelecer um padrão de representação, a fim de promover a integração de dados abertos.

6 CONCLUSÃO

Com a execução da revisão de literatura, foi possível mapear os estudos na CI que tratavam da representação e recuperação dos DGA. Observou-se que a maior parte das publicações analisadas estava em inglês, destacando a relevância e o alcance global do tema abordado. Além disso, constatou-se que todas as publicações selecionadas são recentes, datando dos últimos 12 anos, o que ressalta a atualidade e a importância contínua do assunto em estudo.

O agrupamento realizado para análise dos resultados, mais especificamente por temática, permitiu identificar as principais perspectivas abordadas e interseções. A representação dos DGA emergiu como um elemento-chave na garantia da qualidade desses dados e na satisfação das necessidades dos usuários de sistemas de informação. A temática T1, que envolveu questões sobre ‘Formatos de Publicação / Tipos / Princípios / Padrões / Metadados’, foi a mais frequente,

sendo observada em 20 trabalhos, destacando elementos da representação descritiva e a padronização de metadados.

Diante das abordagens variadas sobre DGA encontradas na literatura, nota-se que poucos estudos focam especificamente nos aspectos de recuperação da informação, ainda que muitos mencionem algum aspecto ou característica dos DGA que estão relacionados ao tema. Foi observado que geralmente a representação está relacionada à publicação, à recuperação e ao acesso, sendo todos eles essenciais para sua reutilização e tornando necessário o aprimoramento dos mecanismos de busca e recuperação, visando facilitar a utilização e a integração desses dados em diferentes contextos.

A representação dos DGA mostra-se como um fator decisivo na interconexão e na interoperabilidade dos dados em um contexto mais amplo, e os vocabulários controlados são instrumentos muito utilizados para esse fim. É possível que estudos futuros aprimorem modelos, padrões ou metodologias específicas que já vêm sendo estudadas para a representação e a recuperação dos DGA, facilitando sua recuperação, seu reuso e sua integração em diversos contextos e aplicações.

Observa-se que os temas analisados apresentam um alinhamento em torno dos DGA. Inicialmente, as discussões concentram-se nas preocupações com a representação e disponibilização dos DGA, destacando como uma boa prática o uso de vocabulários controlados para isso. Há também o desenvolvimento de modelos e *frameworks* a serem aplicados em contextos diversos. Em seguida, abordam discretamente algumas metodologias para a recuperação de dados e, por fim, discutem o uso de métricas para avaliar a qualidade dos dados e a implementação das diferentes fases do ciclo de vida dos DGA.

No entanto, há uma lacuna nas abordagens voltadas ao cidadão; muitos dos estudos enfatizam aspectos técnicos, mas carecem de perspectivas do usuário consumidor, especialmente do usuário não especialista. Estudos futuros poderiam investigar as necessidades desses usuários para avaliar medidas que aprimorem a representação e recuperação dos DGA, considerando melhor as demandas do público, a fim de promover um impacto social.

Em suma, a revisão de literatura ressaltou a importância contínua dos DGA na esfera da CI e de áreas correlatas, proporcionando uma base para futuras pesquisas e aprimoramentos na representação, recuperação e utilização dos DGA, impulsionando, assim, avanços na transparência governamental, na inovação cívica e na ciência cidadã, aplicados a novos modelos de negócio que têm os dados e as informações como principais insumos.

REFERÊNCIAS

AGUNE, R. M.; GREGORIO FILHO, A. S.; BOLLIGER, S. P. Governo aberto SP: disponibilização de bases de dados e informações em formato aberto. In: **CONGRESSO CONSAD DE GESTÃO PÚBLICA**, 3., 2010. Anais eletrônicos... Brasília, 2010. Disponível em: <https://bit.ly/42EO6As>. Acesso em: 10 fev. 2024.

ALMEIDA, T. O vocabulário controlado como instrumento de organização e representação da informação na FINEP. 2011. 151 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal do Rio de Janeiro / Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2011. Disponível em: <http://ridi.ibict.br/handle/123456789/740>. Acesso em: 10 fev. 2024.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2010. Disponível em: <https://encr.pw/ni3Jk.2010> . Acesso em: 10 fev. 2024.

AMO, S. Técnicas de mineração de dados. *In*: Sociedade Brasileira de Computação, Universidade Federal da Bahia. (org.). **Jornadas de atualização em informática**. Salvador: Universidade Federal da Bahia, 2004, v. 2, p. 195-236.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de informação**: conceitos e tecnologia das máquinas de busca. Bookman Editora: Santana, 2013.

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2016.

BERNERS-LEE. **Linked data-design issues**. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 03 nov. 2024.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? *In*: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 9., 2008. São Paulo. Anais...São Paulo: USP, 2008. Disponível em: <https://brapci.inf.br/index.php/res/v/176535>. Acesso em: 10 fev. 2024.

BRASIL. **Governança de Dados: Interoperabilidade**. 2023. Disponível em: <https://www.gov.br/governodigital/pt-br/governanca-de-dados/interoperabilidade>. Acesso em: 15 fev. 2024.

BERTIN, P. R. B. *et al.* A parceria para Governo Aberto como plataforma para o avanço da Ciência Aberta no Brasil. **Transinformação**, Campinas, v. 31, p. e190020, 2019. Disponível em: <https://www.scielo.br/j/tinf/a/hZGKbLBrvv3KjBFNnZ65qDN/>. Acesso em: 10 fev. 2024.

CAFÉ, L.; SALES, R. Organização da informação: Conceitos básicos e breve fundamentação teórica. *In*: Jaime Robredo; Marisa Bräscher (Orgs.). **Passeios no Bosque da Informação**: Estudos sobre Representação e Organização da Informação e do Conhecimento – EROIC. Brasília DF: IBICT, 2010, p. 115-129. Edição eletrônica. Disponível em: <http://www.ibict.br/publicacoes/eroic.pdf>. Acesso em: 10 fev. 2024.

CENTRO DE ESTUDOS SOBRE TECNOLOGIA WEB (CEWEB). NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR. **Guia de incentivo ao re(uso) de dados abertos**. 2016. Disponível em: <https://ceweb.br/guias/incentivo-ao-reuso-de-dados-abertos/>. Acesso em: 10 fev. 2024.

DAVANZO, L.; MOREIRA, W. **O vocabulário controlado como ferramenta do processo de organização e recuperação da informação**. 2014. Disponível em: <https://brapci.inf.br/index.php/res/v/186303>. Acesso em: 10 fev. 2024.

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

FARINELLI, F.; ALMEIDA, M. B.; MELO, S. **O papel das ontologias na interoperabilidade de sistemas de informação**: reflexões na esfera governamental. 2013. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/185252> . Acesso em: 10 fev. 2024.

FERNEDA, E. **Recuperação de informação**: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 137 f. 2003. Tese de Doutorado. Tese (Doutorado em Ciência da Informação e Documentação) – Universidade de São Paulo. São Paulo.

FERNEDA, E. **Introdução aos modelos computacionais de recuperação de informação**. Rio de Janeiro: Editora Ciência Moderna Ltda, 2012.

FINK, A. **Conducting research literature reviews: From the internet to paper**. Los Angeles: Sage publications, 2005.

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. Acesso em: 10 fev. 2024.

GALVÃO, T. F.; PEREIRA, M. G. Revisões sistemáticas da literatura: passos para sua elaboração. **Epidemiologia e Serviços de Saúde**, Brasília, v. 23, n. 1, p. 183-184, mar. 2014. Disponível em: <https://acesse.one/PUUpQg>. Acesso em: 22 nov. 2023.

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

HENNING, P. C. *et al.* GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 389–412, 2019. DOI: 10.19132/1808-5245252.389-412. Acesso em: 10 fev. 2024.

KOLLER, S. H.; COUTO, M. C. P. P.; VON HOHENDORFF, J. **Manual de produção científica**. Porto Alegre: Penso Editora, 2014.

LIMA, J. L. O.; ALVARES, L. Organização e representação da informação e do conhecimento. In: *et al.* (org.). **Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações**. São Paulo: B4, 2012. p. 21-48.

MARCOS-MARTÍN, C.; SORIANO-MALDONADO, S.-L. Reutilización de la información del sector público y open data en el contexto español y europeo. Proyecto Aporta. **Profesional de la información**, [Madrid], v. 20, n. 3, p. 291–297, 2011. DOI: 10.3145/epi.2011.may.07. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2011.may.07>. Acesso em: 10 fev. 2024.

MOOERS, C. N. Zato coding applied to mechanical organization of knowledge. **American documentation**, Hoboken, NJ, v. 2, n. 1, p. 20-32, 1951.

NASCIMENTO SILVA, P. Dados governamentais abertos sobre a COVID-19 no Brasil: avaliação do potencial de reúso. **Em Questão**, Porto Alegre, v. 29, p. 124820, 2023. DOI: [10.19132/1808-5245.29.124820](https://doi.org/10.19132/1808-5245.29.124820).

OKOLI, C.; DUARTE, D. W. A.; MATTAR, J. Guia para realizar uma revisão sistemática de literatura. **EaD em Foco**, [S. l.], v. 9, n. 1, 2019. DOI: 10.18264/eadf.v9i1.748. Acesso em: 22 nov. 2023.

OLIVEIRA, D. ; NASCIMENTO SILVA, P.; MAFRA PEREIRA, F. C. Inteligência Artificial para Recuperação de Dados Abertos: reflexões e proposições para a experiência do usuário no Portal Brasileiro de Dados Abertos. **Brazilian Journal of Information Science: research trends**, Marília, SP, v. 18, p. e024016, 2024. DOI: [10.36311/1981-1640.2024.v18.e024016](https://doi.org/10.36311/1981-1640.2024.v18.e024016). Acesso em: 31 out. 2024.

OPEN GOVERNMENT DATA. **The 8 principles of open government data**. 2007. Disponível em: <https://opengovdata.org/>. Acesso em: 10 fev. 2024.

OPEN GOV DATA (OGD). **Eight principles of open government data**. Califórnia, 2007. Disponível em: https://public.resource.org/8_principles.html. Acesso em: 20 jan. 2024.

OPEN KNOWLEDGE FOUNDATION. **Why open date?**.Disponível em: <https://okfn.org/en/library/why-open-data/>. Acesso em: 01 nov. 2024.

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: 10 fev. 2024.

PESSOA, L. G. de S. B.; SOUSA, M. R. F. de; CARDOSO, T. L. de C. Aplicação de Dados Governamentais Abertos à luz da ciência da informação. **Ciência da Informação**, Brasília, v. 49, n. 3, 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5520>. Acesso em: 10 fev. 2024.

POSSAMAI, A. J.; SOUZA, V. G. Transparência e Dados Abertos Governamentais: Possibilidades e Desafios a Partir da Lei De Acesso À Informação. **Administração Pública e Gestão Social**, Viçosa, v. 12, n. 2, 2020. DOI: 10.21118/apgs.v12i2.5872. Disponível em: <https://periodicos.ufv.br/apgs/article/view/5872>. Acesso em: 10 fev. 2024.

RIBEIRO, C. J. S.; ALMEIDA, R. F. de. Dados abertos governamentais (*open government data*): instrumento para exercício de cidadania pela sociedade. In: Encontro Nacional de Pesquisa em Ciência da Informação, v. 12, p. 2568-2580, 2011. Brasília. **Anais...** Brasília: UNB, 2011. Disponível em: <http://arq.3rengtt.com.br/wp-content/uploads/2015/09/EnancibXII-RibeiroAlmeida.pdf>. Acesso em: 10 fev. 2024.

RIO BRANCO, L. B. P. **Interoperabilidade semântica entre linguagens de indexação para bibliotecas universitárias**. 2020. 142 f. Tese (Doutorado) - Curso de Ciência da Informação, Universidade Estadual Paulista, Marília, 2020. Disponível em: <http://hdl.handle.net/11449/194243>. Acesso em: 10 fev. 2024.

RUSSO, M. **Fundamentos em Biblioteconomia e Ciência da Informação**. Rio de Janeiro: E-papers Serviços Editoriais, 2010. 178 p.

SANT'ANA, R. C. G. Ciclo de vida dos dados e o papel da Ciência da Informação. In: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB)**, 14., Florianópolis, 2013. Anais eletrônicos... Florianópolis: ANCIB, 2013. Disponível em: https://www.brapci.inf.br/_repositorio/2015/12/pdf_a34de5322a_0000013772.pdf. Acesso em: 10 fev. 2024.

SAYÃO, L. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, Belo Horizonte, v. 6, n. 1, 2013. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/300>. Acesso em: 10 fev. 2024.

SEVERINO, A. J. **Metodologia do trabalho científico**. São Paulo: Cortez Editora, 2013.

SETZER, V. W. Dado, informação, conhecimento e competência. **DataGramZero: Revista de Ciência da Informação**, Rio de Janeiro, n. 0, v. 28, 1999. Disponível em: <https://www.ime.usp.br/~vwsetzer/datagrama.html>. Acesso em: 10 fev. 2024.

SINIF, L.; BOUNABAT, B. Approachinganoptimizing open linkedgovernment data portal. In: **Proceedings of the 2nd International Conference on Smart Digital Environment**. 2018. p. 135-139. Disponível em: <https://dl.acm.org/doi/10.1145/3289100.3289122>. Acesso em: 10 fev. 2024.

SILVA, P. N.; PINHEIRO, M. M. K. Métricas para Dados Governamentais Abertos. **Revista Ibero-Americana de Ciência da Informação**, Brasília, DF, v. 11, n. 1, p. 314-332, 2018. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/8431>. Acesso em: 10 fev. 2024.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n.2, p. 161-173, 2006. Disponível em: <https://www.brapci.inf.br/index.php/res/v/33177>. Acesso em: 10 fev. 2024.

VARELA, F. W. **Governo aberto sob a perspectiva da Lei de Acesso à Informação para uma governança sustentável**. 2017. 124 f. Dissertação (Mestrado) - Curso de Ciência Jurídica, Universidade do Vale do Itajaí – Univali, Itajaí, 2017. Disponível em: https://www.univali.br/Lists/TrabalhosMestrado/Attachments/2140/Dissertacao%20Governo%20Aberto%20o_FELIPE_WILDI_VARELA.pdf. Acesso em: 10 fev. 2024.

VAZ, J. C.; RIBEIRO, M. M.; MATHEUS, R. Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no Brasil. **Cadernos PPG-AU/UFBA**, [S. l.], v. 9, n. 1, 2011. Disponível em: <https://periodicos.ufba.br/index.php/ppgau/article/view/5111>. Acesso em: 10 fev. 2024.

W3C BRASIL. **Manual dos dados abertos**: governo / traduzido e adaptado de opendatamanual.org; [cooperação técnica científica entre Laboratório Brasileiro de Cultura Digital e o Núcleo de Informação e Coordenação do Ponto BR (NIC.br)]. São Paulo: Comitê Gestor da Internet no Brasil, 2011. Disponível em: https://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf. Acesso em: 10 fev. 2024.

ZAFALON, Z. R. **Scan for MARC**: Princípios sintáticos e semânticos de registros bibliográficos aplicados à conversão de dados analógicos para o formato MARC21 bibliográfico. 2012. 169f. Tese (Doutorado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista, Marília, 2012. <https://repositorio.unesp.br/handle/11449/103386>. Acesso em: 10 fev. 2024.