



Universidade Estadual Paulista “Júlio de Mesquita Filho”
Faculdade de Filosofia e Ciências
Campus de Marília

JACQUELIN TERESA CAMPEROS REYES

METADADOS NAS INSTRUÇÕES DE GOVERNOS PARA PUBLICADORES DE DADOS

**Marília
2018**



Universidade Estadual Paulista “Júlio de Mesquita Filho”
Faculdade de Filosofia e Ciências
Campus de Marília

JACQUELIN TERESA CAMPEROS REYES

METADADOS NAS INSTRUÇÕES DE GOVERNOS PARA PUBLICADORES DE DADOS

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Informação do Departamento de Ciência da Informação da Universidade Estadual Paulista Faculdade de Filosofia e Ciências como requisito para a obtenção do título de Mestre em Ciência da Informação.

Linha de Pesquisa: Informação e Tecnologia

Orientador: Dr. Ricardo César Gonçalves Sant’Ana

**Marília
2018**

Camperos Reyes, Jacquelin Teresa.
C195m Metadados nas instruções de governos para publicadores de
dados / Jacquelin Teresa Camperos Reyes. – Marília, 2018.
102 f. ; 30 cm.

Orientador: Ricardo César Gonçalves Sant'Ana.
Dissertação (Mestrado em Ciência da Informação) – Universidade
Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, 2018.
Bibliografia: f. 96-102

1. Transparência na administração pública. 2. Metadados - Brasil.
3. Metadados - Colômbia. 4. Metadados – Portugal. 5. Metadados –
Espanha. 6. Tecnologia da informação. I. Título.

CDD 005.72

Ficha catalográfica elaborada por
André Sávio Craveiro Bueno
Bibliotecário
CRB 8/8211

ATESTADO DE APROVAÇÃO - DEFESA

Atestamos que **JACQUELIN TERESA CAMPEROS REYES**, RA nº: PCI160954, RNE nº RNE J232231X, defendeu, no dia 29/01/2018, a dissertação intitulada **METADADOS NAS INSTRUÇÕES DE GOVERNOS PARA PUBLICADORES DE DADOS**, junto ao Programa de Pós Graduação em CIÊNCIA DA INFORMAÇÃO, Curso de Mestrado Acadêmico, tendo sido 'APROVADA'.

Atestamos ainda que a obtenção do título dependerá de homologação pelo Órgão Colegiado competente.

Marília, 29 de janeiro de 2018



Profa. Dra. MARTA LÍGIA POMIM VALENTIM
Coordenadora do PPG em Ciência da Informação

JACQUELIN TERESA CAMPEROS REYES

METADADOS NAS INSTRUÇÕES DE GOVERNOS PARA PUBLICADORES DE DADOS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Informação.

Linha de Pesquisa: Informação e Tecnologia.

Data de defesa: 29/01/2018

Banca Examinadora:

Prof. Dr. Ricardo César Gonçalves Sant'Ana
(UNESP/Marília)
Livre-docente em Sistemas de Informação Gerenciais
(orientador)

Profa. Dra. Plácida L.V. A. da Costa Santos (UNESP/Marília)
Doutorado em Letras (Linguística e Semiótica)

Prof. Dr. Cristian Berrío Zapata (UFPA/Belém)
Doutorado em Ciência da Informação

Realmente, con mi humanidad, utilizando el absolutismo del NUNCA, nunca, hubiese siquiera comenzado este desafío personal y académico.

Por eso, esta conquista la dedico al ser que me brindó la posibilidad de hacerlo: Dios. El Ser que Es, que no tiene y no va tener explicación científica, y que permite que sus hijos humanos se debrucen en otra hija, la ciencia.

AGRADECIMENTOS

A César, mi esposo, porque su sí ante este sueño académico y ante tantos otros, me ayudan a entender un significado más del amor.

A mis queridos padres, Rafael y Mery, por siempre creer en mí y por ser mis modelos de fortaleza, tenacidad y honestidad.

A María Juliana, mi hija, por enseñarme sobre valentía y portugués.

A Juan Diego, mi hijo, por mostrarme que la esperanza ayuda a nacer las ideas y por recordarme la importancia de sonreír.

A mis queridos hermanos, Freddy, Nataly y Jhonatan, porque recibí su amor en la distancia y con sus alegres conversas, sin ellos saberlo me animaban. Y gracias por las fotos semanales de Salomé, Angel, Massimiliano, Adrian e Isaac. Ellas extendían mi batería diaria.

A mis suegros, Luis Eduardo y Emperatriz, porque confiaron en mí, y me enseñaron que los sacrificios con amor saben mejor.

A Doris, cuñada, por apoyarme confiando sin dudar, y con sus visitas a Brasil.

A Juan Carlos y a Edgar Eduardo 'Kini', por animarme con cariño en la certeza que da la academia.

A mis cuñados Alexa, Sonia, Gloria, Julián, Edgar y Erika, por contribuir con pequeñas dichas que nacen en sus lindas familias.

A las tías Nancy Reyes, Celina Bustos, Josefa Bustos y Stella Angarita porque no dejaron un día de orar por todos nosotros.

A Alonso Yáñez y Janeth Castro por animarnos y enseñarnos a tener calma ante todos los desafíos que estábamos viviendo.

Al profesor Ricardo Sant'Ana por interesarse en mi proyecto de investigación, por mantener la permanentemente reflexión tanto académica como ética, y, por obligarme a escribir en portugués desde el comienzo de esta etapa.

A la profesora Plácida Santos por inspirarme no solamente el desarrollo de mis capacidades académicas y profesionales, además, por dar aulas tácitas sobre fortaleza y confianza en el Señor Dios.

A la profesora Raquel Alves, por abrazar la montaña de incertezas que percibí al comenzar la maestría.

A la profesora Silvana Vidotti por mostrarme con nobleza el significado de ser acogedor y afectuoso.

A la profesora Ângela Grossi de Carvalho, porque de ella escuché por primera vez en aula, que los alumnos de un programa de post-graduación somos ante todo humanos, y por ayudarme en la reflexión que definió mis intereses de investigación.

A Cristian Berrio, por orientarme en mi llegada a la Unesp, y por los aportes al desarrollo de esta investigación.

Al profesor Rogério Ramalho por sus pertinentes observaciones en la banca de cualificación de esta pesquisa.

A los profesores Zaira Zafalón, Eduardo Santarém, Fernando Vechiatto, Henry Poncio, Daniel Martínez Ávila y José Augusto Chaves Guimarães, por ayudarme a descubrir poco a poco cada camino posible de pesquisa.

A Sandra Roa y Kamilo Garreta por tener fe en nosotros, confiar en lo desconocido, abrir su hogar, y después, acompañarnos a “matar saudades” de comidas colombianas.

A los amigos Caio Coneglian, Ana Maria Ferreira, Edgar Bisset y Paloma Arraiza por ser las primeras personas que me motivaron a responder con certeza que sí estaba “Gostando do Brasil”, y por ayudarme a recordar que en la vida universitaria se sufre, pero también se goza.

A Diana Aleixo, Elaine Affonso, Elizabete Monteiro y Cristina Nagai, por entender con total precisión lo que estaba pasando por mi cabeza cuando pedía auxilio y me sentía caer en los intentos. Sufrimos juntas el dolor de la academia, pero el amor recibido en familia y con los amigos, sustentó y permitió que la perseverancia no se cansara.

A Felipe Arakaki, por escuchar y entender mis primeros lamentos académicos, y por abrir tanto su amabilidad como los archivos que contenían los textos que estaba precisando conocer.

A Maira Nani, por tener un corazón tan amplio y fuerte que se sale a veces del pecho y hace mover las manos para comenzar a cocinar con amor y ternura.

A Luiza Romanetto, Juliano Ferreira, Ilidio Lobato y Luiz Galeffi por nuestras conversas sobre aspectos más importantes que los que nos asaltaban por las noches de la vida académica: el amor, el cine, la música y las mascotas.

A Liliana Serra, por abrir su corazón y enseñarme que así pasemos por situaciones que sean difíciles, la vida es bella, y vale la pena luchar.

A todos los demás familiares y amigos que desde Colombia torcieron para que esta etapa fuera provechosa para nosotros como de hecho lo está siendo.

A la OEA por el financiamiento de esta pesquisa.

A las comunidades del Camino Neocatecumenal, tanto de Colombia como de Brasil que nos acogieron con tanto cariño y colaboración. Nos han mostrado tantos rostros de Dios que no sabíamos que era posible conocer.

A María de Nazareth, mi maestra del sacrificio silencioso.

A Jesús de Nazareth por ser y estar, por, a pesar de lo que soy, mediar mi salvación, por enseñarme acerca de la nobleza de corazón, y por aumentar mi fe en su Divino Padre.

A Dios, por permitir todo lo anterior.

“La tarea de un hombre [...] es no aflojar el propio ánimo en la consideración de las cosas celestiales y transitar sin inquietarse entre las múltiples preocupaciones, no como un tonto, sino con la prerrogativa de un alma libre que no tiene afecto desordenado a ninguna criatura”. Tomás de Kempis

RESUMO

Gerar valor para a sociedade a partir da abundância de dados governamentais tornou-se imperativo nas estratégias de disponibilização de dados que estão sendo publicados por meio de conjuntos de dados ou datasets. Os datasets, dados tabulados com certa estrutura, constituem um exemplo de reunião de bases de dados que pretendem obter sucesso erguendo-se como catálogos centrais do ponto de vista dos cidadãos, ampliando a visibilidade sobre e das ações da gestão pública. Atingir a estruturação desses recursos informacionais, de forma que auxilie na sua revalorização, é um dos desafios da Ciência da Informação. A questão de investigação é: Como está sendo abordada a aderência ao uso de metadados nas instruções entregues aos publicadores de dados em governos? O objetivo é descrever a aderência ao uso de metadados nos datasets de governos de países, tomando como base o contexto e o marco conceitual apresentados nas instruções para publicadores de dados, encontradas nos sites de dados abertos oficiais dos países analisados. Acredita-se que estudos como este podem fornecer elementos que atuem como subsídios para as estratégias governamentais, atendendo dimensões sociais a partir dos profissionais da informação. Trata-se de pesquisa descritiva, de natureza qualitativa, focada na observação crítica dos documentos que abordam o tratamento descritivo dos datasets governamentais. Utilizam-se como procedimentos a análise bibliográfica e documental, e a definição de estudos de caso nos países Colômbia, Brasil, Espanha e Portugal, abordando o volume de dados e informações mediante a técnica de análise de conteúdo. Percebe-se o esforço realizado pelos detentores dos documentos disponibilizados nos quatro países analisados pela ampla abordagem de conteúdo temático relacionado com o uso experimental dos metadados, dando assim maior importância ao aspecto prático em relação ao teórico, sem desconsiderar a relevância das explicações teóricas. Acredita-se na importância da criação e implementação de perfis de aplicação entre comunidades de países, como o caso do DCAT-AP, criado e recomendado pela comunidade europeia de nações e sugerido pelos sites de dados dos países estudados. Admitem-se inquietações referentes aos processos de publicação de dados de governo e suas relações com outros tópicos de interesse socioeconômico, tais como possíveis vínculos com indicadores de desenvolvimento em países e regiões, sob o prisma de pesquisas originadas a partir da Ciência da Informação.

Palavras-chave: Metadados. Representação de dados. Dados governo. Instruções para publicação de dados. Publicadores de dados. *Datasets*. Colômbia. Brasil. Espanha. Portugal.

ABSTRACT

Generating value to society from the abundance of government data, has become imperative in the strategies of data availability that are being published through datasets. The datasets, tabulated data with a certain structure, are an example of a meeting of databases that aim to be successful setting up as central catalogs from the point of view of the citizens, increasing the visibility on and of the actions of the public management. Accomplishing the structuring of these informational resources, so that this helps in their reevaluation, is one of the challenges of Information Science. The research question is: How is the adherence to the use of metadata in the instructions given to data publishers in South American governments being addressed? The objective is to describe the adherence to the use of metadata in the datasets of governments of South American countries, based on the context and conceptual framework presented in the instructions for data publishers, found on the official open data sites of the analyzed countries. It is believed that studies such as this one can provide elements that act as subsidies for government strategies, addressing social dimensions deriving out of the information professionals. It is a descriptive research, of qualitative nature, focused on the critical observation of the documents that approach the descriptive treatment of the governmental datasets. Bibliographic and documentary analyses are used as methodological procedures, and the definition of case studies in the countries Colombia, Brazil, Spain and Portugal, addressing the volume of data and information through the technique of content analysis. The effort made by the holders of the documents available in the four analyzed countries by the broad thematic content related to the experimental use of the metadata is noticed, thus giving greater importance to the practical aspect in relation to the theoretical, without disregarding the relevance of the theoretical explanations. It is believed that creating and implementing application profiles among communities of countries, such as the DCAT-AP, created and recommended by the European community of nations and suggested by the data sites of the analyzed countries, is important. Concerns referent to the publication processes of government data and their relations with other topics of socio-economic interest are admitted, such as possible linkages with indicators of development in countries and regions, under the prism of research originated from the Information Science.

Keywords: Metadata. Data representation. Government data. Instructions for data publishing. Data publishers. Datasets. Colombia. Brazil. Spain. Portugal.

LISTA DE QUADROS

Quadro 1 – Sites de dados como pontos de entrada dos estudos de caso	30
Quadro 2 – Caracterização do material da análise	35
Quadro 3 – Síntese da técnica a aplicar na pesquisa	36
Quadro 4 – Caracterização do corpus da pesquisa.....	68
Quadro 5 – Resumo de intensidades nas categorias.....	90

LISTA DE FIGURAS

Figura 1 – Características de importância no acesso a dados publicados por governos	21
Figura 2 – Elementos do contexto e foco da pesquisa.....	23
Figura 3 – Esquema geral da pesquisa.....	26
Figura 4 – Fluxograma descoberta de recursos na fase de pré-análise	29
Figura 5 – OURDATA Index do estudo	39
Figura 6 – Esquema gráfico básico do Dublin Core	53
Figura 7 – Modelo de dados do DCAT (Colômbia)	56
Figura 8 – Dados gerais da coleta	64
Figura 9 – Peso por país no total	65
Figura 10 – Gênese de cada site de dados abertos de governo	66
Figura 11 – Apresentação sintética da caracterização do corpus da pesquisa Colômbia	74
Figura 12 – Apresentação sintética da caracterização do corpus da pesquisa Brasil.....	81
Figura 13 – Apresentação sintética da caracterização do corpus da pesquisa Espanha	84
Figura 14 – Modelo de dados do DCAT (Espanha)	85
Figura 15 – Apresentação sintética da caracterização do corpus da pesquisa Portugal	89
Figura 16 – Estados na evolução do modelo de metadados	89
Figura 17 – Ocorrências das categorias com os níveis de intensidade	91

LISTA DE ABREVIATURAS E SIGLAS

ABRAJI	Associação Brasileira de Jornalismo Investigativo
BID	Banco Interamericano de Desenvolvimento
CERN	<i>European Organization for Nuclear Research</i>
CKAN	<i>Comprehensive Knowledge Archive Network</i>
CPCSAE	<i>Comisión Permanente del Consejo Superior de Administración Electrónica</i>
CSAE	<i>Consejo Superior de Administración Electrónica</i>
DANE	<i>Departamento Administrativo Nacional de Estadística</i>
DARPA	<i>Defense Advanced Research Project Agency</i>
DC	<i>Dublin Core</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCAT-AP	<i>Data Catalog Vocabulary Application Profile</i>
DCAP	<i>Dublin Core Application Profile</i>
DCMES	<i>Dublin Core Metadata Element Set</i>
DCMI	<i>Dublin Core Metadata Initiatives</i>
DCMI-TERMS	<i>Metadata Terms</i>
DERI	<i>Digital Enterprise Research Institute</i>
EPMG	Padrão de Metadados para o Governo Eletrônico
FOAF	<i>Friend of a Friend</i>
GLD	<i>Government Linked Data</i>
GODAN	<i>Global Open Data for Agriculture & Nutrition</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IODC	<i>International Open Data Conference</i>
MINTIC	<i>Ministerio de Tecnologías de la Información y las Comunicaciones</i>
MNU	Metadados No Uso
MT	Metadados Teoria
NISO	<i>National Information Standards Organization</i>
OCDE	Organização para a Cooperação e Desenvolvimento Económico
OGP	<i>Open Government Partnership</i>
OKFN	<i>Open Knowledge Foundation</i>
OKI	<i>Open Knowledge International</i>

RDF	<i>Resource Description Framework</i>
SKOS	<i>Simple Knowledge Organization System</i>
TIC	Tecnologias da Informação e Comunicação
VCGE	Vocabulário Controlado do Governo Eletrônico
W3C	<i>World Wide Web Consortium</i>

SUMÁRIO

1 INTRODUÇÃO	19
1.1 Objetivos da dissertação.....	22
1.1.1 Objetivo geral	22
1.1.2 Objetivos específicos	22
1.2 Justificativa	24
1.3 Metodologia	25
1.3.1 Fase de pré-análise.....	28
1.3.2 Fase da exploração do material.....	31
1.3.3 Fase de tratamento dos resultados.....	36
1.3.4 Síntese da técnica a ser aplicada.....	36
1.4 Por que a escolha da Colômbia como pré-teste?	37
1.5 Colômbia, Brasil, Espanha e Portugal	38
2 DEMARCAÇÃO DA REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS	41
2.1 Dado e seu peso na representação	43
3 DADOS ABERTOS NO ÂMBITO GOVERNAMENTAL	45
3.1 Iniciativas estratégicas para o fomento da abertura de dados governamentais	47
3.2 Plataformas como opções pró-implementação de sites de dados abertos governamentais	49
4 MECANISMOS DE REPRESENTAÇÃO NO CONTEXTO WEB	52
4.1 Dublin Core.....	52
4.2 O Friend of a Friend	54
4.3 Simple Knowledge Organization System	54
4.4 Data Catalog Vocabulary	55
4.5 GEONAMES	56
4.6 Desenvolvimento de perfil de aplicação	57
5 RELATO DOS ESTUDOS DE CASO	60
5.1 Resultados da fase pré-análise	60
5.1.1 Colômbia.....	60

5.1.2 Brasil.....	61
5.1.3 Espanha	62
5.1.4 Portugal	63
5.1.5 Dados gerais na coleta do corpus	63
5.2 Resultados da exploração do material	66
5.3 Tratamento dos resultados	73
5.3.1 Colômbia.....	73
5.3.2 Brasil.....	80
5.3.3 Espanha	83
5.3.4 Portugal	88
5.3.5 Inferências no grupo de países	90
6 CONSIDERAÇÕES	93
REFERÊNCIAS	96

1 INTRODUÇÃO

Gerar valor para a sociedade a partir da abundância de dados, neste caso de estudo, dados governamentais, tornou-se imperativo nas estratégias de disponibilização de dados que estão sendo publicados por meio de conjuntos de dados ou *datasets*.

Um dos desafios da Ciência da Informação é transformar essa abundância de dados de modo que permitam além da tomada de decisões, acrescentar valor nas atividades de um determinado negócio (SANTOS, 2010). Assim faz-se necessário considerar a estruturação desses recursos informacionais, dados de governo, como um processo basilar na busca da ampliação e eficiência no acesso e uso desses dados.

McGarry (1999) assevera que os recursos informacionais devem ser estruturados de alguma forma para evitar o seu amorfismo e sua inutilidade. Para ele “A informação deve ser representada para nós de alguma forma e transmitida por algum tipo de canal” (McGARRY, 1999, p. 11).

Para Sant’Ana (2013) diante da demanda do excesso de oferta nos fluxos informacionais digitais, a tomada de decisão no desenvolvimento de meios inovadores de representação desses conteúdos, contribuem com a recuperação e acesso a esses recursos informacionais, coadjuvando na condição primária de convivência e de sobrevivência do fortalecimento das capacidades humano-informacionais.

Os *datasets* são dados tabulados com certa estrutura. Para a comunidade Schema.org, conhecida como uma atividade colaborativa que visa promover, criar e manter esquemas para dados na internet, os *datasets* são uma coleção de recursos informacionais estruturados que “[...] descrevem um ou vários temas de interesse” (SCHEMA, 2017, tradução nossa).

O caso dos *datasets* governamentais constitui um exemplo de reunião de bases de dados que pretendem obter sucesso estruturando-se como catálogos centrais a partir do ponto de vista dos cidadãos, ampliando a visibilidade sobre e das ações da gestão pública, e como “[...] uma forma inteligente de medir e monitorar a saúde das bases públicas disponíveis para a sociedade” (SÃO PAULO, 2015, p. 43).

Conforme Sant’Ana (2009) indica, as Tecnologias da Informação e Comunicação (TIC) podem ser observadas a partir de dois pontos segundo seu uso nas organizações públicas: para uso interno e para uso externo. O uso interno, faz referência ao uso das TIC em processos internos da administração pública e suas organizações. O uso externo

corresponde à utilização das TIC como apoio das relações estabelecidas entre essas organizações e outras chamadas externas, que podem ser individuais (pessoas ou organizações), ou em grupos (indivíduos ou organizações).

Assim, os *datasets*, objeto deste estudo, encaixam-se no contexto do uso externo das TIC, por corresponderem em conjuntos de dados que estão sendo visíveis à sociedade geral, não incluindo aqueles conjuntos de dados com dados internos dos governos, ou das instâncias deles.

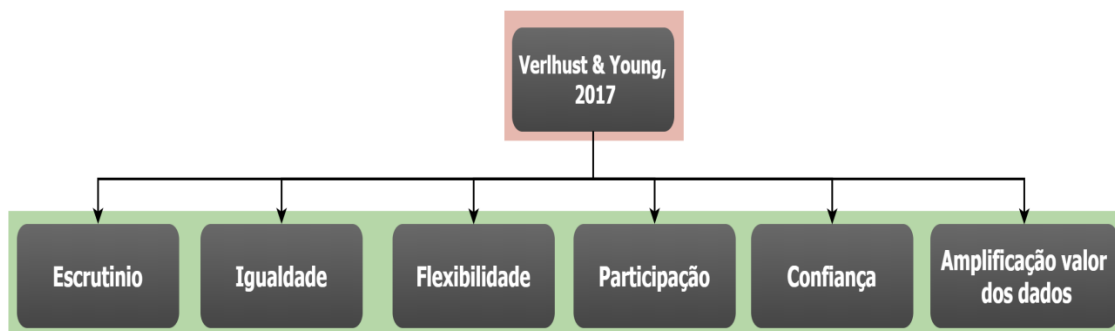
Considerando a representação da informação, quanto estruturação de recursos informacionais, Santos (2017) assevera que a correta representação que possibilita e simplifica a busca e a recuperação desses recursos informacionais, configura-se "[...] como uma das especificidades do fazer da ciência da informação e uma atividade imprescindível no armazenamento, no acesso e na recuperação de informações e de recursos" (SANTOS, 2017, p. 2).

Segundo Santos (2010), uma das características potenciais do ciberespaço, que aglomera ambientes informacionais dinâmicos, é a revalorização dos metadados para a construção de um saber coletivo na representação dos recursos informacionais. Encontram-se nesse contexto os *datasets* que estão, assim, sendo focados com o intuito de aproveitar ao máximo seus conteúdos, isto é, dados disponibilizados por governos.

Interessados em países com economias em desenvolvimento, Verhulst e Young (2017, destaque nosso) elencam seis características como distintivas, que demarcam a importância no acesso e uso de dados publicados a partir de governos, ilustrados ainda na Figura 1:

- a) possibilidade de **escrutínio** pela sociedade;
- b) **igualdade** na distribuição da informação;
- c) **flexibilidade** nos formatos de apresentação para a posterior utilização;
- d) amplo alcance de **participação**, tanto interna quanto na tomada de experiências fora do contexto regional;
- e) aumento da **confiança** por parte do grupo social representado pelo governo; e
- f) **amplificação do valor** dos dados.

Figura 1 – Características de importância no acesso a dados publicados por governos



Fonte: Adaptado de Verlhust e Young (2017).

Para eles esses seis elementos têm uma qualidade similar de importância no concernente a economias em desenvolvimento, por isso a ilustração em um mesmo nível hierárquico.

Nesse contexto, esta dissertação centra a atenção na seguinte inquietação: Como está sendo abordada a aderência ao uso de metadados nas instruções entregues a publicadores de dados em governos?

Segundo Saracevic (1996) o imperativo tecnológico da Ciência da Informação, faz com que ela seja partícipe da transformação da sociedade moderna em sociedade da informação, considerando sua forte dimensão social e humana, a qual ultrapassa a tecnologia.

Acredita-se que estudos como este, emergidos a partir da Ciência da Informação, focados no acompanhamento de processos de publicação de dados, podem fornecer elementos que atuem como subsídios para as estratégias governamentais, atendendo as dimensões indicadas pelo autor supracitado, a partir de uma interseção que pode ser proveitosa tanto para a comunidade científica, quanto para os órgãos oficiais responsáveis pela publicação de dados.

Espera-se contribuir então, no fortalecimento da Ciência da Informação, como asseverado por Borko (1968), ao investigar as propriedades e fluxos da informação, relacionadas aos dados governamentais, junto às forças que governam esses fluxos, emergidas a partir dos mesmos governos, buscando otimizar o armazenamento, recuperação e disseminação da informação.

1.1 Objetivos da dissertação

Nesta seção são apresentados o objetivo geral e os objetivos específicos da pesquisa.

1.1.1 Objetivo geral

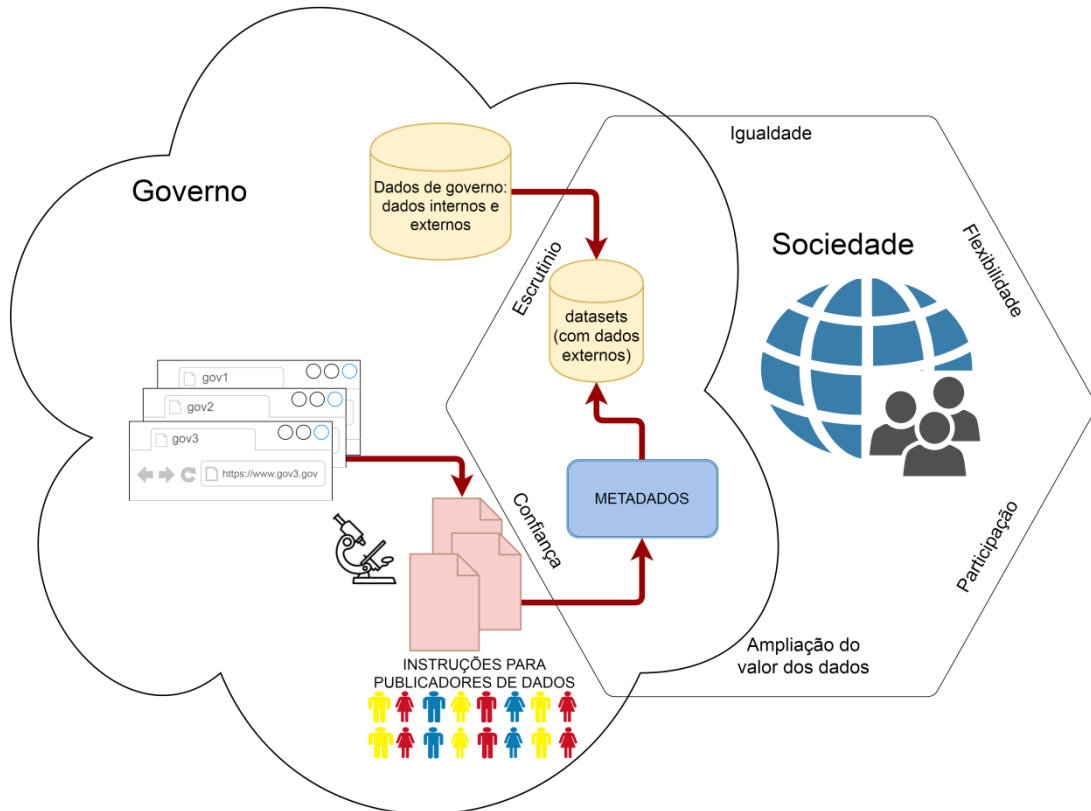
Descrever a aderência ao uso de metadados nos *datasets* de governos de países, tomando como base o contexto e o marco conceitual apresentados nas instruções para publicadores de dados, encontradas nos sites de dados abertos oficiais dos países analisados.

1.1.2 Objetivos específicos

- a) Identificar o referencial teórico que sustenta a proposta desta pesquisa, no que se refere à descrição de *datasets* de governos como recursos informacionais;
- b) recuperar e analisar documentos em sites oficiais que abordam o uso de metadados para a descrição de *datasets* a serem publicados nos sites de dados governamentais;
- c) realizar análise de conteúdo categorizando as informações encontradas nas instruções recuperadas; e,
- d) explanar o resultado dos estudos de caso desenvolvidos em cada site de dados de governo.

Com o fim de contribuir com a elucidação e delimitação da pesquisa, foi concebida a Figura 2, a qual apresenta graficamente os elementos do contexto e do foco deste trabalho investigativo.

Figura 2 – Elementos do contexto e foco da pesquisa



Fonte: Elaborado pela autora.

Percebe-se nesta figura a ilustração da interseção entre o âmbito governamental e a sociedade em geral, na observância dos dados governamentais, especificamente na dimensão externa, obtendo as fontes de informação a partir de sites oficiais, que além de permitir a publicação dos *datasets*, disponibilizam documentos de apoio e instruções para quem quer publicar dados nesses sites.

A Figura 2 expõe o fluxo e as relações existentes entre esses ativos informacionais, *datasets*, e atores de governo, que nesta pesquisa são os publicadores de dados, e, a sociedade.

Nesse cenário, espera-se recuperar instruções disponibilizadas pelo governo aos publicadores de dados e, a partir deste ponto, distinguir aquelas que especificamente atingem a descrição de *datasets*.

No caso dos atores publicadores de dados quanto da sociedade, representam-se com diversos desenhos, salientando que os membros desses grupos são de natureza heterogênea, mesmo que alguns membros estejam dentro do governo e os outros unicamente do lado da sociedade mantêm-se essa heterogeneidade.

O desenvolvimento desta dissertação tem como foco o conteúdo de documentos de apoio que estão disponibilizados nos sites de dados abertos em países da América do Sul e da Europa, os quais poderiam conter as devidas instruções para a publicação de *datasets*, nos atentando no relacionado ao uso de metadados dentro do processo descritivo desses recursos informacionais. Acredita-se que uma interseção entre governo e sociedade se dá mediante os *datasets* e os metadados que descrevem esses recursos informacionais, marcando assim a conjuntura de interesse no desenvolvimento desta pesquisa.

1.2 Justificativa

Nesta contemporaneidade, profissionais de diversas áreas do conhecimento, neste caso da Ciência da Informação, que pesquisam realidades regionais ou de grupos sociais ligados mediante características que os fazem similares, são requeridos de maneira que possam enlaçar a comunidade acadêmica e científica ao desenvolvimento e crescimento integral da sociedade.

Acredita-se, como indicado no estudo do Verhulst e Young (2017), na contribuição que a abertura dos governos tem trazido em casos específicos, a qual não somente pode ser assinalada como apoio na melhora dos índices de transparência dos aparelhos dos Estados. Há outros pontos de referência que tem a ver diretamente com o aprimoramento na qualidade de vida e da economia, situação que norteia o desenvolvimento desta pesquisa ao motivá-la e fundamentá-la com sentido altruísta, fundado a partir da comunidade científica, porém, visando a realidade que não pode ser alheia ao desenvolvimento deste grupo acadêmico que pertence de fato às ciências sociais aplicadas.

Espera-se com este estudo contribuir, concomitante com tantos outros como o de Gurstein (2007), no que diz respeito às condições de acesso as TIC, as quais podem fazer, ou não, com que elas sejam usáveis e úteis para a população. O Gurstein discorre no âmbito de comunidades particularmente distantes ou excluídas no uso das TIC, bem como grupos de indivíduos comuns no meio da heterogeneidade inata, salientando a possibilidade de desenvolver novas formas que permitam uma maior aproximação a sistemas de informação, procurando a diminuição da distância entre elas e a possibilidade do desenvolvimento econômico local, as condições de justiça social e o empoderamento político.

Este estudo pretende aproximar o desenvolvimento acadêmico à realidade de países com similaridades sociais, econômicas, históricas e tecnológicas, visando a melhoria no contexto de publicação de ativos informacionais, como os dados abertos governamentais. Neste sentido, sintetizam-se as seguintes relevâncias:

- a) científica: ao afirmar o objeto social e prático da Ciência da Informação, Ciência Social Aplicada, criando subsídios alicerçados nela, dentro de entornos informacionais contemporâneos que brindam novas possibilidades e abrem horizontes a grupos sociais, bem como o empoderamento na retomada de ativos informacionais que *per se* lhes correspondam e
- b) social: na perspectiva do verdadeiro aproveitamento da conjuntura da publicação de recursos informacionais, os dados abertos em *datasets* governamentais, auxiliando para que a sociedade possa utilizar e reutilizar, enfim, apropriar-se de recursos informacionais que possuam valor para esse conglomerado social, como mais uma possibilidade que contribua com e para o seu desenvolvimento integral.

1.3 Metodologia

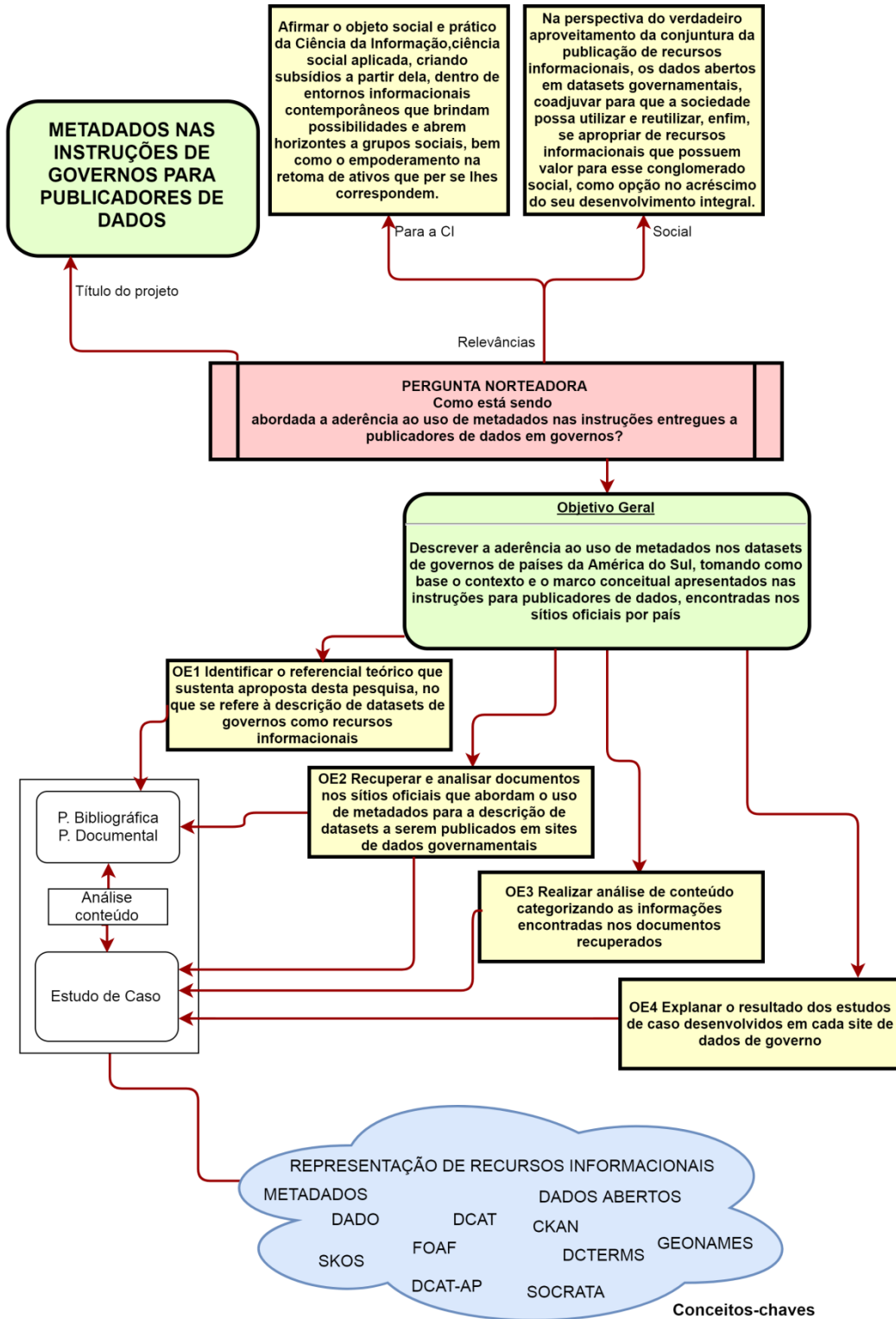
Para atingir o objetivo desta pesquisa, que é de natureza qualitativa, focada na observação crítica dos documentos que abordam o tratamento descritivo dos *datasets* governamentais, adotou-se a pesquisa descritiva, com o intuito de caracterizar o panorama das instruções sugeridas por governos da América do Sul e da Europa aos publicadores de dados. Serão utilizados como procedimentos metodológicos a análise bibliográfica e documental, e a aplicação de estudos de caso por países – para assim delinear o cenário atual no que concerne à representação de recursos informacionais objeto desta pesquisa, *datasets* governamentais.

Optou-se por aproximar o estudo de caso às pesquisas bibliográfica e documental, considerando que uns dos propósitos desse procedimento metodológico são a exploração de situações da vida real e a preservação do caráter unitário dos objetos da pesquisa (GIL, 2002).

A Figura 3 apresenta o esquema geral da pesquisa, considerando a inquietação norteadora junto às relevâncias do projeto nas dimensões científica e social, bem como o

cruzamento dos procedimentos metodológicos junto aos conceitos-chave que orientam o processo.

Figura 3 – Esquema geral da pesquisa



Fonte: Elaborado pela autora.

Os conceitos chaves apresentados no elemento nuvem da figura, foram os tópicos que nortearam a construção do referencial teórico. O intuito do esquema foi auxiliar na guarda da coerência durante todo o processo de desenvolvimento da pesquisa, buscando assim a efetivação dos objetivos específicos e o foco na questão norteadora da dissertação.

O volume de dados e informações obtido durante a realização dos procedimentos metodológicos foi abordado mediante a técnica de análise de conteúdo, aplicada nas fontes obtidas, baseando-se em que o processo de interpretação é fundamentalmente interativo, no qual “[...] o pesquisador elabora pouco a pouco uma explicação lógica do fenômeno ou da situação estudados, examinando as unidades de sentido, as inter-relações entre essas unidades e entre as categorias em que elas se encontram reunidas” (GIL, 2002, p. 90).

Contando que no estudo de caso devem-se redobrar os cuidados tanto no planejamento quanto na coleta e análise dos dados (GIL, 2002), espera-se encontrar na análise de conteúdo o rigor científico desejado na consecução do objetivo geral da pesquisa.

Os dois polos que segue a análise de conteúdo são o desejo de rigor na superação da incerteza e a necessidade de descobrir o conteúdo das mensagens, além das aparências. Considerando esses polos, são duas as funções que Bardin (2010, p. 31) apresenta a esse tipo de análise:

- a) a verificação prudente em função de uma administração da prova, onde uma possível consequência seja uma “afirmação”, mesmo que provisória: “A análise de conteúdo para servir de prova” e a
- b) função heurística, quanto interpretação dos conteúdos para “ver o que dá”.

A autora supracitada conceitua a análise de conteúdo como

um conjunto de técnicas de análise das comunicações visando obter por procedimentos sistemáticos e objetivos de descrição do conteúdo de mensagens indicadores (quantitativos ou não), que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens (BARDIN, 2010, p. 44).

O intuito é, após recuperação dos documentos que contém as indicações e estratégias dos governos alvo da pesquisa, desenvolver uma análise para inferir o

panorama com relação às estratégias que buscam a aderência de metadados nos *datasets* publicados no site de dados dos governos.

O desenvolvimento da análise de conteúdo foi realizado conforme prescrito por Bardin (2010) nessas três fases cronológicas:

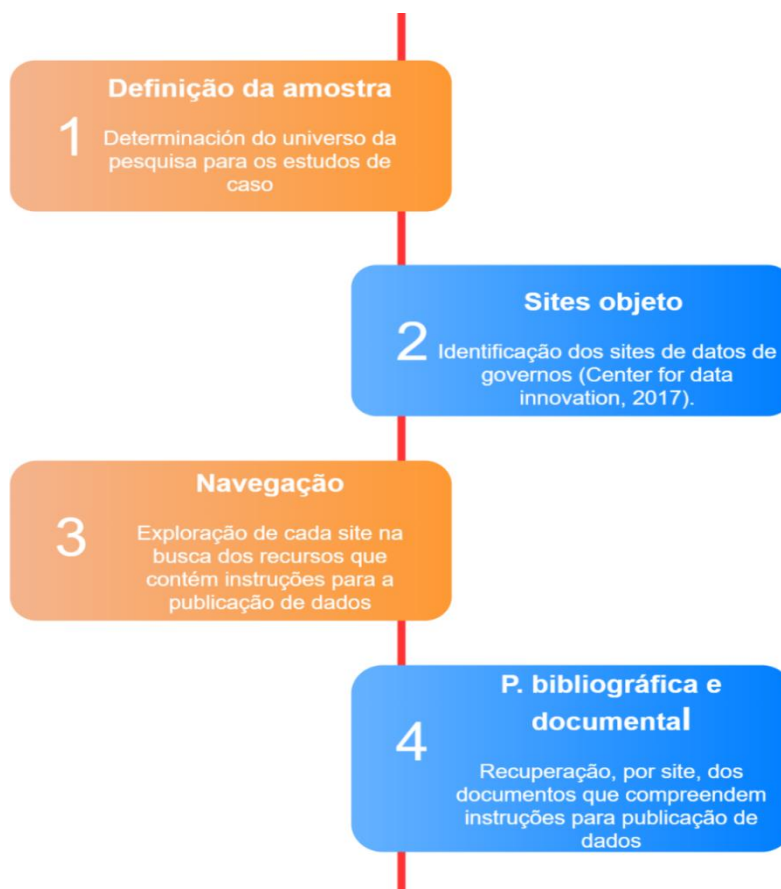
Pré-análise → Exploração do material → Tratamento dos resultados

A seguir, estão descritas cada uma das fases que compõem a análise de conteúdo realizada.

1.3.1 Fase de pré-análise

De acordo com a proposta de Bardin (2010) no que diz respeito ao privilégio que têm os procedimentos exploratórios na fase de pré-análise, a Figura 4 ilustra o método que permite a descoberta dos recursos, documentos com as instruções para publicadores de dados governamentais, os quais vão compor o *corpus* desta pesquisa.

Figura 4 – Fluxograma descoberta de recursos na fase de pré-análise



Fonte: Elaborado pela autora.

No primeiro estágio apresentado na Figura 4, Definição da amostra, optou-se por considerar, em um primeiro momento, a realidade do continente americano, em específico, no tocante às condições geográficas, sociais e econômicas de países do cone sul.

Esses países apresentam similitudes históricas fundamentadas em seus processos de colonização, a disponibilidade dos seus recursos naturais e na forma em que esses países foram inseridos na economia mundial; incluindo as similitudes econômicas e sociais que vêm se configurando desde a época colonial (LUCCI; BRANCO; MENDOÇA, 2013).

Não apenas foi tocada a realidade da América do Sul no alcance da pesquisa. Dois países atuam sob uma relação tanto histórica quanto cultural e política. Nesse ponto, incluem-se como foco na amostra da pesquisa uma dupla de governos da Europa.

Esta pesquisa apontou, em uma primeira instância, à realidade da Colômbia, desenvolvendo um pré-teste com as informações recuperadas nesse país. Logo após da

validação da metodologia, o estudo foi desenvolvido ampliando para os países Brasil, Espanha e Portugal.

Como ponto de partida, considerou-se a pesquisa do *Center for Data Innovation* (2017) acerca do estado dos portais de dados abertos de governos, nesse caso, na América Latina. Dentre outros resultados da pesquisa, são elencados os sites de dados governamentais, os quais constituem os pontos de entrada desta pesquisa, no que diz aos sites dos países definidos para os estudos de caso.

Já para as plataformas dos governos da Espanha e da Portugal, mediante pesquisa exploratória foram reconhecidos os sites correspondentes com publicação de dados abertos oficiais.

O Quadro 1 relaciona cada país objeto deste estudo e o endereço do site de dados abertos correspondente.

Quadro 1 – Sites de dados como pontos de entrada dos estudos de caso

País	Link do site de dados governamental
Colômbia	https://datos.gov.co/
Brasil	http://dados.gov.br/
Espanha	http://datos.gob.es/
Portugal	http://www.dados.gov.pt/pt/inicio/inicio.aspx

Fonte: Adaptado de *Center for Data Innovation* (2017).

Em cada site oficial de dados efetuar-se-á a exploração dos sistemas de organização presentes, considerando que para atingir o objetivo desta pesquisa, as informações relacionadas aos documentos que fornecem as instruções para os processos de publicação de dados, são de fato dirigidas para uma audiência específica.

Com isso, visando a recuperação dos documentos, será efetuada navegação nos esquemas de organização de cada site considerando: a) tópicos; b) orientados a tarefas; e, c) esquemas para audiências específicas (ROSENFELD; MORVILLE; ARANGO, 2015).

No desenvolvimento da pesquisa documental e bibliográfica, o critério de busca dos documentos, objeto da pesquisa, é o de recuperar todos os documentos que têm como propósito capacitar o capital humano que participa da publicação de *datasets*, como fornecedores ou consumidores de dados.

Devido ao embasamento na proposta da Bardin (2010) para o processo de constituição do *corpus* da pesquisa, foram apontadas, como principais, as seguintes regras na distinção dos documentos a serem analisadas:

- a) exaustividade;
- b) representatividade;
- c) homogeneidade e
- d) pertinência.

Uma vez reunido o conglomerado de documentos, realizou-se um refinamento com o intuito de identificar os documentos que efetivamente contém instruções relacionadas à descrição dos *datasets*, e que serão descritos na seção seguinte.

1.3.2 Fase da exploração do material

Como alternativa para a “administração das técnicas no *corpus*” (BARDIN, 2010, p. 128, destaque nosso), realizou-se a exploração do material refinado escolhendo como unidade de registro a análise temática, assumindo que “fazer uma análise temática consiste em descobrir os núcleos de sentido que compõem a comunicação e cuja presença, ou frequência de aparição podem significar alguma coisa para o objetivo analítico escolhido” (BARDIN, 2010, p. 131). Complementa a autora, que o tema, oferece uma unidade de recorte do sentido e não de forma, “[...] que não é fornecida de uma vez por todas, visto que o recorte depende do nível de análise e não de manifestações formais reguladas” (BARDIN, 2010, p. 131).

Para tal, o tema eixo em torno do qual realizou-se a exploração foi **metadados**, agindo como objeto ou referente nesta fase. Nos documentos recuperados, foram identificados os trechos nos quais os autores se referiram ao tratamento descritivo dos *datasets*, considerando as variações pelas línguas português e castelhano (*metadados*, *metadatos*).

Uma vez definido o tema metadados, como unidade de registro, foram estipuladas categorias para representar as abordagens identificadas nos documentos com as instruções para publicadores de dados, a partir de um estudo desenvolvido por Santos et al. (2013), com a finalidade de tratar o material do *corpus*.

Neste processo de codificação, foram definidas as categorias para classificar o conteúdo, objeto da análise, para o posterior tratamento dos dados:

- a) Metadados Teoria (MT): conceitos, fundamentações, técnicas, estudos, explicações e críticas e
- b) Metadados no Uso (MNU): experimentos, relatos de experiência, estudos de caso, recomendações e percepções de uso (SANTOS et al., 2013).

Para Bardin (2010), uma vez definida a unidade de registro, neste caso **o tema**, faz-se necessário estabelecer as regras de enumeração ou modo de contagem. Adotaram-se como regras: a presença ou ausência das categorias supracitadas, e, a intensidade, a observar-se seguindo variações semânticas ou formais ao aparecer os elementos da análise, ou seja, das categorias adotadas. Como auxílio para a avaliação da intensidade da abordagem das unidades de registro temático, considerou-se a intensidade semântica dos verbos, assim como a ênfase e o aprofundamento dos sintagmas nos quais está inserida a unidade de registro que está sendo considerada na análise de conteúdo.

A unidade de contexto da técnica da análise é cada documento que contenha as instruções alvo desta pesquisa.

Em seguida, propõe-se o esquema do Quadro 2 que pretende auxiliar na caracterização dos documentos que vão ser analisados, integrando nele as categorias utilizadas na codificação das unidades de registro. O esquema inclui atributos que descrevem os documentos recuperados, a indicação da lei na qual está fundamentada a estratégia de dados abertos em cada país, as categorias acima indicadas, as regras de contagem, e por fim, os esquemas de metadados indicados nas instruções aos publicadores de dados.

Como atributos (Quadro 2), que contém a proposta de caracterização para os documentos a serem analisados, destacam-se:

- a) nome do documento: obedece ao título com o que documento está identificado;
- b) detentor: nome da entidade ou órgão oficial que elabora o documento a ser analisado e que o fornece como instrução para os processos de publicação de dados abertos;
- c) tipo do documento: natureza do documento segundo a entidade ou órgão oficial que o fornece;
- d) data: data na qual é elaborado ou publicado o documento;
- e) função: segundo o objetivo do documento, bem como a introdução, sintetiza-se a função principal do documento;

- f) fundamentado na lei: neste atributo registra-se a Lei na qual está fundamentado o documento;
- g) metadados teoria: primeira categoria adotada para a codificação dos documentos analisados. Os valores do atributo variam de acordo com o tipo de abordagem identificada:
- Conceitos (C);
 - Fundamentações (F);
 - Técnicas (T);
 - Estudos (E);
 - Explicações (EX);
 - Críticas (CR).

Considerando que a primeira regra de contagem é Presença (ou ausência), o cálculo dela será realizado assumindo que qualquer valor neste atributo corresponde com Presença, enquanto o traço '-' será considerado como Ausência;

- h) intensidade MT: sendo a segunda regra de contagem, será avaliada segundo a intensidade semântica dos sintagmas que contêm a unidade de registro temática em três valores: Baixa (B), Média (M) e Alta (A). Nesta primeira ocorrência do atributo corresponde com a intensidade da abordagem do conteúdo da categoria Metadado Teoria;
- i) metadado no uso: segunda categoria adotada para a codificação dos documentos analisados. Os valores do atributo variam de acordo com o tipo de abordagem identificada:
- Experimentos (EXP);
 - Relatos de experiência (RE);
 - Estudos de caso (EC);
 - Recomendações (R);
 - Percepções de uso (PU).

Como no caso da categoria Metadado Teoria, para a contagem, qualquer valor presente para este atributo corresponde com Presença, enquanto o traço '-' será considerado como Ausência;

- j) intensidade MNU: Este atributo corresponde à intensidade do conteúdo encontrado na categoria Metadado No Uso, avaliada igualmente nos três valores: Baixa (B), Média (M) e Alta (A);
- k) metadados estruturais: este atributo contém os metadados estruturais que sejam indicados nas instruções contidas nos documentos a serem analisados. O atributo se relaciona diretamente com o exposto na seção 2.1, “padrões de estrutura dos dados” (GILLILLAND, 2008) e
- l) metadados descritivos: o atributo contém os metadados como indicado por Gilliland (2008), exposto na seção 2.1 como “padrões de valores dos dados”.

Quadro 2 – Caracterização do material da análise

Nome documento	Detentor	Tipo de documento	Data	Função	Fundamentado em LAI	Metadados Teoria Conceitos (C), Fundamentações (F), Técnicas (T), Estudos (E), Explanações (EX), Críticas (CR)	Intensidade MT Alta (A), Média (M), Baixa (B)	Metadados Uso Experimentos (EXP), Relatos de Experiencia (RE), Estudos de Caso (EC), Recomendações (R), Percepções de Uso (PU)	Intensidade MNU Alta (A), Média (M), Baixa (B)	Metadados Estruturais	Metadados Descritivos
Título com o que documento está identificado	Nome da entidade ou órgão oficial que elabora o documento	Natureza do documento segundo a entidade ou órgão oficial que o fornece	Data de elaboração do documento	Segundo o objetivo do documento, bem como a introdução, sintetiza-se a função principal do documento	Registro, se informado, da Lei na qual está fundamentado o documento	Valores do atributo de acordo com o tipo de abordagem identificada: Conceitos (C), Fundamentações (F), Técnicas (T), Estudos (E), Explanações (EX), Críticas (CR) Considerando que a primeira regra de contagem é Presença (ou ausência), o cálculo dela será realizado assumindo que qualquer valor neste atributo corresponde com Presença, enquanto o traço '-' será considerado como Ausência	Identificação da intensidade semântica dos sintagmas que contém a unidade de registro temática em três valores: Baixa (B), Média (M) e Alta (A), para a categoria Metadados Teoria	Valores do atributo de acordo com o tipo de abordagem identificada nos seguintes valores: Experimentos (EXP), Relatos de experiência (RE), Estudos de caso (EC), Recomendações (R), e, Percepções de uso (PU). Como no caso da categoria Metadado Teoria, para a contagem, qualquer valor presente para este atributo corresponde com Presença, enquanto o traço '-' será considerado como Ausência	Este atributo corresponde à intensidade do conteúdo encontrado na categoria Metadado No Uso, avaliada igualmente nos três valores: Baixa (B), Média (M) e Alta (A)	Contém os metadados estruturais que indicados nas instruções analisadas. O atributo se relaciona diretamente com o exposto na seção 2.1, como "padrões de estrutura dos dados" (GILLILLAND, 2008)	Para o registro dos metadados como indicado por Gilliland (2008), como "padrões de valores dos dados".

Fonte: Elaborado pela autora.

Após da fase de exploração do material, os dados brutos foram examinados de modo que possam se tornar significativos na consecução dos objetivos da pesquisa. Para isso, apresenta-se no item seguinte a proposta de tratamento dos dados obtidos na caracterização dos documentos analisados.

1.3.3 Fase de tratamento dos resultados

Quais são as significações das mensagens contidas nos documentos? Quais são os conteúdos da comunicação escrita nos documentos analisados? Atingindo a teoria exposta por Bardin (2010), quais poderiam ser as subcategorias presentes segundo a proposta da codificação dos documentos e o aprofundamento percebido?

A variável de inferência vai estar em torno ao objetivo geral desta pesquisa, buscando observar a aderência ao uso dos metadados nos documentos que contêm instruções para aqueles usuários, órgãos de governos, envolvidos na publicação de *datasets*.

1.3.4 Síntese da técnica a ser aplicada

Resumindo, o Quadro 3 apresenta a síntese da técnica *Análise de conteúdo* aplicada no desenvolvimento desta pesquisa.

Quadro 3 – Síntese da técnica a aplicar na pesquisa

Pré-análise	Exploração do material	Tratamento dos resultados
Exploração dos sites por país para descoberta de recursos: documentos com instruções para publicadores de dados	Unidade de registro: análise temática Tema: Metadados	Polo a observar: a mensagem nos documentos, no nível significação (BARDIN, 2010).
Regras de distinção dos recursos: <ul style="list-style-type: none"> - regra de exaustividade; - regra de representatividade; - regra de homogeneidade; - regra de pertinência (BARDIN, 2010). 	Regras de enumeração: <ul style="list-style-type: none"> - presença ou ausência das categorias; - intensidade, seguindo variações semânticas ou formais das categorias (BARDIN, 2010). 	Variável de inferência: a aderência ao uso dos metadados nos documentos que contêm instruções dos órgãos de governos para os envolvidos na publicação de <i>datasets</i> .

Fonte: Elaborado pela autora.

Esta seção finaliza a explanação da técnica de análise de conteúdo que foi aplicada na consecução dos objetivos desta pesquisa. Antes de abordar os conceitos norteadores da pesquisa, será exposta brevemente a argumentação para a escolha como estudo de caso da Colômbia, como primeiro país analisado no desenvolvimento deste estudo.

1.4 Por que a escolha da Colômbia como pré-teste?

No contexto mundial da publicação de *datasets* de dados oficiais, a Colômbia tem obtido destaque na implementação de estratégias que visam impulsionar a publicação de dados a partir do governo nacional (ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÓMICO, 2017; GLOBAL EDITORS NETWORK, 2017). No *site* de dados do governo, datos.gov.co, atualmente, estão publicados 7.357 *datasets* (COLOMBIA, 2017) com dados provenientes de diversos órgãos do Estado colombiano.

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) desenvolveu *The OCED OURdata Index*, um marco de trabalho que reúne indicadores na temática de dados no governo aberto. Esse marco permite resumir pontos fortes e pontos fracos, identificando áreas de ação que intervenham às políticas nacionais dos países membros da OCDE (2017).

Na última medição do *OURdata Index*, com dados pesquisados entre novembro e dezembro de 2016, mensuraram-se os esforços desenvolvidos pelos governos para a promoção, uso, reuso e disponibilização de ferramentas para a publicação de dados. Nessa edição a Colômbia aparece em quarto lugar no ranking global, que tem em seus primeiros colocados, a Coreia do Sul, a França e o Japão (ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÓMICO, 2017).

Ademais, a Colômbia obteve destaque na sexta edição dos *Data Journalism Awards*, organizado pela *Global Editors Network*. Das onze categorias propostas pela rede, esse país foi premiado pelo projeto *Rutas del Conflicto*, apresentado na categoria site de Jornalismo de Dados do Ano (GLOBAL EDITORS NETWORK, 2017).

Esses fatos assinalam a seleção desse país como ponto de partida no desenvolvimento desta pesquisa.

Em relação ao período de tempo estipulado, indica-se que a pesquisa recuperou documentos no contexto do primeiro país estudado, a Colômbia, durante o mês de junho de 2017.

1.5 Colômbia, Brasil, Espanha e Portugal

Delineado o começo com a análise do site de dados da Colômbia, definiu-se como amostra completa do estudo os sites de dados dos países Brasil, Espanha e Portugal.

O Brasil não somente é o país sede da realização desta pesquisa. Este país é um dos países pioneiros da América Latina e do mundo na promoção e execução de políticas de Estado, com repercussão e convocatória para outros países, que buscaram atingir os anseios da sociedade no aprimoramento da percepção de transparência do governo, mediante a utilização e o aproveitamento da conjuntura oferecida pela internet, com suas ferramentas e mecanismos tecnológicos que cada vez mais estão ao alcance dessa sociedade que representam.

Como será tratado na seção 3.1, têm-se desenvolvido fatos com destaque global onde o Brasil, junto a outros Estados, manifestou e convocou à comunidade global a movimentação para afrontar os desafios comuns no que concerne aos índices de corrupção que preocupam os cidadãos e os governos.

A Colômbia e o Brasil têm fortes laços com duas nações do continente europeu: Espanha e Portugal, dois países que têm similaridades históricas, culturais e linguísticas com os dois primeiros, que configuram interesse de estudo dentro do escopo desta pesquisa.

Tanto o governo de Portugal quanto o da Espanha fizeram parte da convenção¹ da OCDE, em 1960, realizada com o intuito de buscar a prosperidade e o fortalecimento econômico de países, bem como o aprimoramento do bem-estar geral neles.

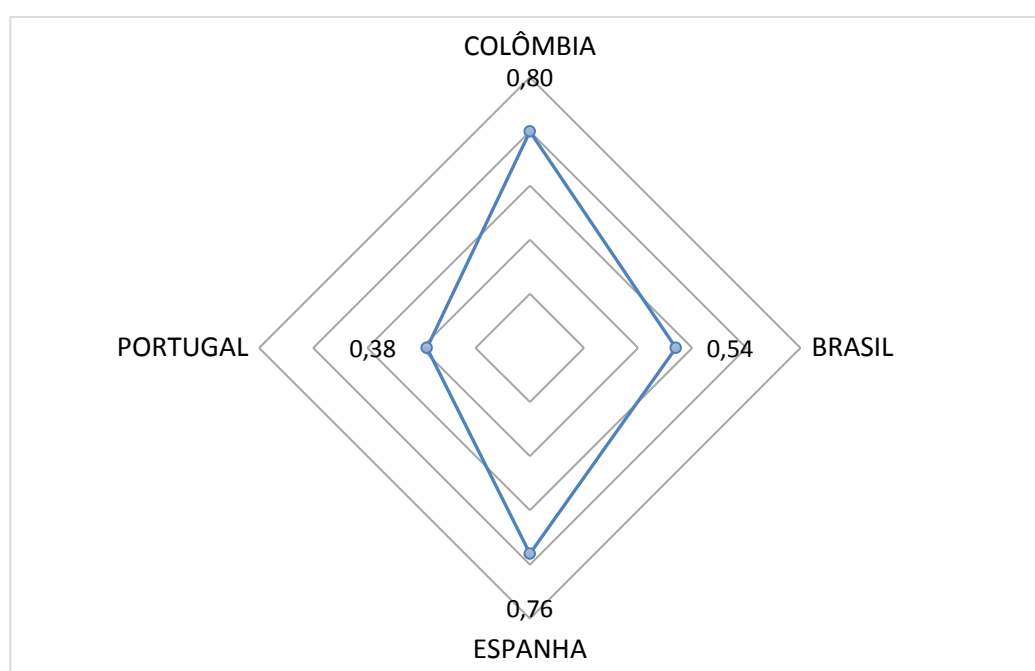
Já no caso da Colômbia e do Brasil, até este momento (2018), nenhum deles é membro pleno dessa organização. O pedido de adesão da Colômbia foi admitido em 2013, sendo considerado pela OCDE como um país com o processo de adesão “em estado de discussão”; enquanto o Brasil, ante a própria manifestação do interesse, espera pelo início da discussão da adesão na organização. No entanto, o Brasil é considerado pela OCDE como um parceiro chave, particularidade que compartilha com a Índia, Indonésia, China

¹ Os governos reunidos em dezembro de 1960 foram os da Áustria, Bélgica, Canada, Dinamarca, França, Alemanha, Grécia, Islândia, Irlanda, Itália, Luxemburgo, Holanda, Noruega, Portugal, Espanha, Suécia, Suíça Turquia, Reino Unido, Estados Unidos (ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO, 2018).

e África do Sul (ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO, 2018).

O índice desenvolvido pela OCDE, índice composto, considera três aspectos como são: Disponibilidade dos dados; Acessibilidade aos dados; e, Suporte governamental para o reuso dos dados. O índice tem uma magnitude que vai até o algarismo 1, que é calculado seguindo a metodologia explicitada na documentação da OCDE (2017). A Figura 5 apresenta o *OURdata Index* para os quatro países da pesquisa, cuja

Figura 5 – *OURDATA Index* do estudo



Fonte: Adaptado de OCDE (2017a, 2017b).

Como indicado na seção precedente, a *performance* da Colômbia teve destaque, mesmo sendo um país em processo de adesão. Observa-se uma pontuação maior em relação aos outros três países, classificados na ordem decrescente da seguinte forma: Colômbia, Espanha, Brasil e, Portugal.

Essa avaliação proposta pela OCDE analisa três áreas, a saber: disponibilidade dos dados, acessibilidade aos dados, e suporte governamental para o reuso dos dados (ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO, 2017a).

Outro país dos determinados para esta pesquisa foi também reconhecido na edição 2017 dos *Data Journalism Awards*. Além da Colômbia, o projeto Ctrl+X proveniente do

Brasil, recebeu prêmio na categoria *Pequenas Redações Jornalísticas*. Este projeto é realizado pela Associação Brasileira de Jornalismo Investigativo (ABRAJI), e “[...] mapeia as tentativas de censura a veículos de mídia no Brasil” (ASSOCIAÇÃO BRASILEIRA DE JORNALISMO INVESTIGATIVO, 2017).

Em relação as datas da coleta dos documentos com instruções dos países Brasil, Espanha e Portugal, realizadas nos meses de novembro e dezembro de 2017.

Assim, encerra-se a etapa de a escolha dos quatro países para desenvolvimento deste estudo, o qual atinge um dos vários aspectos relacionados à publicação de dados abertos de governo.

2 DEMARCAÇÃO DA REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS

Esta seção apresenta fundamentos teóricos da pesquisa, introduzindo definições que partem do geral ao específico em relação aos recursos informacionais no contexto de processos de representação.

Assim, ao referenciar o ‘recurso informacional’, partindo do processo mental de abstração dentro de um determinado campo do conhecimento, concebe-se como um objeto, analógico ou digital, tangível ou intangível, que é portador de informação, ou seja, que tem valor informacional. O valor informacional desse objeto caracteriza esse processo de abstração e expõe suas manifestações estruturadas. (LE COADIC, 2004; REITZ, 2004; SANTOS; SIMIONATO; ARAKAKI, 2014).

Ao considerá-lo tanto recurso quanto objeto, faz-se correlação com Gilliland (2008) que acrescenta comportamentos possíveis de serem realizados neles, tais como ser administrados ou manipulados por atores humanos ou máquinas.

Incluindo um último elemento, no contexto de um campo do conhecimento, os recursos informacionais requerem, para manifestar seu valor informacional, estruturas analógicas ou digitais, por precisarem da apresentação de suas características mediante um “tratamento descritivo específico para otimizar sua busca, acesso e recuperação” (SANTOS; SIMIONATO; ARAKAKI, 2014, p. 148).

Dentro dessa necessidade de estruturação e considerando o contexto dinâmico que envolve os recursos informacionais na contemporaneidade, especificamente no âmbito digital, McGarry (1999) afirma que esses recursos devem ser estruturados para dar-lhes forma e utilização factível nesse âmbito.

Configura-se a representação dos recursos informacionais como “[...] uma atividade imprescindível no armazenamento, no acesso e na recuperação de informações e de recursos, e sua evolução e transformações afetam o cotidiano de milhares de profissionais” (SANTOS, 2017, p. 2).

Simionato (2015, p. 72) aponta que “se os recursos não estiverem descritos adequadamente ao tipo de ambiente inserido, ocorrerá problemas, e conseqüentemente, o usuário não recuperará o que deseja”.

Os metadados são apontados como elementos que permitem representar as características dos recursos informacionais, nomeados como objetos de informação por Gilliland (2008) quando assevera que tais objetos apresentam três características, a

saber: conteúdo, contexto e estrutura, as quais podem e devem ser retratadas mediante metadados.

A autora supracitada diz que os recursos informacionais vão se descrever mediante metadados, partindo do fato de que o objeto pode conter informação “intrínseca” do recurso, passando pelos aspectos extrínsecos associados a ele, aspectos que respondam questões tais como quem, que, por que, onde, como, até, finalmente, abordar o relativo às associações dentro ou entre recursos, mediante informações intrínsecas, extrínsecas ou das duas (GILLILAND, 2008).

A ciência da computação utiliza o termo metadado referenciando objetos de sistemas de bases de dados e/o sistemas de informação, utilizando-o para o encapsulamento da informação que permite descrever qualquer objeto ou documento, tanto em formato analógico quanto digital (ZENG; QIN, 2008).

As três características indicadas pela Gilliland (2008), correlaciona-se à heurística inerente da ciência da computação (ZENG; QIN, 2008) no ponto das associações ao interior dos recursos informacionais ou entre recursos, o qual é materializado na capacidade de compartilhamento que fornecem os metadados (NATIONAL INFORMATION STANDARDS ORGANIZATION - NISO, 2017).

Seguido ao fato de considerar a definição clássica literal baseada na etimologia da palavra metadado, “dado sobre dado”, vem a questão de se possível encontrar metadados em todos lados. Faz-se evidente essa presença generalizada dos metadados, “metadados em todas partes” no caso apresentado pela organização acima citada, acontecido em 2013, com a familiaridade que chegou a ter o termo metadado nos Estados Unidos pela cobertura da mídia na coleta de dados da Agencia Nacional de Segurança em relação as ligações domésticas, alcançando, nessa data, uma ampla caracterização delas.

No ambiente da internet a influência dos metadados foca no acesso e na recuperação da informação, sendo deste modo, considerados elementos basilares no tratamento descritivo da informação, já que garantem a descrição e a representação padronizada de recursos informacionais (ALVES; SANTOS, 2013).

Essa padronização origina estruturas conhecidas como padrões de metadados,

[...] estruturas de descrição constituídas por um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade gerando uma

representação unívoca e padronizada que possa ser utilizada para recuperação da mesma (ALVES, 2010, p. 47-48).

Com os padrões de metadados espera-se garantir a permuta de informações entre agentes ou organizações, processo alvo da representação informacional, pela diminuição da carga de trabalho na descrição de recursos ou objetos informacionais (FUSCO, 2010).

Ademais, "[...] um sistema de informação deve priorizar a estrutura interna do domínio de conhecimento" (SIMIONATO; SANTOS, 2013, p. 3), considerando as necessidades próprias dos sujeitos informacionais que com ele vão interagir. Neste contexto, devem ser estabelecidos padrões de metadados que melhor atendam esse domínio para a correta representação dos recursos informacionais. Destaca-se ainda que, segundo as particularidades de um determinado domínio existe a possibilidade de utilizar vários padrões de metadados.

Os padrões de metadados estão organizados nas seguintes tipologias (GILLILAND, 2008):

- a) padrões de estrutura de dados: categorias ou *containers* dos dados que compõem um registro;
- b) padrões de valores dos dados: termos, nomes, e outros valores usados como conjuntos de elementos dos metadados;
- c) padrões de conteúdo dos dados: diretrizes para o formato e sintaxe dos valores dos dados e
- d) padrões de intercâmbio de dados: formatos de codificação para intercâmbio de dados mediante máquinas.

Identificados conceitos gerais, na seguinte seção vai se abordar um recurso informacional em particular, o dado.

2.1 Dado e seu peso na representação

Setzer (1999) direciona o conceito de dado ao de entidade matemática, isto é, puramente sintático, uma sequência de símbolos quantificáveis, indicando a associação direta a um conjunto de elementos finitos.

Este componente somente sintático implica uma baixa carga semântica (SANTOS; SANT'ANA, 2013), o qual traz como consequência um esforço adicional na representação do dado

já que não bastam elementos que o descrevam como um todo e que propiciem sua recuperação. São necessários, ainda, elementos que permitam a sua interpretação por quem os acesse, com informações que detalhem sua estrutura e possibilitem a interpretação de cada atributo que os compõe (SANT'ANA, 2017, p. 4).

Adota-se o dado como uma

unidade de conteúdo necessariamente relacionada a determinado contexto e composta pela tríade entidade, atributo e valor, de tal forma que, mesmo que não esteja explícito o detalhamento sobre contexto do conteúdo, ele deverá estar disponível de modo implícito no utilizador, permitindo, portanto, sua plena interpretação (SANTOS; SANT'ANA, 2013, p. 205).

Independentemente das habilidades e competências que tenham os usuários que vão fazer uso dos dados, considerando que eles estão vinculados explicitamente com um determinado contexto, dito contexto deve-se manifestar diretamente na sua representação, na busca do ganho na recuperação desses recursos informacionais. Deve-se garantir a disponibilização do recurso informacional, o dado, de um modo implícito para o utilizador, no entanto implique maiores esforços na representação por parte dos detentores, procurando assim, o favorecimento da sua visibilidade, utilização e reutilização (SANT'ANA, 2017).

Vão ser referidos a continuação, focando os objetivos desta pesquisa, dados de tipo abertos, no âmbito governamental.

3 DADOS ABERTOS NO ÂMBITO GOVERNAMENTAL

Dado aberto, segundo a *Open Knowledge International* (OKI), quer dizer dado acessível por qualquer pessoa ou agente, e utilizável sem restrição em seu tipo de uso e compartilhamento (OKI, 2017). A OKI foi fundada em 2004 como rede mundial de pessoas que trabalham pela abertura de dados e informações para transformá-los em conhecimento. Distinguem-se dados abertos científicos, de governo, de empresas privadas, e de universidades.

Para Peset, Ferrer-Sapena e Subirats-Coll (2011) dados abertos referem-se aos dados surgidos dentro de um movimento que promove a liberação de dados a partir das instituições públicas, com a particularidade de estarem representados de tal forma que seja válida sua reutilização.

Os dados abertos estão sendo observados como uma grande oportunidade dos governos no aprimoramento da sua relação com os cidadãos, tanto pela ampliação na disseminação de dados e informações a partir dos Estados, quanto na possibilidade de atender pedidos dessa natureza por parte de grupos sociais que buscam caminhos para suprir necessidades particulares.

Contudo, o primeiro uso do termo “aberto” para dados e informações foi no contexto da pesquisa científica, trabalhado nessa perspectiva desde a metade do século passado pelo Centro de Dados Mundial (INTERNATIONAL COUNCIL FOR SCIENCE, 2017), com o objetivo de reunir esforços e evitar a duplicidade de atividades nos caminhos percorridos pelos pesquisadores das ciências exatas.

Neste âmbito, a OCDE, considerando a conjuntura oferecida pela internet, e ciente de que além do uso da tecnologia, mediante a criação de políticas públicas, afirma ser possível o estabelecimento de procedimentos que busquem aumentar o retorno dos investimentos públicos da pesquisa científica nos países membros da organização. A OCDE criou, em 2007, treze princípios para os dados dentro do contexto da pesquisa científica (ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO, 2007).

O primeiro princípio determinado pela OCDE é *Openness*, onde foi explicitada a importância da abertura dos dados para dar um acesso em igualdade de termos à comunidade científica com o menor custo, com fácil acesso e agilidade.

Na atualidade, o termo “aberto” geralmente está sendo aplicado para o conteúdo proveniente tanto dos governos como do setor privado, e usado em vários domínios do conhecimento: dados em todos os setores das economias, sociais, geográficos, meteorológicos, turísticos, dentre outros, próprios da função exercida pelos governos nas sociedades que representam e dirigem. Esses dados estão sendo disponibilizados em marcos de desenvolvimento de políticas de dados abertos nacionais e regionais, com determinadas características técnicas para a sua reutilização, buscando incrementar a transparência no acesso, sem custo, por parte dos cidadãos em geral.

Quanto ao uso e aos benefícios dos dados abertos, Shintaku, Duque e Suaiden (2015) identificam como aspecto determinante, que dados advindos de resultados de pesquisas sejam usados e reusados em processos de difusão e geração da informação, sem gerar emolumentos que sejam assumidos pelos usuários.

No âmbito desta pesquisa, o caso de dados abertos gerados pelos governos, desde 1998 o Parlamento Europeu mediante a Diretiva 2003/98 (PARLAMENTO EUROPEU, 2003), assinalava o caminho que pretendia iniciar, dentro do contexto da emergente sociedade da informação, no que diz respeito ao fortalecimento da institucionalidade dos governos membros, cuidando para que essa força começasse pelo empoderamento da sociedade, no acesso a esse tipo de recursos informacionais.

A diretiva supracitada indica que a disposição geral dos recursos nos governos, “[...] não só relativos aos atos políticos, mas também aos processos judiciais e administrativos [...]” (PARLAMENTO EUROPEU, 2003), constitui um instrumento fundamental da democracia dos governos. Ela considerava a evolução da sociedade da informação na região e visava uma maior influência dela na vida dos cidadãos, mediante novos meios de acesso aos assuntos públicos, e esteve motivada pela idealização de uma “harmonização das regras”, para a reutilização de documentos e seus conteúdos, buscando estabelecer nos estados membros um “conjunto de regras mínimas aplicáveis” com formas práticas e fáceis aos cidadãos.

No contexto americano, o fato que tem recebido, midiaticamente, indicação de gênese na abertura de dados de governo, aconteceu no ano 2007, quando nos Estados Unidos, em Sebastopol, reuniram-se ativistas do mundo na busca do aprimoramento do acesso e uso de dados governamentais como mecanismo para melhorar a transparência e lidar com a corrupção (*OPEN GOVERNMENT PARTNERSHIP* – OGP, 2017). Nesse encontro discutiram-se considerações legais, tecnológicas e de licenciamento dos dados

disponibilizados por governos e pelas empresas privadas que administram dinheiro público, destacando que tal iniciativa foi realizada sem intervenção de governo algum.

Assim, na busca do fortalecimento da atuação da sociedade por meio da visibilidade das ações governamentais, os dados abertos estão sendo disponibilizados em sites, administrados por governos, mediante *datasets*. Salientando-os como parte do alvo desta dissertação, os *datasets*, assumem-se como recursos informacionais que conglomeram dados com uma certa estrutura acerca de algum tópico ou tema.

3.1 Iniciativas estratégicas para o fomento da abertura de dados governamentais

Apresenta-se na sequência a descrição do ambiente global ao qual pertencem iniciativas que pretendem promover a mobilização de ainda mais estratégias de abertura de dados dentro e fora dos governos, relatando alguns fatos políticos e estratégicos que contextualizam três iniciativas baseadas no canal de comunicação da internet, desenvolvidas dentro do marco dos dados abertos: a *Open Government Partnership*, a *Global Open Data for Agriculture & Nutrition* e, a *International Open Data Conference*.

A *Open Government Partnership* nasceu, em 2011, como uma iniciativa de oito países preocupados por fazer de seus governos verdadeiros órgãos a serviço dos cidadãos que representam. Os países fundadores procuravam estabelecer diretrizes para criar políticas de alcance global que buscassem a eficiência e transparência dos governos, fazendo público ante a sociedade o fato de que estes governos tinham a vontade de converterem-se em fornecedores de dados e informações, visando maior visibilidade, principalmente nas questões referentes a maior efetividade política e diminuição da corrupção (BRASIL, 2011).

Esta iniciativa, Parceria de Governo Aberto (OGP pelo nome no inglês *Open Government Partnership*), buscou estabelecer compromissos concretos de parte dos governos que estivessem interessados em estimular a disseminação aberta de suas ações, com o fim de encorajar os cidadãos, mediante a promoção, dentre outras ações, da disponibilização, uso e reuso de dados abertos governamentais (OGP, 2017).

A primeira reunião da OGP, cujos participantes foram países de vários continentes, Estados Unidos, Reino Unido, Noruega, Filipinas, Indonésia, África do Sul, Brasil e México, aconteceu em 2011, e agiu dando o primeiro passo para uma iniciativa de governos parceiros em prol do uso e reuso de ativos informacionais oficiais, estratégia que foi

oficializada ao mesmo tempo em que acontecia a 66^a Assembleia Geral das Nações Unidas (BRASIL, 2011).

Em um primeiro momento o Brasil presidiu, junto aos Estados Unidos, a OGP, e voltaram-se junto ao México os países latino-americanos que se oficializaram como pioneiros na estratégia de abertura de ativos informacionais de governos. Na atualidade, 75 países são membros da OGP (2017).

Assim, foi criado o portal opengovpartnership.org como apoio à estratégia, sendo o meio de difusão das informações, convocações e discussões dos processos desenvolvidos pelos países membros, das formas de ingressar na parceria, e como meio de divulgação para a sociedade em geral.

No contexto do Brasil um dos compromissos adquiridos no plano da primeira OGP concretizou-se em março de 2012: a disponibilização do site de dados abertos governamentais. Ainda, segundo o Tribunal de Contas da União (BRASIL, 2015), o Estado brasileiro viu-se motivado para tal pela procura da progressão da:

- a) transparência na gestão pública;
- b) possibilidade na contribuição da sociedade com serviços inovadores ao cidadão;
- c) aprimoramento na qualidade dos dados governamentais;
- d) viabilização de novos negócios e
- e) por obrigatoriedade de lei.

Em quanto isso, continuando com fatos políticos, no ano de 2012, na Cimeira do G8², que aconteceu no retiro presidencial *Camp David*, pertencente ao governo dos Estados Unidos, os presidentes do país anfitrião, da Rússia, do Reino Unido, do Canadá, da França, da Alemanha, da Itália e do Japão assumiram entre outras responsabilidades, o compromisso diante da crise humanitária que acontecia no planeta, em especial na África por conta da insegurança alimentar e do desenvolvimento não sustentável. O G8 determinou nessa data que, a partir de seus países membros iriam compartilhar de forma aberta, irrestrita e oportuna, dados relacionados à Agricultura e Nutrição, com o objetivo de serem usados em favor do progresso e da melhoria dos processos agrícolas, pecuários, nutricionais, e, de desenvolvimento sustentável, nos países menos avançados nestes aspectos (GODAN, 2017).

² Reunião de chefes de Estado das principais democracias industriais, na qual são abordados temas políticos, económicos, energéticos e de segurança (UNIVERSITY OF TORONTO, 2017).

A primeira decisão foi a criação de uma conferência internacional que abordasse os dados agrícolas em formato aberto. Assim, em abril de 2013 realizou-se a primeira conferência com sede no Reino Unido. Nesse evento os governos participantes se comprometeram a elaborar os planos de ação em cada país, como apoio à iniciativa citada.

Como subproduto desta primeira reunião criou-se a iniciativa *Global Open Data for Agriculture & Nutrition* (GODAN), e o seu site foi disponibilizado no endereço godan.info em outubro do mesmo ano.

A estratégia GODAN age como pivô do nível governamental e de organizações geradoras, administradoras ou propulsoras dos dados abertos em favor da segurança alimentar e nutricional com ênfase no desenvolvimento sustentável, mediante a junção de esforços entre os seus membros.

Na sequência, derivado das iniciativas anteriormente descritas para a gestão das conferências internacionais sobre dados abertos, ampliando o foco para além dos dados agrícolas, foi criado o site opendatacon.org, com o objetivo de se estabelecer como um ponto de encontro para a comunidade global debater e estudar o futuro dos dados abertos. Os administradores do site têm publicado informações relacionadas às conferências realizadas nos anos de 2015, em Otawa, Canadá e de 2016, em Madrid, Espanha; divulgando os eventos atuais e convidando à comunidade para participação no próximo evento que acontecerá no ano 2018, na Argentina.

3.2 Plataformas como opções pró-implementação de sites de dados abertos governamentais

As seguintes plataformas tecnológicas estão sendo utilizadas como ferramentas nas fases de planejamento e implementação de estratégias relacionadas com a publicação de dados abertos por governos estatais ou mesmo regionais.

Iniciando com a plataforma para a administração de sites de dados abertos SOCRATA, corresponde a uma plataforma para gestão de dados e serviços em nuvem dirigida aos órgãos oficiais tais como governos de cidades, estados e países. SOCRATA é um produto comercial, que requer o pagamento dos custos relacionados a sua implementação e manutenção (SOCRATA, 2017).

Essa empresa foi criada em 2007, por Kevin Merritt, e encontra-se localizada nos Estados Unidos. O seu objetivo principal é estabelecer uma plataforma para organizações

públicas que lhes permita administrar e compartilhar facilmente os dados governamentais. Suas funcionalidades incluem opções de visualização e análise dos dados com a finalidade de oferecer entendimento e aproximação a esses recursos informacionais. Alguns dos clientes atuais da SOCRATA são o site de dados do Banco Interamericano de Desenvolvimento (BID)³, o governo da Colômbia⁴, e sites de dados de cidades como New York⁵ e Chicago⁶ nos Estados Unidos (SOCRATA, 2017).

Por ser uma plataforma paga, a SOCRATA fornece uma implementação mais ágil, o qual supõe um menor tempo em sua implementação. Isto pela centralidade da equipe técnica, ou seja, que oferece rapidez da operacionalização da solução informática, embora seja associada também com uma arriscada e alta dependência desta organização que fornece a plataforma (WORLD BANK GROUP, 2015).

Outra plataforma para a administração de sites de dados abertos é a *Comprehensive Knowledge Archive Network* (CKAN), a qual é uma plataforma gratuita que permite criar sites para a publicação de dados abertos, plataforma *Open Source*, isto é, com código fonte disponível. Ela é desenvolvida e promovida pela *Open Knowledge Foundation* (OKFN), e possibilita administrar e publicar catálogos de dados e *datasets*. Pelo fato do CKAN ser desenvolvida pela OKFN, a fundação recebe contribuições de uma comunidade de desenvolvedores que contribuem com extensões para aprimorar ou customizar o CKAN (OKFN, 2017).

O público alvo desta plataforma são governos nacionais ou locais, e também as instituições de pesquisa e organizações que coletam grande quantidade de dados. Na atualidade alguns dos usuários do CKAN são os governos do México⁷, dos Estados Unidos⁸, do Canadá⁹, do Brasil¹⁰, e também o portal europeu de dados (COMPREHENSIVE KNOWLEDGE ARCHIVE NETWORK, 2017).

Sendo uma plataforma gratuita, não restringe funcionalidades primordiais na administração de dados, tais como funções de busca facetadas e diversas ferramentas para a visualização dos dados pelos utilizadores, tanto internos como externos. Sua

³ Cf. <https://data.iadb.org/>.

⁴ Cf. <https://datos.gov.co/>.

⁵ Cf. <https://data.ny.gov/>.

⁶ Cf. <https://data.cityofchicago.org/>.

⁷ Cf. <https://datos.gob.mx/>.

⁸ Cf. <https://www.data.gov/>.

⁹ Cf. <http://open.canada.ca/en>.

¹⁰ Cf. <http://dados.gov.br/>.

arquitetura modular permite desenvolver características adicionais como a integração com as plataformas Google+, Twitter, Facebook, acompanhamento mediante RSS, funções geoespaciais e coleta de dados *harvesting* (COMPREHENSIVE..., 2017).

A facilidade de cada organização publicadora de dados poder administrar o CKAN, requer um esforço maior nos aspectos técnicos do seu processo de implementação e manutenção, desde os processos próprios nas adaptações internas até a gestão com a comunidade de desenvolvedores da OKFN. No entanto, cabe destacar que ao adotar o CKAN como plataforma, inicia-se a aproximação com toda a comunidade CKAN o que traz como vantagem a possibilidade de que novas instâncias possam ser desenvolvidas a partir de contribuições de outras instâncias (WORLD BANK GROUP, 2015).

4 MECANISMOS DE REPRESENTAÇÃO NO CONTEXTO WEB

Nesta seção serão abordados os mecanismos tecnológicos que auxiliam processos de representação usados no âmbito de *datasets* governamentais. Percebe-se que esses mecanismos têm adquirido relevância pelo fato de fazer parte do grupo de recomendações do consórcio *World Wide Web Consortium* (W3C).

O W3C é um consórcio internacional de organizações e do público em geral, que atua no desenvolvimento de padrões para a web¹¹. Foi fundado em 1994, por Tim Berners-Lee em conjunto com a *European Organization for Nuclear Research* (CERN¹²), com a colaboração da *Defense Advanced Research Project Agency* (DARPA), dos Estados Unidos e da Comunidade Europeia. O consórcio movimenta-se visando o cumprimento de sua missão: “guiar a web para seu máximo potencial por meio do desenvolvimento de protocolos e pautas que assegurem o crescimento da web” (W3C, 2017).

O W3C pretende alcançar a sua missão mediante a criação de padrões denominados por ele como recomendações, sendo um dos seus pilares o desenvolvimento de padrões abertos, não proprietários, buscando evitar assim a fragmentação do mercado e da web em si.

Alguns dos mecanismos que vêm sendo utilizados no âmbito da representação de recursos informacionais em sites de dados governamentais serão descritos nas próximas seções.

4.1 Dublin Core

Em 1995, com o acréscimo na quantidade de recursos no contexto digital oferecido pela internet, foi criado o padrão de metadados *Dublin Core*. Segundo Alves e Santos (2013) o Dublin Core é um padrão de propósito geral, simples e estruturado, criado com o intuito de promover a descoberta de recursos na internet.

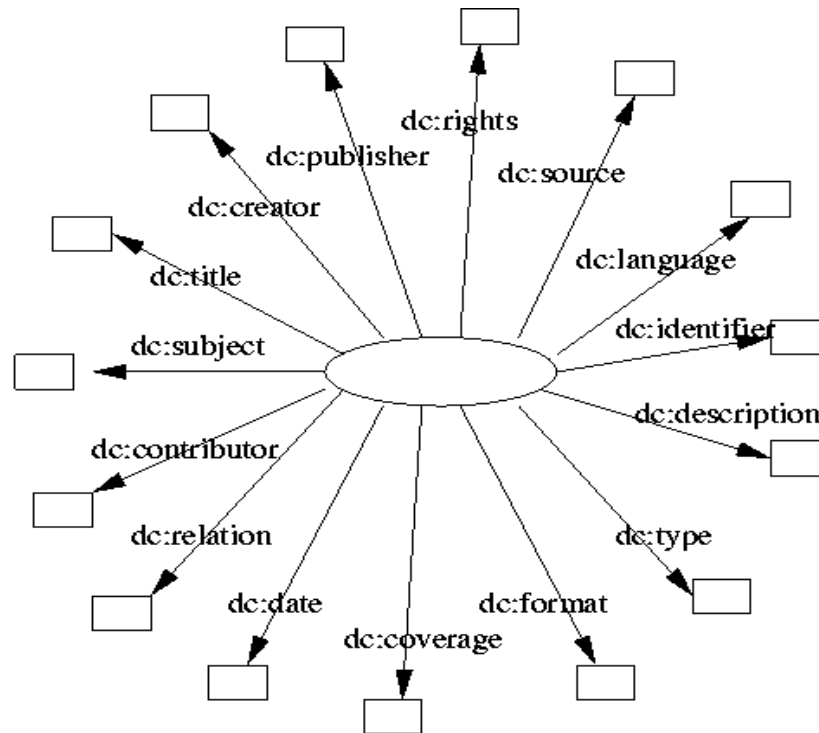
Este padrão foi estabelecido e é mantido pela *Dublin Core Metadata Initiatives* (DCMI), organização aberta que contribui com a inovação no *design* de metadados e as melhores práticas no seu âmbito (DUBLIN CORE METADATA INITIATIVE, 2017).

¹¹ Sistema hipertextual que funciona baseada na internet.

¹² A sigla corresponde ao antigo nome oficial *Conseil Européen pour la Recherche Nucléaire*.

O *Dublin Core* é composto por um conjunto de elementos de metadados, nomeado pela DCMI como um “[...] vocabulário de quinze propriedades para usar na descrição de recursos.” (DUBLIN CORE METADATA INITIATIVE, 2017, tradução nossa). A Figura 6 apresenta os elementos do Dublin Core.

Figura 6 – Esquema gráfico básico do *Dublin Core*



Fonte: DCMI (2017).

Esses quinze elementos são conhecidos como o *Dublin Core Metadata Element Set* (DCMES¹³). Esse conjunto de metadados faz parte do grande conjunto de vocabulários de metadados e especificações da DCMI, nomeado como o *DCMI Metadata Terms* (DCMI-TERMS). Ou seja, na DCMI, o DCMES não é o único conjunto de elementos: os demais elementos da iniciativa (vocabulários, termos, e outros elementos sintáticos) compõem o conglomerado da iniciativa e estão planejados para serem compatíveis entre si na construção de perfis de aplicação (DUBLIN CORE METADATA INITIATIVE, 2017).

A seção 4.5 traz uma aproximação aos perfis de aplicação, que apesar de não ser o foco deste estudo, permite melhor contextualizar a importância dos elementos que o compõem.

¹³ Conjunto de elementos de metadados do *Dublin Core*.

4.2 O Friend of a Friend

A *National Information Standards Organization* (NISO, 2017), organização sem fins lucrativos que identifica, desenvolve, mantém e publica padrões técnicos para a administração de informação no contexto digital (NISO, 2017), define o *Friend of a Friend* (FOAF) como um vocabulário para uso na internet, que fornece metadados descritivos a respeito de pessoas e organizações. A sua primeira versão surgiu em abril de 2005 e na atualidade já está sendo reutilizado por 362 vocabulários no entorno web (LINKED OPEN DATA VOCABULARIES, 2017).

O FOAF encontra-se codificado no *Resource Description Framework* (RDF), e o seu core focaliza-se na descrição de propriedades que especificam como as pessoas e as organizações, comunidades, e/ou objetos, se relacionam entre si no contexto da internet.

4.3 Simple Knowledge Organization System

O Sistema de Organização do Conhecimento Simples, tradução na língua portuguesa de *Simple Knowledge Organization System* (SKOS), é definido como um vocabulário, já recomendado pelo W3C, que fornece um modelo de dados para a disponibilização e compartilhamento de sistemas de organização do conhecimento na Web. O SKOS encontra-se estruturado no RDF e permite representar taxonomias, tesouros, listas de cabeçalhos de assunto, esquemas de classificação e outras formas de organização do conhecimento (W3C, 2017).

A primeira versão do SKOS surgiu em 2004, sendo postulado como recomendação ante o W3C e, finalmente, se converteu em recomendação desse consórcio em agosto de 2009 (LINKED OPEN DATA VOCABULARIES, 2017).

O SKOS tem o intuito de favorecer a reutilização e interoperabilidade entre os vocabulários controlados. Ele foi criado como uma opção para a transição dos sistemas de organização de conhecimento tradicionais para o ambiente da internet, permitindo a representação de elementos semânticos de recursos informacionais, “SKOS fornece um elo entre o formalismo lógico rigoroso das linguagens de ontologias e o mundo caótico, informal e pouco estruturado de ferramentas de colaboração baseadas na Web” (RAMALHO, 2015, p. 68).

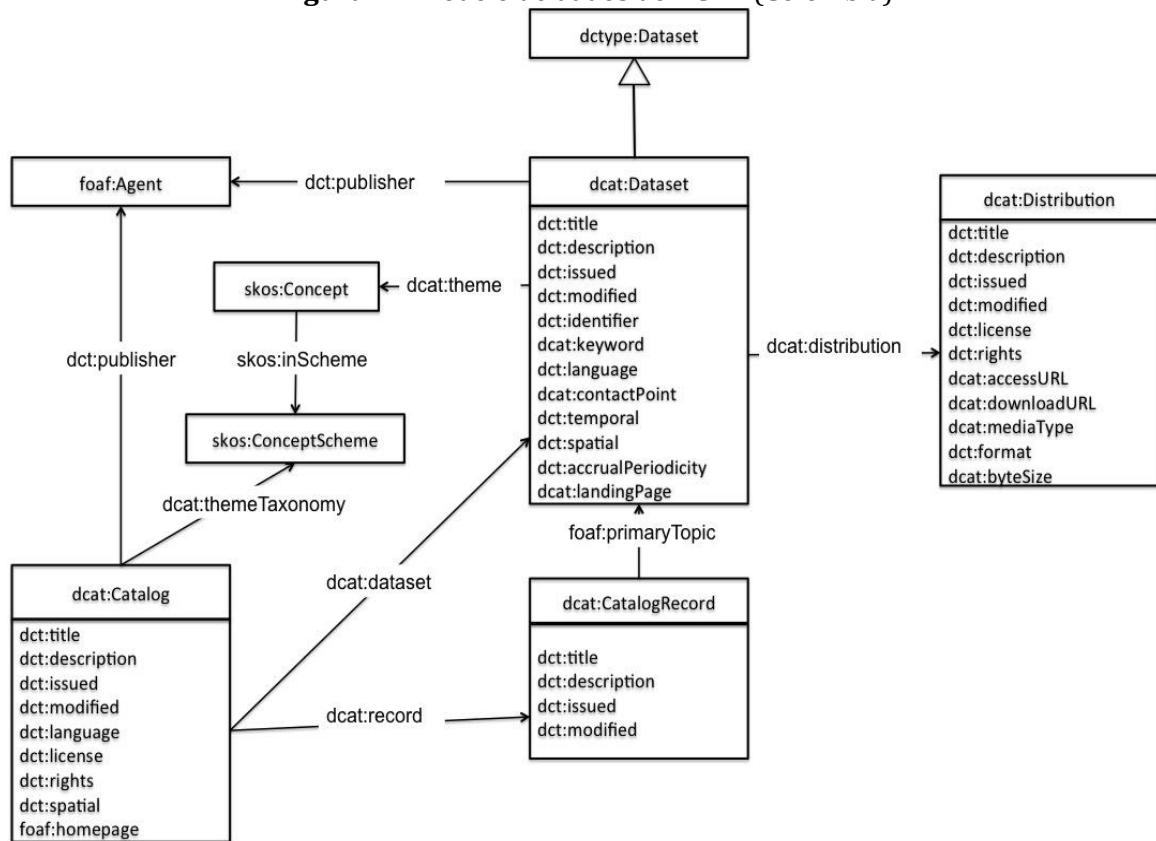
Na atualidade, comunidades que compartilham foco em determinados domínios do conhecimento estão adotando o SKOS como mecanismo para representar seus vocabulários na busca da interoperabilidade com outras instâncias aproveitando o potencial da rede internet.

4.4 Data Catalog Vocabulary

O *Data Catalog Vocabulary* (DCAT) é “um vocabulário RDF criado para facilitar a interoperabilidade entre catálogos de dados” (W3C, 2017). Foi criado no *Digital Enterprise Research Institute* (DERI), padronizado no grupo de trabalho *Government Linked Data* (GLD) do W3C, consórcio que na atualidade fornece as recomendações e atualizações acerca deste vocabulário.

O DCAT é usado para descrever *datasets* em catálogos de dados, buscando melhorar a descoberta de recursos, a disponibilização descentralizada de catálogos e a curadoria digital (W3C, 2017). Ainda, incorpora metadados de outros vocabulários que já existem, como nos casos do *Dublin Core* e o FOAF. A Figura 7, apresenta o modelo de dados do vocabulário DCAT.

Figura 7 – Modelo de dados do DCAT (Colômbia)



Fonte: W3C (2017).

Percebem-se no modelo elementos inseridos a partir de outros esquemas, quanto maioritariamente o Dublin Core, mas também do SKOS e FOAF.

4.5 GEONAMES

O GEONAMES é um banco de dados geográficos, disponível para uso gratuito, que atualizado diariamente fornece dados que descrevem lugares, acidentes geográficos, divisões político-administrativas, dentre outras entidades localizáveis geograficamente (GEONAMES, 2017).

Existe a possibilidade de obter características das entidades geolocalizáveis tais como nomes em várias línguas; coordenadas geográficas; população; se for o caso, altitude, fuso horário, etc. O GEONAMES utiliza sistemas de codificação para países conforme normas internacionais e sistemas de codificação próprios que permitem identificar classes dos lugares, como também descrever classificações de subtipos de divisões administrativas (GEONAMES, 2017).

Os dados do GEONAMES são fornecidos por todos os continentes do mundo, mediante instituições tanto públicas como privadas, tendo como exemplo delas o Instituto Brasileiro de Geografia e Estatística (IBGE), do Brasil; o *Departamento Administrativo Nacional de Estadística* (DANE), da Colômbia; *The European Environment Agency*, da Comunidade Europeia, dentre outras.

O intuito é manter um banco de dados atualizado que proporcione dados estruturados que possam agir como elos na linkagem de dados na web.

4.6 Desenvolvimento de perfil de aplicação

“O uso de vocabulários de esquemas internacionalmente reconhecidos, como o caso do Dublin Core e do FOAF, facilita a interoperabilidade entre sistemas.” (ARAKAKI, 2016, p. 35). No entanto, as organizações ou os entornos onde eles vão agir, necessitam de modelos de dados que apoiem essas necessidades de interoperabilidade, prática que implica métodos para o estabelecimento e a concepção desses modelos de dados, que por sua vez, vão se utilizar desses vocabulários nos seus processos descritivos (MALTA, 2014).

Ainda, esses modelos de dados vêm a abordar a questão da inexistência de um padrão de metadados que funcione para várias comunidades em geral, por isso, Coyle e Baker (2009, tradução nossa): “Quando se trata de metadados, um tamanho não serve para todos”.

Considerando os domínios do conhecimento, diversos e dinâmicos por natureza, os seus parâmetros de ordem, correspondem às “formas de representação construídas para descrever o conteúdo e a forma da informação corporificada, registrada, aquela passível de tratamento, armazenamento, acesso, recuperação, uso, reuso e preservação” (SANTOS, 2017, p. 4).

Os perfis de aplicação vêm como um auxílio na necessidade de adaptação dos conjuntos de metadados existentes, sendo que esses metadados agem como parâmetros de ordem dentro de um domínio de conhecimento determinado.

Os parâmetros de ordem como formas de representação, encontram-se em ambientes informacionais, integrando assim sistemas informacionais dinâmicos. Esse dinamismo se dá porque os parâmetros de ordem auxiliam e são utilizados nas interações próprias dos padrões de comportamento. Esses padrões de comportamento são as

diversas manifestações da sociedade, onde as interações estão permanentemente mudando, pelos indivíduos e as necessidades que as compõem e dinamizam (SANTOS, 2017).

É por isso que apesar de existirem conjuntos de metadados construídos de forma clara, eles têm um alcance limitado a um domínio de conhecimento, ou comunidades de prática como nomeadas por Duval et al. (2002), devido às necessidades descritivas específicas que abrangem esses metadados e pelas aplicações particulares, as quais tendem a ser diversas (COYLE; BAKER, 2009).

Neste contexto, surgem os perfis de aplicação para facilitar a interação entre distintos conjuntos de metadados. Eles são então, “um conjunto de metadados selecionados de um ou mais esquemas de metadados e combinados em um esquema composto” (DUVAL et al., 2002, tradução nossa). O seu propósito é adaptar esquemas já existentes tendo como premissa manter a interoperabilidade entre os novos esquemas criados e os que têm dado fundamento a esses novos esquemas (DUVAL et al., 2002, tradução nossa).

Além disso, os perfis de aplicação apresentam a forma como esses novos esquemas de metadados podem ser usados em um contexto que tenha as suas próprias regras de uso.

Um perfil de aplicação é então “um constructo genérico para desenhar registros de metadados” (BAKER; COYLE, 2013). Eles correspondem a um método para expressar padrões de metadados que sejam independentes das plataformas e que descrevam a estrutura e o conteúdo dos dados. Os perfis de aplicação usam vocabulários RDF já existentes, e especificam como seus elementos se restringem e se empacotam num contexto colaborativo (BAKER; COYLE, 2013).

Souto (2014) apresentou um perfil de aplicação específico para ser usado em um contexto de descrição arquivística, que atingisse uma norma internacional específica desse domínio do conhecimento. Esse perfil de aplicação procurou a descrição de unidades arquivísticas utilizando um vocabulário próprio comum, baseado no modelo RDF.

A própria DCMI mantém diretrizes para a criação de perfis de aplicação alicerçadas em seus conjuntos de metadados. O site da iniciativa apresenta orientações para a criação de perfis de aplicação *Dublin Core*, explicando quais os componentes significativos e o

processo do desenvolvimento desses perfis, intitulados por eles como *Dublin Core Application Profile* (DCAP) (DUBLIN CORE METADATA INITIATIVE, 2017).

5 RELATO DOS ESTUDOS DE CASO

Finalizada a análise de conteúdo nas instruções recuperadas por cada país, as seções a seguir apresentam o resultado em cada uma das fases propostas para a obtenção dos objetivos da pesquisa. Individualizou-se a situação de cada país estudado, para posteriormente discutir acerca do panorama global.

Na sequência, seguindo a metodologia proposta, serão desdobrados os resultados da pesquisa.

5.1 Resultados da fase pré-análise

Devido à configuração dos estudos de caso, que buscam a caracterização de cada país, pormenorizou-se cada descoberta de documentos:

5.1.1 Colômbia

Realizou-se navegação pelo site de dados datos.gov.co, no qual, por meio das indicações dos seus sistemas de organização, optou-se pela opção do menu nomeada *Herramientas*¹⁴ (Ferramentas). Essa funcionalidade do site dá acesso a documentos de diferentes tipos que descrevem o desenvolvimento de atividades que envolvem a gestão de *datasets*. Os documentos encontram-se distribuídos nas categorias: *Aprender sobre datos* (Aprender sobre dados); *Usar los datos* (Usar os dados); *Publicar conjuntos de datos* (Publicar conjuntos de dados); *Desarrollar usando los datos* (Desenvolver usando dados); e, *Preguntas Frecuentes* (Perguntas frequentes).

Após leitura da descrição de cada categoria elencada, observou-se que a categoria *Aprender sobre datos* está diretamente relacionada com os objetivos desta pesquisa, disponibilizando documentos em arquivos PDF classificados em: *Conferencias* (Conferências); *Estudios* (Estudos); *Guías* (Guias); *Manual* (Manual); e, *Presentación* (Apresentação). Todos os documentos têm a intenção de formar o capital humano que participa da gestão de *datasets*, seja como fornecedores ou consumidores de dados.

Nas subcategorias *Guías* e *Manual*, foram localizados onze documentos que contêm informação acerca de processos relacionados especificamente com a publicação de

¹⁴ Cf. <https://herramientas.datos.gov.co/es/content/aprender-sobre-datos-0>.

datasets. Considerando que as regras definidas para a distinção dos documentos foram as de exaustividade; representatividade; homogeneidade e pertinência (BARDIN, 2010). Após leitura seletiva, aplicando esses critérios de distinção, determinou-se considerar os seguintes documentos, os quais atendem o objetivo da pesquisa quanto a representação dos *datasets*:

- a) *Guía de datos abiertos en Colombia* (COLOMBIA, 2016a);
- b) *Guía para la estandarización de Datos Abiertos* (COLOMBIA, 2016b);
- c) *Manual de datos abiertos en Colombia para Publicador* (COLOMBIA, 2016c), e
- d) *Manual de datos abiertos en Colombia para Administrador* (COLOMBIA, 2016d).

5.1.2 Brasil

O site de dados do governo do Brasil, dados.gov.br, tem no *footer menu* (menu inferior) as opções: O Portal; Comunidades; Manuais; e, Saiba Mais. Dentre elas, a opção Manuais, ocupa-se de fornecer recursos para a gestão de *datasets* no portal brasileiro, abrangendo recomendações originadas no W3C.

Os recursos encontram-se dispostos na forma de lista, visando indicar, com o título de cada recurso, o público ao qual está dirigido. Realizando leitura seletiva nos 14 recursos elencados nesse ponto do site, identificaram-se dois recursos que fornecem instruções diretas no que se refere à descrição de *datasets*, e, portanto, contribuem na efetivação desta pesquisa.

Seguindo a proposta para a distinção dos documentos anunciada na metodologia, determinou-se considerar os seguintes documentos:

- a) *Cartilha Técnica para Publicação de Dados Abertos no Brasil* (BRASIL, 2012) e
- b) *Manual de Catalogação no Portal Brasileiro de Dados Abertos* (BRASIL, 2016b).

5.1.3 Espanha

O governo da Espanha disponibiliza os dados abertos do setor público no site datos.gob.es. No ano 2009 o governo desse país, por meio da iniciativa Aporta¹⁵, dá começo às ações que buscam motivar a disponibilização de recursos informacionais oficiais, e um dos resultados da iniciativa é a implementação do site de dados abertos desse país ibérico.

Sendo as opções do menu principal do site: *Inicio* (Início), *Iniciativa Aporta* (Iniciativa Aporta), *Catálogo de datos* (Catálogo de dados), *Impacto* (Impacto), *Interactúa* (Interage), e, *Actualidad* (Atualidade), observou-se na exploração, que a opção *Interactúa*, compreende recursos informacionais concernentes à gestão de *datasets*.

Para tal acesso é preciso escolher a opção *Documentación* (Documentação), que por sua vez tem nos tipos de documentos as opções: *Informes y estudios* (Informes e estudos), *Guías* (Guias), *Materiales formativos* (Materiais formativos), *Estrategias RISP* (*Reutilización de la Información del Sector Público*) (Estratégias RISP), e, *Normativa* (Normativa).

Foram considerados somente os documentos que disponíveis na opção *Guías*, observando em um primeiro momento que a maior parte dos documentos disponibilizados no site da Espanha estão simultaneamente em formatos de arquivo PDF, DOCX, ODT, tendo em algumas ocasiões versão em castelhano e no inglês. Percebeu-se ainda que embora o menu do site tenha possibilidade de escolher dentre os idiomas: Espanhol, Inglês, Catalão, Galego e Euskera, com o recurso de alterar a aparência do site, no entanto o conteúdo dos documentos se mantém em castelhano, ou em inglês em alguns casos.

Distinguiram-se nesse ponto do site de dados, 14 documentos relacionados com a gestão de *datasets*, dos quais oito corroboram o interesse deste estudo:

- a) *Manual de Usuario: Perfil Organismo Público* (ESPAÑA, 2017a);
- b) *Plantilla RDF/XML para la federación de conjuntos de datos* (ESPAÑA, 2017b);
- c) *Orientaciones y garantías en los procedimientos de anonimización de datos personales* (ESPAÑA, 2016b);

¹⁵ Cf. <http://datos.gob.es/es/acerca-de-la-iniciativa-aporta>.

- d) *Orientaciones sobre la protección de datos en la reutilización de la información del sector público* (ESPAÑA, 2016b);
- e) *Guía de aplicación de la Norma Técnica de Interoperabilidad* (ESPAÑA, 2016c);
- f) *Como publicar datos abiertos de manera rápida y sencilla* (con CKAN) (ESPAÑA, 2015); e,
- g) *Guía de aplicación del Real Decreto 1495/2011* (ESPAÑA, 2012).

5.1.4 Portugal

A Agência para a Modernização Administrativa (AMA) é um órgão vinculado ao governo de Portugal, encarregado da gestão do site de dados dessa república, dados.gov.pr. Após navegação para a descoberta dos recursos, percebeu-se que o site de dados de Portugal não tem uma opção específica que contenha documentação sobre processos de publicação de *datasets*.

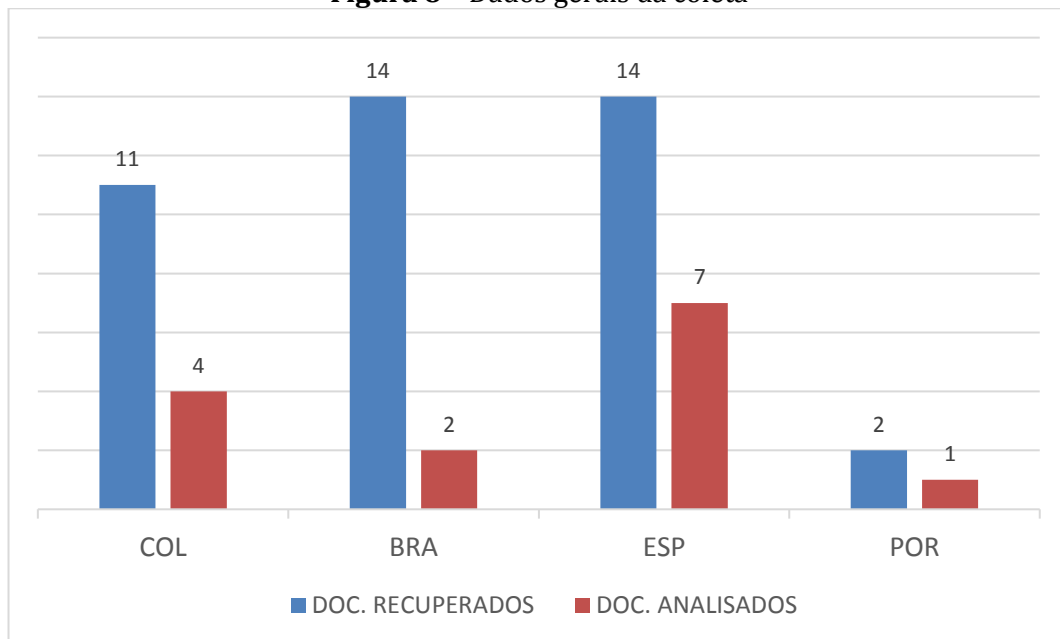
Existem dois recursos documentais no site, os quais foram disponibilizados mediante a divulgação de uma notícia datada de novembro de 2016, e apenas são acessíveis nos links contidos na respectiva nota informativa.

Desses recursos, somente um se enquadra na descrição de *datasets*: Guia Dados abertos (PORTUGAL, 2016).

5.1.5 Dados gerais na coleta do *corpus*

Após coletar todos os documentos que compõem o *corpus* da pesquisa em cada site dos países selecionados, têm-se alguns dados gerais a destacar. Na descoberta de recursos foram recuperados 41 documentos que contém instruções generalizadas, dos quais foram analisados efetivamente 14 documentos, por conter instruções específicas para a descrição de *datasets*. Assim, do total de documentos disponibilizados, 35% contém informações específicas para publicadores de dados acerca dos processos descritivos de seus recursos informacionais tipo *dataset*.

Na Figura 8 indicam-se para cada país os valores correspondentes aos documentos recuperados e analisados.

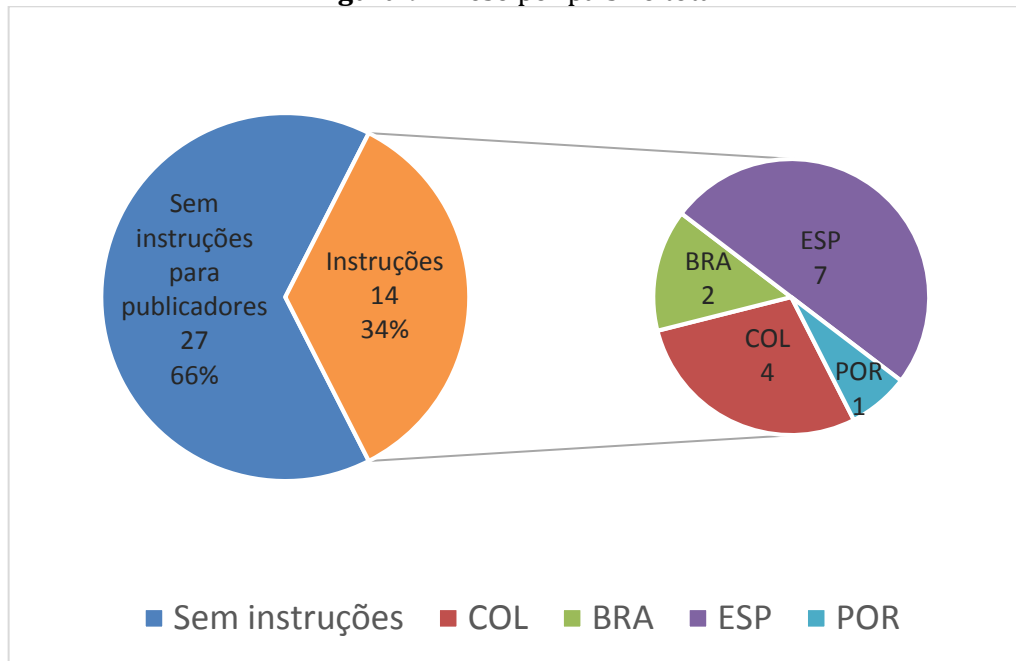
Figura 8 – Dados gerais da coleta

Fonte: Dados da pesquisa.

Utilizaram-se para cada país as siglas: Colômbia (COL), Brasil (BRA), Espanha (ESP), e Portugal (POR).

Para cada país foi especificada a quantidade de documentos que foram recuperados, e posteriormente, abordados mediante a técnica de análise de conteúdo proposta, observando que o país com maior número de documentos dispostos com instruções para os publicadores de dados é a Espanha, seguida, na ordem, por Colômbia, Brasil e Portugal.

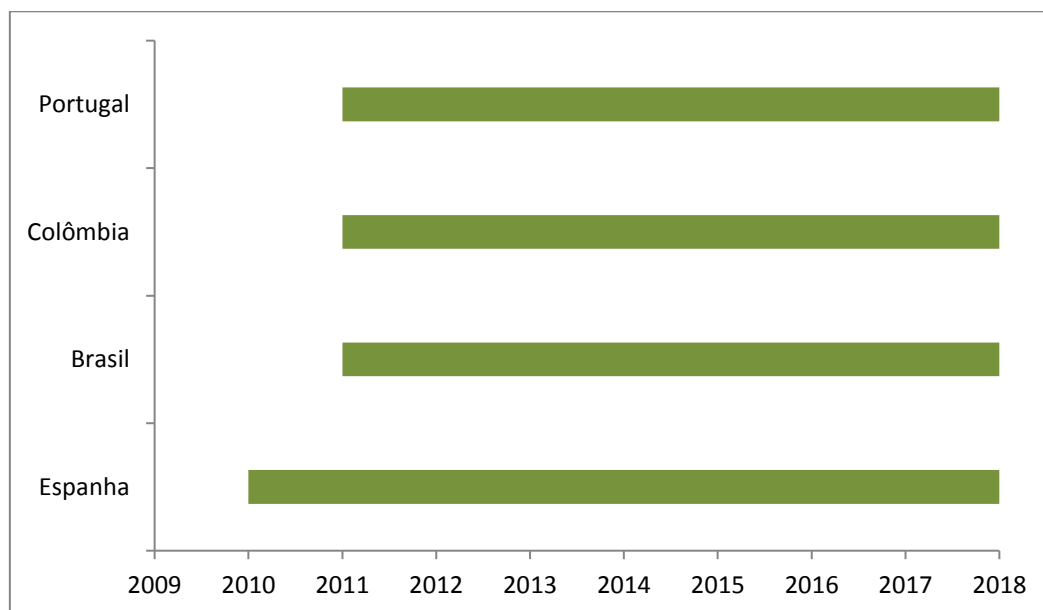
A Figura 9 auxilia na observação dos resultados percentuais com relação aos valores globais da pesquisa, no que se refere às quantidades de documentos que compuseram o *corpus* da pesquisa.

Figura 9 – Peso por país no total

Fonte: dados da pesquisa.

Apresenta-se a distribuição por país dos documentos que serão analisados, observando que a maior quantidade de documentos foi recuperada no site de dados da Espanha, correspondendo a 50% dos documentos objeto deste estudo. Na sequência, apresentam-se Colômbia com 28,6%; Brasil com 14,3% e Portugal com apenas 7,1% dos documentos que serão analisados por conter instruções específicas acerca da descrição de *datasets*.

Ainda, em relação à quantidade total de documentos recuperados nesta fase de descoberta de recursos, é notória a pouca documentação disponibilizada no site de dados de Portugal. Na Figura 10 é possível observar a data de início de disponibilização do site de dados de cada país, objeto deste estudo.

Figura 10 – Gênese de cada site de dados abertos de governo

Fonte: Dados da pesquisa.

Observando que a disponibilização dos sites de dados governamentais dos países estudados aconteceu quase que de maneira concomitante, e ainda, estão no ar até a presente data (2018), chama a atenção a diferença relacionada com a quantidade de instrumentos que o site de dados de Portugal está disponibilizando, número bastante inferior no que se refere a quantidade publicada pelos outros três países analisados neste estudo.

Na seção seguinte, procedeu-se com a revisão do *corpus* da pesquisa, acima citado.

5.2 Resultados da exploração do material

Acatando o procedimento metodológico proposto para esta pesquisa, realizou-se a gestão do *corpus* da pesquisa (BARDIN, 2010) com a exploração do material. Nos documentos elencados na pré-análise, identificou-se o conteúdo da unidade temática **metadados**, com o intuito de distinguir os núcleos de sentido presentes nos documentos com as instruções para publicadores de dados.

Aplicou-se de forma sistemática, conforme a proposta detalhada na metodologia, a observação dos trechos em que foi abordado o tratamento descritivo dos *datasets*, no caso da Colômbia, considerando neste caso o termo *metadatos*, pela variação na língua castelhana.

Com relação à operação adotada para esta pesquisa, codificação para o tratamento do *corpus* da pesquisa, foram observadas as categorias definidas a partir do estudo de Santos et al. (2013), detalhadas na seção 1.2.2, buscando classificar e organizar os documentos analisados, visando “o elo entre os dados dos textos e a teoria” (BARDIN, 2010, p. 129), teoria discutida no referencial teórico deste documento.

De acordo com Bardin (2010), uma vez definida a unidade de registro, neste caso **unidade temática**, faz-se necessário estabelecer as regras de enumeração ou o modo de contagem. Como regras, nesta etapa, foram adotadas a presença ou ausência das categorias: Metadados Teoria e Metadados no Uso; e, a intensidade dessas abordagens, observando segundo a intensidade semântica dos sintagmas nos elementos da análise, ou seja, das categorias adotadas.

Como resultado desta exploração do material, apresenta-se no Quadro 4 a sinopse dos dados brutos obtidos pela aplicação sistemática do procedimento metodológico proposto.

Quadro 4 – Caracterização do *corpus* da pesquisa

Nome do documento	Detentor	Tipo de documento	Data	Função	Fundamenta-se em LAI	Metadados Teoria Conceitos (C), Fundamenta- ções (F), Técnicas (T), Estudos (E), Explicações (EX), Críticas (CR)	Intensidad e Alta (A), Média (M), Baixa (B)	Metadados Uso Experimentos (EXP), Relatos de experiência (RE), Estudos de caso (EC), Recomendações (R), Percepções de uso (PU)	Intensidad e Alta (A), Média (M), Baixa (B)	Metadado s Estruturais	Metadados Descritivos
<i>Guía de datos abiertos en Colombia</i>	MINTIC – Colômbia	Guia	2016	Fornece diretrizes e boas práticas para o desenvolvimento de estratégias de dados abertos	Lei 1712 de 2014, Lei de transparência e acesso à informação	C	B	R PU	B B	-	-
<i>Guía para la estandarización de datos abiertos</i>	MINTIC – Colômbia	Guia	2016	Fornece recomendações e alinhamentos para a padronização de dados abertos	-	C EX	B M	R PU	M B	DCAT	GEONAMES
<i>Manual de datos abiertos en Colombia para publicador</i>	MINTIC – Colômbia	Manual	2016	Fornece instruções para o desenvolvimento de: gestões no <i>front-end</i> para informação da sua organização, e personalização de funcionalidades de gestão nos <i>datasets</i>	-	-	-	R	A	-	-

Nome do documento	Detentor	Tipo de documento	Data	Função	Fundamenta-se em LAI	Metadados Teoria Conceitos (C), Fundamentações (F), Técnicas (T), Estudos (E), Explicações (EX), Críticas (CR)	Intensidade Alta (A), Média (M), Baixa (B)	Metadados Uso Experimentos (EXP), Relatos de experiência (RE), Estudos de caso (EC), Recomendações (R), Percepções de uso (PU)	Intensidade Alta (A), Média (M), Baixa (B)	Metadados Estruturais	Metadados Descritivos
<i>Manual Datos abiertos en Colombia para Administrador</i>	MINTIC – Colômbia	Manual	2016	Fornecer instruções para o desenvolvimento de: gestões no <i>front-end</i> como o CRUD ¹⁶ de dados da organização, e personalização de funcionalidades de gestão nos <i>datasets</i> e da interface do site	-	-	-	R	M	-	-
Cartilha Técnica para Publicação de Dados Abertos no Brasil	Ministério do Planejamento Orçamento e Gestão - Brasil	Cartilha	2012	Fornecer contexto inicial e diretrizes básicas para publicar <i>datasets</i> . Introduz conceitos fundamentados na Lei de acesso à informação	Unicamente menciona a Lei de Acesso à Informação 12.527/2011, sem especificações	C	B	-	-	-	Vocabulário VCGE

¹⁶ Operações *Create, Read, Update e Delete* (CRUD)

Nome do documento	Detentor	Tipo de documento	Data	Função	Fundamenta-se em LAI	Metadados Teoria Conceitos (C), Fundamentações (F), Técnicas (T), Estudos (E), Explicações (EX), Críticas (CR)	Intensidade Alta (A), Média (M), Baixa (B)	Metadados Uso Experimentos (EXP), Relatos de experiência (RE), Estudos de caso (EC), Recomendações (R), Percepções de uso (PU)	Intensidade Alta (A), Média (M), Baixa (B)	Metadados Estruturais	Metadados Descritivos
Manual de Catalogação no Portal Brasileiro de Dados Abertos	Ministério de Planejamento desenvolvimento e gestão - Brasil	Manual	2016	Apoiar órgãos do Poder Executivo Federal, e demais instituições públicas, na catalogação dos dados abertos a publicarem no site de dados do governo brasileiro	Indica a publicação do Decreto nº 8.777/2016, que instituiu a Política de Dados Abertos, fundamentado na LAI, sem dar especificidades sobre dita lei	-	-	PU	B	-	-
<i>Manual de Usuario: Perfil Organismo Público</i>	<i>Ministerio de Hacienda y Función Pública, Ministerio de Energía, Turismo y Agenda Digital Iniciativa Aporta, Red.es - Espanha</i>	Manual	2017	Manual criado especificamente para órgãos do Estado espanhol, envolvidos na publicação de dados.	-	EX	A	R RE PU	B B A	DCAT	Vocabulários SKOS próprios e recomendações para o valor dos metadados, que chama como "Recursos de estilo"
<i>Plantilla RDF/XML para la federación de conjuntos de datos</i>	<i>Ministerio de Hacienda y Función Pública, Ministerio de Energía, Turismo y Agenda Digital - Espanha</i>	Template	2017	Fornece template RDF/XML com estrutura para os dados de catálogos, datasets e distribuições.	-	EX	A	-	-	DCAT	Vocabulários SKOS próprios e recomendações que chama como "Recursos de estilo"

Nome do documento	Detentor	Tipo de documento	Data	Função	Fundamenta-se em LAI	Metadados Teoria Conceitos (C), Fundamentaões (F), Técnicas (T), Estudos (E), Explicações (EX), Críticas (CR)	Intensidade Alta (A), Média (M), Baixa (B)	Metadados Uso Experimentos (EXP), Relatos de experiência (RE), Estudos de caso (EC), Recomendações (R), Percepções de uso (PU)	Intensidade Alta (A), Média (M), Baixa (B)	Metadados Estruturais	Metadados Descritivos
<i>Orientaciones y garantías en los procedimientos de anonimización de datos personales</i>	<i>Agencia Española de protección de datos - Espanha</i>	Guia ¹⁷	2016	Entregar diretrizes que coadjuvem na proteção de dados pessoais	-	-	-	PU	B	-	-
<i>Orientaciones sobre la protección de datos en la reutilización de la información del sector público</i>	<i>Agencia Española de protección de datos - Espanha</i>	Guia	2016	Fornecer diretrizes de proteção de dados pessoais no contexto da reutilização de informação pública	Lei 37/2007	-	-	PU	B	-	-
<i>Guía de aplicación de la Norma Técnica de Interoperabilidad</i>	<i>Ministerio de Hacienda y Función Pública, @dministración electrónica, Esquema Nacional de Interoperabilidad - Espanha</i>	Guia	2016	No contexto do <i>Esquema Nacional de Interoperabilidad</i> , fundamentado na lei de acesso à informação pública, o governo fornece um documento explicativo da aplicação da <i>Norma Técnica de Interoperabilidad</i> , entregando detalhes técnicos e procedimentais.	Lei 37/2007	EX T	A A	R	A	DCAT	SKOS para áreas temáticas Taxonomia de órgãos públicos Taxonomia de abrangência geográfica

¹⁷ Ante a ausência da especificação do tipo do documento no conteúdo dele, assume-se o termo guia pelas características configuradas no documento.

Nome do documento	Detentor	Tipo de documento	Data	Função	Fundamentação em LAI	Metadados Teoria Conceitos (C), Fundamentações (F), Técnicas (T), Estudos (E), Explicações (EX), Críticas (CR)	Intensidade Alta (A), Média (M), Baixa (B)	Metadados Uso Experimentos (EXP), Relatos de experiência (RE), Estudos de caso (EC), Recomendações (R), Percepções de uso (PU)	Intensidade Alta (A), Média (M), Baixa (B)	Metadados Estruturais	Metadados Descritivos
<i>Como publicar datos abiertos de manera rápida y sencilla (con CKAN)</i>	Ministerio de Hacienda y Administraciones Públicas, Ministerio de Industria, Energía y Turismo, Iniciativa Aporta, Red.es - Espanha	Guia	2015	Guia rápido para implementar abertura de dados em plataforma CKAN.	-	C	B	R	B	-	-
<i>Guía de aplicación del Real Decreto 1495/2011</i>	Minis. Industria, Energía y Turismo, Minis. de Hacienda y Administraciones Públicas, @administración electrónica - Espanha	Guia	2012	Guia de aplicação com generalidades sobre a implementação da lei de reutilização da informação em Espanha.	Lei 37/2007	C	B	R PU	A A	DCAT	Taxonomias, Sistemas SKOS
Guia Dados abertos	Agência para a modernização administrativa AMA - Portugal	Guia	2016	Guia que apresenta conceitos básicos sobre abertura de dados nos governos, junto a aspectos sociais e técnicos relacionados.	Lei 46/2007	C	B	-	-	DCAT	-

Fonte: Elaborado pela autora.

Na seguinte seção, alusiva ao tratamento dos resultados, apresentam-se as intuições ou inferências em relação à aderência percebida nos documentos analisados, a respeito da unidade de registro temática Metadados e que contém instruções para publicadores de dados nos quatro países estudados.

5.3 Tratamento dos resultados

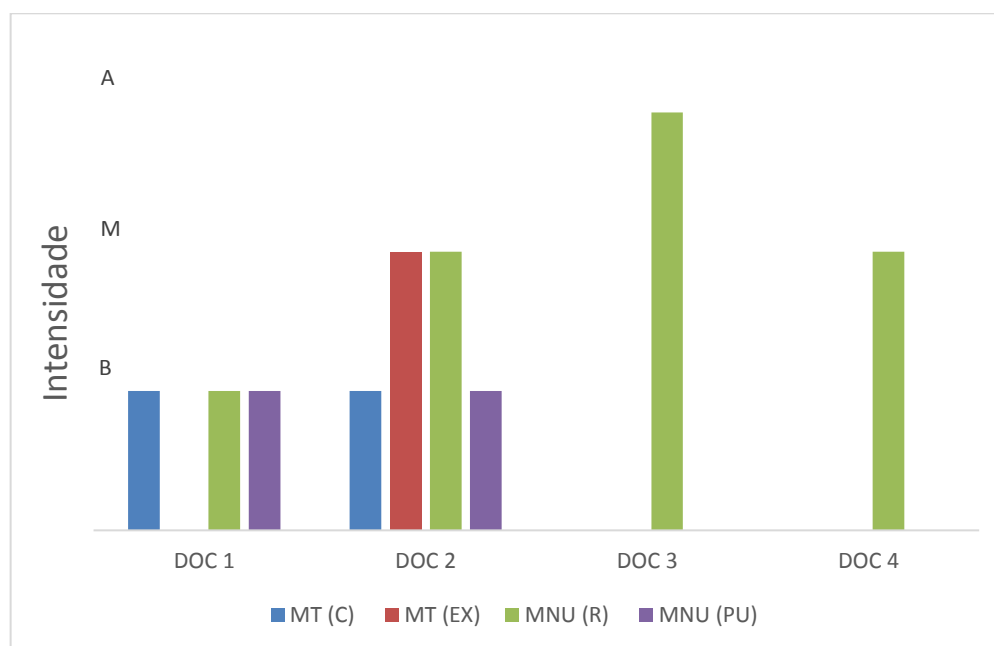
Conforme indica Bardin (2010, p. 127), os resultados brutos devem tratar-se “de maneira a serem significativos e válidos”. Para tal, os resultados obtidos na exploração do material podem ser observados de maneira sintetizada mediante figuras (-gráficos), retornando às unidades do estudo de casos (quatro países analisados), e contribuindo na compreensão acerca da aderência à representação dos recursos informacionais alvos desta pesquisa, *datasets* com dados de governo.

5.3.1 Colômbia

A Figura 11 compendia a categorização dos documentos analisados no contexto colombiano. Com fim de uma apresentação sucinta, assumiram-se as seguintes siglas para referenciar os quatro documentos analisados, que foram também referenciados na seção 5.1: DOC 1: *Guía de datos abiertos en Colombia* (COLOMBIA, 2016a); DOC 2: *Guía para la estandarización de Datos Abiertos* (COLOMBIA, 2016b); DOC 3: *Manual de datos abiertos en Colombia para Publicador* (COLOMBIA, 2016c); e, DOC 4: *Manual de datos abiertos en Colombia para Administrador* (COLOMBIA, 2016d).

Nesta figura que contém os resultados da fase da exploração do material conforme o procedimento metodológico proposto utilizou-se a sigla MT como referência à primeira categoria proposta para a contagem da categoria da unidade de registro temática, neste caso Metadados Teoria, como também, a sigla MNU, para representar o resultado obtido na categoria Metadados no Uso.

Figura 11 – Apresentação sintética da caracterização do *corpus* da pesquisa Colômbia



Fonte: Dados da pesquisa.

Assim, no eixo horizontal estão referenciados tanto a identificação dos documentos analisados na Colômbia, quanto as categorias observadas na análise de conteúdo. Já no eixo vertical representam-se as intensidades na abordagem das categorias temáticas em conformidade com: Alta (A), Média (M), e Baixa (B).

A maioria do conteúdo explorado foi classificado na categoria **Metadados no Uso**, embora somente um dos documentos analisado tenha sido considerado como Alto nível de intensidade; no entanto, com ocorrências em todos os documentos recuperados e analisados, os quais discutem as indicações que os publicadores de dados devem seguir como parte do processo descritivo dos *datasets*.

Para a categoria **Metadados Teoria**, foram estabelecidos dois documentos que apontaram nesse sentido à unidade de registro; contudo, em ambos os casos, unicamente chegou-se ao nível de intensidade Baixo.

Refletindo sobre a variável de inferência, que está intrinsecamente relacionada ao objetivo geral desta pesquisa, foram percebidas as seguintes interpretações a partir do conjunto de documentos analisados.

Comparando as intensidades observadas nas abordagens das duas categorias de análise, evidencia-se que o *corpus* de pesquisa abrange predominantemente as Recomendações, variável (R), no relacionado ao processo de descrição dos *datasets*, ou seja, há uma maior intensidade nos conteúdos que abordam Metadados no Uso (MNU), no

tocante às Recomendações. Nessa categoria (MNU), apenas um documento aborda a percepção de uso dos metadados (PU). É o caso do documento *Guía de Datos Abiertos en Colombia*, no qual, evidencia-se a possibilidade de, ao ter dados corretamente descritos, tê-los referenciados ou enlaçados com outros sites da Colômbia que disponibilizam dados do governo (COLOMBIA, 2016a).

Em relação ao tópico recomendações no uso dos metadados, observa-se a presença de diretrizes técnicas no contexto da publicação de dados abertos, embasando-se nas diretivas e recomendações do consórcio W3C. A Colômbia indica explicitamente acolher o embasamento produzido por esse consórcio. No documento *Guía para la Estandarización de Datos Abiertos*, na seção dos Metadados, indica-se que “*para la descripción de esta información se toma como referencia el estándar DCAT*” (COLOMBIA, 2016b, p. 18). O DCAT, como abordado no referencial teórico é uma recomendação do consórcio W3C.

O documento supracitado fornece indicações ao respeito do processo de publicação dos *datasets* e destaca a importância da descrição apropriada dos *datasets* para garantia de sucesso no acesso e na recuperação de dados. Além disso, refere-se aos metadados como elementos enriquecedores dos dados contidos nos *datasets*, apresentando um esquema em que é possível observar a ligação dos metadados, com os objetivos que estes devem alcançar durante o processo de acesso aos *datasets* e às distribuições associadas a eles (COLOMBIA, 2016b).

Os benefícios assinalados por esse governo em relação ao uso dos metadados estão relacionados com interoperabilidade, economia de esforços nos desenvolvimentos de projetos de software, uso e reuso de dados, e, estratégias de gestão e acompanhamento por meio de sistemas de indicadores padronizados (COLOMBIA, 2016b).

A partir da ocorrência do tópico percepção do uso na categoria Metadado no Uso, destaca-se ainda que, neste guia, mesmo que sem profundidade, são expostos casos de sites de dados abertos em países que utilizaram o "vocabulário" DCAT (W3C, 2017), bem como outros esquemas para o tratamento descritivo dos *datasets*, apresentando os modelos usados pelos Estados Unidos e Espanha, chegando até a descrever o uso do perfil de aplicação do vocabulário DCAT – DCAT-AP. Este perfil de aplicação está sendo indicado pela Comissão Europeia para ser utilizado por seus países membros (PORTAL EUROPEU DE DADOS, 2017).

Para a Comissão Europeia o desenvolvimento de um perfil de aplicação como o DCAT-AP, reveste-se de importância pelo alicerce que fornece na busca da interoperabilidade entre as diversas fontes de dados armazenados nos sites de dados abertos no contexto europeu.

Ele é o perfil de aplicação que utiliza o site do Portal de dados europeu.

A Comissão Europeia, assevera que nas fontes de dados das administrações públicas, com sistemas de administração eletrônica, existem ainda necessidades insatisfeitas partindo do olhar da interoperabilidade. Estudos mostram dificuldades, tanto de organizações quanto dos cidadãos, para encontrar e reutilizar recursos informacionais, e, portanto, dificuldades para aprimorar consideravelmente o valor dos dados que já estão disponíveis (PORTAL EUROPEU DE DADOS, 2017).

A Comissão Europeia, considerando os *datasets* existentes, com dados externos do setor público, acredita que o perfil de aplicação DCAT-AP respaldaria o intercâmbio, descobrimento e a reutilização dos dados publicados (PORTAL EUROPEU DE DADOS, 2017).

Aponta o Portal europeu de dados (2017) que os perfis de aplicação, como o DCAT-AP, ao usar esquemas para a definição de registros de metadados, apresentam dois benefícios:

- Os editores de dados aumentam a visibilidade dos dados e, portanto, aumentam a reutilização;
- Os usuários que reutilizam dados, podem buscar entre plataformas sem enfrentar as dificuldades ocasionadas pelo uso de esquemas isolados ou pelas diferentes línguas.

A documentação da Colômbia exemplifica, mesmo que de maneira sucinta, o DCAT-AP como uma alternativa para um desenvolvimento futuro que poderia trazer maior efetividade no uso do site de dados abertos desse país (COLOMBIA, 2016b).

Destaca-se a intensidade observada na abordagem da categoria Metadados no Uso, no tópico Recomendações, onde percebe-se como “Alta” no documento *Manual Datos Abiertos en Colombia para Publicador*, já que é feita uma ampla e suficiente descrição dos metadados que os usuários publicadores devem registrar durante o processo de publicação dos *datasets*. O documento traz uma descrição extensiva do processo.

A primeira frase que abre a seção do processo de descrição é enfática: “Ao encontrarmos localizados nos metadados é muito importante ter em conta a seguinte informação para ter segurança do correto preenchimento da informação e segurar o desdobre do *dataset*” (COLOMBIA, 2016c, p. 17, tradução nossa). Ressalta-se a pertinência do conteúdo do documento com a função para a qual ele foi criado.

Não obstante, em nível de intensidade Médio, identifica-se o documento *Manual Datos Abiertos en Colombia para Administrador*, que apresenta a unidade de registro temática mediante indicações com recomendações sobre o uso dos metadados nos *datasets*, assinalando a possibilidade da customização de alguns deles, quanto à criação segundo a necessidade do publicador (COLOMBIA, 2016d).

Ainda, observa-se a falta de precisão linguística ao usar a palavra metadado para referir-se aos atributos das interfaces gráficas em dois dos documentos analisados. Este evento poderia levar a erros nas interpretações por parte dos usuários alvos dos documentos analisados. O caso foi observado nos documentos *Manual Datos Abiertos en Colombia para Publicador* e *Manual Datos Abiertos en Colombia para Administrador* (COLOMBIA, 2016c; COLOMBIA, 2016d).

Mais uma imprecisão apresenta-se nos sintagmas em que o documento *Guía para la Estandarización de Datos Abiertos* menciona os termos “*Metadatos descriptivos*” e “*Metadatos estructurales*”. Observou-se que o documento relaciona “*Metadatos Descriptivos*” ao conjunto de metadados que descreve cada *dataset*, bem como “*Metadatos estructurales*” aos conjuntos de metadados que vão descrever as distribuições de cada *dataset*.

Analogicamente, essa abordagem exposta no documento colombiano daria para associar tanto aos “*Metadatos Descriptivos*” quanto aos “*Metadatos estructurales*” descritos no referencial teórico deste estudo como “padrão de estrutura dos dados” (GILLILAND, 2008). O diferencial está na descrição do recurso, que no primeiro caso é o *dataset*, em si mesmo e, no segundo, a distribuição do *dataset*.

Em relação à proposta da Gilliland (2008) para “padrões de valores dos dados”, no conteúdo do documento *Guía para la Estandarización de Datos Abiertos*, esse tipo de padrão é chamado simplesmente como “valores” dos metadados ou “identificadores”, no caso dos dados de natureza geográficos (COLOMBIA, 2016b).

A categoria analisada Metadados Teoria teve uma menor contagem na regra Presença nos documentos analisados, já que somente dois documentos apresentaram uma aproximação; apesar da Baixa intensidade nos tópicos: Conceitos e Explicações.

No caso do primeiro documento analisado, *Guía de Datos Abiertos en Colombia*, o tratamento descritivo dos *datasets* é abordado sob o conceito de "documentação dos dados" (COLOMBIA, 2016a, p. 21, tradução nossa) para descrever os aspectos básicos dos *datasets* mediante a necessidade de "classificar e melhorar os resultados da busca e recuperação dos dados" (COLOMBIA, 2016a, p. 21, tradução nossa).

Nesse tópico, percebe-se consenso com a definição de metadados de Alves e Santos (2013, p. 43) quando afirmam que são "dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação". Porém, nota-se que a abordagem teórica exposta no documento necessita de uma profundidade que possa fornecer clareza aos usuários que vão fazer uso desse documento.

Embora o documento *Guía para la Estandarización de Datos Abiertos* desenvolvido pela Colômbia apresente aos usuários uma descrição detalhada de todos os termos das entidades utilizados no modelo de dados do DCAT, observa-se a ausência da descrição das relações e das cardinalidades presentes entre as entidades do modelo. Acredita-se na importância da descrição não somente dos termos, mas também, das relações e das cardinalidades presentes entre as entidades do modelo com o intuito de permitir a aproximação dos usuários aos modelos de dados próprios dos seus entornos (COLOMBIA, 2016b).

Percebe-se ainda, que não é oferecido aos usuários alvo dos documentos *Manual Datos Abiertos en Colombia para Publicador* e *Manual Datos Abiertos en Colombia para Administrador* nenhuma aproximação teórica ou algum modelo de dados no qual possa se visualizar a modelagem ou alguma abstração do domínio dos *datasets* (COLOMBIA, 2016c, 2016d).

No que se refere a saber se os documentos com instruções para os publicadores de dados indicam especificamente os metadados a serem utilizados na descrição dos *datasets*, e considerando a imprecisão da linguagem percebida, para os metadados descritivos ("padrões de estrutura dos dados", conforme discutido no referencial teórico) o governo da Colômbia recomenda a utilização do DCAT, citando o modelo de dados criado pelo W3C. Nessa representação gráfica do "vocabulário" que propõe o W3C, é

possível perceber que além de metadados próprios do DCAT, são utilizados metadados de vários sistemas de representação como o *Dublin Core*, o SKOS e o FOAF (COLOMBIA, 2016b).

Lembrando a escolha da Colômbia no que diz respeito às diretrizes para os processos descritivos dos *datasets*, vale a pena esclarecer que o W3C recomenda o uso de metadados descritivos, “padrões de estruturas dos dados”, tanto nos catálogos, quanto nos *datasets* e em suas distribuições. Em relação ao esquema específico que poderia ser utilizado pelos publicadores de dados, o W3C recomenda o uso do DCTERMS e do DCAT, chamando-os como vocabulários¹⁸ (W3C, 2017).

Já no caso dos valores para os metadados, “padrões de valores dos dados”, os documentos apresentam como proposta o uso de categorias para descrever os assuntos dos *datasets*, identificando em um primeiro momento 22 itens que estão relacionadas com os setores primários da economia na Colômbia, dos quais se espera que forneçam uma gama de possibilidades composta pelo maior número de temas a serem publicados (COLOMBIA, 2016b).

Neste contexto, são sugeridos ainda:

- a) padronização dos valores dos dados correspondentes às localizações geográficas, indicando o uso do GEONAMES;
- b) padronização dos nomes das entidades e órgãos oficiais desse país, propondo a criação de uma "Taxonomia de Órgãos Publicadores", que seria um diretório comum de entidades orgânicas (COLOMBIA, 2016b, p. 63, tradução nossa), processo que deveria ser liderado pelo governo central desse país e
- c) dados sobre os processos de qualidade que foram desenvolvidos durante o processo de publicação, e que afetaram os conjuntos de dados. Neste caso, não é direcionado o uso de algum metadado específico que aborde a qualidade nos *datasets*.

No final, indicam que a Colômbia optou por utilizar a solução oferecida pela plataforma SOCRATA para a administração do seu site de dados abertos.

Em 2015, em um estudo realizado pelo *World Bank Group*, foram avaliadas em paralelo as plataformas CKAN e SOCRATA, observando detalhadamente as características

¹⁸ O W3C utiliza o termo "vocabulário" no lugar de "padrão de metadados" para referir-se a esquemas recomendados na catalogação de recursos informacionais como os *datasets* (W3C, 2017).

de cada uma para a futura implementação na Colômbia. Na oportunidade, foram examinados os atributos técnicos, tais como: linguagem de programação, tipos de licença, tipo de código fonte, além das características funcionais das plataformas e da forma de se colocar em funcionamento. Ao final do estudo desenvolvido optou-se pela adoção da SOCRATA (WORLD BANK GROUP, 2015).

O *Manual Datos abiertos en Colombia para Administrador* indica que os metadados solicitados pelo site nos processos de criação de datasets, correspondem com essa solução implementada, que ainda, permite a customização de metadados segundo as necessidades dos publicadores de dados (COLOMBIA, 2016d; WORLD BANK GROUP, 2015).

5.3.2 Brasil

Seguindo a dinâmica proposta para o estudo de caso por país, encontrou-se no site de dados do governo federal do Brasil, dois documentos publicados que chegam a compor o *corpus* da pesquisa.

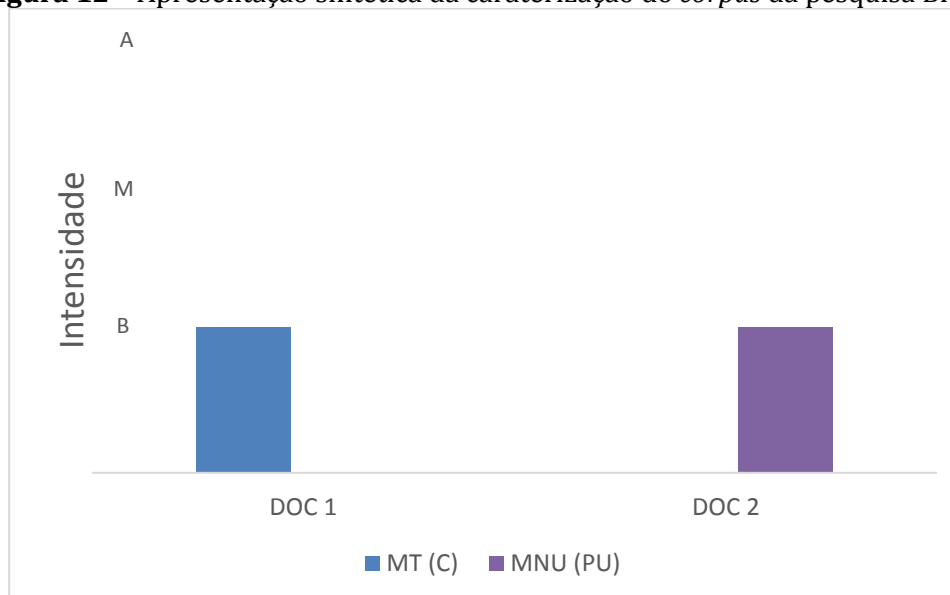
O primeiro, a Cartilha Técnica para Publicação de Dados Abertos no Brasil (BRASIL, 2012), é um documento elementar que contém o contexto inicial sobre a temática dos dados abertos, e diretrizes básicas dirigidas aos órgãos do governo brasileiro que desejam publicar *datasets*. Introduce conceitos sobre dados abertos, valor dos dados abertos, informações técnicas relevantes e generalidades acerca da gestão política que o governo brasileiro tem desenvolvido no marco da lei de acesso à informação no Brasil.

Já no o segundo documento analisado, Manual de Catalogação no Portal Brasileiro de Dados Abertos (BRASIL, 2016b), trata-se de um manual desenvolvido para apoiar, órgãos do poder executivo federal e demais instituições públicas, especificamente na catalogação dos dados abertos a publicarem no site de dados do governo brasileiro. Com frequência este documento, faz referência ao supracitado no parágrafo anterior, indicando os metadados que devem ser utilizados no processo de catalogação e, também, a possibilidade de adoção de outros metadados desejáveis. Em geral, o documento descreve o processo administrativo da catalogação de recursos informacionais tipo *datasets*, dos órgãos de governo.

Em seguida, realizadas as análises de conteúdo atingindo as regras de contagem propostas, chegou-se à síntese apresentada na Figura 12, na qual é mantida a organização dos eixos do gráfico de barras exposta na Figura 11, sendo DOC 1 utilizado para

representar a Cartilha Técnica para Publicação de Dados Abertos no Brasil (BRASIL, 2012); e, DOC 2, Manual de Catalogação no Portal Brasileiro de Dados Abertos (BRASIL, 2016b).

Figura 12 – Apresentação sintética da caracterização do *corpus* da pesquisa Brasil



Fonte: Dados da pesquisa.

Chama a atenção que dos 14 documentos recuperados, disponíveis no site de dados, na seção com recursos preparados para auxiliar na gestão de *datasets*, somente dois continham efetivamente instruções para publicadores de dados. Percebeu-se nesse ponto, que vários dos documentos dispostos no site contemplam, de maneira concomitante, alguns aspectos relacionados à gestão de *datasets*, ou seja, o conteúdo dos recursos em ocasiões apresenta-se repetitivo ou referenciando conteúdo já abordado nos dois documentos representados na Figura 12.

Cada documento analisado aborda, de maneira exclusiva, somente uma das categorias propostas neste estudo. Tem-se então, na regra de contagem Presença, uma aparição no que tange à temática **Metadados Teoria**, na abordagem Conceitos (C), e, uma aparição na temática **Metadados no Uso**, com Percepções de Uso (PU). Ambas as categorias foram explanadas com um nível de intensidade Baixo.

É um panorama equilibrado, olhando desde o ponto de vista das categorias que foram escolhidas como foco desta análise, mas que, porém, poderão estar abordadas com um maior nível de aprofundamento no que diz respeito às fundamentações, exemplificações, até melhores orientações práticas, de modo que os usuários

publicadores possam observar e aproximarem-se às melhores técnicas na descrição de *datasets*.

No *corpus* de documentos analisados no site de dados do governo do Brasil, foi possível observar que não está sendo explicitamente indicado o uso de um esquema de metadados estrutural, e, aliás, têm divergências no relacionado ao uso de elementos de metadados indicados no processo descritivo dos *datasets*.

Nesse ponto, percebeu-se nas instruções que o site de dados brasileiro segue a proposta da plataforma CKAN, porém, nos dois documentos analisados, não há uma aproximação explícita para alguma estrutura unificada para a descrição de *datasets*.

Apresentam-se sequências para a inclusão de elementos de metadados, que têm diferenças em alguns elementos desses metadados. Observou-se ainda uma mudança na ordem de inclusão dos metadados, o que poderia gerar confusão para os usuários publicadores que estão se baseando nessas instruções.

Para ilustrar, no contexto do Brasil, o DOC 1, faz uma clara distinção entre dois tipos de metadados para a descrição de *datasets*: “obrigatórios e desejáveis”. Os metadados obrigatórios incluem os seguintes elementos: Título, Descrição, Catálogo Origem, Órgão responsável, Categoria VCGE¹⁹, e, Recursos (BRASIL, 2012, p. 7).

Por sua vez, o DOC 2 (BRASIL, 2016b) indica que no momento da criação de *datasets* os elementos a serem considerados durante o processo de catalogação, sem distinguir a obrigatoriedade ou não, são: Título, Descrição, Etiquetas, Licença, Organização, e, Visibilidade.

Observa-se neste cenário que na Cartilha, DOC 1, o intuito do governo brasileiro é a correta descrição dos recursos informacionais, aproximando-se de uma estratégia descritiva própria, e que norteia o contexto do governo eletrônico brasileiro. Por outro lado, o Manual, DOC 2, apresenta os elementos descritivos pertencentes à plataforma CKAN, que é a utilizada pelo site de dados do governo do Brasil (COMPREHENSIVE..., 2017). Explica-se neste ponto, o porquê das sutis diferenças na ordem e na indicação dos elementos de metadados entre esses dois documentos com instruções.

Um fato que sustenta a inferência acima indicada é a solicitude de ingresso da categoria do Vocabulário Controlado do Governo Eletrônico (VCGE), “Categoria VCGE” (BRASIL, 2012, p. 7) que a Cartilha, DOC 1, faz.

¹⁹ Vocabulário controlado do governo eletrônico (BRASIL, 2016a).

O VCGE é uma lista hierarquizada de termos com assuntos do governo, usada para indexar informações no governo federal, buscando fortalecer a interface com a sociedade e fornecer uma ferramenta de gestão (BRASIL, 2016a; 2016b).

Outro mecanismo federal criado para fortalecer o entorno do governo eletrônico no Brasil, é o Padrão de Metadados para o Governo Eletrônico (e-PMG). O e-PMG foi fundamentado no padrão Dublin Core, mantendo para si os 15 elementos básicos do padrão internacional, acrescentando cinco elementos mais por considerar-se necessários no contexto do governo eletrônico brasileiro (BRASIL, 2014).

Observou-se que nos documentos recuperados e analisados não há menção acerca do padrão e-PMG, sendo ele criado para a descrição de recursos informacionais de governo. Este fato aponta que, mesmo que o governo do Brasil tenha trabalhado com um mecanismo de representação para seus recursos informacionais, no caso do e-PMG, ele não está sendo ligado à plataforma de dados abertos de governo.

Uma boa prática observada nas instruções brasileiras é que durante o processo de catalogação, se faz a inclusão de um dicionário de dados que permite descrever as colunas nas distribuições dos *datasets*. Este recurso vem como arquivo adjunto às distribuições dos *datasets* (BRASIL, 2016b). Mesmo que o Manual não utilize o termo metadados nesse ponto, é claramente uma indicação da importância manifesta de um processo descritivo no entorno estudado.

O item que site de dados brasileiro nomeia como “dicionário de dados” corresponde com aquele que site de dados colombiano denomina de “metadados estruturais”. Esses dois termos são aplicados para descrever a estrutura nas distribuições dos *datasets*.

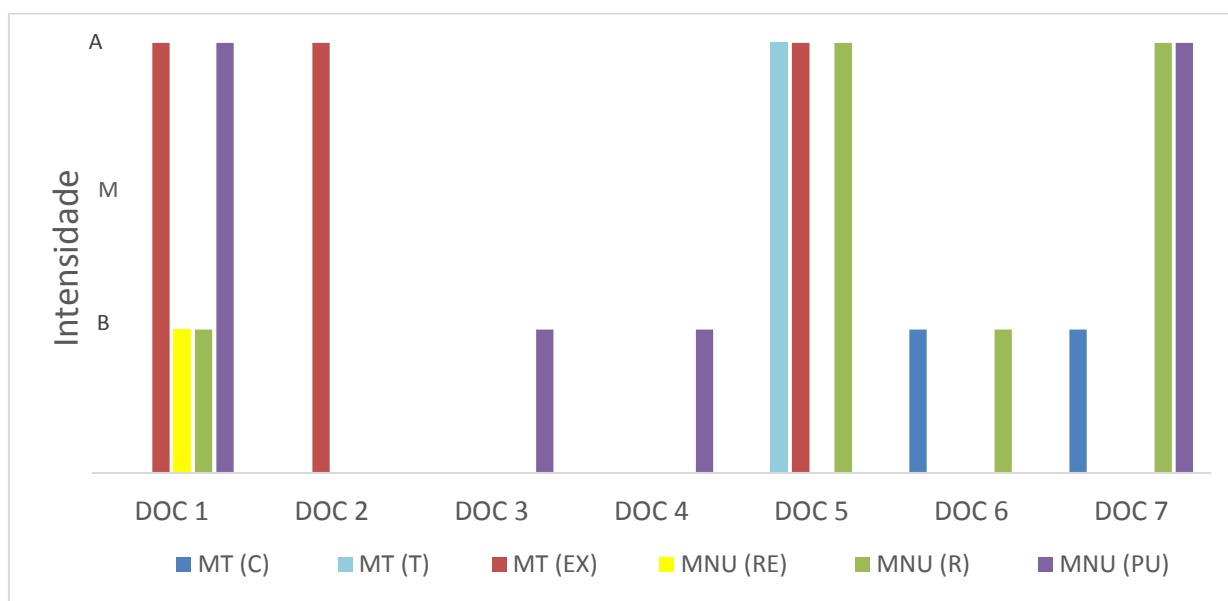
5.3.3 Espanha

Um número significativo de documentos, em relação aos dois casos anteriores (Colômbia e Brasil), foi analisado no contexto da Espanha, totalizando sete documentos com instruções nos quais foram identificadas seis categorias temáticas, merecendo destaque frente aos casos da Colômbia, com quatro categorias; e, ao do Brasil, com duas.

A nomenclatura adotada, na Figura 13, para identificar os documentos no eixo horizontal é: DOC 1: *Manual de Usuario: Perfil Organismo Público* (ESPAÑA, 2017a); DOC 2: *Plantilla RDF/XML para la federación de conjuntos de datos* (ESPAÑA, 2017b); DOC 3:

Orientaciones y garantías en los procedimientos de anonimización de datos personales (ESPAÑA, 2016b); DOC 4: *Orientaciones sobre la protección de datos en la reutilización de la información del sector público* (ESPAÑA, 2016a); DOC 5: *Guía de aplicación de la Norma Técnica de Interoperabilidad* (ESPAÑA, 2016c); DOC 6: *Como publicar datos abiertos de manera rápida y sencilla (con CKAN)* (ESPAÑA, 2015); e, DOC 7: *Guía de aplicación del Real Decreto 1495/2011* (ESPAÑA, 2012).

Figura 13 – Apresentação sintética da caracterização do *corpus* da pesquisa Espanha



Fonte: Dados da pesquisa.

Nota-se que as duas grandes categorias aparecem com o mesmo número de ocorrências: **Metadados Teoria** e **Metadados no Uso**, ambas com três subcategorias.

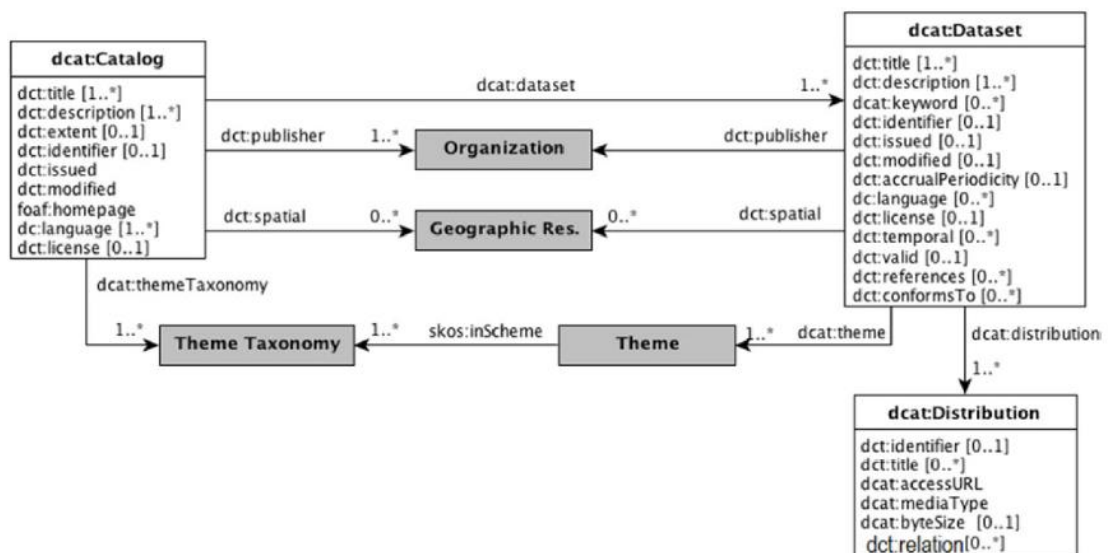
No caso das temáticas tratadas para a categoria **Metadados Teoria**, registrou-se em duas, das três, uma intensidade Alta: Técnicas e Explicações. Enquanto isso, em **Metadados no Uso**, tanto Recomendações quanto Percepções de Uso têm também ocorrência com Alta intensidade. Ou seja, das seis categorias tratadas nas instruções espanholas, quatro foram abordadas com um alto nível de detalhamento.

O tratamento que as instruções fazem das temáticas na categoria **Metadados Teoria**, foi percebido com um nível de profundidade altamente proveitoso para os publicadores de dados. Exemplificando este fato, observa-se que no DOC 5, *Guía de aplicación de la Norma Técnica de Interoperabilidad* (ESPAÑA, 2016c), há uma descrição ampla e completa do modelo de dados recomendado para ser utilizada no tratamento

descritivo dos recursos informacionais na Espanha. Este recurso com instruções aborda três categorias temáticas, todas com um nível de intensidade Alto: MT (T) Técnicas, MT (T) Explicações, e, MNU (R) Recomendações.

Indica-se no documento citado, o uso do modelo DCAT. O guia apresenta graficamente os tipos de relações entre entidades, cardinalidade, e características dos atributos em conformidade com as entidades do modelo recomendado pelo governo espanhol. Esta descrição pormenorizada outorga clareza para os usuários publicadores. No corpo do documento é apresentado o modelo DCAT exibido na Figura 14.

Figura 14 – Modelo de dados do DCAT (Espanha)



Fonte: España (2016c).

O guia especifica a inclusão de relações das entidades do DCAT com esquemas de conceitos SKOS, utilizados para descrever órgãos do governo, recursos geográficos, e taxonomias²⁰ de temas. Aliás, o documento indica que esses tipos de mecanismos de representação, SKOS e taxonomias, são utilizados como metadados para o valor ou conteúdo, isto é, metadados descritivos (ESPAÑA, 2016c).

²⁰ As taxonomias são um tipo de esquemas de codificação de valores que pertencem ao tipo de metadados padrões de valores dos dados (ALVES; SANTOS, 2013).

Outros elementos contidos no guia supracitado são a descrição dos *namespaces*²¹ que utiliza o modelo DCAT e o detalhamento do dicionário de dados acerca desse modelo de descrição (ESPAÑA, 2016c).

O dicionário de dados descreve as entidades do modelo DCAT, a saber, catálogo, *dataset*, e, distribuição, incluindo os elementos de metadados com suas descrições, especificando se utiliza elementos dos modelos FOAF, SKOS ou Dublin Core. Ainda, aponta sobre a obrigatoriedade dos elementos, a permissão de múltiplos valores, para finalmente indicar o tipo de valores que são permitidos para os elementos, assinalando especificamente, se vai permitir algum esquema de metadados descritivos no seu valor (ESPAÑA, 2016c).

Além do DOC 5, outros três documentos com instruções assinalam o uso do vocabulário DCAT como modelo de dados para a descrição de *datasets*. Trata-se do DOC 1: *Manual de Usuario: Perfil Organismo Público* (ESPAÑA, 2017a); DOC 2: *Plantilla RDF/XML para la federación de conjuntos de datos* (ESPAÑA, 2017b), e, DOC 7: *Guía de aplicación del Real Decreto 1495/2011* (ESPAÑA, 2012).

Um recurso que apresenta quatro categorias temáticas, tanto de **Metadados Teoria** quanto de **Metadados no Uso**, é o DOC 1, *Manual de Usuario: Perfil Organismo Público* (ESPAÑA, 2017a), destacando que o conteúdo tenta manter equilíbrio entre as abordagens teóricas e o uso dessas teorias tratadas. No relacionado com **Metadados no Uso**, destaca-se, neste documento, a possibilidade de cada órgão publicador incluir relatos de aplicações e da reutilização de dados baseados nos *datasets* que estão sendo publicados por eles.

Este documento enfatiza, ainda, na correta catalogação dos *datasets*, o que contribui na obtenção de benefícios pela geração de estatísticas, que permitem retroalimentar a gestão da plataforma de dados da Espanha. Em *Cuadros de mando*, no site de dados são indicadas várias alternativas que permitem observar o uso, por parte da sociedade, dos dados publicados por cada órgão (ESPAÑA, 2017a, p. 75).

A plataforma de dados da Espanha é a única, das quatro analisadas, que possui dois documentos com instruções que atingem a categoria temática Explicações, de Metadados Teoria, em nível de intensidade Alto. É o caso do documento DOC 2, que explicita detalhadamente o processo descritivo de *datasets*, com os elementos de metadados do

²¹ Alternativa tecnológica que permite distinguir elementos, instancias, entre diferentes vocabulários (LINKED OPEN VOCABULARIES, 2017).

vocabulário DCAT, auxiliando o usuário publicador com comentários e observações para cada elemento descritivo (ESPAÑA, 2017b).

No relacionado com aspectos de privacidade no site de dados, o governo espanhol fornece também documentos com instruções para publicadores, mediante o tratamento da temática Percepções de Uso (PU), outorgando atenção aos elementos de metadados como aspectos determinantes a serem considerados no processo de anonimização e de proteção de dados pessoais. Tais documentos são o DOC 3: *Orientaciones y garantías en los procedimientos de anonimización de datos personales* (ESPAÑA, 2016b) e o DOC 4: *Orientaciones sobre la protección de datos en la reutilización de la información del sector público* (ESPAÑA, 2016a).

O DOC 4 discorre sobre a importância dos metadados como elementos a serem fornecidos que, além de entregues em aberto e processáveis por máquina, estejam "com os níveis mais elevados de precisão e desagregação" (ESPAÑA, 2016a, p. 11, tradução nossa).

Em relação a indicação do uso de metadados descritivos, embora o DOC 1 não contenha uma indicação específica sobre algum padrão, observa-se em seu conteúdo, o indício do uso de taxonomias para tal fim. Indica-se o uso do campo `dcat:themeTaxonomy`, que pertence ao modelo DCAT, e que nesse ponto está direcionando para a URL <http://datos.gob.es/kos/sector-publico/sector>, a qual vai conter valores esperados que referenciam categorias ou temáticas para os *dataset*.

Além disso, é perceptível que as organizações que estão sujeitas à publicação de dados encontram-se normalizadas na plataforma, de modo a manter a coerência interna no sistema no relacionado aos organismos publicadores.

Em resumo, o DOC 1 indica o uso de três recursos a serem empregados como elementos de conteúdo em alguns dos elementos descritivos: órgãos publicadores²², âmbitos geográficos²³ e temas²⁴.

Nos documentos espanhóis, percebe-se a ênfase no que tange aos processos recomendados nas diretrizes, de modo que sejam aplicados na publicação de cada órgão

²² Cf. <http://datos.gob.es/apidata/catalog/publisher>

²³ Cf. <http://datos.gob.es/apidata/catalog/spatial>

²⁴ Cf. <http://datos.gob.es/apidata/catalog/theme>

oficial, mesmo em seus sites particulares, já que são diretrizes que implementam o marco legal do país, enfatizando a reutilização da informação.

Assim, o alvo desta pesquisa, documentos com instruções para publicadores de dados no âmbito da Espanha, foca em publicadores de todos os órgãos de governo que disponibilizam recursos informacionais tanto no site de dados do país, quanto nos sites de dados de cada instância do governo.

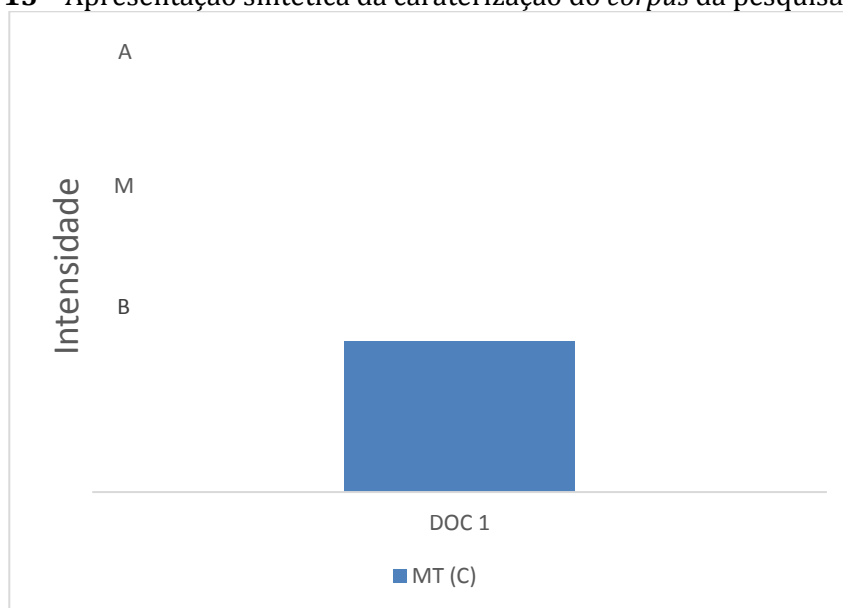
Percebe-se ainda unicidade no *corpus* da pesquisa proveniente da administração espanhola. O trabalho conjunto realizado entre órgãos oficiais que tratam diretamente dos assuntos do governo eletrônico, tais como o *Consejo Superior de Administración Electrónica* (CSAE) e a *Comisión Permanente del Consejo Superior de Administración Electrónica* (CPCSAE), ratificam esta afirmação. A CPCSAE especificamente tem como uma das suas tarefas cuidar da harmonização dos serviços de reutilização da informação, diretamente no que se refere à manutenção dos metadados no âmbito dos *datasets* de dados de governo (ESPAÑA, 2012).

Em geral, observa-se maior intencionalidade, por parte de organismos do governo espanhol envolvidos na publicação de dados, em fornecer subsídios mediante indicações para distintos tipos de usuários e processos associados à descrição de *datasets* governamentais.

5.3.4 Portugal

Este país ibérico aportou o menor número de documentos a ser analisado sob o prisma desta pesquisa. Somente dois documentos estão disponibilizados no site de dados de Portugal, dos quais um deles foi objeto da análise de conteúdo por ter relação com o objetivo a desenvolver.

Figura 15 – Apresentação sintética da caracterização do *corpus* da pesquisa Portugal

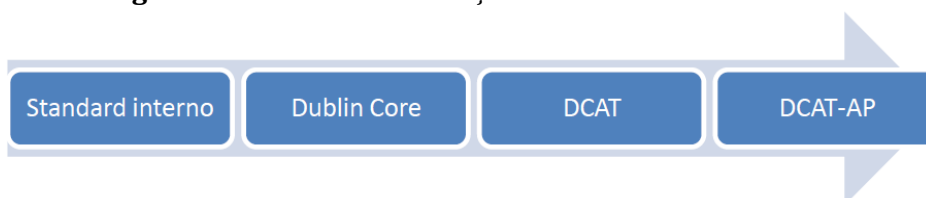


Fonte: Dados da pesquisa.

Observa-se na Figura 15 que apenas uma categoria temática foi abordada neste *corpus*, **Metadados Teoria** Conceitos (C), em um nível de baixa intensidade. No DOC 1: Guia dados abertos (PORTUGAL, 2016), o governo português se refere de modo sucinto aos metadados como elementos descritivos, de suma importância para a reutilização da informação e a interoperabilidade.

O documento detalha de modo conciso a evolução de modelos de metadados que vêm sendo utilizados por países nos seus processos descritivos de recursos informacionais de governo. A Figura 16 mapeia os elementos que consideram nessa evolução, implicando nos processos de normalização, padronização, até maturidade dos órgãos que estão implementando estratégias de publicação de dados.

Figura 16 – Estados na evolução do modelo de metadados



Fonte: Portugal (2016).

O documento faz descrição de cada fase nessa evolução, com as características, vantagens e inconvenientes, observando-se a inclusão do modelo de dados DCAT como opção na representação de *datasets* governamentais. Contudo, o documento, mesmo indicando claramente o uso do vocabulário DCAT, como modelo a seguir na descrição dos

datasets governo, está vulnerável na descrição desse modelo de dados, das entidades, relações, cardinalidades e demais características, de modo a oferecer maior detalhamento ao usuário publicador.

Como no caso da Colômbia, o Portugal faz menção do perfil de aplicação DCAT-AP, contextualizando-o e indicando ele como uma opção a implementar que atendendo o nível de maturidade do país, poderia contribuir com a interoperabilidade entre os países do continente europeu.

5.3.5 Inferências no grupo de países

Já na fase final da pesquisa, espera-se refletir em torno às cifras totais que compendiam as ocorrências das categorias temáticas objetos de estudo.

Como indicado na seção 1.3.2, a unidade temática objeto deste estudo, **Metadados**, foi abordada mediante categorias agrupadas em dois itens, **Metadados Teoria** e **Metadados no Uso**, onde a primeira inclui como valores Conceitos, Fundamentações, Técnicas, Estudos, Explicações, e, Críticas. Metadados no Uso trata Experimentos, Relatos de Experiência, Estudos de caso, Recomendações, e, Percepções de Uso. São 11 categorias no total.

O Quadro 5 permite visualizar as categorias que apareceram nos documentos com instruções, junto aos níveis de intensidade com que elas foram desenvolvidas, e, as quantidades que surgiram na análise.

Quadro 5 - Resumo de intensidades nas categorias

CATEGORIAS	NÍVEIS DE INTENSIDADE		
	B	M	A
MT. CONCEITOS	6	0	0
MT. EXPLANAÇÕES	0	1	3
MT. TÉCNICAS	0	0	1
MNU. RECOMENDAÇÕES	3	2	3
MNU. PERCEP. USO	5	0	2
MNU. RELAT. EXPERIÊNCIA	0	0	1
Totais	14	3	10

Fonte: Dados da pesquisa.

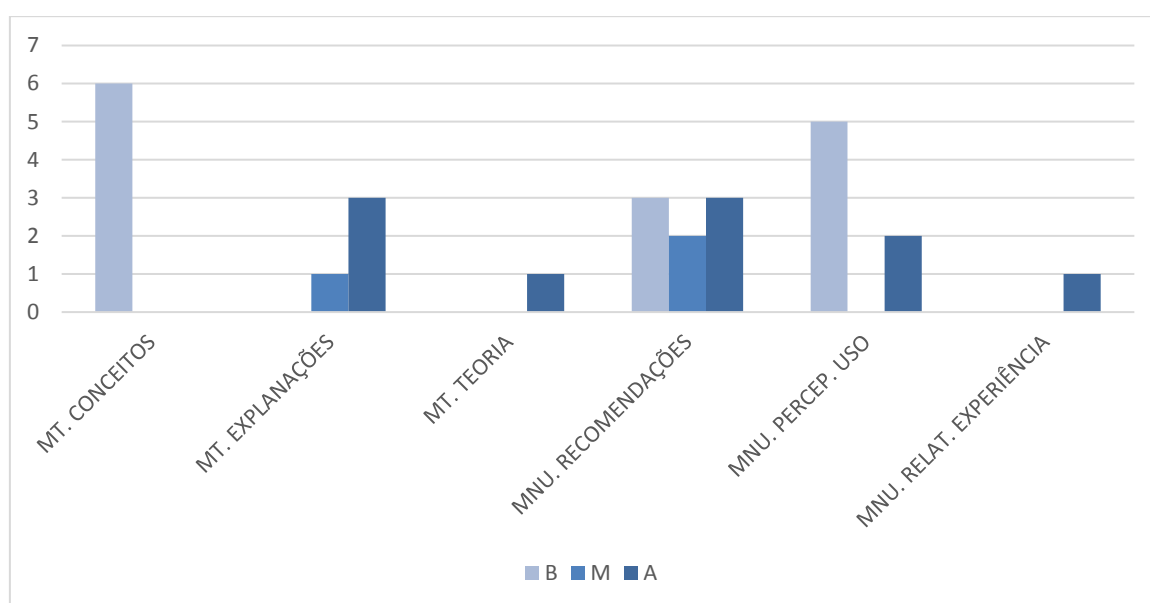
No Quadro 5 percebe-se que das onze categorias possíveis, seis foram registradas segundo a primeira regra de contagem proposta, Aparição, observando que é mantido o

equilíbrio entre a abordagem Teórica e do Uso, com três categorias de **Metadados Teoria** e três de **Metadados no Uso**.

Discernindo sobre a segunda regra de contagem, Intensidade, ditas categorias têm na maioria dos casos abordagens em nível de intensidade Baixo, algarismo 14, seguido de perto por intensidade Alta, com o algarismo 10, dos quais, 9 documentos provêm da Espanha.

A Figura 17 permite observar a quantidade de documentos que trataram cada categoria registrada e o nível de intensidade com que foi abordado.

Figura 17 – Ocorrências das categorias com os níveis de intensidade



Fonte: Dados da pesquisa.

Os documentos analisados apresentam referências às categorias, indicadas na figura, com níveis de intensidade frequentemente Baixos e Médios, porém, é preciso destacar que das seis categorias, em cinco delas, pelo menos um documento foi abordado com intensidade Alta.

Partindo do ponto de vista dos esquemas de metadados recomendados nas instruções, observou-se que a Colômbia, a Espanha e Portugal fazem menção acerca da utilização do modelo DCAT como esquema proposto para a descrição dos *datasets*. Enquanto isso, o Brasil não indica explicitamente a utilização de um modelo de dados.

Nesse ponto, dito país opta na documentação por abordar a descrição dos *datasets* seguindo unicamente o que para tal fim tem a plataforma de CKAN.

Na sequência, em relação às plataformas de dados que utilizam os sites de dados analisados, foi possível constatar que a Colômbia utiliza a SOCRATA²⁵, enquanto o Brasil e a Espanha optaram pela solução oferecida na CKAN²⁶²⁷. Já no caso de Portugal, o site de dados funciona sobre uma plataforma própria, desenvolvida e mantida pela Agência para a Modernização Administrativa (AMA).

Essas escolhas de plataforma deverão de modo particular desenvolver estratégias na busca da interoperabilidade entre os recursos informacionais de cada país, auxiliando-se, dentre outros aspectos, nos que atingem a descrição dos *datasets*, incluindo o desenvolvimento de competências nos usuários publicadores de dados.

²⁵ Cf. <http://estrategia.gobiernoenlinea.gov.co/623/w3-article-9407.html>

²⁶ Cf. <http://dados.gov.br/pagina/cartilha-publicacao-dados-abertos>

²⁷ Cf. <http://datos.gob.es/es/noticia/datosgobes-inicia-sus-pasos-en-beta>

6 CONSIDERAÇÕES

Em primeira instância, percebe-se o esforço realizado pelos detentores nos quatro países analisados ao abordar preferencialmente a categoria “Metadados no Uso”, dando assim maior importância ao aspecto prático da abordagem em relação ao teórico. Sem desconsiderar a relevância das explicações teóricas, a análise dos documentos apresenta mais detalhes procedimentais na descrição de *datasets* mediante o uso de metadados. Há uma maior quantidade absoluta nos documentos abordados nessa categoria, 16, em número, quanto 11 na categoria “Metadados Teoria”.

Acredita-se no sentido da cobertura teórica que, atingida com justo equilíbrio, poderia permitir uma melhor compreensão das vantagens da correta representação dos *datasets* quando recursos informacionais.

Nesse sentido teórico, considera-se a importância da ampla descrição, tanto das entidades que compõem os modelos de dados, bem como das relações e das cardinalidades presentes entre as entidades do modelo, aproximando, assim, os usuários aos modelos de dados.

Em seguida, inferiu-se, no relacionado aos documentos espanhóis, que o anseio pela aproximação aos usuários está norteando a elaboração das instruções ao aportar a maior quantidade de documentos disponibilizados, maior número de categorias abordadas, e maior intensidade no tratamento outorgado a cada componente temático.

Percebeu-se a intenção do governo espanhol no desenvolvimento das competências informacionais que precisam os usuários publicadores, não somente ao corroborar a alta intensidade na maioria das abordagens realizadas, mas também ao apreciar a engrenagem que o governo pretende atingir entre diversos órgãos oficiais na busca de interoperabilidade dos seus recursos informacionais.

Destaca-se ainda no país ibérico, a concorrência de várias instâncias do governo na elaboração de documentos com instruções, ao ser identificada a interação de 7 órgãos nos quais há dois ministérios: *Ministerio de Hacienda y Función Pública*, e *Ministerio de Energía Turismo y Agenda Digital*, outorgando interdisciplinaridade na abordagem dos aspectos relacionados

Na outra ponta, a estratégia descrita nos documentos do site de dados da Colômbia poderia considerar o aprimoramento da abordagem conceitual no tratamento descritivo dos *datasets*, visando clarificar o entendimento aos usuários publicadores do site.

Contudo, observou-se uma tendência positiva na apresentação dos conteúdos analisados, encontrando-se em um ponto médio no desenvolvimento das instruções aos publicadores de dados. Seria interessante que o grupo de usuários publicadores da Colômbia aproveitara o corpo de documentos que tem disponibilizado o governo na sua plataforma de dados, o qual percebe-se aceitável em seu conteúdo ao apresentar quatro categorias com intensidades médias e altas.

Em relação à falta de precisão linguística em alguns trechos dos documentos analisados nesse país, recomenda-se o aprimoramento no uso de termos adequados que evitem a confusão e os erros nas interpretações por parte dos usuários alvos desses recursos. Uma possível solução seria a fundamentação com referenciais teóricos, de modo que os usuários possam consultar e elucidar o exposto nos documentos.

Enquanto isso, o Brasil apresenta uma plataforma de dados de governo com diversas opções que pretendem desenvolver as competências dos usuários publicadores de dados. Assim, percebe-se a intenção no aprimoramento das competências e habilidades dos usuários ao oferecer várias opções documentais, até com recursos didáticos para acompanhar o processo de publicação.

Contudo, os documentos que conformaram o *corpus* da pesquisa deverão passar por processos de melhoramento no referente ao marco conceitual e ao uso dos metadados, no tocante aos processos descritivos dos *datasets*. Isto, a fim de entregar ferramentas conceituais e práticas para aprimorar nos publicadores o nível e a percepção teórica, bem como a compreensão, mediante exposição de casos de sucesso, das vantagens de um correto processo descritivo de *datasets*.

Por fim, considera-se que a documentação oferecida pela AMA em Portugal, encontrando-se escassa em número, poderia ser ampliada tanto em quantidade quanto na intensidade do tratamento das áreas temáticas relacionadas com a descrição dos recursos informacionais, os *datasets*.

Uma vez que foram encontrados variedade de recursos para a descrição de ativos informacionais como os *datasets*, a saber, DCAT, Dublin Core, FOAF, SKOS, etc., seria interessante estudar as implicações no uso de esquemas onde eles se misturem, visando-as sob o prisma da interoperabilidade.

Acredita-se na importância da criação e implementação de perfis de aplicação entre comunidades de países, como o caso do DCAT-AP, criado e recomendado pela Comissão europeia. Esta comunidade está convencida de que a partir do aprimoramento

da infraestrutura tecnológica digital há como obter maior desenvolvimento, tanto em cada país como na comunidade, destinando “fundos europeus estruturais e de investimento” para o fortalecimento das capacidades institucionais das administrações públicas. E, nesses esforços institucionais e tecnológicos, cria-se e recomenda-se frequentemente o uso de modelos que garantam a interoperabilidade, tendo como exemplo o perfil de aplicação DCAT-AP, que vem sendo incluído nas suas estratégias de execução (EUR-LEX, 2017).

Nota-se que a busca do fortalecimento das capacidades humano-informacionais (SANT’ANA, 2013) poderia passar por estudos que atingem um dos aspectos desse tipo de capacidades, o registro dos recursos informacionais, a fim de contribuir, como asseverado pelo Le-Coadic (1994), no referente ao objeto social imperante e necessário das pesquisas na área da Ciência da Informação. Neste caso em particular, como o autor supracitado indica, “a demanda premente da tecnologia da informação” assinala caminhos de pesquisa dentro das práticas sociais, que estejam vestidos de utilidade e eficácia (LE COADIC, 1994, p. 21).

Os resultados, obtidos após a análise e interpretação dos dados, permitem visualizar o panorama geral da aderência ao uso de metadados nos processos de publicação de dados de governo. Paralelamente, admitem-se inquietações referentes aos processos de publicação de dados de governo e suas relações com outros tópicos de interesse, como os socioeconômicos, possíveis vínculos com indicadores de desenvolvimento em países e regiões, bem como a reciprocidade dessas estratégias nos grupos sociais, e a efetividade transformada em fatos que demonstrem o acréscimo da transparência nos governos, tudo isso na perspectiva de pesquisas originadas a partir da Ciência da Informação.

REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE JORNALISMO INVESTIGATIVO. **CTRL+X**. 2017. Disponível em: <<http://www.ctrlx.org.br/#/infografico>>. Acesso em: 21 dez. 2017.
- ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2010.
- ALVES, R. C. V.; SANTOS, P. L. V. A. da C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.
- ARAKAKI, F. A. **Linked Data: ligação de dados bibliográficos**. 146 f. 2016. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciência, Marília, 2016.
- BAKER, T.; COYLE, K. **Application profiles as an alternative to owl ontologies**. 2013. Disponível em: <https://github.com/dcmi/archive/blob/master/mediawiki_wiki/APvOWL_Lisbon.md>. Acesso em: 15 out. 2017.
- BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 2010.
- BORKO, H. Information science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, jan. 1968. Disponível em: <<http://pt.scribd.com/doc/39023567/Information-Science-What-is-It>>. Acesso em: 17 set. 2017.
- BRASIL. Governo Eletrônico. **Padrão de Metadados do Governo Eletrônico – ePMG**. 2014. Disponível em: <<https://www.governoeletronico.gov.br/eixos-de-atuacao/governo/interoperabilidade/padrao-de-metadados-do-governo-eletronico-epmg>>. Acesso em: 10 maio 2017.
- BRASIL. Governo Eletrônico. **Vocabulário Controlado do Governo Eletrônico – VCGE**. 2016a. Disponível em: <<https://www.governoeletronico.gov.br/eixos-de-atuacao/governo/interoperabilidade/vocabulario-controlado-do-governo-eletronico-vcge>>. Acesso em: 10 nov. 2017.
- BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Cartilha técnica para publicação de dados abertos no Brasil v1.0**. 2012. Disponível em: <<http://dados.gov.br/pagina/cartilha-publicacao-dados-abertos>>. Acesso em: 7 nov. 2017.
- BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. **Manual de catalogação no Portal Brasileiro de Dados Abertos**. Brasília, DF, 2016b. Disponível em: <<http://wiki.dados.gov.br/GetFile.aspx?File=%2fManuais%2fManual-de-Cataloga%C3%A7%C3%A3o-v1.0.pdf>>. Acesso em: 7 nov. 2017.
- BRASIL. Ministério da Transparência e Controladoria-Geral da União. **Jorge Hage participa, nos Estados Unidos, de evento que consolida parceria para governo**

aberto. Brasília, DF, 11 set. 2017. Disponível em:

<<http://www.cgu.gov.br/noticias/2011/07/jorge-hage-participa-nos-estados-unidos-de-evento-que-consolida-parceria-para-governo-aberto>>. Acesso em: 3 set. 2017.

BRASIL. Tribunal de contas da união. **5 motivos para a abertura de dados na administração pública.** 2015. Disponível em:

<<http://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A8182A24F0A728E014F0B36E7016F34&inline=1>>. Acesso em: 10 set. 2017.

CENTER FOR DATA INNOVATION. **The State of Open Data Portals in Latin America.** 2017. Disponível em:

<<https://www.datainnovation.org/2017/07/the-state-of-open-data-portals-in-latin-america/>>. Acesso em: 18 jul. 2017.

COMPREHENSIVE KNOWLEDGE ARCHIVE NETWORK. 2017. Disponível em: <<https://ckan.org/>>. Acesso em: 10 jul. 2017.

COLOMBIA. **Guía de datos abiertos en Colombia.** 2016a. Disponível em:

<https://herramientas.datos.gov.co/sites/default/files/articles-8248_Guia_Apertura_Datos_0.pdf>. Acesso em: 6 maio 2017.

COLOMBIA. **Guía para la estandarización de Datos Abiertos.** 2016b. Disponível em:

<https://herramientas.datos.gov.co/sites/default/files/Guia_Estandarizacion_DatosAbiertos_final.pdf>. Acesso em: 6 maio 2017.

COLOMBIA. **Manual de datos abiertos en Colombia para publicador.** 2016c.

Disponível em:

<https://herramientas.datos.gov.co/sites/default/files/CO_417_MNL_PUBLICADOR_0.pdf>. Acesso em: 6 maio 2017.

COLOMBIA. **Manual de datos abiertos en Colombia para administrador.** 2016d.

Disponível em:

<https://herramientas.datos.gov.co/sites/default/files/CO_417_MNL_ADMINISTRADOR.pdf>. Acesso em: 6 maio 2017.

COLOMBIA. Gobierno Digital Colômbia. **Datos abiertos.** 2017. Disponível em:

<<https://www.datos.gov.co/browse?sortBy=newest>>. Acesso em: 10 maio 2017.

COYLE, K. BAKER, T. Guidelines for Dublin core application profile (working draft).

Dublin Core Metadata Initiative, 18 maio 2009. Disponível em:

<<http://dublincore.org/documents/profile-guidelines/>>. Acesso em: 30 out. 2017.

DUBLIN CORE METADATA INICIATIVE. 2017. Disponível em:

<<http://dublincore.org/about/>>. Acesso em: 10 out. 2017.

DUVAL, E. et al. Metadata principles and practicalities. **D-lib Magazine**, v. 8, n. 4, 2002.

Disponível em: <<http://www.dlib.org/dlib/april02/weibel/04weibel.html>>. Acesso em: 10 out. 2017.

ESPAÑA. Iniciativa Aporta. **Como publicar datos abiertos de manera rápida y sencilla (con CKAN)**. 2015. Disponível em: <<http://datos.gob.es/es/documentacion/guia-para-publicar-datos-abiertos-de-manera-rapida-y-sencilla-con-ckan>>. Acesso em: 28 nov. 2017.

ESPAÑA. Iniciativa Aporta. **Manual de usuario: perfil organismo público**. 2017a. Disponível em: <<http://datos.gob.es/es/documentacion/guias-de-datosgobes>>. Acesso em: 28 nov. 2017.

ESPAÑA. Iniciativa Aporta. **Plantilla RDF/XML para la federación de conjuntos de datos**. 2017b. Disponível em: <<http://datos.gob.es/es/documentacion/guias-de-datosgobes>>. Acesso em: 28 nov. 2017.

ESPAÑA. Agencia Española de Protección de Datos. **Orientaciones sobre la protección de datos en la reutilización de la información del sector público**. 2016a. Disponível em: <<http://datos.gob.es/es/documentacion/orientaciones-sobre-la-proteccion-de-datos-en-la-reutilizacion-de-la-informacion-del>>. Acesso em: 28 nov. 2017.

ESPAÑA. Agencia Española de Protección de Datos. **Orientaciones y garantías en los procedimientos de anonimización de datos personales**. 2016b. Disponível em: <<http://datos.gob.es/es/documentacion/orientaciones-y-garantias-en-los-procedimientos-de-anonimizacion-de-datos-personales>>. Acesso em: 28 nov. 2017.

ESPAÑA. Ministerio de Hacienda y Administraciones Públicas. **Guía de aplicación de la Norma Técnica de Interoperabilidad**. 2016c. Disponível em: <<http://datos.gob.es/es/documentacion/guia-de-aplicacion-de-la-norma-tecnica-de-interoperabilidad-de-reutilizacion-de>>. Acesso em: 28 nov. 2017.

ESPAÑA. Ministerio de Industria, Energía y Turismo, Ministerio de Hacienda y Administraciones Públicas. **Guía de aplicación del Real Decreto 1495/2011**. 2012. Disponível em: <<http://datos.gob.es/es/documentacion/guia-de-aplicacion-del-real-decreto-14952011>>. Acesso em: 28 nov. 2017.

EUR-Lex. **Document 52017DC0134**. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions European Interoperability Framework – Implementation Strategy. Mar. 2017. Disponível em: <<http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52017DC0134>>. Acesso em: 15 set. 2017.

FUSCO, E. **Modelos conceituais de dados como parte do processo da catalogação: perspectiva de uso dos FRBR no desenvolvimento de catálogos bibliográficos digitais**. 2010. 249 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2010.

GEONAMES. 2017. Disponível em: <<http://www.geonames.org/>>. Acesso em: 1 out. 2017.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GILLILAND, A. J. Setting the stage. In: BACA, M. (Ed.) **Introduction to metadata**. 2nd ed. Los Angeles: Getty Research Institute, 2008, p. 1-19.

GLOBAL EDITORS NETWORK. **Data Journalism Awards**. 2017. Disponível em: <<https://www.datajournalismawards.org/>>. Acesso em: 21 jul. 2017.

GLOBAL OPEN DATA FOR AGRICULTURE & NUTRITION. **About GODAN**. 2017. Disponível em: <<http://www.godan.info/es/about>>. Acesso em: 31 ago. 2017.

GURSTEIN, M. **What is community informatics?** Milan: Polimetrica, 2007.

INTERNATIONAL COUNCIL FOR SCIENCE. 2017. Disponível em: <<https://www.icsu-wds.org/>>. Acesso em: 10 out. 2017.

LE COADIC, Y-F. **A ciência da informação**. Brasília: Brinquet de Lemos, 2004.

LINKED OPEN VOCABULARIES. 2017. Disponível em: <<http://lov.okfn.org/>>. Acesso em: 1 nov. 2017.

LUCCI, E. A.; BRANCO, A. L.; MENDOÇA, C. **Território e sociedade no mundo globalizado**. 2. ed. São Paulo: Saraiva, 2013.

MALTA, M. C. **Contributo metodológico para o desenvolvimento de perfis de aplicação no contexto da Web Semântica**. 2014. 398 f. Tese (Doutoramento em Tecnologias e Sistemas de Informação) – Universidade de Minho, Escola de Engenharia, Minho, 2014.

McGARRY, K. **O contexto dinâmico da informação: uma análise introdutória**. Brasília: Brinquet de Lemos, 1999.

NATIONAL INFORMATION STANDARDS ORGANIZATION. 2017. Disponível em: <<http://www.niso.org/home/>>. Acesso em: 15 out. 2017.

OPEN GOVERNMENT PARTNERSHIP. 2017. Disponível em: <<https://www.opengovpartnership.org/about/about-ogp>>. Acesso em: 30 ago. 2017.

OPEN KNOWLEDGE FOUNDATION. 2017. Disponível em: <<https://okfn.org/opendata/how-to-open-data/>>. Acesso em: 12 out. 2017.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. **OCDE Principles and guidelines for access to research data from public funding**. Paris: OECD Publishing, 2007. Disponível em: <<http://www.oecd.org/science/sci-tech/38500813.pdf>>. Acesso em: 25 set. 2017.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. **Members and partners**. 2018. Disponível em: <<http://www.oecd.org/about/membersandpartners/>>. Acesso em 2 jan. 2018.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO.
Government at a Glance 2017. Paris: OECD Publishing, 2017a. Disponível em:
<http://dx.doi.org/10.1787/gov_glance-2017-en>. Acesso em: 20 jul. 2017.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO.
Panorama de las Administraciones Públicas América Latina y el Caribe 2017:
Brasil. Paris: OECD Publishing, 2017b. Disponível em: <<http://www.oecd.org/gov/alc-brasil.pdf>>. Acesso em: 22 dez. 2017.

PARLAMENTO EUROPEU. Diretiva 2003/98/CE do Parlamento Europeu e do Conselho,
17 nov. 2003. Relativa à reutilização de informações do sector público. **Jornal Oficial**,
Bruxelas, n. L 345, p. 90-96, 31 dez. 2003. Disponível em: <<http://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32003L0098&rid=1>>. Acesso
em: 19 ago. 2017.

PESET, F.; FERRER-SAPENA, A.; SUBIRATS-COLL, I. Open data y Linked open data: su
impacto en el área de bibliotecas y documentación. **El profesional de la información**,
Barcelona, v. 20, n. 2, p. 165-173, mar./abr. 2011. doi: 10.3145/epi.2011.mar.06.

PORTAL EUROPEU DE DADOS. **Preparing data**. 2017. Disponível em:
<<https://www.europeandataportal.eu/pt/providing-data/goldbook/preparing-data>>.
Acesso em: 20 jul. 2017.

PORTUGAL. **Dados.gov**. 2016. Disponível em:
<<http://www.dados.gov.pt/pt/noticias/guia-dados-abertos.aspx#sthash.YbTEwKOP.fCIW0V2O.dpbs>>. Acesso em: 16 dez.2017.

REITZ, J. M. **Online dictionary for library and information science**: ODLIS. 2004.
Disponível em: <https://www.abc-clio.com/ODLIS/odlis_about.aspx>. Acesso em: 28
dez. 2017.

RAMALHO, R. A. S. Análise do modelo de dados SKOS: sistema de organização do
conhecimento simples para a web. **Informação & Tecnologia**, João Pessoa, v. 2, n. 1, p.
66-79, 2015. Disponível em:
<<http://periodicos.ufpb.br/index.php/itec/article/view/25995>>. Acesso em: 1 out.
2017.

ROSENFELD, L; MORVILLE, P; ARANGO, J. **Information Architecture**: For the Web and
Beyond. Sebastopol: O'Really, 2015.

SÃO PAULO (Estado). **Guia de dados abertos**. 2015. Disponível em:
<<http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/02/Book-Web-Guia-de-Dados-Abertos.pdf>>. Acesso em: 26 jun. 2017.

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

SANT'ANA, R. C. G. Prefácio. In: ALVES, R. C. V.; SANTOS, P. L. V. A. da C. **Metadados no
domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013. p. 7-9.

SANT'ANA, R. C. G. Reflexões sobre a representação no ciclo de vida dos dados. In: ENCONTRO DE REPRESENTAÇÃO DOCUMENTAL, 1., 2017, São Carlos. **Anais...** São Carlos: UFSCar, 2017. Disponível em:

<<http://www.telescopium.ufscar.br/index.php/enredo/enredo/paper/viewFile/109/96>>. Acesso em: 10 out. 2017.

SANTOS, P. L. V. A da C. **Catálogo revisitada**: sua história e contemporaneidade. 2010. 248 f. Tese (Livre-Docência) –Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2010.

SANTOS, P. L. V. A. da C. et al. Mapeamento do termo tecnologia em periódicos da CI no escopo do GT8: informação e tecnologia. In: ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis: ANCIB, 2013. Disponível em:

<<http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/201>>. Acesso em: 8 jul. 2017.

SANTOS, P. L. V. A da C. Representação documental no comportamento de ambientes informacionais. In: ENCONTRO DE REPRESENTAÇÃO DOCUMENTAL, 1., 2017, São Carlos. **Anais...** São Carlos: UFSCar, 2017. Disponível em:

<<http://www.telescopium.ufscar.br/index.php/enredo/enredo/paper/viewFile/104/105>>. Acesso em: 2 out. 2017.

SANTOS, P. L. A. da C.; SANT'ANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, Brasília, v. 42, n. 2, p. 199-209, 2013. Disponível em:

<<http://revista.ibict.br/ciinf/article/view/1382/1560>>. Acesso em: 2 jun. 2017.

SANTOS, P. L. A. da C.; SANT'ANA, R. C. G. Transferência da Informação: análise para valoração de unidades de conhecimento. **DataGramZero**, Rio de Janeiro, v. 3, n. 2, abr. 2002. Disponível em:

<<http://basessibi.c3sl.ufpr.br/brapci/index.php/article/download/7482>>. Acesso em: 4 set. 2017.

SANTOS, P. L. V. A. da C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Informação & Informação**, Londrina, v. 19, n. 1, p. 146-163, 2014.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em ciência da informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>>. Acesso em: 12 maio 2017.

SCHEMA. 2017. Disponível em: <<http://schema.org/Dataset>>. Acesso em: 14 ago. 2017.

SETZER, V. W. Dado, informação, conhecimento e competência. **DataGramZero**, Rio de Janeiro, n. 0, dez. 1999. Disponível em:

<<http://www.brapci.ufpr.br/brapci/index.php/article/download/14562>>. Acesso em: 25 abr. 2017.

SHINTAKU, M.; DUQUE, C.; SUAIDEN, E. J. Análise da adesão às tendências da ciência pelos repositórios institucionais brasileiros. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 6, n. 2, p. 148-169, 2015. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/17969>>. Acesso em: 10 jun. 2017.

SIMIONATO, A. C. **Modelagem conceitual DILAM**: princípios descritivos de arquivos, bibliotecas e museus para o recurso imagético digital. 2015. 200 f. Tese (Doutorado em Ciência da Informação) –Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2015.

SIMIONATO, A. C.; SANTOS, P. L. V. A da C. Descrição de recursos imagéticos digitais: apresentação de um modelo conceitual. **Anales de Documentación**, Murcia, v. 16, n. 2, 2013. Disponível em: <<http://dx.doi.org/10.6018/analesdoc.16.2.179261>>. Acesso em: 30 jun. 2017.

SOCRATA. 2017. Disponível em: <<https://socrata.com/>>. Acesso em: 10 jul. 2017.

SOUTO, D. V. B. **A aplicação do modelo RDF na descrição arquivística**: em foco a Norma ISAD(G). 2014. 180 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual de Londrina, Londrina, 2014.

UNIVERSITY OF TORONTO. **G7 Information Centre**. 2017. Disponível em: <<http://www.g8.utoronto.ca/>>. Acesso em 30 dez. 2017.

VERHULST, S. G.; YOUNG, A. **Open data in developing economies toward building an evidence**: toward building an evidence base on what works and how. New York: Governance Lab, New York University, 2017. Disponível em: <<http://odimply.org/files/odimply-developing-economies.pdf>>. Acesso em: 31 ago. 2017.

WORLD BANK GROUP. **Comparativo plataformas comparativo plataformas de datos abiertos**. 2015. Disponível em: <<https://herramientas.datos.gov.co/sites/default/files/Comparativo%20Plataformas%20Datos%20Abiertos%20Banco%20Mundial2015.pdf>>. Acesso em: 6 maio 2017.

WORLD WIDE WEB CONSORTIUM. 2017. Disponível em: <<https://www.w3.org/TR/vocab-dcat/>>. Acesso em: 10 jul. 2017.

ZENG, M. L.; QIN, J. **Metadata**. New York: Neal-Schuman Publishers, 2008.