

**ENANCIB 2022**

PORTO ALEGRE | UFRGS | PPGCIN

XXII Encontro Nacional de Pesquisa em Ciência da Informação • ENANCIB
Porto Alegre • 07 a 11 de novembro de 2022**XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB****ISSN 2177-3688****GT-2 – Organização e Representação do Conhecimento****ORGANIZAÇÃO DA INFORMAÇÃO EM APIS DE DADOS GOVERNAMENTAIS ABERTOS: UM BREVE PANORAMA BRASILEIRO****ORGANIZATION OF INFORMATION IN OPEN GOVERNMENT DATA APIS: A BRIEF BRAZILIAN OVERVIEW****Patrícia Nascimento Silva. UFMG.****Benildes Coura Moreira dos Santos Maculan. UFMG.****Elisângela Cristina Aganette. UFMG.****Modalidade: Trabalho Completo**

Resumo: A organização é um elemento fundamental para recuperação da informação. No contexto dos dados governamentais abertos (DGA) essa questão é potencializada, visto que o reuso é o maior objetivo da publicação dos DGA. Este estudo investigou a organização da informação nos conjuntos DGA disponibilizados em *Application Programming Interface* (API) do governo brasileiro. A metodologia envolveu uma busca exploratória nos instrumentos que compõem a política de DGA, selecionou uma amostra de dois conjuntos de dados de cada órgão que disponibiliza DGA em APIs e analisou a documentação das APIs considerando quatro critérios que envolvem técnicas e métodos essenciais para a organização da informação. Foi identificado que as APIs são utilizadas por apenas 6% dos órgãos, as documentações apontam algumas técnicas de organização da informação, mas com baixa periodicidade e de forma individualizada entre os órgãos. Conclui-se que o uso de APIs para disponibilização de DGA ainda é tímido e a definição de padrões de organização junto à política da DGA pode ser uma estratégia para aumentar a interoperabilidade e reuso dos DGA no Brasil.

Palavras-Chave: Organização da Informação. Dados Governamentais Abertos. API.

Abstract: Organization is a fundamental element for information retrieval. In the context of open government data (OGD) this issue is magnified, since reuse is the main objective of the publication of OGD. This study investigated the organization of information in the OGD sets made available in Application Programming Interface (API) of the Brazilian government. The methodology involved an exploratory search in the instruments that make up the OGD policy, selected a sample of two data sets from each agency that makes OGD available in APIs, and analyzed the documentation of the APIs considering four criteria involving techniques and methods essential for the organization of information. It was identified that APIs are used by only 6% of the agencies, the documentation points to some techniques for organizing information, but with low frequency and in an individualized way among the agencies. We conclude that the use of APIs to make OGD available is still timid and the definition of organization standards together with the OGD policy may be a strategy to increase the interoperability and reuse of OGD in Brazil.



Keywords: Information organization. Open Government Data. API.

1 INTRODUÇÃO

Dados governamentais abertos (DGA) ou *open government data* são dados abertos relativos ao governo, disponibilizados para livre utilização por qualquer cidadão (AGUNE et al., 2010). O maior objetivo dos DGA é seu reuso, em diversos contextos, inclusive em cenários diferentes do qual o dado foi publicado. Muitas são as possibilidades e oportunidades relacionadas aos DGA, que ganham cada vez mais destaques em diversos países e fomenta novos modelos de negócio.

O termo *DGA* entrou em destaque a partir de 2007, após a publicação de um conjunto de princípios, por um grupo de especialistas, em Sebastopol, Califórnia, sendo referido como "*8 Open Government Data Principles*" ou "*Sebastopol Principles*¹". Os princípios foram considerados um parâmetro inicial para avaliação de iniciativas de DGA (OPEN GOV DATA, 2007). Após a definição dos princípios foram estabelecidas três leis que regem os DGA, por David Eaves, e Timothy John Berners-Lee sugeriu um esquema de cinco estrelas para publicação de dados abertos. As três métricas possuem recomendações importantes, mas percebe-se que a falta de definições em relação à publicação dos dados pode influenciar diretamente na organização e recuperação dos DGA.

A Política de dados abertos do Brasil foi instituída em 2016, por meio do Decreto Nº 8.777/2016, contudo, desde a Constituição de 1988, o acesso à informação já era previsto. A Lei de Acesso à Informação (LAI), Lei Nº 12.527/2011, foi um dos marcos da política de DGA no Brasil, alavancando práticas de transparência orçamentária, acesso público à informação e participação social, até a publicação do Decreto Nº 8.777/2016. Por meio deste último Decreto foram formalizadas questões em relação ao planejamento e publicação dos DGA pelos órgãos da Administração Pública Federal (APF), porém muitas definições ainda são recomendações e boas práticas, ficando a cargo dos órgãos estabelecerem suas prioridades e estratégias para a publicação dos dados. Destaca-se que a política de DGA ainda é fragmentada, por envolver diversos instrumentos como: manuais, padrões, Leis, Decretos, Resoluções e Acórdãos, que possuem relação direta com a publicação de DGA, mas não estão centralizados ou mapeados, de alguma maneira, a nível institucional (SILVA, 2022).

¹ <http://www.opengovdata.org/home/8principles>.



Com a evolução dos ambientes digitais, os grandes volumes de dados e o aumento da complexidade tecnológica, o compartilhamento de dados e metadados é uma tendência mundial, incentivada pelos princípios *Findable, Accessible, Interoperable* e *Reusable* (FAIR), inicialmente no campo dos dados de pesquisa (WILKINSON; DUMONTIER et al., 2016), e também pelos princípios dos DGA no campo dos dados abertos. Assim, acessar e utilizar conjuntos de dados de alta qualidade e interoperáveis é essencial em muitos cenários, mas torna-se também um desafio que demanda esforços de diferentes áreas (LIMA, 2020) e que ganha cada vez mais importância no contexto da organização da informação (FABER; LAMPRECHI, 2022).

A utilização de *Application Programming Interface* (API) é cada vez mais comum em portais de dados abertos e possibilita o processamento por máquina, um dos princípios dos que se tornou condição indispensável para a recuperação dos DGA, frente ao grande volume e as constantes atualizações. Contudo, além das ferramentas tecnológicas, o reuso dos DGA perpassa também por questões que envolvam a organização da informação, a ligação dos dados e sua interoperabilidade. Assim, a partir da problematização exposta, a pesquisa pretende responder: Como os DGA brasileiros publicados em APIs estão organizados? O objetivo do trabalho é investigar a organização da informação nos conjuntos de dados disponibilizados em APIs e para isso buscou-se, especificamente: (i) identificar os conjuntos de dados do governo brasileiro que são disponibilizados em APIs e (ii) analisar a documentação das APIs sob a óptica da organização da informação. Destaca-se que a pesquisa é parte de um projeto em andamento, justifica-se para mapear as práticas utilizadas pelos órgãos, e visa fomentar a discussão em relação à organização dos DGA em consonância com a evolução das ferramentas tecnológicas e a evolução da política de DGA no Brasil na comunidade da Ciência da Informação.

2 ORGANIZAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS

Os DGA são dados abertos no contexto governamental, ou seja, referem-se ao setor público, tornando a informação livremente disponível em formatos abertos e permitindo vias de acesso públicas que facilitem a exploração (KALAMPOKIS; TAMBOURIS; TARABANIS, 2011). Neste sentido, DGA contextualizados se assemelham ao conceito de informação, que no âmbito da Ciência da Informação é entendida como:



[...] uma produção fenomenicamente social que tem por finalidade dinamizar a inter-comunicação humana e promover exposições e descobertas para construção do conhecimento através de interações entre sujeito/autor e sujeito/usuário por meio de dados (plano físico e histórico-social dos sujeitos da informação), mensagens (no plano abstrativo) e atividades documentais (plano material), que favorecem predicativos hermenêuticos aos sujeitos da informação e resultam na apreensão e apropriação pelo sujeito/usuário efetivando um caráter de compreensão (ESPÍNDOLA et al., 2015, p. 150).

Esse caráter de compreensão pelo público está estreitamente ligado ao fato de que a informação deve ser disponibilizada de modo que faça sentido ao usuário, ou seja, que seja significativa e de valor para ele. Desta forma, pode-se dizer que a organização da informação é um “processo que envolve a descrição física e de conteúdo de objetos informacionais e que tem como produto a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico” (BRASCHER; CAFÉ, 2008, p. 5).

A organização dos DGA iniciou com a definição de princípios e boas práticas, perpassando por modelos, métricas e metodologias para publicação e avaliação de DGA que geralmente se apoiavam em legislações (SILVA, 2018). Características elementares dos dados abertos, observadas nos princípios dos DGA e no modelo cinco estrelas, já evidenciavam a organização da informação e estes instrumentos continuam sendo referência para a área. Os oito princípios de DGA estabeleceram recomendações de melhores práticas sobre como os governos deveriam publicar os dados abertos na Internet. Conforme W3C (2016), os oito princípios definem que:

1. Os dados devem estar completos: todos os dados devem estar disponíveis;
2. Os dados devem ser primários: os dados são apresentados tais como coletados na fonte, com o maior nível possível de granularidade e sem agregação ou modificação;
3. Os dados devem ser atualizados: os dados devem ser disponibilizados tão rapidamente quanto necessário à preservação do seu valor;
4. Os dados devem estar acessíveis: os dados devem ser disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades;
5. Os dados devem ser processáveis por máquina: os dados devem estar razoavelmente estruturados de modo a possibilitar processamento automatizado;



6. O acesso deve ser não discriminatório: os dados devem estar disponíveis para todos, sem exigência de requerimento ou cadastro;
7. Os formatos de dados devem ser não proprietários: os dados devem estar disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo;
8. Os dados devem ser livres de licença: os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas à privacidade, segurança e privilégios de acesso são permitidas.

O modelo cinco estrelas foi criado por Berners-Lee para publicação de dados abertos, tendo como referência conceitos de dados abertos conectados ou *Linked Open Data*. A proposta de Berners-Lee é que cada nova estrela alcançada torne os dados progressivamente mais poderosos e mais fáceis para as pessoas utilizarem (OPEN KNOWLEDGE BRASIL, 2013). De acordo com a Open Knowledge Brasil (2013), as estrelas são definidas como: Uma Estrela: os dados estão disponíveis na *Web*, independentemente de formato, sob uma licença aberta (por exemplo: um documento PDF sob uma licença aberta); Duas estrelas: a condição anterior mais disponível como dados estruturados legíveis por máquina (por exemplo: um arquivo Excel ao invés de uma imagem escaneada de uma tabela); Três estrelas: todas as anteriores mais utilizar um formato não proprietário (por exemplo: um arquivo CSV ao invés de um Excel); Quatro estrelas: todas as anteriores mais utilizar URIs bem desenhadas para identificar as coisas, então as pessoas podem referenciá-las; Cinco estrelas: todas as anteriores mais ligar seus dados com dados de outras pessoas para prover contexto.

Os oito princípios dos DGA e o modelo das cinco estrelas destacam a necessidade de automação, integração, acesso e recuperação dos DGA, o que envolve o uso de métodos, técnicas e recursos. Assim, do ponto de vista da organização da informação, técnicas envolvendo a dimensão descritiva, compreendendo elementos relativos à forma do documento e a dimensão temática, voltada aos conteúdos informacionais (CAFÉ; SALES, 2010), podem contribuir com a organização dos DGA. Do ponto de vista tecnológico, a API é um dos recursos que permitem implementar o acesso por máquina, de maneira padronizada e facilitada (SAUDATE, 2021). O uso de APIs na publicação de DGA vai de encontro com os princípios e objetivos dos DGA. Para tanto, a documentação de uma API deve indicar como a



informação foi organizada, assim como os parâmetros de comunicação para que os dados possam ser recuperados prontamente.

3 MÉTODO

Esta pesquisa caracteriza-se como um estudo descritivo e exploratório com abordagem qualitativa e quantitativa. Utilizou-se a análise documental nos instrumentos que compõem a política de DGA do Brasil, para se extrair um reflexo objetivo da fonte original, a localização, identificação, organização e avaliação das informações contidas no documento, além da contextualização dos fatos em determinados momentos (MOREIRA, 2005). Quanto aos procedimentos metodológicos, esses foram realizados em três etapas, detalhas a seguir.

Etapla 1 - Seleção dos instrumentos da política de DGA: Atualmente a política de DGA é composta por três principais instrumentos de monitoramento, publicação e disponibilização: o Painel de Monitoramento de Dados Abertos (PMDA)², criado e mantido pela Controladoria Geral da União em 2017, o Portal Brasileiro de Dados Abertos³ (PBDA) e o Portal da Transparência⁴ (SILVA, 2022). Inicialmente foram realizadas buscas no PMDA para identificar conjuntos de dados publicados em APIs. No entanto, verificou-se que esse instrumento exibe apenas os conjuntos de dados publicados por órgãos, sem detalhar métodos ou formatos de publicação. Assim foi necessário, realizar uma busca em outro instrumento, especificamente no PBDA. Ressalta-se que o Portal da Transparência, que também é um instrumento da política, não foi utilizado nesta pesquisa, pois para acessar os dados disponibilizados em sua API⁵ era necessária a autenticação com uma conta GOV BR. A autenticação exigida está em desacordo com o sexto princípio dos DGA (acesso não discriminatório), assim este instrumento foi considerado fora do escopo desta pesquisa.

Etapla 2 - Buscas no Portal Brasileiro de Dados Abertos: Os filtros existentes no PBDA foram analisados, assim como as informações existentes sobre a publicação de conjuntos de dados em APIs. Foi identificado somente o filtro “Formato” com o tipo “API”. As buscas no PBDA foram realizadas entre abril e maio de 2022 e retornaram três órgãos que disponibilizam DGA em APIs: Banco Central do Brasil (528 ocorrências), Ministério da Economia (8

² <http://paineis.cgu.gov.br/dadosabertos/index.htm>

³ <https://dados.gov.br/>

⁴ <https://www.portaltransparencia.gov.br/>

⁵ <https://api.portal.datransparencia.gov.br/swagger-ui.html#/>



ocorrências) e Ministério da Saúde (3 ocorrências). Desta forma, o escopo da análise foi restringido aos DGA dos três órgãos encontrados: Banco Central do Brasil, Ministério da Economia e Ministério da Saúde.

Etapas 3 - Seleção dos dados: Diante dos resultados apresentados no PBDA, foram selecionados, aleatoriamente, dois conjuntos de dados de cada órgão que disponibiliza DGA em APIs. Por entender que a estrutura implementada para o desenvolvimento de uma API de um órgão é única para todos os conjuntos de dados publicados, optou-se por selecionar somente dois conjuntos de cada órgão, uma vez que o fundamento lógico para organização da informação tende a ser o mesmo, alterando somente o tipo de dado.

Para cada conjunto de dados foi analisada a documentação e demais informações disponibilizadas no portal de dados abertos de cada órgão, a fim de recuperar informações sobre a construção das APIs utilizadas. Destaca-se que nesta pesquisa todas as informações analisadas são provenientes da documentação disponibilizada. O acesso e a coleta de dados direto na API não será objeto deste estudo. A amostra de conjuntos de dados selecionados e as principais documentações consultadas são apresentadas no Quadro 1.

Quadro 1: Amostra de conjuntos de dados selecionados

Órgão	Conjunto de dados	Documentação da API no PBDA	Documentação da API no órgão responsável
Banco Central do Brasil	(1) Dólar comercial (venda e compra) - cotações diárias	https://dados.gov.br/dataset/dolar-americano-usd-todos-os-boletins-diarios	https://olinda.bcb.gov.br/olinda/servico/PTAX/versao/v1/documentacao
	(2) Tarifas Bancárias - por Segmento e por Instituição	https://dados.gov.br/dataset/tarifas-bancarias-por-segmento-e-por-instituicao	https://olinda.bcb.gov.br/olinda/servico/Informes_ListaTarifasPorInstituicaoFinanceira/versao/v1/documentacao
Ministério da Economia	(3) Dados Abertos de Custos com Pensionistas do Governo Federal	https://dados.gov.br/dataset/dados-de-custos-pensionistas-do-governo-federal	https://apidatalake.tesouro.gov.br/docs/custos/
	(4) Dados Abertos de Custos do Governo Federal - Demais Itens	https://dados.gov.br/dataset/gestao-por-centros-de-custos-demais-custos	
Ministério da Saúde	(5) Notificações de Síndrome Gripal	https://dados.gov.br/dataset/casos-nacionais	https://opendatasus.saude.gov.br/dataset/notificacoes-de-sindrome-gripal-api-elasticsearch
	(6) Campanha Nacional de Vacinação contra Covid-19	https://dados.gov.br/dataset/covid-19-vacinacao	https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao

Fonte: Elaborado pelas autoras.

Para a análise da documentação das APIs foram definidos os seguintes critérios:



1. Existência de dicionário de dados e/ou informações sobre os metadados utilizados: verificar a existência de um documento com a descrição dos atributos disponibilizados na API de forma detalhada e clara para o usuário e com informações relativas à organização dos DGA.
2. Utilização de vocabulário controlado ou aderência ao Vocabulário Controlado do Governo Eletrônico (VCGE): verificar se os termos utilizados pertencem a algum vocabulário ou se seguem o VCGE, que é o padrão indicado pelo governo brasileiro.
3. Formatos de disponibilização dos dados e centralização de informações: verificar em quais formatos de arquivo a API disponibiliza os conjuntos de dados e se utiliza frameworks para centralização e gestão dos dados.
4. Utilização de padrões de metadados e possíveis ligações com outros conjuntos de dados: verificar informações sobre padrões utilizados em integração com outros conjuntos ou bases de dados.

Os critérios de análise foram definidos considerando informações sobre a organização dos DGA e sua interoperabilidade com outros conjuntos de dados e elementos da Perspectiva Técnica da Métrica DGABr⁶ que avalia o potencial de reúso dos DGA no Brasil. Elementos comuns da arquitetura de uma API não foram avaliados neste estudo, pois são padrões do recurso utilizado e já contribuem para a organização da informação como: URIs para identificação única do conjunto de dados, definição de parâmetros de entrada e saída e requisições de acesso para processamento automatizado por máquina.

4 RESULTADOS e DISCUSSÃO

Quanto à análise quantitativa, em abril de 2022, o PBDA possuía 11.478 conjuntos de dados publicados, sendo 528 conjuntos em formato API, ou seja, apenas 4,6% dos conjuntos de dados são disponibilizados em APIs. E ainda, dos 50 órgãos que publicam DGA, conforme listado no PBDA, somente três utilizam API, ou seja, 6%. Esses dados revelam que a utilização de APIs, tanto pelo número de órgãos, quanto pelo quantitativo de conjuntos de dados disponibilizados, ainda é muito baixa. No que se refere à análise qualitativa das APIs, os resultados são apresentados no Quadro 2.

⁶A DGABr é uma métrica brasileira composta por dimensões qualitativas e quantitativas que avaliam o potencial de reúso dos DGA. A versão completa está publicada em: <https://www.dgabr.com.br>.

**Quadro 2: Análise das documentações das APIs**

Órgão	Conjunto de dados	Documentação/Dicionário de dados	Vocabulário controlado	Padrões e formatos	Dados ligados
Banco Central do Brasil	(1)	Sim	Não	Padrão: Odata Formato: json, xml, text/csv, text/plain e text/HTM	Não informado.
	(2)	Sim	Não		
Ministério da Economia	(3)	Sim. O dicionário de dados não detalha o formato exato dos dados e usa códigos de outros sistemas.	Não	Formato: Json (agrupados por unidade organizacional, em sua menor granularidade)	Não. Somente referência a sistemas internos.
	(4)				
Ministério da Saúde	(5)	Sim. Dicionário de dados com formato, regra, categorias, cardinalidade e obrigatoriedade.	Não. Usa linguagem natural com termos específicos da área.	Formato: Json	Não. Somente compartilha a metadados (códigos IBGE).
	(6)	Sim. Mas com breve descrição dos conceitos.			

Fonte: Elaborado pelas autoras.

A API do Banco Central disponibiliza cotações e taxas, com periodicidade diária, que inclui os conjuntos (1) e (2). Ao consultar as documentações das APIs do Banco Central foram identificadas duas versões desenvolvidas com o framework Swagger⁷ e uma versão desenvolvida com o padrão OASIS Open Data Protocol (OData)⁸.

A API do Ministério da Economia, intitulada “Custos API”, inclui vários conjuntos de dados sobre custos do governo federal, inclusive os conjuntos (3) e (4) que consistem em séries temporais disponibilizadas desde janeiro de 2015 com granularidade temporal mensal. A API foi construída com o framework Swagger e é um serviço de dados abertos do Tesouro Nacional. Os atributos são organizados em dimensões que é uma arquitetura geralmente utilizada em armazém de dados (*Data warehouse*).

A API do Ministério da Saúde disponibiliza dados sobre a pandemia de COVID-19 (notificações, vacinação) com granularidade temporal diária, cobertura nacional e

⁷ Framework utilizado em diversas APIs no Brasil que permite a descrição, consumo e visualização de serviços para implementação de uma API com arquitetura Representational State Transfer (REST).

⁸ Protocolo com padrão da International Organization for Standardization e da International Electrotechnical Commission (ISO/IEC) que define um conjunto de práticas recomendadas para criar e consumir Web APIs.



granularidade geográfica por município. Observou-se que apesar de existirem duas fontes de informações para os conjuntos de dados (5) e (6) os *endpoint* das APIs são os mesmos e foram desenvolvidos com a ferramenta Elasticsearch⁹.

Para os conjuntos de dados observados neste estudo, a existência de documentação ou dicionário de dados foi um critério atendido por todos. O detalhamento dos atributos e respectivos metadados é descrito na documentação, de modo completo em alguns casos, mas predomina informações bem objetivas e sintetizadas apresentando relação direta com os conceitos da área de negócio dos DGA. Saudate (2021) aponta que uma boa API deve ser o mais auto explicativa possível, ter uma boa documentação e ser retrocompatível, ou seja, permitir versionamento e acesso a versões anteriores.

Foi identificado que as APIs do Banco Central e do Ministério da Saúde possuem versionamento e no caso dos conjuntos de dados (5) e (6), que envolviam informações sobre a pandemia de COVID-19, essa característica foi essencial. No início da pandemia haviam poucas informações sobre a doença e mudanças importantes na ficha de notificação, como a inclusão de novos campos e alterações de regras de negócio, foram implementadas ao longo do tempo (DATASUS, 2021). Desta forma, alterações nos metadados puderam ser implementadas de maneira rápida, sem impactar em integrações e coletas já desenvolvidas em outros sistemas que reusavam os DGA.

O uso de vocabulário controlado não foi identificado em nenhuma documentação. Apesar de os conceitos e atributos estarem inteligíveis em linguagem natural, há muitas definições técnicas e específicas, como visualizado nos conjuntos de dados do Ministério da Saúde, que podem gerar diferentes interpretações e dúvidas para os usuários comuns. Conforme destacado por Moreira, Fujita e Santos (2016), os vocabulários controlados desempenham um importante papel na redução dos eventuais ruídos presentes na comunicação entre os sistemas e seus usuários e atuam como instrumentos auxiliares indispensáveis aos processos de organização, representação e recuperação da informação.

Conforme a legislação da política de DGA, entre 2005 e 2009 foram criados os padrões que hoje, atualizados, prevalecem para normalizar a gestão de dados na APF. São eles:

⁹ A Elasticsearch é uma ferramenta para construção de APIs que permite realizar buscas por texto completo e análise dos dados. Ela é altamente escalável e suporta grande volume de dados (DEV MEDIA, 2022).



Padrões de interoperabilidade em Governo Eletrônico¹⁰ (ePING), Modelo de Acessibilidade de Governo Eletrônico¹¹ (eMAG), Padrão de Metadados do Governo Eletrônico¹² (e-PMG) e Padrões Web de Governo Eletrônico¹³ (ePWG). Em 2010 foi criado o Vocabulário Controlado de Governo Eletrônico¹⁴ (VCGE) e, em 2012 criaram os Vocabulários e Ontologias do Governo Eletrônico (e-VoG¹⁵) que permitiam criar uma interoperabilidade semântica para APF. Em 2012 foi criada a Infraestrutura Nacional de Dados Abertos (INDA) e formalizado o e-PING como um padrão oficial a ser utilizado na APF (POSSAMAI, 2016). Contudo, apesar dos padrões existentes, para os conjuntos de dados analisados não foi identificada nenhuma referência a eles na documentação. É importante destacar que em consulta realizada nos sites dos padrões foram observadas poucas atualizações desde a data de criação do documento até hoje.

Nas documentações analisadas não foram identificadas informações claras sobre a utilização de padrões de metadados, exceto para os conjuntos de dados (1) e (2) que utilizam o padrão OData. O uso de padrões de metadados, normas e modelos informacionais permitem a integração e o intercâmbio de dados (e informações) entre sistemas (ALVES; SOUZA, 2007). A falta de um padrão único de metadados para os DGA dificulta a integração dos dados com outros sistemas, uma vez que esse trabalho de conversão até poderá ser implementado, mas com o uso de várias técnicas da computação que podem demandar grandes esforços. De acordo com Santarem Segundo (2015), grande parte dos conjuntos de dados brasileiros disponibilizados publicamente não foram estruturados seguindo esquemas e padrões de metadados reconhecidos internacionalmente. Contudo, o autor destaca que o uso de esquemas de metadados, para atender as necessidades de um conjunto específico de dados, é muito comum, porém, torna difícil a interoperabilidade e a recuperação dos dados. Uma solução para os DGA poderia ser o uso de modelos adaptados de metadados conhecidos como *Application Profiles*.

¹⁰ <https://eping.governoeletronico.gov.br/>

¹¹ <https://emag.governoeletronico.gov.br/>

¹² https://www.gov.br/governodigital/pt-br/governanca-de-dados/PMGVersao1_1.pdf

¹³ <https://www.gov.br/governodigital/pt-br/acessibilidade-digital/padroes-web-em-governo-eletronico>

¹⁴ https://www.gov.br/governodigital/pt-br/governanca-de-dados/vcge_2_1_0.pdf

¹⁵ <https://wiki.dados.gov.br/Ontologia-Estrutura-Organizacional.ashx>



Apesar de as APIs terem sido desenvolvidas em *frameworks* diferentes foi identificado que o formato json prevalece como saída de dados e a arquitetura REST, baseada em requisições Hyper Text Transfer Protocol (HTTP), predomina como padrão de comunicação. A ligação dos dados com outros conjuntos não foi localizada, somente o compartilhamento de poucos códigos e conceitos com outros sistemas como: os códigos IBGE para identificar ocupações (função) e municípios e códigos do sistema SIORG¹⁶.

Apesar da ausência de dados ligados, o compartilhamento de metadados, a utilização do formato aberto json e a arquitetura REST cooperam com a interoperabilidade. Porém, em relação à ligação dos dados e a interoperabilidade, além da questão técnica, envolvendo formatos e configurações, no contexto dos DGA é essencial um tratamento semântico para que o reuso seja factível. A interoperabilidade semântica está ligada a capacidade de trocar e reusar conteúdos contextualizados semanticamente e reside na estruturação da informação (atribuição de metadados) por meio do controle de vocabulário, a partir de linguagens estruturadas que considerem relações sintáticas e semânticas (MARTINS, 2013). O uso de ontologias também não foi identificado, mas poderiam contribuir com a interoperabilidade e a ligação dos dados, ao definir um conjunto de termos, regras de combinação e relacionamentos (MORAIS; AMBROSIO, 2007) envolvendo os conceitos específicos presentes nos conjuntos de dados do Ministério da Saúde e do Banco Central, por exemplo.

Na análise das documentações também foi observado, além das APIs, soluções de *data warehouse* e armazenamento em nuvem, técnicas e recursos da computação utilizados para lidar com grandes volumes de dados, demonstrando modernização e capacidade técnica dos órgãos. É importante destacar que estes recursos tecnológicos apóiam a organização, recuperação e reutilização dos DGA, por permitir o acesso por máquina em dados já estruturados, contudo é através da organização da informação que será possível fornecer acesso a esse conhecimento estruturado (ALVARES, 2012). Deste modo, a evolução tecnológica precisa estar alinhada com métodos e técnicas da organização para que o acesso à informação e seu reuso seja, de fato, perceptível a sociedade.

5 CONSIDERAÇÕES FINAIS

¹⁶ Sistema de Organização e Inovação Institucional do Governo Federal.



Após dez anos da LAI e seis anos da instituição da política de dados abertos no Brasil, percebe-se uma tímida evolução na forma de disponibilização dos DGA brasileiros. A tecnologia tem contribuído consideravelmente com a organização da informação, no que diz respeito à manipulação e processamento dos conjuntos de DGA, no entanto, a complexidade e o grande volume de dados evidenciam a importância em definir métodos e padrões de organização e a consequente recuperação dos DGA.

Esta pesquisa realizou um breve mapeamento sobre a organização dos DGA em APIs do governo brasileiro, com o intuito de iniciar as discussões sobre essa temática. O controle terminológico e a interoperabilidade estão consonância com os princípios de publicação dos DGA e a ligação dos dados abertos, porém ainda são realidades distantes nos DGA analisados. Nas documentações foram identificadas poucas técnicas de organização, algumas individualizadas e outras em desenvolvimento, mas espera-se que mudanças e melhorias sejam implementadas à medida que os órgãos alcancem maturidade no uso das ferramentas e na gestão dos dados.

Conclui-se que o uso de APIs para disponibilização de DGA ainda é tímido. A utilização de técnicas da Biblioteconomia e da Ciência da Informação, juntamente com a Ciência da Computação, pode ser um caminho possível para a evolução da política de DGA no Brasil, que precisa incluir, oficialmente, métodos e técnicas a serem seguidas e não somente recomendações, a fim de oferecer ações concretas para interoperabilidade dos dados e fomentar o reuso dos DGA.

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, III, Brasília, 2010. **Anais [...]**. Disponível em: http://www.prefeitura.sp.gov.br/cidade/secretarias/upload/controladoria_geral/arquivos/c3_tp_governo%20aberto%20sp%20disponibilizacao%20de%20bases%20de%20dados.pdf. Acesso em: 18 maio 2022.

ALVARES, L. (Org.). **Organização da informação e do conhecimento**: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4 Editores, 2012. 148 p.

ALVES, Maria das Dores Rosa; SOUZA, Marcia Izabel Fugisawa. Estudo de correspondência de elementos metadados: DUBLIN CORE e MARC 21. **Revista Digital de Biblioteconomia & Ciência da Informação**, v. 4, n. 2, 2007. Disponível em:



https://bdm.unb.br/bitstream/10483/7299/1/2013_SuelendaSilvadosSantos.pdf. Acesso em: 19 abril 2022.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento?. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 28 de setembro a 1 de outubro de 2008, Universidade de São Paulo (USP), São Paulo/SP. **Anais [...]**. São Paulo: USP, 2008.

CAFÉ, Lígia; 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. Brasília DF: IBICT, 2010. 335 p.

DATASUS. Nova versão e-SUS Notifica. 2021. Disponível em: <https://datasus.saude.gov.br/wp-content/uploads/2021/08/eSUS-Notificav7-1.pdf>

DEVMEDIA. O que é Elasticsearch? Conheça o Elasticsearch e porque ele é uma poderosa ferramenta para realização de buscas e análise de grandes volumes de dados. 2022. Disponível em: <https://www.devmedia.com.br/o-que-e-elasticsearch/40207>. Acesso em: 15 maio 2022.

ESPÍNDOLA, P. L.; SALM JUNIOR, J. F. ; ROSA, F. ; JULIANI, J. P. Data governance applied to information science: analysis of a scientific data system for the health area. **RDBCI: Rev. Digit. Bibliotecon. Cienc. Inf.**, Campinas, v. 16, n. 3, p. 274-298, set./dez. 2018.

FABER, M., LAMPRECHT, D. The data set knowledge graph: Creating a linked open data source for data sets. **Quantitative Science Studies**, v. 2, n. 4, 2021. Disponível em: <https://direct.mit.edu/qss/article/2/4/1324/108051/The-data-set-knowledge-graph-Creating-a-linked>. Acesso em: 31 mar. 2022.

KALAMPOKIS, E.; TAMBOURIS, E.; TARABANIS, K. A classification scheme for open government data: towards linking decentralised data. **International Journal of Web Engineering and Technology**, v.6, n.3, p.266-285, 2011. Disponível em: <https://dl.acm.org/citation.cfm?id=1999591>. Acesso em: 20 jan. 2022.

LIMA, G. N. B. O. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em Ciência da Informação**, v. 25, n. Especial, p. 57-97, 2020. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/135734>. Acesso em: 19 abril 2022.

MARTINS, Maria Catarina Barradas. **Indexação e controlo da terminologia em bibliotecas do ensino superior politécnico em Portugal**: o sistema no instituto politécnico de Portalegre. Salamanca: Ediciones Universidad de Salamanca, 2013.

MORAIS, E. A. M., AMBROSIO, A. P. L. Ontologias: conceitos, usos, tipos, metodologias, ferramentas e linguagens. Relatório Técnico do Instituto de Informática da Universidade Federal de Goiás, n. 1, v. 7, 2007. Disponível em: https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-07.pdf. Acesso em: 19 abril 2022.

MOREIRA, Sonia Virgínia. **Análise documental como método e como técnica**. In: DUARTE, Jorge; BARROS, Antonio (Org.). Métodos e técnicas de pesquisa em comunicação. São Paulo: Atlas, 2005. p. 269-279.



MOREIRA, W.; FUJITA, M. S. L.; SANTOS, L. B. P. D. Vocabulário controlado para arquivos universitários: a experiência do vtarq-unesp. **Scire: representación y organización del conocimiento**, v. 22, n. 2, p. 107-112, 2016. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/168125>. Acesso em: 14 ago. 2022.

OPEN GOV DATA. Eight principles of open government data. Califórnia, 2007. Disponível em: https://public.resource.org/8_principles.html. Acesso em: 15 maio 2022.

OPEN KNOWLEDGE BRASIL. **Maturidade em Dados Abertos**: Entenda as 5 Estrelas. 2013. Disponível em: <https://ok.org.br/noticia/maturidade-em-dados-abertos-entenda-as-5-estrelas/>. Acesso em: 20 jan. 2018.

POSSAMAI, A. J. **Dados abertos no Governo Federal brasileiro**: desafios de transparência e interoperabilidade. 2016. 313 f. Tese (Doutorado em Ciência Política) Universidade Federal do Rio Grande do Sul, Instituto de Filosofia e Ciências Humanas, Programa de Pós-Graduação em Ciência Política, Porto Alegre, BR-RS. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/156363/001015755.pdf?sequence=1>. Acesso em: 19 abril 2022.

SANTAREMSEGUNDO, J. E. Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais. In: Encontro Nacional de Pesquisa em Ciência da Informação – Enancib, 16., 2015, João Pessoa. Informação, memória e patrimônio: do documento às redes. **Anais [...]**. João Pessoa: Enancib, 2015. Disponível em: <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/paper/view/3149>

SAUDATE, A. **APIs REST**. Seus serviços prontos para o mundo real. [s.l.]: Casa do Código, 2021.

SILVA, P. N. Acesso à informação no Brasil: política de dados governamentais abertos. In: GERALDES, E. et al. (orgs.). **Dez anos da lei de acesso à informação**: limites, perspectivas e desafios. São Paulo: INTERCOM, 2022. 748p. Disponível em: https://10anoslai.com/wp-content/uploads/2022/05/Livro_DALAI_digital_FINAL_17.05.2022.pdf. Acesso em: 19 abril 2022.

SILVA, P. N. **Dados governamentais abertos**: métricas e indicadores de reúso. 2018. 322f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: <https://repositorio.ufmg.br/handle/1843/BUBD-AYNG4U>. Acesso em: 19 abril 2022.

W3C. **Dados Abertos Governamentais**. World Wide Web Consortium Escritório Brasil. 2016. Disponível em: <<http://www.w3c.br/divulgacao/pdf/dados-abertos-governamentais.pdf>>. Acesso em: 21 jan. 2022.

WILKINSON, M., DUMONTIER, M., AALBERSBERG, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Sci Data** 3. 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 08 de abril de 2022.