

Metadados para descrição de *datasets* e recursos informativos do "Portal Brasileiro de Dados Abertos"

Fábio Mosso Moreira

***Doutorando pelo Programa de Pós Graduação em
Ciência da Informação da Universidade Estadual
Paulista, Campus de Marília.***

Ricardo César Gonçalves Sant'Ana

***Docente da Universidade Estadual Paulista,
Campus de Tupã. Docente do Programa de Pós-
Graduação em Ciência da Informação da
Universidade Estadual Paulista, Campus de
Marília. Doutorado em Ciência da Informação pela
Universidade Estadual Paulista, Campus de
Marília. Livre-docente em Sistemas de
Informações Gerenciais pela Universidade
Estadual Paulista, Campus de Tupã.***

Plácida Leopoldina Ventura Amorim da Costa Santos

***Docente no Programa de Pós-Graduação em
Ciência da Informação da Universidade Estadual
Paulista, Campus de Marília. Doutorado em
Semiótica e Linguística Geral pela FFLCH/USP.
Livre-docente em Catalogação pela Universidade
Estadual Paulista, Campus de Marília.***

Zaira Regina Zafalon

***Docente na Universidade Federal de São Carlos
(UFSCar). Doutora em Ciência da Informação pela
Universidade Estadual Paulista, Campus de
Marília.***

<http://dx.doi.org/10.1590/1981-5344/2947>

O Governo brasileiro e os fundadores da Open Government Partnership adotaram, em 2011, a declaração de Governo Aberto, anunciando planos de ação para tornar conjuntos de dados acessíveis baseado nos princípios de Dados Governamentais Abertos, o que resultou na implantação do Portal Brasileiro de Dados Abertos. Por meio de um sistema de recuperação é possível realizar buscas neste website, identificando fontes de dados sobre saúde, transporte, segurança pública, educação, gastos governamentais e processo

eleitoral. Considerando que o compartilhamento de dados pode se tornar uma atividade complexa devido à baixa carga semântica deste tipo de informação, durante a busca por datasets governamentais na Web podem ser encontrados problemas como acesso parcial, superficial e de difícil compreensão dos dados. A descrição eficiente dos recursos informativos e a representação adequada dos conjuntos de dados podem amenizar estes fatores, e contribuir com a obtenção de resultados relevantes no processo de recuperação. Levanta-se a questão sobre quais elementos de metadados os serviços de disponibilização de dados governamentais estão utilizando para representar seus recursos informativos. Para tanto, esta pesquisa tem como objetivo estudar o processo de recuperação de dados no Portal Brasileiro de Dados Abertos analisando os elementos de metadados utilizados na representação dos datasets.

Palavras-chave: *Metadados; Recuperação; Dados Governamentais; Dataset; Recursos Informativos.*

Metadata for description of *datasets* and information resources of "Portal Brasileiro de Dados Abertos"

The Brazilian government and the founders of the Open Government Partnership adopted in 2011 the Declaration of Open Government, announcing action plans to make datasets accessible according with the eight principles of Open Government Data, which resulted in the implementation of the "Portal Brasileiro de Dados Abertos". Through an information retrieval system, the users can search datasets about health, transportation, public safety, education, government spending and electoral process. Share government data in Web can become a complex activity and the low semantic level of this type of information can causes problems during the search like superficial access and difficult understanding of datasets. An Efficient description of information resources and an appropriate representation of the datasets can mitigate these problems and contribute to achieving relevant results in the retrieval process. This research brings the question about which elements of metadata are using by government services to represent its information

resources. In this sense, the aim of this research is study the process of data sharing in the "Portal Brasileiro de Dados Abertos" and identifying the metadata used for the representation of datasets.

Keywords: *Metadata; Retrieval; Government Data; Dataset; Information Resources.*

Recebido em 01.10.2016 Aceito em 07.06.2017

1 Introdução

A Internet é uma tecnologia que pode viabilizar maior interação entre a administração pública e a sociedade, tanto para os cidadãos que buscam por informações para participação do desenvolvimento das políticas públicas quanto para o Governo no fornecimento de serviços informativos mais eficientes.

Sant'Ana (2009) ressalta a importância dos governos utilizarem-se de recursos tecnológicos para a disponibilização de informações sobre suas ações, fornecendo dados sobre: planos e programas governamentais; orçamentos, arrecadação e fontes de receita; licitações, contratos e seus aditivos; e dados sobre funcionários municipais e município. O acesso a este conteúdo pode permitir maior acompanhamento das ações do poder público por parte da sociedade, gerando maior confiança no processo político.

A responsabilidade das administrações públicas em disponibilizar informações sobre suas ações foi intensificada com a promulgação da Lei Federal nº 12.527/2011. A Lei de Acesso à Informação (LAI) adicionou novas obrigações às instituições públicas, reforçando o dever do Estado de garantir o Direito de acesso à informação de forma transparente, clara e em linguagem de fácil compreensão, dispondo também sobre a obrigação da divulgação de informações em sítios oficiais da rede mundial de computadores – Internet (BRASIL, 2011).

No Brasil, uma das primeiras iniciativas no sentido de se utilizar a Internet como meio para disponibilização de dados governamentais ocorreu em 1994, com o Portal da Transparência da Controladoria-Geral da União (BRASIL, 2017a), que trazia informações sobre a distribuição de recursos financeiros do Governo Federal, como transferências e aplicações diretas (PLATT NETO; CRUZ; VIEIRA, 2006). Com esta ação, o país começou a incorporar os preceitos de Governo Eletrônico, adotando a Internet como tecnologia chave para divulgação de informações governamentais e promoção do acompanhamento público de modo proativo.

O Governo Eletrônico envolve tanto questões tecnológicas quanto organizacionais, mesclando fatores relacionados com a utilização de recursos de informática e o tratamento da efetividade operacional da

gestão pública. Os objetivos perseguidos são: transparência, acessibilidade, prestação de contas, serviços públicos eficientes, políticas públicas e envolvimento dos usuários. Ao incorporá-los, os órgãos governamentais buscam eficiência, eficácia e melhoria na qualidade das informações disponibilizadas (ANDERSEN; HENRIKISEN, 2006).

Concomitante à evolução do Governo Eletrônico no país, o governo brasileiro, no ano de 2011, e os governos de Estados Unidos, Reino Unido, Tailândia, Filipinas, Noruega, África do Sul e México, fundaram uma iniciativa internacional para difundir e incentivar globalmente práticas relacionadas à transparência governamental, acesso à informação pública e participação social. Os fundadores da *Open Government Partnership (OPG)* adotaram a declaração de Governo Aberto, anunciando planos de ação para tornar conjuntos de dados acessíveis atendendo aos princípios de dados abertos (CONTROLADORIA GERAL DA UNIÃO, 2016).

Como resultado do compromisso firmado no âmbito do OPG, no primeiro Plano de Ação do Governo Aberto, o Brasil implantou o Portal Brasileiro de Dados Abertos (BRASIL, 2017b), disponibilizado para que pessoas possam encontrar dados e informações públicas. O Portal atua como um grande catálogo digital que facilita a busca e uso de dados publicados pelos órgãos governamentais, buscando proporcionar ao cidadão um melhor entendimento sobre o processo governamental, o acesso aos serviços públicos, o controle das contas públicas e a participação do cidadão no planejamento das políticas públicas (BRASIL, 2016).

Por meio de um sistema de recuperação é possível realizar buscas no Portal Brasileiro de Dados Abertos, identificando fontes de dados sobre saúde suplementar, sistemas de transporte, segurança pública, educação, gastos governamentais e processo eleitoral.

Considerando que o compartilhamento de dados pode se tornar uma atividade complexa devido à baixa carga semântica deste tipo de informação¹, durante a busca por conjuntos de dados governamentais podem ser encontrados problemas relacionados com o acesso parcial, superficial e de difícil compreensão das bases, o que acaba inviabilizando possíveis análises e interpretações.

A descrição eficiente dos recursos informativos e a representação adequada dos conjuntos de dados podem amenizar estes fatores, contribuindo com a obtenção de resultados relevantes no processo de recuperação. Para Moura e Fernandes (2001), a representação de um recurso informacional busca a padronização e normalização no tratamento e recuperação da informação, e tem como objetivo fornecer um conjunto de elementos para a descrição de recursos informativos; obter resultados mais precisos no processo de busca; e estabelecer padrões de metadados, modelos e protocolos, e sua integração com os mecanismos de busca.

¹ Elemento básico, formado por signo ou conjunto finito de signos que não contém, intrinsecamente, um componente semântico, mas somente elementos sintáticos (SANTOS; SANT'ANA, 2002).

Na administração pública é fundamental que os documentos eletrônicos, principalmente aqueles que contém conjunto de dados disponibilizados pelos órgãos governamentais, sejam estruturados e organizados de tal forma que facilite sua recuperação, introduzindo níveis de normalização no desenvolvimento, processamento, armazenamento, recuperação e conservação, de forma que assegure a interoperabilidade da informação no ciclo completo de compartilhamento (MARTINES USERO, 2007).

Neste contexto, levanta-se a questão sobre quais elementos de metadados os serviços de disponibilização de dados governamentais estão utilizando para representar seus recursos informativos. Para tanto, esta pesquisa tem como objetivo estudar o processo de recuperação de dados no Portal Brasileiro de Dados Abertos, de modo a analisar os elementos de metadados utilizados na representação dos *datasets*.

Para estruturar esta questão e atender aos objetivos propostos, a Ciência da Informação tem um importante papel na medida em que se torna fundamental proporcionar a representação adequada do conteúdo informacional para garantir resultados relevantes no processo de recuperação dos conjuntos de dados. Segundo Hawkins (2001), a Ciência da Informação estuda os processos de geração, organização, **representação**, processamento, distribuição, comunicação e uso da informação, no contexto da busca por informações para atender necessidades informativas.

Adotou-se, como procedimento metodológico, a pesquisa qualitativa de caráter exploratório, realizada a partir de uma coleta realizada no Portal Brasileiro de Dados Abertos. Antecedente à coleta, realizou-se uma pesquisa teórica, para contextualizar o tema envolvendo a utilização de tecnologias no compartilhamento de dados governamentais, e aspectos relacionados à representação de recursos informativos na Web.

A partir de uma coleta de dados realizada no Portal Brasileiro de Dados Abertos, foi possível traçar um cenário da disponibilização de dados governamentais na Web, estudando o processo de recuperação dos recursos informativos e a maneira como são apresentados os elementos de metadados que os representam. Segundo Sant'Ana (2013), o processo de recuperação de dados pode ser entendido como a fase do Ciclo de Vida dos Dados em que ocorre ações relacionadas à disponibilização dos dados para consulta e visualização

No total foram analisados vinte e cinco datasets (Quadro 1), localizados por meio da realização de pesquisas utilizando o sistema de busca interno do Portal Brasileiro de Dados Abertos. Para a realização das pesquisas, foram utilizados cinco termos de busca (Agricultura, Saúde, Infraestrutura, Educação e Segurança), definidos de modo que, no escopo final, constassem recursos informativos abrangendo segmentos distintos trabalhados pelo Governo. Após a realização da pesquisa utilizando cada um dos termos de busca, foram selecionados os cinco primeiros resultados listados pelo sistema de recuperação para que fosse analisada

a apresentação dos elementos de metadados utilizados na descrição do dataset.

Quadro 1 - Escopo de datasets localizados no Portal Brasileiro de Dados Abertos

Segmento	Título dos datasets	Hiperlink
Agricultura	Aprovação Sistema BNDES Setor CNAE Agricultura e Pesca 2015	http://dados.gov.br/dataset/bndes_ap_2015
	Indicadores sobre Brasil Sem Miséria	http://dados.gov.br/dataset/brasil-sem-miseria
	Indicadores sobre Infraestrutura Hídrica - Agricultura Irrigada	http://dados.gov.br/dataset/infraestrutura-hidrica-agricultura-irrigada
	Indicadores sobre PRONAF - Programa Nacional de Fortalecimento da Agricultura Familiar	http://dados.gov.br/dataset/pronaf-programa-nacional-de-fortalecimento-da-agricultura-familiar
	Indicadores sobre Seguro da Agricultura Familiar - SEAF	http://dados.gov.br/dataset/seguro-da-agricultura-familiar-seaf
Educação	Indicadores sobre Educação no Campo	http://dados.gov.br/dataset/educacao-no-campo
	Indicadores sobre Ensino Básico - Mais Educação	http://dados.gov.br/dataset/ensino-basico-mais-educacao
	Microdados do Censo da Educação Superior	http://dados.gov.br/dataset/microdados-do-censo-da-educacao-superior
	Microdados do Exame Nacional de Desempenho de Estudantes - Enade	http://dados.gov.br/dataset/microdados-do-exame-nacional-de-desempenho-de-estudantes-enade
	Taxa de analfabetismo funcional do Brasil de 2001 a 2009	http://dados.gov.br/dataset/taxa-de-analfabetismo-funcional-2001-2009
Infraestrutura	Indicadores sobre PAC Equipamentos	http://dados.gov.br/dataset/pac-equipamentos
	Infraestrutura Turística - 1º Balanço do PAC 2015	http://dados.gov.br/dataset/infraturistica

Segmento	Título dos datasets	Hiperlink
	Manifestações de Usuários Registrados na Agência Nacional de Aviação Civil - ANAC	http://dados.gov.br/dataset/manifestacoes-de-usuarios-registrados-na-anac
	Movimento dos Aeroportos Administrados pela Infraero	http://dados.gov.br/dataset/movimento-dos-aeroportos-administrados-pela-infraero
	Projetos da Carteira COSIPLAN	http://dados.gov.br/dataset/mpog_cosiplan2
Saúde	Distribuição de Agentes Comunitários de Saúde – ACS	http://dados.gov.br/dataset/acs1
	Indicadores sobre Saúde Complementar	http://dados.gov.br/dataset/saude-complementar
	Indicadores sobre Saúde da Família	http://dados.gov.br/dataset/saude-da-familia
	Limites das regiões de saúde	http://dados.gov.br/dataset/limites_regiao_saude
	Unidades Básicas de Saúde - UBS	http://dados.gov.br/dataset/unidades-basicas-de-saude-ubs
Segurança	Indicadores sobre Aprimoramento das Atividades de Segurança Pública	http://dados.gov.br/dataset/aprimoramento-das-atividades-de-seguranca-publica
	Indicadores sobre Força Nacional de Segurança Pública	http://dados.gov.br/dataset/forca-nacional-de-seguranca-publica
	Indicadores sobre Segurança Pública	http://dados.gov.br/dataset/seguranca-publica
	Postos e delegacias do Departamento de Polícia Federal	http://dados.gov.br/dataset/postos-e-delegacias-do-dpf
	Segurança Pública com Cidadania	http://dados.gov.br/dataset/mpog_seguranca_publica

Fonte: Os autores.

A análise dos *datasets* identificados foi realizada por meio do método de observação direta, direcionada na identificação dos principais elementos de metadados descritivos que eram apresentados durante o processo de recuperação. A identificação dos elementos de metadados foi embasada pelo Padrão de Metadados do Governo Eletrônico (e-PMG 1.1), na qual foram verificadas a definição dos metadados deste padrão, servindo para direcionar a análise realizada no Portal.

Os dados obtidos foram sistematizados por meio de tabelas (Tabela 1 e Tabela 2), na qual os rótulos de linha referem-se aos elementos de metadados identificados, e os rótulos de coluna referem-se aos *datasets* analisados. A tabela foi completa conforme os elementos eram identificados no conteúdo do *dataset*, verificando-se a frequência com que determinado elemento de metadado era apresentado no escopo total dos *datasets* analisados, e quais dos elementos identificados podiam estar associados aos elementos do e-PMG 1.1.

Para relacionar os elementos de metadados utilizados na representação dos *datasets* verificados com os elementos de metadados do e-PMG 1.1, aplicou-se o método *Crosswalk*. Segundo Baca (2008), o *Crosswalk* é um produto visual e textual do processo de mapeamento de elementos que mostra o relacionamento e equivalência entre dois ou mais formatos de metadados. São utilizados para comparar elementos de metadados entre padrões distintos, e avaliar o grau de interoperabilidade e possibilidades de conversão da estrutura de dados dos recursos informativos de um formato para outro, mostrando as relações de similaridade e diferença entre os padrões. O *Crosswalk* também permite ao sistema de busca pesquisar campos com o mesmo conteúdo ou similar em diferentes fontes de dados dando suporte à interoperabilidade semântica.

O artigo está estruturado em quatro seções: esta introdução, contendo uma contextualização geral do objeto de estudo e objetivos da pesquisa, acompanhada da descrição metodológica e de coleta de dados; uma segunda seção contendo conceitos levantados durante a pesquisa teórica e uma subseção descrevendo o e-PMG 1.1; uma terceira seção contendo os resultados obtidos a partir da análise dos dados coletados; e a quarta, e última seção, contendo as considerações finais identificadas durante a realização da pesquisa.

2 Metadados para recuperação de recursos informativos na Web

A *Web* pode ser considerada a maior coleção de documentos eletrônicos mundiais, com conteúdo alimentado dinamicamente por bases de dados e *inputs* dos usuários. Entretanto, a maior parte da informação disponível possui qualidade variável, e o desafio consiste em encontrar informações relevantes neste universo.

Para facilitar a localização de informações relevantes na *Web*, Baca (2008) ressalta que é necessária a descrição eficiente dos recursos para dar sentido à vasta quantidade de informação disponível, o que implica a representação e a apresentação adequada dos documentos.

A expansão das unidades de informação e o crescimento exponencial de acervos documentais em ambientes digitais trouxeram a necessidade de reaver processos para satisfazer a demanda informacional de diferentes públicos. Segundo Zafalon (2012), a produção e a proliferação de informação em ambientes digitais exigem a adoção de

padrões para a representação dos documentos eletrônicos de modo que a recuperação e o acesso estejam garantidos pela identificação unívoca do recurso.

Representar recursos informativos é atividade fundamental para o estabelecimento do processo comunicativo entre unidades de informação e usuários, principalmente em uma plataforma digital como a *Web*. Esta representação envolve a comunicação, aos usuários, dos documentos disponíveis de modo individualizado, reunidos com base em suas características de conteúdo ou de suas expressões e obras (ZAFALON, 2012).

Um documento tem a função de materializar uma informação, mas não necessariamente vincular a mesma a um suporte físico. Os documentos armazenados em um dispositivo eletrônico, apesar de não serem físicos, são entidades identificadas e estruturadas, contendo textos, gráficos, sons, imagens ou qualquer outra classe de informação que possa ser armazenada, editada, extraída e compartilhada entre sistemas de tratamento de informação (MARTINES USERO, 2007).

No processo de representação, aspectos como a integridade, a clareza, a precisão, a lógica e a consistência dos conteúdos dos documentos a serem representados devem ser resguardados, tendo em vista sempre o usuário e as necessidades por ele demonstrada (MEY, 1995).

Para Garrido Arilla (1999² *apud* ZAFALON, 2012), é importante que seja feita uma representação codificada e econômica de um documento: representação pois reproduz certas características do documento; codificada porque identifica o documento a que se propôs identificar; e econômica porque busca proporcionar informação de um documento em um mínimo espaço e com o mínimo de símbolos possíveis.

Segundo Santos e Correa (2009), para representar existe a necessidade de se estabelecerem esquemas e regras de descrição padronizadas, no que se refere à forma e ao conteúdo das manifestações informativas.

Para Foulonneau e Riley (2008) alguns fatores são decisivos na escolha dos padrões para representação de recursos informativos: a instituição (cultura organizacional, rotinas institucionais); o padrão dos elementos representativos; os documentos a serem representados; e os objetivos do projeto (para atender às necessidades identificadas).

Representar recursos informativos de forma padronizada só é possível adotando convenções sintáticas e semânticas aplicadas por meio de padrões de estrutura de metadados descritivos e padrões de conteúdo (FERNANDES, 2005). A sintaxe fica por meio da ordem dos elementos dispostos para a representação, e a semântica baseia-se no significado estabelecido entre os elementos representados e o documento representado (ZAFALON, 2012).

² GARRIDO ARILLA, M. R. Procedimientos automáticos de creación y transformación de los registros bibliográficos. *Anales de Documentación*, n. 4, p. 127-137, 2001.

Baca (2008) ressalta que os padrões de metadados servem para modelar os atributos mais importantes do objeto que descreve, onde cada componente do padrão é delimitado por etiquetas que identificam o significado de cada pedaço de informação presente no objeto. Com base neste papel executado pelos metadados, para esta pesquisa define-se metadados como sendo o conjunto de atributos que apresentam o conteúdo informacional de um recurso que pode estar em meio eletrônico ou não (ALVES, 2005).

No meio eletrônico, Martines Usero (2007) destaca que os metadados podem auxiliar na representação de recursos informativos por meio da convergência de formatos de codificação, da organização de conhecimento segundo estruturas que permitem o usuário atingir seus objetivos, e melhora na recuperação da informação, rotulando conteúdos para facilitar sua localização por meio de sistemas de busca.

Segundo Baca (2008), existem alguns padrões de metadados específicos para descrever recursos informativos em um meio eletrônico, sendo os principais: Dublin Core³ (conjunto de 15 elementos de informação que podem ser utilizados para descrever genericamente uma variedade de recursos) e o RDF⁴ (desenvolvido pelo W3C para codificar a descrição de recursos de modo que computadores possam 'entender', compartilhar e processar informação de modo útil e acrescentar semântica nos metadados).

A utilização de padrões de metadados pode ajudar os sistemas de busca na *Web* a responder consultas que exigem a extração e combinação de dados de múltiplas fontes, permitindo o acesso integrado a estes serviços. Serviços integrados de busca de dados têm o objetivo de fornecer ao usuário acesso uniforme às diversas fontes, as quais podem ser autônomas, heterogêneas e distribuídas. O objetivo principal é evitar que o usuário tenha que localizar as fontes de dados, interagir com cada uma isoladamente e combinar manualmente os dados provenientes de múltiplas fontes (LEVY, 2000).

O objeto de estudo analisado por esta pesquisa pode ser considerado um serviço de recuperação de conteúdo na *Web* que fornece acesso uniforme às fontes de dados providas de distintos órgãos governamentais. Portanto, é válido considerar a utilização de um padrão específico de metadados para representação de recursos informativos aplicados na esfera pública. Descreve-se então, o Padrão Metadados do Governo Eletrônico (e-PMG 1.1), cujos elementos e diretrizes de aplicação auxiliaram na identificação dos elementos utilizados para representação das bases de dados disponibilizadas no Portal Brasileiro de Dados Abertos, verificando os metadados apresentados durante o processo de recuperação.

De acordo com o Comitê Executivo do Governo Eletrônico (CEGE, 2014), o e-PMG é resultado da política de Infraestrutura Nacional de

³ Mais informações sobre o Dublin Core podem ser obtidas em: <<http://dublincore.org>>. Acesso em: 12 set. 2017.

⁴ Mais informações sobre o RDF podem ser obtidas em: <<https://www.w3.org/RDF>>. Acesso em: 12 set. 2017.

Dados Abertos (INDA) e estabelece um conjunto mínimo de elementos contendo dados necessários para recuperação e gerenciamento de informações governamentais, estabelecendo semântica entre os elementos, qualificadores e a forma de utilizá-los para descrição de conteúdos no domínio público. O e-PMG 1.1 é composto por vinte elementos e suas respectivas definições: quinze deles baseados no Padrão Dublin Core (DC), e outros cinco elementos adicionais identificados como necessários para o contexto do Governo Eletrônico brasileiro.

Os elementos que compõem o e-PMG 1.1 são: Abrangência, Assunto, Colaborador, Contexto jurídico-administrativo, Criador, Data, Descrição, Destinação, Destinatário, Direitos, Fonte, Formato, Identificador, Idioma, Localização, Preservação, Publicador, Relação, Tipo e Título.

Com relação à implantação do padrão, cada organização deve decidir quais recursos informativos requerem os metadados e-PMG 1.1 prioritariamente. Sua implementação pode ser feita de modo gradativo, agregando coleções de recursos até reduzir a granularidade, cabendo à organização decidir quanto à necessidade de descrever os recursos já existentes (CEGE, 2014).

No guia específico do e-PMG (BRASIL, 2009), para cada elemento, são descritos os seguintes dados:

- a. Nome do elemento: termo específico que nomeia o elemento;
- b. Identificador: termo atribuído de acordo com o padrão original Dublin Core;
- c. Definição: descrição do elemento;
- d. Obrigatoriedade: grau de uso do elemento, sendo este obrigatório, obrigatório se aplicável ou opcional;
- e. Objetivo: aplicabilidade do elemento;
- f. Comentários: informações adicionais para compreensão e uso do elemento;
- g. "Não-confundir com": apontando quais outros elementos geralmente pode gerar dúvida no momento da aplicação;
- h. Qualificadores: que tornam o significado do elemento mais específico; exemplos (situações de aplicação dos elementos);
- i. Sintaxe HTML: forma de uso do elemento na página HTML;
- j. Esquemas: especificando a forma do elemento;
- k. Mapeamento: correlação com outros padrões de metadados tradicionais.

O Quadro 2 sintetiza os dados identificados no guia do e-PMG 1.1, destacando os seguintes atributos dos elementos: nome do elemento,

definição do elemento, obrigatoriedade e qualificadores (título e descrição).

Quadro 2 – Elementos do e-PMG 1.1 e seus principais atributos

Nome	Definição	Obrigatoriedade	Qualificadores	
			Título	Descrição
Abrangência	Extensão espacial e temporal do recurso	Obrigatória, se aplicável	abrangência.espacial	Identificação de localização geográfica
		Obrigatória, se aplicável	abrangência.temporal	Indicação de datas como início e fim
Assunto	Palavras-chave ou termos que representam o conteúdo do recurso	Obrigatória	assunto.categoria	Pode ser utilizado como valor um termo do Vocabulário Controlado do Governo Eletrônico (BRASIL, 2017c) que esteja associado ao conteúdo do recurso
		Opcional	assunto.palavra-chave	Pode ser utilizado um termo de qualquer outro vocabulário-controlado
		Opcional	assunto.pessoa	Indica quando o recurso trata de uma pessoa específica
Colaborador	Entidade que contribuiu para a elaboração do conteúdo do recurso	Opcional	colaborador.papel	Indica o papel desempenhado pelo colaborador na produção do recurso informacional, como tradutor, fotógrafo, relator, revisor
Contexto Jurídico-Financeiro	Legislação e às normas que regulam e autorizam o funcionamento e o desempenho das atividades da organização responsável pela produção e/ou disponibilização do recurso	Opcional	contexto.estatutoDeAutorização	Indica a legislação que autoriza a criação do recurso
Criador	Refere-se à entidade originalmente responsável pela criação e manutenção do recurso	Obrigatória	criador.autor	Entidade responsável pela criação do conteúdo intelectual do recurso
		Obrigatória	criador.arquivístico	Entidade responsável pela manutenção dos recursos e documentos arquivísticos
Data	Refere-se a uma data associada à ocorrência de um evento no ciclo de vida	Obrigatória	data.aceitação	Indica quando o recurso foi aceito
		Obrigatória	data.alteração	Indica quando o recurso foi alterado
		Obrigatória	data.criação	Indica quando o recurso foi criado

Nome	Definição do recurso	Obrigatoriedade	Qualificadores	
			Título	Descrição
		Obrigatória	data.capturado	Indica quando o recurso foi integrado por um sistema informatizado
		Obrigatória	data.disponibilidade	Indica quando o recurso foi disponibilizado para acesso
		Obrigatória	data.publicação	Indica quando o recurso foi publicado
		Obrigatória	data.frequênciaDeAtualização	Indica a frequência com que o recurso é atualizado
		Obrigatória	data.próximaVersão	Indica quando haverá a substituição do recurso por um mais atualizado
		Obrigatória	data.recebimento	Indica quando o recurso deu entrada na organização
		Obrigatória	data.submissão	Indica quando o recurso foi submetido à uma apreciação
		Obrigatória	data.validade	Indica o período máximo que um recurso está válido para consulta
Descrição	Refere-se ao resumo e explicação sucinta do conteúdo do recurso	Opcional	descrição.resumo	Consiste da exposição sucinta do conteúdo do recurso
		Opcional	descrição.sumário	Indica a lista de subunidades do conteúdo do recurso
Destinação	Refere-se aos aspectos decisórios quanto ao encaminhamento dos documentos arquivísticos para guarda permanente ou eliminação	Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.dataDeAuto-Remoção	Indica a data de remoção automática do recurso
		Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.ação	Descreve a ação a ser realizada no recurso quando determinada condição for alcançada
		Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.prazoDeGuarda	Indica o prazo em que o recurso deve ser mantido em guarda
		Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.condições	Indica o evento que dispara a contagem do prazo de guarda

Nome	Definição	Obrigatoriedade	Qualificadores	
			Título	Descrição
		Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.data	Indicando a data em que uma ação de destinação deve ser efetivada
		Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.responsável	Indica a identidade responsável pela autorização da destinação
		Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.nota	Comentários adicionais sobre a ação de destinação
		Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.localDeDestinação	Indica o local previsto para destinação de um recurso exportado
		Obrigatória para documentos arquivísticos e Opcional para os demais	destinação.estágioDeExportação	Informa sobre o andamento da exportação do recurso
Destinatário	Refere-se à entidade para quem a informação contida no documento foi dirigida	Obrigatória se aplicável	destinatário.cópia	Indica o nome das entidades para quem foi encaminhado uma cópia do recurso
Direitos	Refere-se às restrições e às permissões sobre o acesso e o uso do recurso	Opcional	direitos.classificaçãoDoGrauDeSigilo	Indica se o recurso é ostensivo ou sigiloso
		Opcional	direitos.descritor	Indica o significado de uma marcação de proteção
		Opcional	direitos.dataDeExpiraçãoDoGrauDeSigilo	Indica quando expira o grau de sigilo do recurso
		Opcional	direitos.custodiador	Indica o usuário ou função com poderes de gestão sobre o recurso
		Opcional	direitos.listaDeAcessoDosGrupos	Indica o nome de um grupo que têm direito a ter acesso ao recurso

Nome	Definição	Obrigatoriedade	Qualificadores	
			Título	Descrição
Fonte	Refere-se à identificação do recurso que deu origem ao recurso atual	Opcional	Não possui qualificadores	Não possui qualificadores
Formato	Refere-se ao conjunto de características físicas e lógicas que definem a apresentação do recurso	Opcional	formato.formatoDeArquivo	Indica o formato de arquivo em que o conteúdo do recurso está registrado
		Opcional	formato.dimensão	Indica o tamanho físico e/ou capacidade do recurso
		Opcional	formato.mídia	Indicando o tipo de material físico no qual o recurso está persistido
Identificador	Aponta a referência inequívoca do recurso em um determinado contexto	Obrigatória	identificador.idDoSistema	Indica o número gerado automaticamente e atribuído pelo sistema
		Opcional	identificador.códigoDeClassificação	Indica a categoria atribuída para o documento com base em um plano de classificação
		Obrigatória se aplicável	identificador.versão	Indica a versão ou edição do recurso, importante para verificar a existência de versões anteriores
Idiomas	Refere-se à língua em que é expresso o conteúdo do documento	Opcional	Não possui qualificadores	Não possui qualificadores
Localização	Refere-se ao lugar de armazenamento atual do recurso (podendo ser a URL para acesso ao recurso no repositório)	Opcional	Não possui qualificadores	Não possui qualificadores
Preservação	Refere-se às informações necessárias para apoiar o acesso, uso e preservação de longo prazo de um recurso	Opcional	preservação.formatoOriginal	Indica nome do formato em que o recurso foi criado
		Opcional	preservação.dependênciaDeHardware	Indica o hardware necessário para acessar o recurso
		Opcional	preservação.dependênciaDeSoftware	Indica o software e versões necessárias para acesso ao recurso
		Opcional	preservação.dependênciaDeSistemaOperacional	Indica o sistema operacional e versões necessárias para acesso ao recurso
		Opcional	preservação.resolução	Descreve a resolução em

Nome	Definição	Obrigatoriedade	Qualificadores	
			Título	Descrição
				formato digital
		Opcional	preservação.dataDaAção	Indicando a data e hora em que a ação de preservação foi realizada
		Opcional	preservação.tipoDeAção	Indica a ação de preservação realizada
		Opcional	preservação.descriçãoDaAção	Descreve detalhes específicos da ação de preservação, incluindo o estado original, as mudanças realizadas e os motivos
		Opcional	preservação.ferramentasDaAção	Descreve as ferramentas utilizadas na ação de preservação
		Opcional	preservação.próximaAção	Indica a próxima ação ou avaliação da preservação pela qual o recurso deve passar
		Opcional	preservação.dataDaPróximaAção	Indica quanto haverá próxima ação ou avaliação da preservação pela qual o recurso deve passar
Publicador	Refere-se à entidade responsável por tornar o recurso disponível	Obrigatória se aplicável	Não possui qualificadores	Não possui qualificadores
Relação	Diz respeito à referência ou associação que o recurso descrito possui perante outro recurso existente	Opcional	relação.conformidadeCom	Indica a referência a um padrão estabelecido pelo qual o recurso está em conformidade
		Opcional	relação.descriçãoDaRelação	Destaca a relação entre dois ou mais recursos
		Opcional	relação.éFormatoDe	Indica a referência ao formato do recurso que deu origem ao recurso descrito
		Opcional	relação.éParteDe	Indica partes físicas ou lógicas do recurso referenciado
		Opcional	relação.éReferenciadoPor	Indica se o recurso descrito é citado ou apontado por outro recurso
		Opcional	relação.éRequeridoPor	Indica se o recurso descrito é requerido para apoiar a utilização ou manter a coerência do conteúdo de outro recurso relacionado
		Opcional	relação.éSubstituídoPor	Indica se o recurso descrito foi substituído por algum outro recurso

Nome	Definição	Obrigatoriedade	Qualificadores	
			Título	Descrição
		Opcional	relação.éVersãoDe	Indica se o recurso descrito é uma edição, versão, ou adaptação de outro recurso
		Opcional	relação.razãoDaTarjaOuTruncamento	Descreve a razão para a publicação de um truncamento, como privacidade ou sigilo
		Opcional	relação.referência	Indica se o recurso descrito cita ou aponta outro recurso relacionado
		Opcional	relação.requer	Indica se o recurso descrito necessita de outro recurso para assegurar sua inteligibilidade
		Opcional	relação.substitui	Indica se o recurso substitui um recurso anterior
		Opcional	relação.tarjaOuTruncamento	Indica se o recurso descrito possui uma versão com alguma parte de conteúdo removida
		Opcional	relação.temFormatoDe	Indica se o recurso possui outros formatos
		Opcional	relação.temParteDe	Indica se o recurso descrito tem parte de outro recurso relacionado
		Opcional	relação.temVersãoDe	Indica se o recurso descrito tem outras versões ou edições
Tipo de Recurso	Refere-se à natureza ou gênero do conteúdo presente no recurso	Obrigatória se aplicável	Não possui qualificadores	Não possui qualificadores
Título	Refere-se à uma palavra, frase ou grupo de caracteres que nomeia um recurso e que pode corresponder à sua denominação oficial	Obrigatória, mas Opcional para título alternativo	título.títuloAlternativo	Título substituto ou outra forma de designar um recurso, podendo ser o título oficial em outro idioma

Fonte: Adaptação de CEGE (2014).

3 Resultados

Para compor os resultados desta pesquisa, foram estudados os *datasets* disponibilizados pelo Portal Brasileiro de Dados Abertos, sendo analisada a representação dos recursos informativos por meio da

identificação dos elementos de metadados apresentados durante o processo de recuperação realizado por meio de pesquisas via buscador interno do Portal.

Na Figura 1 é apresentado um esquema com a estrutura padrão dos *datasets* analisados no Portal Brasileiro de Dados Abertos, destacando os principais componentes do conteúdo da página.

A maioria dos elementos de metadados descritivos do *dataset* está apresentada na área localizada na parte inferior do recurso. Contudo, no topo, verificam-se metadados como Título e Descrição. Para poder associar os elementos de metadados identificados nos *datasets* com os elementos do e-PMG 1.1 utilizou-se a técnica *Crosswalking*, adotando critérios como a observação da terminologia dos metadados apresentados, a verificação do conteúdo referente aos valores dos elementos e as diretrizes de aplicação verificadas no guia oficial do e-PMG 1.1.

O Título do *dataset* pode ser associado ao elemento "título" do e-PMG 1.1, de aplicação obrigatória. Este metadado foi verificado em todos os *datasets* analisados durante a pesquisa e é utilizado para denominar um recurso e permitir sua identificação.

O conjunto de texto localizado logo abaixo do Título descreve o conteúdo presente naquele *dataset*, e a partir do *Crosswalk* pode ser associado ao elemento "descrição" do e-PMG 1.1 e seu qualificador "descrição.resumo", de aplicação opcional. Este metadado foi verificado na descrição de todos os *datasets* analisados durante a pesquisa, e é utilizado para realizar uma exposição sucinta do conteúdo do recurso.

Os termos que resumem o assunto abordado por cada *dataset*, localizados após a lista de recursos informativos disponíveis no *dataset*, a partir do *Crosswalk* podem ser associados ao elemento "assunto" do e-PMG 1.1 e seu qualificador "assunto.palavra-chave", de aplicação opcional. Este metadado foi verificado na descrição de todos os *datasets* analisados durante a pesquisa, e é utilizado para descrever de forma precisa os termos de interesse presentes no recurso informativo utilizando um vocabulário controlado, no caso, o Vocabulário Controlado do Governo Eletrônico.

Figura 1 – Estrutura padrão dos *datasets* no Portal Brasileiro de Dados Abertos



Fonte: Os autores.

Na área localizada na parte inferior da página (ilustrado na Figura 1) foram verificados um conjunto de elementos de metadados descrevendo informações adicionais do recurso. Contudo, nem todos os elementos estavam apresentados na descrição dos *datasets* analisados (a frequência com que cada elemento de metadado constava na página pode ser observada na Tabela 1).

O elemento "Autor", a partir do *Crosswalk*, pode ser associado ao elemento "criador" do e-PMG 1.1 e seu qualificador "criador.autor", de aplicação obrigatória. Este metadado foi verificado na descrição de todos os *datasets* analisados durante a pesquisa e apresentava valores referentes ao nome de entidades responsáveis pela criação do conteúdo intelectual disponível pelo *dataset* (ex: Secretária Básica de Saúde).

O elemento "Mantenedor", a partir do *Crosswalking*, pode ser associado ao elemento "criador" do e-PMG 1.1 e seu qualificador "criador.produtor", de aplicação obrigatória. Este metadado foi verificado na descrição de todos os *datasets* analisados durante a pesquisa e apresentava valores referentes ao nome de entidades responsáveis pela manutenção e proteção dos recursos contendo o conteúdo intelectual disponível pelo *dataset* (ex: Gabinete do Ministro).

O elemento "Fonte", a partir do *Crosswalking*, pode ser associado ao elemento "fonte" do e-PMG 1.1, de aplicação opcional. Este metadado foi verificado na descrição de 96% dos *datasets* analisados durante a pesquisa e tem como função identificar uma fonte específica que deu origem ao recurso atual. Este elemento apresentava como valor um *hiperlink* apontando para o portal da entidade originária do *dataset* (ex: <http://pgi.gov.br/pgi/>).

O elemento "Órgão – Poder" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 92% dos *datasets* analisados durante a pesquisa e apresentava como valor um termo indicando o nível de poder relacionado às entidades responsáveis pelo *dataset* (ex: Executivo).

O elemento "Órgão – Esfera" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 88% dos *datasets* analisados durante a pesquisa e apresentava como valor um termo indicando a esfera relacionada às entidades responsáveis pelo *dataset* (ex: Federal).

O elemento "Granularidade geográfica", a partir do *Crosswalking*, pode ser associado ao elemento "abrangência" do e-PMG 1.1 e seu qualificador "abrangência.espacial", de aplicação obrigatória se aplicável. Este metadado foi verificado na descrição de 88% dos *datasets* analisados durante a pesquisa e apresentava como valor termos indicando nível de detalhamento espacial do conjunto de dados presente no *dataset* analisado (ex: estadual).

O elemento "Frequência de atualização", a partir do *Crosswalking*, pode ser associado ao elemento "data" do e-PMG 1.1 e seu qualificador "data.frequênciaDeAtualização", de aplicação obrigatória. Este metadado foi verificado na descrição de 52% dos *datasets* analisados durante a

pesquisa e apresentava como valor termos indicando frequência de atualização do *dataset* (ex: anual, mensal, diário).

O elemento "Assuntos", a partir do *Crosswalking*, pode ser associado ao elemento "assunto" do e-PMG 1.1 e seu qualificador "assunto.categoria", de aplicação obrigatória. Este metadado foi verificado na descrição de 28% dos *datasets* analisados durante a pesquisa e apresentava como valores termos apontando categorias específicas relacionadas ao assunto contido naquele *dataset* (ex: Unidade Básica de Saúde).

O elemento "VCGE" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 28% dos *datasets* analisados durante a pesquisa e complementava os termos do elemento "Assuntos" com um hiperlink apontando para o *website* no qual são disponibilizadas as relações de termos do Vocabulário Controlado do Governo Eletrônico (ex: Unidade básica de saúde [<http://vocab.e.gov.br/2011/03/vcge#unidade-basica-saude>]).

O elemento "Spatial" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 16% dos *datasets* analisados durante a pesquisa e apresentava como valores termos estruturados em formato de coordenadas geográficas (ex: "type": "Polygon", "coordinates": [[[[-74.0046,-33.7411],[-34.7929,-33.7411],[-34.7929,5.27271],[-74.0046,5.27271],[-74.0046,-33.7411]]]]).

O elemento "Atualidade" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 16% dos *datasets* analisados durante a pesquisa e apresentava como valor a data indicando a última atualização do *dataset* (ex: 14-04-2016).

O elemento "Cobertura geográfica" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 8% dos *datasets* analisados durante a pesquisa e apresentava como valor termos identificando a localização do conjunto de dados presente no *dataset* (ex: Brasil).

O elemento "Granularidade Temporal", a partir do *Crosswalking*, pode ser associado ao elemento "abrangência" do e-PMG 1.1 e seu qualificador "abrangência.temporal", de aplicação obrigatória se aplicável. Este metadado foi verificado na descrição de 8% dos *datasets* analisados durante a pesquisa e apresentava como valor termos indicando o nível de detalhamento temporal do conjunto de dados presente no *dataset* (ex: anual).

Na Tabela 1 estão sistematizados os elementos de metadados descritivos identificados nos *datasets* observados no Portal Brasileiro de Dados Abertos, acompanhados pelo grau de frequência com que cada elemento apresentou-se durante o processo de recuperação. Para cada elemento identificado, a tabela aponta o elemento de metadado do e-PMG 1.1 no qual o mesmo pode ser associado, assim como o grau de obrigatoriedade de sua aplicação conforme disposto no guia oficial.

Tabela 1 – Elementos de metadados descritivos do *dataset* e relação com elementos de metadados do e-PMG

Metadados do <i>Dataset</i>	Elementos do e-PMG 1.1	Obrigatoriedade	Presença em <i>datasets</i>
Título do <i>Dataset</i>	título	Obrigatório	100%
Descrição do <i>Dataset</i>	descrição.resumo	Opcional	100%
Palavras-chave	assunto.palavra-chave	Opcional	100%
Autor	criador.autor	Obrigatório	100%
Mantenedor	criador.produtor	Obrigatório	100%
Fonte	fonte	Opcional	96%
Órgão – Poder	Não encontrado	Não encontrado	92%
Órgão – Esfera	Não encontrado	Não encontrado	88%
Granularidade Geográfica	abrangência.espacial	Obrigatório, se aplicável	56%
Frequência de atualização	data.frequênciaDeAtualização	Obrigatório	52%
Assuntos	assunto.categoria	Obrigatório	28%
VCGE	Não encontrado	Não encontrado	28%
Spatial	Não encontrado	Não encontrado	16%
Atualidade	Não encontrado	Não encontrado	16%
Cobertura Geográfica	Não encontrado	Não encontrado	8%
Granularidade Temporal	abrangência.temporal	Obrigatório, se aplicável	8%

Fonte: Os autores.

Observa-se, na Tabela 1, que elementos como "Título", "Descrição", "Palavras-chave", "Autor" e "Mantenedor", constavam em todos os *datasets* analisados. Contudo, elementos como "Fonte", "Órgão – Poder", "Órgão – Esfera", "Granularidade geográfica", "Frequência de atualização", "Assuntos", "VCGE", "Spatial", "Atualidade", "Cobertura geográfica" e "Granularidade temporal", não constavam em todos os *datasets* analisados, mesmo quando definida a sua obrigatoriedade.

Por meio dos *datasets* presentes no Portal Brasileiro de Dados Abertos são disponibilizados recursos informativos contendo arquivos com o conjunto de dados e/ou serviços de visualização. Ao selecionar um recurso presente em determinado *dataset* o usuário é direcionado para uma página contendo outros elementos de metadados descritivos relacionados especificamente ao arquivo contendo os dados, ou seja, ao recurso informativo, e não ao *dataset* no qual o mesmo é disponibilizado.

Na Figura 2, é apresentada a estrutura padrão dos recursos informativos contidos nos *datasets* acessíveis pelo Portal Brasileiro de Dados Abertos. Na área inferior da página, apresentam-se os metadados descritivos do recurso informativo e, no topo da página, localizam-se outros elementos como "Título" e "Descrição" do conjunto de dados. Além da apresentação de elementos de metadados descritivos, na área central, é possível visualizar o conjunto de dados presentes no arquivo.

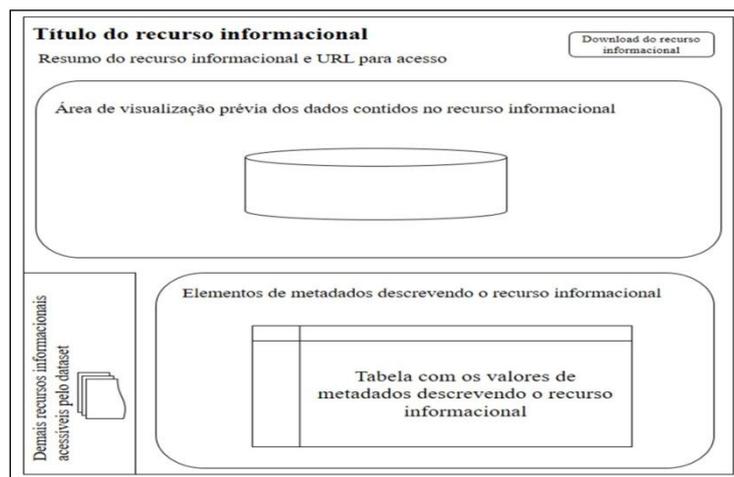
O elemento "Título", a partir do *Crosswalking*, pode ser associado ao elemento "título" do e-PMG 1.1, de aplicação obrigatória. Este metadado foi verificado na descrição de todos os recursos informativos

disponibilizados nos *datasets* analisados e é utilizado para denominar um recurso permitindo sua localização.

O conjunto de texto complementando o "Título" no topo da página e descrevendo o conteúdo dos dados presentes no arquivo de forma resumida, a partir do *Crosswalking*, pode ser associado ao elemento "descrição" do e-PMG 1.1 e seu qualificador "descrição.resumo", de aplicação opcional. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e é utilizado para realizar uma exposição sucinta do conteúdo.

Com relação aos elementos de metadados apresentados na área inferior do recurso informativo, o elemento "Última atualização" não pode ser associado a nenhum elemento e-PMG 1.1. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor a data indicando a última atualização do arquivo contendo o conjunto de dados (ex:24/Setembro/2016).

Figura 2 – Estrutura padrão dos recursos informativos nos *datasets* acessíveis pelo Portal Brasileiro de Dados Abertos



Fonte: Os autores.

O elemento "Criado", a partir do *Crosswalking*, pode ser associado ao elemento "data" do e-PMG 1.1 e seu qualificador "data.criação", de aplicação obrigatória. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor a data referente à data de criação do recurso contendo o conjunto de dados (ex: 24/Setembro/2016).

O elemento "Formato", a partir do *Crosswalking*, pode ser associado ao elemento "formato" do e-PMG 1.1 e seu qualificador "formato.formatoDeArquivo", de aplicação opcional. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor a denominação do formato de arquivo contendo o conjunto de dados (ex: XML, JSON, PDF).

O elemento "Licença" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor termos indicando qual a licença utilizada para acesso ao arquivo contendo o conjunto de dados (ex: Licença Aberta para Bases de Dados do Open Data Commons).

O elemento "format" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor o mesmo resultado do elemento "Formato", referindo-se à denominação do formato de arquivo contendo o conjunto de dados (ex: XML, JSON, PDF).

O elemento "id", a partir do *Crosswalking*, pode ser associado ao elemento "identificador" do e-PMG 1.1 e seu qualificador "identificador.idDoSistema", de aplicação obrigatória. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor um conjunto de caracteres gerados automaticamente pelo sistema para identificação unívoca do recurso (ex: 91b8af2e-07a1-4952-932d-13d941d82733).

O elemento "mime type" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor o mesmo valor dos elementos "Formato" e "format", indicando o formato de arquivo contendo o conjunto de dados (ex: XML, JSON, PDF).

O elemento "resource group id", a partir do *Crosswalking*, pode ser associado ao elemento "identificador" do e-PMG 1.1 e seu qualificador "identificador.códigoDeClassificação", de aplicação opcional. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor um conjunto de caracteres atribuídos pelo sistema possivelmente para identificação do recurso em um plano de classificação específica para grupos (ex: e7e5e1e1-b35f-4525-95cd-069df00222e5).

O elemento "revision id" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor um conjunto de caracteres atribuídos pelo sistema para identificação do recurso (ex: b5f53d60-5397-4b67-af09-25f74e1d4767).

O elemento "state" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de todos os recursos informativos disponibilizados nos *datasets* analisados durante a pesquisa e apresentava como valor uma indicação da condição de acesso ao recurso (ex: *active*, *inactive*).

O elemento "position" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 92% dos

recursos informativos disponibilizados nos *datasets* analisados durante a pesquisa e apresentava como valor um número inteiro indicando referência sequencial da ordem dos arquivos acessíveis em determinado *dataset* (ex: *position 2*).

O elemento "*can be previewed*" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 84% dos recursos informativos disponibilizados nos *datasets* analisados durante a pesquisa e apresentava como valor um número booleano indicando a possibilidade de pré-visualizar o recurso ou não (ex: 0 ou 1).

O elemento "*created*" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 84% dos recursos informativos disponibilizados nos *datasets* analisados e apresentava como valor a quantidade de dias decorrentes da data de criação do arquivo contendo o conjunto de dados (ex: há mais de um ano, 20 dias).

O elemento "*webstore url*", a partir do *Crosswalking*, pode ser associado ao elemento "localização" do e-PMG 1.1, de aplicação opcional. Este metadado foi verificado na descrição de 56% dos recursos informativos disponibilizados nos *datasets* analisados e apresenta como valor um endereço eletrônico apontando para o local de armazenamento do recurso.

O elemento "*mime type inner*" não pode ser associado a nenhum elemento do e-PMG 1.1. Este metadado foi verificado na descrição de 24% dos recursos informativos disponibilizados nos *datasets* analisados. Assim como os elementos "Formato", "*format*" e "*mime type*", este metadado também apresenta como valor a denominação referente ao formato de arquivo contendo o conjunto de dados (XML, JSON, PDF).

O elemento "*resource type*", a partir do *Crosswalking*, pode ser associado ao elemento "tipo de recurso" do e-PMG 1.1, de aplicação obrigatória se aplicável. Este metadado foi verificado na descrição de 24% dos recursos informativos disponibilizados nos *datasets* analisados e apresenta como valor a natureza do tipo de conjunto de dados (ex: *file*, *map*).

O elemento "*size*", a partir do *Crosswalking*, pode ser associado ao elemento "formato" do e-PMG 1.1 e seu qualificador "formato.dimensão", de aplicação opcional. Este metadado foi verificado na descrição de 4% dos recursos informativos disponibilizados nos *datasets* analisados e como valor recebia indicação de tamanho físico ou capacidade do recurso (ex: 105 *bytes*).

Tabela 2 – Metadados descritivos dos recursos informativos acessíveis
pelos datasets e relação com elemento do e-PMG

Metadados dos recursos informativos	Elementos do e-PMG 1.1	Obrigatoriedade	Presença em datasets
Título	título	Obrigatório	100%
Resumo	descrição.resumo	Opcional	100%
Última atualização	Não encontrado	Não encontrado	100%
Criado	data.criação	Obrigatório	100%
Formato	formato.formatoDeArquivo	Opcional	100%
Licença	Não encontrado	Não encontrado	100%
format	Não encontrado	Não encontrado	100%
Id	identificador.idDoSistema	Obrigatório	100%
mimetype	Não encontrado	Não encontrado	100%
resource group id	identificador.códigoDeClassificação	Opcional	100%
revision id	Não encontrado	Não encontrado	100%
state	Não encontrado	Não encontrado	100%
position	Não encontrado	Não encontrado	92%
can be previewed	Não encontrado	Não encontrado	84%
created	Não encontrado	Não encontrado	84%
Webstore url	localização	Opcional	56%
mimetype inner	Não encontrado	Não encontrado	24%
resource type	tipo de recurso	Obrigatório, se aplicável	24%
size	formato.dimensão	Opcional	4%

Fonte: Os autores.

Observa-se na Tabela 2 que elementos como "Título", "Resumo do conjunto de dados", "URL", "Última atualização", "Criado", "Formato", "Licença", "format", "id", "mime type", "resource group id", "revision id" e "state", constavam em todos os recursos informativos acessíveis nos datasets. Contudo, elementos como "position", "can be previewed", "created", "webstore url", "mime type inner", "resource type" e "size" não constavam em todos os recursos informativos acessíveis pelos datasets.

Alguns metadados identificados, que descreviam os recursos informativos disponibilizados nos datasets analisados, ao serem associados com elementos do e-PMG 1.1 podem ser considerados ambíguos, como nos casos de "Criado" e "created", onde ambos os metadados referem-se à data de criação do arquivo, diferenciando-se, apenas, na forma de representação dos valores. Outro caso que também pode gerar ambiguidade refere-se aos elementos "Formato", "format", "mime type" e "mime type inner", que também recebiam o mesmo valor referente à descrição do formato de arquivo.

4 Considerações finais

O uso da Web pode ampliar os horizontes do compartilhamento de informações governamentais, viabilizando maior interação entre administração pública (no fornecimento de serviços mais eficientes) e

sociedade (na busca de conteúdos para participação e controle das ações do Governo).

Esta pesquisa descreve o cenário da disponibilização de dados governamentais na *Web*, analisando a descrição dos *datasets* e recursos informativos acessíveis pelo Portal Brasileiro de Dados Abertos, sendo possível verificar a frequência com que os elementos de metadados eram apresentados durante o processo de recuperação.

Por meio de um *Crosswalk* abrangendo os elementos de metadados identificados nos *datasets* e nos recursos informativos disponibilizados pelo Portal Brasileiro de Dados Abertos, e associando-os aos elementos de metadados do e-PMG 1.1, foi possível apontar relações de similaridade e diferença entre eles.

Alguns pontos relacionados à apresentação dos elementos de metadados descrevendo os *datasets* e os recursos informativos analisados merecem ser destacados, como a inconsistência na apresentação de elementos que puderam ser associados a metadados do e-PMG 1.1 de aplicabilidade obrigatória ("Frequência de atualização" e "Assuntos" na descrição dos *datasets*) e obrigatória se aplicável ("Granularidade Geográfica" e "Granularidade Temporal" na descrição dos *datasets* e "*resource type*" na descrição dos recursos informativos).

Com base nestes pontos apresentados, questiona-se o porquê da não utilização do e-PMG 1.1 de forma completa no Portal Brasileiro de Dados Abertos, uma vez que este padrão é específico para a descrição de recursos governamentais. Entende-se que deve haver maiores esforços por parte das organizações governamentais na consolidação da aplicação deste padrão, visto que a padronização da representação dos recursos informativos impacta diretamente no processo de recuperação da informação em ambientes abertos.

Referências

ALVES, R. C. V. *Web semântica: uma análise focada no uso de metadados*. 2005. 180f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005.

ANDERSEN, K. V.; HENRIKISEN. H. Z. E-government maturity models: extension of the Layne and Lee model. *Government Information Quarterly*, v. 23, n. 1, p. 236-248, 2006. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0740624X05000973>>. Acesso em: 20 fev. 2015.

BACA, M. *Introduction to metadata*. 2 ed. Los Angeles: Getty Research Institute, 2008.

BRASIL. Comitê Executivo de Governo Eletrônico. *Padrão de metadados do governo eletrônico: e-PMG*. 2009. Disponível em: <<http://www.governoeletronico.gov.br/documentos-e-arquivos/e-PMG%20V0.pdf>>. Acesso em: 24 abr. 2016.

BRASIL. Presidência da República. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. *DOU*, Brasília, ed. extra, 18 nov. 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm>. Acesso em: 24 abr. 2016.

BRASIL. *Sobre o dados.gov.br*. Governo Federal, 2016. Disponível em: <<http://dados.gov.br/sobre/>>. Acesso em: 24 abr. 2016.

BRASIL. Ministério da Transparência; Controladoria Geral da União. *Portal da transparência*. 2017a. Disponível em: <<http://transparencia.gov.br>>. Acesso em: 1 maio 2017.

BRASIL. Secretaria de Tecnologia da Informação; Ministério do Planejamento, Desenvolvimento e Gestão. *Portal brasileiro de dados abertos*. 2017b. Disponível em: <<http://dados.gov.br>>. Acesso em: 1 maio 2017.

BRASIL. *Repositório de vocabulários e ontologias do governo eletrônico*. 2017c. Disponível em: <<http://vocab.e.gov.br/>>. Acesso em: 1 maio 2017.

COMITÊ EXECUTIVO DO GOVERNO ELETRÔNICO (CEGE). *Padrão de metadados do Governo Eletrônico (e-PMG 1.1)*. CEGE, 2014. Disponível em: <http://www.governoeletronico.gov.br/biblioteca/arquivos/e-PMG_1.1_versao_1_1>. Acesso em: 24 abr. 2016.

CONTROLADORIA GERAL DA UNIÃO (CGU). *Esclarecimentos sobre atuação e funcionamento da Parceria para Governo Aberto (OGP) no Brasil*. CGU, 2016. Disponível em: <<http://www.governoaberto.cgu.gov.br/no-brasil/perguntas-frequentes>>. Acesso em: 24 abr. 2016.

FERNANDES, A. M. R. *Inteligência artificial: noções gerais*. Florianópolis: VisualBooks, 2005.

FOULONNEAU, M.; RILEY, J. Choosing metadata standards for a digital library project. In: FOULONNEAU, M.; RILEY, J. *Metadata for digital resources: implementation, systems design and interoperability*. Oxford: Chandos, 2008. p. 13-28.

HAWKINS, D.T. Information science abstracts: tracking the literature of information science. Part 1: definition and map. *Journal of the American Society for Information Science and Technology*, v. 52, p. 44-54. 2001. Disponível em: <[http://onlinelibrary.wiley.com/doi/10.1002/1532-2890\(2000\)52:1%3C44::AID-ASI1057%3E3.0.CO;2-6/epdf](http://onlinelibrary.wiley.com/doi/10.1002/1532-2890(2000)52:1%3C44::AID-ASI1057%3E3.0.CO;2-6/epdf)>. Acesso em: 4 abr. 2016.

LEVY, A. Y. Logic-based techniques in data integration. In: MINKER, J. (Ed.) *Logic-based artificial intelligence*. Norwell, MA: Kluwer Academic Publisher, 2000. p. 575-595. Disponível em:

<<http://rakaposhi.eas.asu.edu/cse494/notes/levy-di00.pdf>>. Acesso em: 14 abr. 2016.

MARTINES USERO, J. A. *Nuevas tecnologías para nuevas bibliotecas: desarrollo de servicios de información electrónica*. Buenos Aires: Alfagrama Ediciones, 2007.

MEY, E. S. A. *Introdução a catalogação*. Brasília: Briquet de Lemos, 1995.

MOURA, A. M. de C.; FERNANDES, C. S. A Metadata approach to represent and visualize sites on the web. In: INTERNATIONAL WORKSHOP ON INFORMATION INTEGRATION TECHNOLOGIES ON THE WEB, n.1, Itaipava, 2001. *Anais eletrônicos...* Itaipava, RJ, 9-11 abril 2001. Disponível em: <<http://www.lbd.dcc.ufmg.br/bdbcomp/servlet/Trabalho?id=1649>>. Acesso em: 20 abr. 2016.

PLATT NETO, O. A. P.; CRUZ, F. da; VIEIRA, A. L. Transparência das contas públicas: um enfoque no uso da internet como instrumento de publicidade na UFSC. *Revista Contemporânea de Contabilidade*, v. 1, n. 5, p. 135-146, 2006. Disponível em: <<https://periodicos.ufsc.br/index.php/contabilidade/article/view/1143>>. Acesso em: 6 abr. 2016.

SANT'ANA, R. C. G. *Tecnologia e gestão pública municipal: mensuração da interação com a sociedade*. São Paulo: Cultura Acadêmica Editora - UNESP, 2009. v. 1.

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: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/284/319>>. Acesso em: 30 mar. 2015.

SANTOS, P. L. A. da C.; SANT'ANA, R. C. G. Transferência da Informação: análise para valoração de unidades de conhecimento. *DataGramaZero*, v. 3, n. 2, abr., 2002. Disponível em: <http://www.dgz.org.br/abr02/Art_02.htm>. Acesso em: 6 mar. 2015.

SANTOS, P. L. V. A. C.; CORRÊA, R. M. R. *Catalogação: trajetória para um código internacional*. Niterói: Intertexto, 2009.

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.