



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Campus de Marília

CLAYTON MARTINS PEREIRA

Framework para publicação de metadados geoespaciais semânticos
seguindo as boas práticas de *Linked Data*

Marília
2023

CLAYTON MARTINS PEREIRA

***Framework* para publicação de metadados geoespaciais semânticos
seguindo as boas práticas de *Linked Data***

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências, da Universidade Estadual Paulista (UNESP), Campus de Marília, como parte dos requisitos para a obtenção do título de Doutor em Ciência da Informação.

Área de Concentração: Informação, Tecnologia e Conhecimento.

Orientador: Prof. Dr. José Eduardo Santarem Segundo

Marília
2023

P436f

Pereira, Clayton Martins

Framework para publicação de metadados geoespaciais semânticos
seguindo as boas práticas de Linked Data / Clayton Martins Pereira. --
Marília, 2023

179 f. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Filosofia e Ciências, Marília

Orientador: José Eduardo Santarem Segundo

1. Ciência da informação. 2. Metadados. 3. Dados ligados. 4. Dados
geoespaciais. 5. Web semântica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de
Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CLAYTON MARTINS PEREIRA

***Framework para publicação de metadados geoespaciais semânticos
seguindo as boas práticas de *Linked Data*.***

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), como requisito parcial para a obtenção do título de Doutor em Ciência da Informação.

Área de concentração: Informação, Tecnologia e
Conhecimento

Linha de pesquisa: Informação e Tecnologia

BANCA EXAMINADORA

Prof. Dr. José Eduardo Santarem Segundo
UNESP – Marília
Orientador

Prof. Dr. Edberto Ferneda
UNESP – Marília

Prof. Dr. Caio Saraiva Coneglian
UNESP – Marília

Profa. Dra. Ana Carolina Simionato Arakaki
UFSCar – São Carlos

Dra. Lúbia Vinhas
INPE/MCTI – São José dos Campos

Marília, 15 de dezembro de 2022.

À minha sobrinha Maya Martins, nascida em 05 julho de 2019,
a quem espero venha acreditar que ‘a educação transforma’.

À Sua Majestade, Rainha Elizabeth II
(1926 – 2022)

AGRADECIMENTOS

Aos meus pais, Julio e Dalva, por todo esforço, dificuldades enfrentadas e, principalmente, por terem renunciado a muitos planos para que eu pudesse me dedicar aos estudos até o final de minha formação técnica.

Ao Prof. Dr. Nei Yoshihiro Soma, do Instituto Tecnológico de Aeronáutica (ITA), pelo apoio e incentivo recebidos durante todo o período em que servi àquela renomada instituição de ensino da Força Aérea Brasileira para mudar de rumo profissional e enveredar pela carreira acadêmica, fazendo com que pudesse encontrar minha verdadeira vocação.

Ao Dr. Gerald Jean Francis Banon, do Instituto Nacional de Pesquisas Espaciais, pelo apoio, incentivo e contribuições a este trabalho.

Ao Prof. Dr. Luis Antônio Tozzi, da Faculdade de Tecnologia (FATEC) de São José dos Campos, pelo incentivo recebido durante minha segunda graduação e, posteriormente, quando de minha atuação como docente naquela instituição.

Ao Prof. Dr. Ricardo Magnus Osório Galvão e à Profa. Dra. Maria Virgínia Alves, respectivamente ex-Diretor e ex-Chefe de Gabinete do Instituto Nacional de Pesquisas Espaciais (INPE), onde trabalho atualmente, pela autorização e liberação para que eu pudesse realizar este curso de Doutorado.

À toda equipe do Serviço de Informação e Documentação do INPE, em especial à ex-chefe Me. Silvia de Castro Marcelino, por todo apoio, compreensão e incentivo.

Ao meu orientador, Prof. Dr. José Eduardo Santarem Segundo, e aos colegas do NEWSDA – Núcleo de Estudos em Web Semântica e Dados Abertos, pela parceria e contribuições para que este trabalho pudesse se concretizar.

Por fim, ao meu cão *July Pinscher*, companheiro inseparável durante grande parte dessa empreitada.

“O homem, quanto mais preenche os espaços mentais com as ideias do bem, mediante o estudo, a ação ou a reflexão, mais aumenta a sua capacidade e conquista mais amplos recursos para o progresso.”

Joanna de Ângelis (1985, p. 14)

RESUMO

Os metadados são atributos de recursos informacionais, com a finalidade de identificá-los e fornecer informações detalhadas a seu respeito, de forma que possam ser rapidamente encontrados e processados pelos usuários. No caso dos metadados que descrevem recursos (*datasets*) que podem ser referenciados espacialmente em relação à superfície da Terra, os chamados ‘metadados geoespaciais’, esses devem transmitir ainda outras informações específicas a respeito da posição precisa daquele recurso, chamadas de ‘extensão espacial’. Isso aumenta consideravelmente o espaço necessário para armazenamento dos *datasets*, bem como torna seu processamento mais complexo, o que exige maior alocação de recursos computacionais. Diante dessa complexidade, e dos altos custos para produzir e manter enormes volumes de dados geoespaciais, é que diversos governos e entidades se lançaram na concepção das chamadas ‘Infraestruturas de Dados Espaciais (IDEs)’, que têm se popularizado como uma importante solução para permitir a interoperabilidade de dados geoespaciais. No entanto, nas plataformas de IDEs de geração atual, as buscas aos metadados disponibilizados nos respectivos servidores de catálogo são sintáticas, ou seja, as palavras-chave fornecidas pelos usuários são utilizadas de forma literal, quer sejam isoladas (termo) ou agrupadas (frase), sem considerar o contexto ou significado específico delas. Assim, podem ser retornados resultados que não são de interesse dos usuários, ou ainda resultados relevantes deixarem de ser encontrados devido à falta de desambiguação de termos com vários sentidos e por não incorporar todos os significados e sinônimos desses termos na busca. Diante disso, as tecnologias e ferramentas da *Web Semântica*, e as boas práticas de *Linked Data*, tem se mostrado eficazes para promover a interoperabilidade de dados em nível semântico, em larga escala, ao usar a arquitetura da *Web* atual, que é constituída de padrões totalmente abertos. Assim, o presente trabalho pretende contribuir para a concepção das IDEs de próxima geração por meio de sua integração com as tecnologias e ferramentas da *Web Semântica*, o que possibilitaria a recuperação e reuso de *datasets* geoespaciais tanto por humanos quanto por máquinas. Para isso, é proposto um *framework* que estabelece um fluxo organizacional para orientar a execução das etapas de um processo de publicação de metadados geoespaciais semânticos na *Web* seguindo as boas práticas de *Linked Data*. Trata-se de uma pesquisa teórica, de natureza aplicada, de abordagem qualitativa e do tipo exploratória, que emprega como métodos a revisão bibliográfica e a pesquisa documental, nas quais os conteúdos obtidos são analisados por meio da aplicação do método de ‘Análise de Conteúdo’. O *framework* proposto foi validado em uma prova de conceito, na qual foram executados cada um dos procedimentos que o compõem, utilizando como amostra um registro de metadados extraído do catálogo de metadados da IDE ‘TerraBrasilis/INPE’. A partir da prova de conceito foi possível constatar a viabilidade de implementação de cada procedimento proposto no *framework*, dentre os quais o *crosswalk* para conversão de padrões de metadados, a estruturação das triplas RDF com o uso do vocabulário DCAT-2, e o enriquecimento semântico dos metadados por meio de sua ligação à *Web* de dados.

Palavras-chave: *Web* semântica; dados geoespaciais; metadados; infraestruturas de dados espaciais; dados ligados.

ABSTRACT

Metadata are attributes of information resources, with the purpose of identifying them and providing detailed information about them, so that they can be quickly found and processed by users. In the case of metadata describing resources (datasets) that can be spatially referenced relative to the Earth's surface, the so-called 'geospatial metadata', it must also convey other specific information about the precise position of that resource, called 'spatial extent'. This considerably increases the space required for storing the datasets, as well as making their processing more complex, which requires a greater allocation of computational resources. Given this complexity, and the high costs to produce and maintain huge volumes of geospatial data, is that several governments and entities have launched in the design of the so-called 'Spatial Data Infrastructures (SDIs)', which have become popular as an important solution to enable the interoperability of geospatial data. However, in the current generation SDI platforms, the searches to the metadata made available in the respective catalog servers are syntactic, i.e., the keywords provided by users are used in a literal way, either isolated (term) or grouped (phrase), without considering their specific context or meaning. Thus, results may be returned that are not of interest to the users, or even relevant results may not be found due to the lack of disambiguation of terms with multiple meanings and by not incorporating all the meanings and synonyms of these terms in the search. Therefore, Semantic Web technologies and tools, and Linked Data best practices have proven to be effective in promoting large-scale semantic data interoperability by using the current Web architecture, which consists of completely open standards. Thus, this work aims to contribute to the design of next generation SDIs through their integration with Semantic Web technologies and tools, which would enable the retrieval and reuse of geospatial datasets by both humans and machines. To this end, a framework is proposed that establishes an organizational flow to guide the execution of the steps of a semantic geospatial metadata publishing process on the Web following Linked Data best practices. This is theoretical research, of applied nature, of qualitative approach and exploratory type, that employs as methods the bibliographic review and the documentary research, in which the obtained contents are analyzed through the application of the 'Content Analysis' method. The proposed framework was validated in a proof of concept, in which each of the procedures that compose it were executed, using as sample a metadata record extracted from the metadata catalog of the SDI 'TerraBrasilis/INPE'. From the proof of concept, it was possible to verify the feasibility of implementing each procedure proposed in the framework, including the crosswalk to convert metadata standards, the structuring of RDF triples using DCAT-2 vocabulary, and the semantic enrichment of metadata through its link to the Web of data.

Keywords: semantic Web; geospatial data; metadata; spatial data infrastructure; linked data.

LISTA DE FIGURAS

Figura 1 – Componentes de uma IDE.	19
Figura 2 – Arquitetura de uma IDE da geração atual.....	21
Figura 3 – Caracterização da pesquisa.	27
Figura 4 – Fluxograma dos procedimentos da revisão de literatura.	30
Figura 5 – Exemplos de conjuntos de metadados.	48
Figura 6 – Exemplo de dados geoespaciais.....	53
Figura 7 – Exemplo dos formatos de representação matricial e vetorial.	55
Figura 8 – Categorias de metadados do padrão CSDGM.	59
Figura 9 – Esquema UML do padrão de metadados ISO 19115.....	60
Figura 10 – Seções da ISO 19115 adotadas pelo Perfil MGB.	63
Figura 11 – Dinâmica dos componentes de uma IDE.	68
Figura 12 – Relação entre níveis de IDEs e níveis organizacionais.....	69
Figura 13 – Arquitetura conceitual de uma IDE.	70
Figura 14 – Interações de descoberta de metadados em IDEs.	71
Figura 15 – Arquitetura do Serviço de Catálogo para a Web (CSW).	73
Figura 16 – Exemplo de requisição e resposta do Serviço de Catálogo para a Web (CSW).74	
Figura 16 – Representação de uma tripla RDF na forma de grafo e seu código XML.....	82
Figura 17 – Comparação entre as estruturas de um <i>triplestore</i> e de um SGBD relacional...87	
Figura 18 – Diagrama de elementos do vocabulário RDFS.....	92
Figura 19 – Visão geral do esquema do vocabulário DCAT-2.	98
Figura 20 – Esquema de classificação de 5 estrelas para dados na Web.....	102
Figura 21 – Estados inicial e atual da nuvem de dados ligados (<i>LOD Cloud</i>).....	104
Figura 22 – Fluxo Organizacional para publicação de Dados.	108
Figura 23 – Esquema para publicação de conjuntos de dados na Web.	112
Figura 24 – Visão geral do <i>framework</i> proposto neste trabalho.	118
Figura 25 – Integração do <i>framework</i> a uma IDE.	119
Figura 26 – Visão da camada de ‘Coleta e Conversão’ do <i>framework</i>	120
Figura 27 – Boas práticas seguidas na camada de ‘Coleta e Conversão’ do <i>framework</i>	125
Figura 28 – Visão da camada de ‘Enriquecimento Semântico’.	126
Figura 29 – Boas práticas seguidas na camada de ‘Ligação’ do <i>framework</i>	128
Figura 30 – Visão da camada de ‘Disponibilização’ do <i>framework</i>	129

Figura 31 – Boas práticas seguidas na camada de ‘Disponibilização’ do <i>framework</i>	131
Figura 32 – Trechos do registro de metadados geoespaciais coletado para a validação.....	132
Figura 33 – Conexão do conjunto de metadados do <i>framework</i> à nuvem LOD.	147
Figura 34 – Carregamento do conjunto de triplas RDF na plataforma de <i>triplestore</i>	148
Figura 35 – Consulta SPARQL sobre o conjunto de triplas carregado no <i>triplestore</i>	149
Figura 36 – Fontes de dados alcançáveis a partir dos resultados da consulta SPARQL.....	149

LISTA DE QUADROS

Quadro 1 – Métodos, técnicas e fontes de pesquisa empregados neste trabalho.	28
Quadro 2 – Definição do tema e das palavras-chave que representam a pesquisa.	31
Quadro 3 – Definição das fontes e das restrições das buscas a serem efetuadas.	31
Quadro 4 – <i>Script</i> para unificação e remoção de duplicidades nos resultados das buscas.	32
Quadro 5 – Equação <i>InOrdinatio</i> para ranqueamento de artigos científicos.	33
Quadro 6 – Critérios para avaliação dos artigos selecionados para leitura do texto completo...	35
Quadro 7 – Categorias de metadados.....	49
Quadro 8 – Exemplo de metadados, em formato XML, estruturados no padrão Dublin Core...	50
Quadro 9 – Elementos do padrão de metadados <i>Dublin Core Metadata Element Set</i>	51
Quadro 10 – Comparação entre os formatos de representação matricial e vetorial.....	55
Quadro 11 – Seções de metadados do padrão ISO 19115.	61
Quadro 12 – Pacotes de elementos do Perfil MGB 2.0.	64
Quadro 13 – Amostra de metadados geoespaciais codificados em GML/XML.....	66
Quadro 14 - Dimensões e elementos de um modelo para concepção de uma IDE.....	68
Quadro 16 – Exemplos de URIs para identificação e descrição de um recurso.	80
Quadro 17 – Declaração de <i>namespaces</i> em uma página <i>Web</i>	81
Quadro 18 – Exemplos de serialização de dados RDF em diferentes sintaxes.....	83
Quadro 19 – Exemplo de consulta SPARQL.....	89
Quadro 20 – <i>Triplestores</i> compatíveis com a extensão GeoSPARQL.	90
Quadro 21 – Principais elementos do SKOS.	93
Quadro 22 – Termos do Dublin Core Metadata Element Set.	95
Quadro 23 – Classes e propriedades do vocabulário FOAF.	96
Quadro 24 – Outros vocabulários para descrição de dados em RDF.....	99
Quadro 25 – Exemplo de aplicação do vocabulário WGS84.	105
Quadro 26 – Boas práticas para a publicação de <i>Linked Data</i>	106
Quadro 27 – Lista de verificação para publicação de <i>Linked Data</i>	107
Quadro 28 – Etapas e tarefas que compõem o fluxo organizacional para publicação de dados.	109
Quadro 29 – Relação das boas práticas do W3C para publicação de dados na <i>Web</i>	111
Quadro 30 - Boas práticas para publicação de dados espaciais na <i>Web</i>	113
Quadro 31 – Correspondência entre elementos no <i>crosswalk</i> ISO 19115 para <i>Dublin Core</i> ...	115
Quadro 32 – Exemplo de arquivo de mapeamento RDB para RDF na linguagem R2RML. ...	122
Quadro 33 – Amostra das triplas RDF geradas no <i>dump</i> da RDB mapeada no Quadro 32.	123

Quadro 34 – Sugestões de elementos de metadados que podem ser ligados à <i>Web</i> de dados..	127
Quadro 35 – Consulta CSW para coleta automatizada da amostra de registro de metadados..	135
Quadro 36 – <i>Crosswalk</i> entre elementos dos modelos de dados ISO 19115 e RDF utilizando o vocabulário DCAT-2.	135
Quadro 37 – Propriedades da classe <i>dcatalog:Catalog</i> a serem acrescentadas ao registro de metadados obtido do <i>crosswalk</i>	137
Quadro 38 – Registro de metadados em RDF/XML entregue pelo procedimento de conversão de modelo de dados.	137
Quadro 39 – Conjunto de metadados RDF convertido para a sintaxe de serialização <i>Turtle</i> ...	140
Quadro 40 - Elementos de metadados RDF a serem conectados à <i>Web</i> de dados.....	143
Quadro 41 - Conjunto de metadados RDF enriquecido com ligações à nuvem LOD.	144

LISTA DE TABELAS

Tabela 1 – Valores do Fator de Impacto e da equação <i>InOrdinatio</i> de cada artigo.	37
Tabela 2 – Resultados da análise do título e do resumo de cada artigo.	37
Tabela 3 – Resultados da análise das seções de introdução e conclusão de cada artigo.	38
Tabela 4 – Resultados da leitura e análise dos textos completos.	39
Tabela 5 – Artigos selecionados para compor o portfólio de trabalhos relacionados.	40
Tabela 6 – Comparação entre trabalhos relacionados.	45

LISTA DE ABREVIATURAS E SIGLAS

CKAN	<i>Comprehensive Knowledge Archive Network</i>
CONCAR	Comissão Nacional de Cartografia
CRS	<i>Coordinate Reference System</i>
CSDGM	<i>Content Standard for Digital Geospatial Metadata</i>
CSV	<i>Comma Separated Values</i>
CSW	<i>Catalog Service for the Web</i>
DC	<i>Dublin Core</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCMES	<i>Dublin Core Metadata Element Set</i>
DCMI	<i>Dublin Core Metadata Initiative</i>
DDL	<i>Data Definition Language</i>
DML	<i>Data Modification Language</i>
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FGDC	<i>U.S. Federal Geographic Data Committee</i>
FOAF	<i>Friend Of A Friend Vocabulary</i>
GEMET	<i>General Multilingual Environmental Thesaurus</i>
GML	<i>Geographic Markup Language</i>
GPS	<i>Global Positioning System</i>
HTTP	<i>HiperText Transfer Protocol</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IDE	Infraestrutura de Dados Espaciais
INDE	Infraestrutura Nacional de Dados Espaciais
INPE	Instituto Nacional de Pesquisas Espaciais
IoT	<i>Internet of Things</i>
ISO	<i>International Organization for Standardization</i>
JCR	<i>Journal Citation Report</i>
LCNAF	<i>Library of Congress Name Authority File</i>
LCSH	<i>Library of Congress Subject Headings</i>
LOD	<i>Linked Open Data</i>
MGB	Metadados Geoespaciais Brasileiros
NEWSDA	Núcleo de Estudos em Web Semântica e Dados Abertos

OGC	<i>Open Geospatial Consortium</i>
OWL	<i>Ontology Web Language</i>
PDF	<i>Portable Document Format</i>
RDB	<i>Relational DataBase</i>
RDBMS	<i>Relational DataBase Management System</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>RDF Schema Vocabulary</i>
SGBD	Sistema Gerenciador de Banco de Dados
SIG	Sistema de Informações Geográficas
SJR	<i>Scimago Journal Rank</i>
SKOS	<i>Simple Knowledge Organization System</i>
SOA	<i>Service Oriented Architecture</i>
SPARQL	<i>Sparql Protocol And Rdf Query Language</i>
SQL	<i>Structured Query Language</i>
SRS	<i>Spatial Reference System</i>
TGN	<i>The Getty Institute Thesaurus of Geographical Names</i>
TIC	Tecnologia da Informação e Comunicação
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
VCGE	Vocabulário Controlado de Governo Eletrônico
VIAF	<i>Virtual International Authority File</i>
VoID	<i>Vocabulary of Interlinked Datasets</i>
W3C	<i>World Wide Web Consortium</i>
WCS	<i>Web Coverage Service</i>
WFS	<i>Web Feature Service</i>
WKT	<i>Well-Know Text</i>
WMS	<i>Web Map Service</i>
WoS	<i>Web of Science</i>
XML	<i>eXtensible Markup Language</i>
XSLT	<i>eXtensible Stylesheet Language for Transformation</i>

SUMÁRIO

1 INTRODUÇÃO.....	17
1.1 PROBLEMA DA PESQUISA	21
1.2 TESE E HIPÓTESE	23
1.3 OBJETIVOS.....	23
1.3.1 Geral	24
1.3.2 Específicos.....	24
1.4 SUJEITO, JUSTIFICATIVA E CONTRIBUIÇÕES DA PESQUISA	24
1.5 METODOLOGIA	26
1.5.1 Metodologia da Revisão de Literatura	29
2 REVISÃO DE LITERATURA E TRABALHOS RELACIONADOS.....	36
2.1 REVISÃO DE LITERATURA	36
2.2 TRABALHOS RELACIONADOS	41
3 DADOS E METADADOS GEOESPACIAIS	46
3.1 METADADOS	46
3.1.1 Padrões e vocabulários para a estruturação de metadados	49
3.2 DADOS GEOESPACIAIS	52
3.2.1 Formatos de representação de dados geoespaciais.....	54
3.3 METADADOS GEOESPACIAIS	56
3.3.1 Padrões e perfis para metadados geoespaciais	58
3.3.2 Codificação de metadados geoespaciais.....	65
3.4 INFRAESTRUTURA DE DADOS ESPACIAIS (IDE).....	67
3.4.1 Serviço de Catálogo de Metadados Geoespaciais para a Web (CSW).....	72
4 WEB SEMÂNTICA E <i>LINKED DATA</i>	76
4.1 EVOLUÇÃO DA WEB SEMÂNTICA.....	76
4.2 TECNOLOGIAS E FERRAMENTAS DA WEB SEMÂNTICA.....	79
4.2.1 Identificador Uniforme de Recursos (URI - <i>Uniform Resource Identification</i>) ...	79
4.2.2 Modelo de dados RDF (<i>Resource Description Framework</i>).....	81
4.2.3 Sintaxes para serialização de dados RDF	82
4.2.4 Armazenamento e consulta de dados no modelo RDF.....	86
4.2.5 Vocabulários para descrição de dados em RDF	91

4.3 DADOS LIGADOS (<i>LINKED DATA</i>).....	100
4.3.1 Princípios de <i>Linked Data</i>	101
4.3.2 Dados Abertos Ligados (LOD)	102
4.3.3 Dados Geoespaciais Ligados (<i>Linked GeoData</i>).....	104
4.3.4 Boas Práticas para Publicação de Dados como <i>Linked Data</i>	106
4.4 INTEGRAÇÃO DE INFRAESTRUTURAS DE DADOS ESPACIAIS COM A WEB SEMÂNTICA.....	109
4.4.1 Boas práticas da OGC e W3C para Dados Espaciais na <i>Web</i>	110
4.4.2 Publicação de Catálogos de Metadados Geoespaciais como <i>Linked Data</i>	113
5 <i>FRAMEWORK</i> PARA PUBLICAÇÃO DE METADADOS GEOESPACIAIS SEMÂNTICOS NA <i>WEB</i>	117
5.1 VISÃO GERAL DO <i>FRAMEWORK</i>	117
5.2 DETALHAMENTO DO <i>FRAMEWORK</i>	120
5.2.1 Camada de Coleta e Conversão	120
5.2.2 Camada de Enriquecimento Semântico.....	126
5.2.3 Camada de Disponibilização	129
5.3 VALIDAÇÃO DO <i>FRAMEWORK</i>	131
5.3.1 Descrição da amostra.....	132
5.3.2 Execução da prova de conceito	134
5.3.3 Análise dos resultados	152
6 CONSIDERAÇÕES FINAIS	156
6.1 Trabalhos Futuros.....	159
REFERÊNCIAS	160
ANEXO A	171

1 INTRODUÇÃO

A imensa quantidade de dados disponibilizados na *Web* traz à tona a necessidade de melhorar sua busca e recuperação pelos usuários que, antes de acessarem um determinado recurso, como por exemplo um conjunto de dados (*dataset*) contendo alguns milhares de registros, precisam verificar se aquele recurso realmente atende suas necessidades informacionais. Para isso surgem os metadados como atributos de recursos informacionais, tanto físicos quanto digitais, com a finalidade de identificá-los e fornecer informações detalhadas a seu respeito, de forma que possam ser rapidamente encontrados e processados pelos usuários, quer sejam humanos ou máquinas.

Neste contexto, um *dataset* pode ser representado pela agregação de ‘dados’ e ‘metadados’, onde ‘dados’ expressa o conteúdo dos dados em si, e ‘metadados’ expressa a informação contextual sobre os dados, de tal forma que “a informação contida ou comunicada por um *dataset* é equivalente a uma agregação dos significados a ele atribuídos pelo usuário que irá consumi-lo.” (FURNER, 2016, p. 302, tradução nossa).

Neste sentido, os metadados devem fornecer informações adicionais que ajudem os usuários de dados a melhor entenderem o significado dos *datasets* descritos e sua estruturação, além de esclarecer outras questões relativas aos direitos e termos de licença, à organização que os gerou, à qualidade, aos métodos de acesso e à periodicidade de atualização (W3C; OGC, 2017).

No caso dos metadados que descrevem recursos (*datasets*) que podem ser referenciados espacialmente em relação à superfície da Terra, os chamados ‘metadados geoespaciais’, estes devem transmitir, além das informações mencionadas anteriormente, outras informações específicas a respeito da posição precisa daquele recurso, chamadas de ‘extensão espacial’, tais como o tipo de representação, as geometrias fornecidas, o sistema de referência de coordenadas utilizado, dentre outras.

Este tipo de dado que contém um componente de localização geográfica (dado geoespacial) é cada vez mais comum em nosso cotidiano. Com o surgimento dos sistemas de GPS (*Global Positioning System*), dos dispositivos móveis e da recente “Internet das Coisas” (IoT – *Internet of Things*), que operam em função de tal componente, percebeu-se, na última década, um notável crescimento na geração e consumo de dados geoespaciais, dando origem ao termo ‘*big data* geoespacial’.

No campo da Ciência e Tecnologia, a evolução dos sistemas de monitoramento (ambiental, clima, cobertura e uso da terra, etc.) baseados em imagens de satélites, bem como as atividades de ensino e pesquisa nas áreas das geociências tem sido capazes de gerar, em curtos espaços de tempo, enormes volumes de dados geoespaciais, principalmente em decorrência do uso intensivo de Sistemas de Informação Geográfica (SIG), que são ferramentas que facilitam a tarefa de combinar diferentes camadas de dados geoespaciais para a elaboração de mapas, com a finalidade de visualização de informações para a tomada de decisão.

Neste sentido, de acordo com Saracevic (1996, p. 43),

[...] uma vez que a ciência e a tecnologia são críticas para a sociedade (por exemplo, para a economia, saúde, comércio, defesa) é também crítico prover os meios para o fornecimento de informações relevantes para indivíduos, grupos e organizações envolvidas com a ciência e a tecnologia, já que a informação é um dos mais importantes insumos para se atingir e sustentar o desenvolvimento em tais áreas.

Assim, os dados geoespaciais podem cumprir esse papel de ser um meio para o fornecimento de informações relevantes em ciência e tecnologia, se diferenciando dos demais tipos de dados pelo fato de conterem elementos de geolocalização e de referência a outros dados, de forma que podem ser combinados em camadas sobre um mapa com a finalidade de dar maior sentido à informação ali apresentada.

Quanto ao enquadramento dos metadados e conjuntos de dados (*datasets*) como recursos de informação, Buckland (1991, p. 352, tradução nossa, grifo nosso) entende que “o termo ‘informação’ também é utilizado atributivamente para objetos, tais como **dados** e documentos, que são referidos como ‘informação’ porque são considerados como sendo informativos [...]”.

Por outro lado, os elementos de geolocalização aumentam consideravelmente o espaço necessário para armazenamento dos *datasets* gerados, bem como torna seu processamento mais complexo, o que exige maior alocação de recursos computacionais. Essas características diferenciadas demandam maiores cuidados quanto à descrição deste tipo de recurso de informação, uma vez que falhas ou incompletude nas descrições podem gerar desperdício de tempo e de uso de rede para um usuário que venha a recuperar um *dataset* que não atenda sua necessidade informacional.

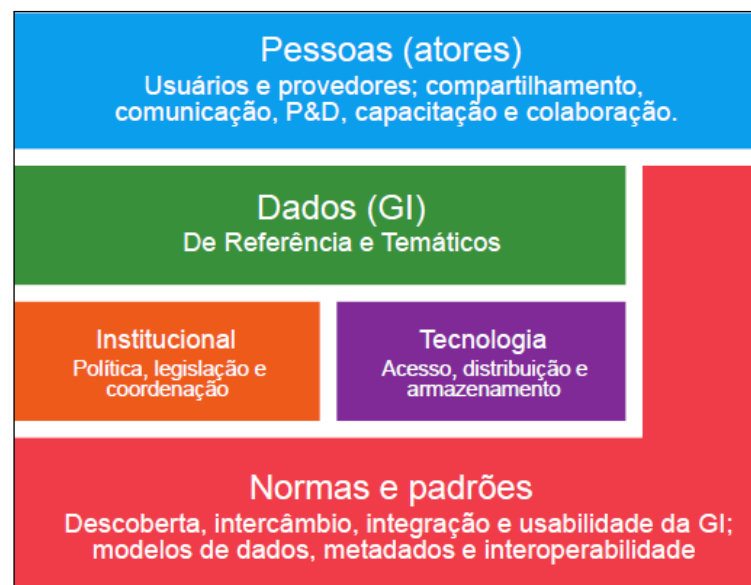
Diante da complexidade e dos altos custos para produzir e manter enormes volumes de dados geoespaciais, somados à crescente demanda pelo compartilhamento público e interoperabilidade entre os diversos provedores, é que, desde o início da década de 1990, diversos governos e entidades se lançaram na concepção e implementação das chamadas ‘Infraestruturas de Dados Espaciais (IDEs)’.

As IDEs são *frameworks* que integram um conjunto de tecnologias, políticas, arranjos institucionais, padrões e acordos que facilitam a busca, o acesso e a integração de diferentes tipos de dados (geo)espaciais (BRASIL, 2010). Dessa forma, uma IDE deve permitir que seus usuários, por meio do acesso a seu ‘geoportal’ na Web, possam encontrar, visualizar, reusar e interoperar conjuntos de dados geoespaciais, provenientes de diferentes produtores de dados, para atender suas necessidades. Nesse sentido, de acordo com Brasil (2010, p.15):

[...] uma IDE pode ser entendida como um conjunto de serviços que oferecem uma série de funcionalidades úteis e interessantes para uma comunidade de usuários de dados geoespaciais. Se antes a ênfase era nos dados que o usuário poderia acessar, agora a ênfase recai nos variados “usos” que podem ser feitos desses dados.

Para Davis Jr., Souza e Borges (2005), o objetivo principal das IDEs é oferecer serviços de acesso à informação geográfica (IG, ou *Geographical Information – GI*), com base em grandes catálogos de metadados, não importando para o usuário o local, meio ou estrutura física onde os dados são armazenados. A Figura 1 apresenta os componentes de uma IDE.

Figura 1 – Componentes de uma IDE.



Fonte: Brasil (2010)

Conforme mostrado na Figura 1, as IDEs têm como componentes centrais pessoas (atores ou *stakeholders*), dados geoespaciais (de referência e temáticos), instituições públicas ou privadas (por meio de políticas, legislação e coordenação) e tecnologias (de acesso, distribuição e armazenamento), todos apoiados em normas e padrões (modelos de dados, metadados, etc.).

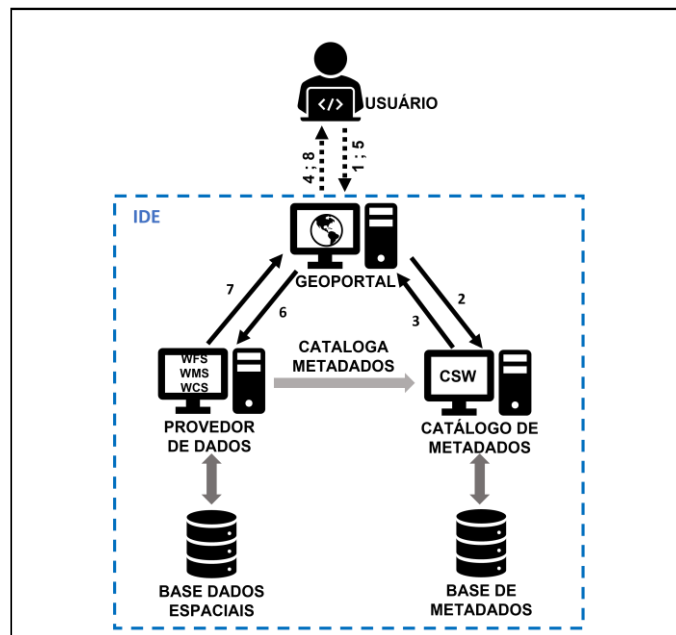
No Brasil, desde a criação da Infraestrutura Nacional de Dados Espaciais - INDE, por meio do Decreto nº 6.666¹, de 27 de novembro de 2008, todos os órgãos e entidades do poder executivo federal devem descrever, nessa plataforma, os metadados dos conjuntos de dados geoespaciais produzidos em suas atividades de pesquisa e desenvolvimento. Atualmente, instituições como IBGE, INPE, agências reguladoras, Embrapa, Fiocruz, entre outras, possuem IDEs próprias, integradas à INDE, possibilitando a busca, recuperação e visualização de conjuntos de dados geoespaciais por intermédio da disponibilização de serviços de catálogos de metadados.

Nas IDEs atuais, que são baseadas em serviços *Web*, os usuários fazem a solicitação e obtêm os metadados e respectivos *datasets* geoespaciais de que necessita por meio de um agente intermediário, normalmente implementado em um servidor *Web* chamado ‘geoportal’. Este, por sua vez, efetua as requisições de metadados a outro servidor *Web* que operacionaliza um catálogo de metadados descritores dos *datasets* e serviços oferecidos pela IDE por meio de um serviço *Web*, padronizado pelo *Open Geospatial Consortium* (OGC), chamado ‘*Catalog Service for the Web* (CSW)’. Neste serviço os metadados requisitados são comumente retornados no formato *eXtensible Markup Language* (XML).

Com os metadados em mãos, os usuários podem então solicitar ao geoportal os *datasets* geoespaciais desejados, os quais serão requisitados aos respectivos provedores de dados por meio de outro serviço *Web* específico, também padronizado pelo OGC, a depender do tipo de dado requisitado. Os *datasets* são, em geral, disponibilizados pelos provedores no formato *Geographic Markup Language* (GML), uma adaptação do formato XML específica para dados geoespaciais (BRASIL, 2010). A Figura 2 ilustra essa arquitetura.

¹ Decreto Nº 6.666, de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo federal, a Infraestrutura Nacional de Dados Espaciais - INDE, e dá outras providências.

Figura 2 – Arquitetura de uma IDE da geração atual.



Fonte: Elaborado pelo Autor.

O fluxo de informação representado na Figura 2 é o seguinte: (1) o usuário acessa o geoportal pela Web (via protocolo HTTP – *HyperText Transfer Protocol*) e solicita os metadados referentes aos *datasets* desejados; (2) o geoportal, por sua vez, faz a requisição de metadados ao servidor de catálogo de metadados por meio de um serviço Web chamado CSW; (3) o servidor de catálogo retorna ao geoportal os metadados requisitados em formato XML; (4) o geoportal disponibiliza ao usuário os metadados solicitados; (5) o usuário então solicita os *datasets* desejados ao geoportal; (6) o geoportal requisita ao provedor de dados, por meio de um serviço Web específico (*Web Feature Service* – WFS, *Web Map Service* – WMS, ou *Web Coverage Service* – WCS), os *datasets* geoespaciais; (7) o provedor de dados envia ao geoportal os *datasets* requisitados; e (8) o geoportal disponibiliza ao usuário os *datasets* geoespaciais que foram solicitados.

1.1 PROBLEMA DA PESQUISA

Na arquitetura das IDEs de geração atual, as buscas (por meio do serviço CSW) aos metadados disponibilizados nos servidores de catálogo das IDEs são sintáticas, ou seja, baseadas em palavras-chave. Neste tipo de busca, as palavras-chave fornecidas pelo usuário são utilizadas de forma literal, quer sejam isoladas (termo) ou agrupadas (frase), sem considerar o contexto ou significado específico delas, na composição da respectiva expressão de busca.

Dessa forma, se um usuário fizer uma busca utilizando as palavras-chave ‘incêndio’ e ‘floresta’, devem ser retornados metadados referentes aos *datasets* geoespaciais que tragam a indicação geográfica de áreas em diferentes tipos de florestas onde tenham ocorrido, ou estejam ocorrendo, focos de incêndio. No entanto, entre os resultados, podem ser apresentados também metadados referentes a outros *datasets* que tragam a indicação geográfica de incêndios (de qualquer tipo, urbano ou rural) em municípios (ou distritos, bairros, etc.) chamados ‘Floresta’ (atualmente no Brasil há dois municípios com este nome, sendo um no estado do Paraná e outro no estado de Pernambuco), o que não é de interesse do usuário. Além disso, podem ser ignorados nos resultados os metadados referentes a indicação geográfica de áreas com detecção de ‘queimada’ ou de ‘foco de calor’, por não serem associadas automaticamente ao termo ‘incêndio’.

Outro problema enfrentado pela atual geração de IDEs é a dificuldade do usuário em encontrar e recuperar os metadados que deseja diretamente por meio dos buscadores da *Web*, uma vez que os conteúdos dos servidores de catálogo das IDEs não são indexados pelos atuais mecanismos de busca existentes na *Web*. Isso se deve à arquitetura baseada em serviços *Web* das IDEs, que adotam protocolos e padrões especificados pela OGC, fazendo com que os dados e metadados geoespaciais fiquem na chamada ‘*deep Web*’, ou ‘*Web* não indexada’.

Assim, um desafio para a próxima geração de IDEs é a integração dessas plataformas com as tecnologias da *Web Semântica*, ou seja, a concepção de IDEs baseadas em dados ligados (*Linked Data*). Para isso, as tecnologias da *Web Semântica* (RDF, Ontologias, SPARQL, etc.) tem se mostrado eficazes para promover interoperabilidade de dados em nível semântico, ao mesmo tempo em que as boas práticas de *Linked Data* visam promover tal interoperabilidade semântica em larga escala ao usar a arquitetura da *Web* atual, que é constituída de padrões totalmente abertos, especificados pelo W3C. Neste sentido, a OGC, após conduzir experimentos sobre a evolução das IDEs, concluiu que um fator crítico para o sucesso de longo prazo dessas infraestruturas vem a ser a interoperabilidade semântica (CAMBOIM; SLUTER, 2013).

Diante dessa problematização, o presente trabalho pretende contribuir para a concepção das IDEs de próxima geração por meio de sua integração com as tecnologias e ferramentas da *Web Semântica*, de forma a facilitar a descoberta de metadados geoespaciais, em nível semântico, diretamente na *Web* por meio de buscadores, sem a necessidade de acessar um geoportal. Tal integração possibilitaria a recuperação e reuso de *datasets* geoespaciais tanto por humanos quanto por máquinas e agentes de software. Para isso, este trabalho busca responder as seguintes questões:

- a) Como efetuar a conversão de modelo de dados e de formato de serialização dos metadados geoespaciais catalogados em uma IDE, de forma a torná-los interoperáveis em nível semântico e legíveis também por máquinas e agentes de software na *Web*?
- b) Quais elementos de metadados geoespaciais poderiam ser ligados a outras fontes de dados na *Web*, de forma a enriquecê-los semanticamente e possibilitar a descoberta de informações adicionais pelos consumidores desses metadados?
- c) Como disponibilizar os conjuntos de metadados geoespaciais enriquecidos semanticamente, de forma a possibilitar sua indexação e facilitar sua consulta e recuperação na *Web*?

1.2 TESE E HIPÓTESE

Para responder as questões de pesquisa formuladas, este trabalho defende a **tese** de que ‘a integração das IDEs com a *Web Semântica* pode ser alcançada por meio da publicação de seus catálogos de metadados geoespaciais na *Web* seguindo as boas práticas de *Linked Data*’.

Neste sentido, a **hipótese** levantada neste trabalho é a de que ‘se um fluxo organizacional puder ser estabelecido e validado para orientar a execução das etapas de um processo de publicação na *Web* dos metadados catalogados em IDEs seguindo as boas práticas de *Linked Data*, então estes metadados geoespaciais semânticos poderão ser indexados e descobertos diretamente na *Web* pelos usuários, quer sejam humanos ou máquinas’.

Com isso, as IDEs terão seus registros de metadados geoespaciais publicados na *Web*, em formato semântico, possibilitando que sejam descobertos e recuperados pelos usuários (humanos ou máquinas) por meio dos mecanismos de busca na *Web*, ou diretamente por meio de uma URI HTTP e de consultas semânticas nas quais seja feita a desambiguação de palavras com vários sentidos e a incorporação de todos os significados e sinônimos dos termos fornecidos pelo usuário, de forma a expandir a consulta formulada. Além disso, os metadados obtidos podem apresentar elementos na forma de dados ligados, que podem ser explorados e levar a dados mais atualizados ou resultados não obtidos inicialmente.

1.3 OBJETIVOS

Considerando a tese e hipótese apresentadas, este trabalho tem como objetivo geral, e objetivos específicos, os descritos nas subseções a seguir:

1.3.1 Geral

Propor um *framework* que estabeleça um fluxo organizacional para orientar a execução das etapas de um processo de publicação na *Web*, em formato semântico, dos metadados geoespaciais catalogados em Infraestruturas de Dados Espaciais (IDEs), seguindo as boas práticas de *Linked Data*.

1.3.2 Específicos

Os objetivos específicos deste trabalho são:

- I. Apresentar os conceitos de dados e metadados geoespaciais, bem como a arquitetura e os componentes de uma Infraestrutura de Dados Espaciais, com ênfase no Serviço de Catálogo de Metadados para a *Web* (CSW);
- II. Conceituar e descrever as tecnologias e ferramentas da *Web Semântica* e as boas práticas de *Linked Data* e de dados espaciais na *Web*;
- III. Construir um *framework* conceitual para a publicação de metadados geoespaciais semânticos na *Web*, descrevendo as etapas e respectivos procedimentos que as compõem;
- IV. Validar o *framework* por meio da realização de uma prova de conceito na qual serão publicados os metadados geoespaciais catalogados na IDE de um provedor de dados geoespaciais.

1.4 SUJEITOS, JUSTIFICATIVA E CONTRIBUIÇÕES DA PESQUISA

Diante do problema de pesquisa apresentado, bem como a tese, hipótese, e os objetivos deste trabalho, é possível identificar como **sujeitos** desta pesquisa os provedores de dados geoespaciais, tais como o Instituto Brasileiro de Geografia e Estatística (IBGE), o Instituto Nacional de Pesquisas Espaciais (INPE), a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), dentre diversas outras Instituições Governamentais que mantêm IDEs em operação, devidamente integradas à Infraestrutura Nacional de Dados Espaciais (INDE) coordenada pelo IBGE.

Embora as IDEs e o *Linked Data* sejam bastante diferentes em relação a suas arquiteturas, fluxos de trabalho e padrões aplicados, eles podem ser considerados complementares (WIEMANN; BERNARD, 2016). Neste sentido, a integração das IDEs com

as tecnologias e ferramentas da *Web Semântica* poderá facilitar a descoberta e recuperação dos metadados e *datasets* geoespaciais por usuários não-especialistas, os quais ainda encontram dificuldades para satisfazerem suas necessidades de informação geoespacial nas IDEs atuais, principalmente quando precisam descobrir e combinar dados de diferentes fontes externas (interoperabilidade).

Assim, justifica-se que essa integração contribuirá para uma mudança no foco das IDEs de próxima geração, que passarão a ser orientadas para os usuários. E as boas práticas de *Linked Data* serão essenciais para essa mudança, ao possibilitar a redução do custo de integração de dados e simplificar a criação de produtos de informação geoespacial na *Web*. Por conta disso, a elaboração de um *framework* para orientar um processo de publicação de metadados geoespaciais semânticos na *Web* poderá ajudar diversas organizações (governos, instituições acadêmicas e de pesquisa, entre outras) a promoverem tal integração.

O *framework* proposto neste trabalho poderá ser aplicado às IDEs que mantenham catálogos de metadados implementados sob o serviço CSW, com registros formatados no padrão ISO 19115, codificados em XML.

É esperado que, com a aplicação do *framework* proposto, a ligação dos valores de diversos elementos de metadados com a *Web* de dados, por meio do *Linked Data*, permita que o usuário possa verificar a acurácia e proveniência deles, além de explorar outros dados relacionados que não estavam previstos em sua busca inicial, o que poderá contribuir para um melhor atendimento de suas necessidades de informação.

A contribuição deste trabalho para as pesquisas na área de *Web Semântica*, em especial as conduzidas pelo Núcleo de Estudos em *Web Semântica* e Dados Abertos – NEWSDA, está em promover a publicação e interoperabilidade semântica de metadados geoespaciais, os quais poderão ser gerados e consumidos não somente por humanos, mas também por máquinas e agentes de software na *Web*, sendo enriquecidos por meio da conexão entre diversos conjuntos de dados. Marcondes (2016, p. 65) afirma que a “[...] a Ciência da Informação sempre se preocupou com a questão do controle semântico dos recursos informacionais, através das linguagens documentárias.”. Neste sentido, a CI tem estudado, em especial, as relações semânticas em ambientes digitais, que são a capacidade de computadores realizarem inferências automáticas a partir dos *links* entre os nós de uma rede semântica, de forma a descobrirem conhecimento não diretamente explicitado.

Além disso, este trabalho contribui para a Ciência da Informação, especificamente para a linha de pesquisa ‘Informação e Tecnologia’ do PPGCI/UNESP, no avanço em questões relativas à representação e recuperação da informação por meio de metadados. A aplicação das

ferramentas e tecnologias da *Web Semântica* permitirão que os metadados passem a descrever um recurso de informação considerando seu contexto e significado, ligando-o a outros recursos semelhantes ou complementares na *Web* de dados. Para isso, as boas práticas de *Linked Data*, ao possibilitarem a estruturação e contextualização dos elementos de metadados através do uso de vocabulários e de URIs com *links* para outras fontes de dados, permitirão melhorar a eficiência e eficácia dos processos de descoberta e recuperação da informação geoespacial a fim de atender às necessidades informacionais dos usuários.

Tal propósito vai ao encontro do entendimento de Borko (1968, p. 5, tradução nossa, grifo nosso) de que a Ciência da Informação:

[...] é uma ciência interdisciplinar que investiga as propriedades e comportamento da informação, as forças que governam os fluxos e os usos da informação, e **as técnicas, tanto manual quanto mecânica, de processamento da informação, visando sua armazenagem, recuperação, e disseminação ideal.**

Apesar de, à primeira vista, a publicação de metadados na *Web*, objeto desta pesquisa, parecer um problema relacionado à disseminação de dados geoespaciais, na verdade se trata de um problema de recuperação da informação, uma vez que o real objetivo da publicação de metadados geoespaciais semânticos na *Web* é facilitar sua descoberta e recuperação com o intuito de atender plenamente a necessidade de informação do usuário. Neste sentido, Mooers (1951, p. 25, tradução nossa) esclarece que a recuperação da informação “[...] engloba os aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregados para o desempenho da operação.”.

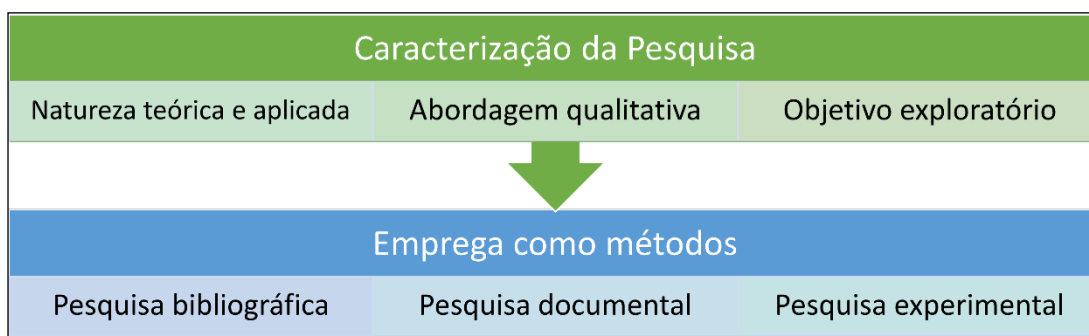
Por fim, para a área da Geoinformação, este trabalho contribui para a concepção das IDEs de próxima geração ao propor um *framework* para integrá-las com a *Web Semântica*, com o intuito de ampliar a descoberta e acesso aos conjuntos de dados geoespaciais nelas disponibilizados. Isso levará a um melhor aproveitamento, pela sociedade, dos recursos de informação disponibilizados pelas IDEs, principalmente por usuários não especialistas, de forma a justificar os elevados investimentos para manutenção deste tipo de plataforma.

1.5 METODOLOGIA

Uma vez definidos problema e objetivo, segue-se para o planejamento e escolha do tipo de pesquisa, passo este fundamental para o delineamento do estudo, visando o alcance do objetivo proposto e, por conseguinte, responder as questões de pesquisa (SANTOS et al., 2018).

Neste sentido, a consecução do objetivo geral deste trabalho, conforme ilustrado na Figura 3, é resultado de uma pesquisa de natureza teórica e aplicada, de abordagem qualitativa e do tipo exploratória, que emprega como métodos a revisão bibliográfica e a pesquisa documental, nas quais os conteúdos obtidos são analisados por meio da aplicação do método de ‘Análise de Conteúdo’ proposto por Bardin (1979). A seguir, é apresentada uma descrição detalhada e fundamentada dos procedimentos metodológicos aplicados neste trabalho.

Figura 3 – Caracterização da pesquisa.



Fonte: Elaborado pelo autor.

Quanto à natureza, de acordo com Gil (2010), a pesquisa teórica e aplicada “objetiva gerar conhecimentos para aplicação prática e dirigidos à solução de problemas específicos”, o que vai ao encontro do objetivo geral deste trabalho, que é propor um modelo para publicação de metadados geoespaciais semânticos na *Web*, que possa ser aplicado para responder às questões de pesquisa apresentadas.

No que se refere à abordagem, trata-se de uma pesquisa qualitativa que, segundo Silveira e Córdova (2009) “[...] busca explicar o porquê das coisas, exprimindo o que convém ser feito, mas não quantificam os valores [...] pois os dados são não-métricos (suscitados e de interação) e se valem de diferentes abordagens.”.

Com relação ao tipo de pesquisa foi definida a pesquisa exploratória, que tem como objetivo reunir dados, informações e padrões sobre um problema de pesquisa com pouco ou nenhum estudo anterior (MUELLER, 2007). A pesquisa exploratória tem ainda como propósito proporcionar maior familiaridade com o problema, buscando torná-lo mais explícito. Trata-se de um tipo de pesquisa bastante flexível em seu planejamento, de forma a levar em consideração os mais variados aspectos envolvidos no problema em estudo.

Como métodos empregados na pesquisa exploratória, para possibilitar a coleta e análise de dados com o propósito de aprofundar o entendimento do problema, delimitá-lo e possibilitar sua fundamentação teórica, foi realizada uma pesquisa bibliográfica por meio da busca e coleta

da bibliografia relevante e de trabalhos já realizados sobre o tema em estudo, a partir das bases de dados *Web of Science (WoS)*, *Scopus*, *Scielo* e *Brapi*, no período de 2011 a 2021 (dez anos anteriores), no qual o tema ‘integração de Infraestruturas de Dados Espaciais (IDE) com a *Web Semântica*’ passou a ser mais intensamente abordado na comunidade científica. Além da pesquisa bibliográfica, foi realizada ainda uma pesquisa documental, onde foram coletados para análise manuais e documentação das ferramentas e tecnologias da *Web Semântica* a serem aplicadas, além dos relatórios de boas práticas para dados geoespaciais na *Web*.

Uma vez coletados os dados e informações por meio das pesquisas bibliográfica e documental, estes serão analisados utilizando o método da Análise de Conteúdo proposto por Bardin (1979). Trata-se de uma técnica de organização e análise de dados coletados na pesquisa qualitativa. Dentre as diversas modalidades de análise de conteúdo (lexical, de expressão, de relações, temática e de enunciação), a análise temática (ou categorial) é considerada a mais simples e apropriada para pesquisas qualitativas, sendo a empregada neste trabalho. Essa modalidade de análise trabalha com o conceito de tema, relacionando sua frequência de citação em um texto à uma afirmação a respeito de determinado assunto, o qual recebe um determinado peso.

Como pesquisa experimental, na segunda etapa deste trabalho, uma vez construído o *framework* proposto, este será validado por meio de uma prova de conceito no qual serão publicados os metadados geoespaciais catalogados na IDE de um provedor de dados geoespaciais a fim de confirmar as hipóteses levantadas inicialmente.

A seguir, o Quadro 1 resume a metodologia apresentada nesta seção, ao relacionar os métodos, técnicas e fontes de pesquisa empregados para a consecução de cada objetivo específico do trabalho.

Quadro 1 – Métodos, técnicas e fontes de pesquisa empregados neste trabalho.

OBJETIVO ESPECÍFICO	MÉTODOS	TÉCNICAS	FONTES
Apresentar os conceitos de dados e metadados geoespaciais, bem como a arquitetura e os componentes de uma Infraestrutura de Dados Espaciais.	Pesquisa Bibliográfica; Pesquisa Documental.	Análise de conteúdo (Bardin).	Artigos em periódicos e conferências, livros, relatórios, notas e normas técnicas, especificações, dissertações e teses relacionados ao tema da pesquisa.
Conceituar e descrever as tecnologias e ferramentas da <i>Web Semântica</i> e as boas práticas de <i>Linked Data</i> e de dados espaciais na <i>Web</i> .			
Construir um <i>framework</i> conceitual para a publicação de metadados geoespaciais semânticos na <i>Web</i> , descrevendo as etapas e respectivos procedimentos.	Pesquisa Experimental.	Construção de <i>framework</i> .	Material coletado e analisado nas pesquisas bibliográfica e documental.

Validar o <i>framework</i> desenvolvido por meio de uma prova de conceito na qual serão publicados os metadados geoespaciais catalogados na IDE de um provedor de dados geoespaciais.		Prova de conceito para validação do <i>framework</i> .	Metadados geoespaciais catalogados na IDE de um provedor de dados geoespaciais.
---	--	--	---

Fonte: Elaborado pelo autor.

A seguir é apresentada a metodologia empregada na revisão da literatura deste trabalho, a qual consistiu em um levantamento e seleção de artigos publicados em periódicos, ou anais de conferências, ambos indexados em bases de dados, que sejam relacionados ao tema ‘publicação de metadados geoespaciais na *Web Semântica*’, com a finalidade de compor um portfólio de trabalhos relacionados, bem como determinar o atual estágio de evolução do tema.

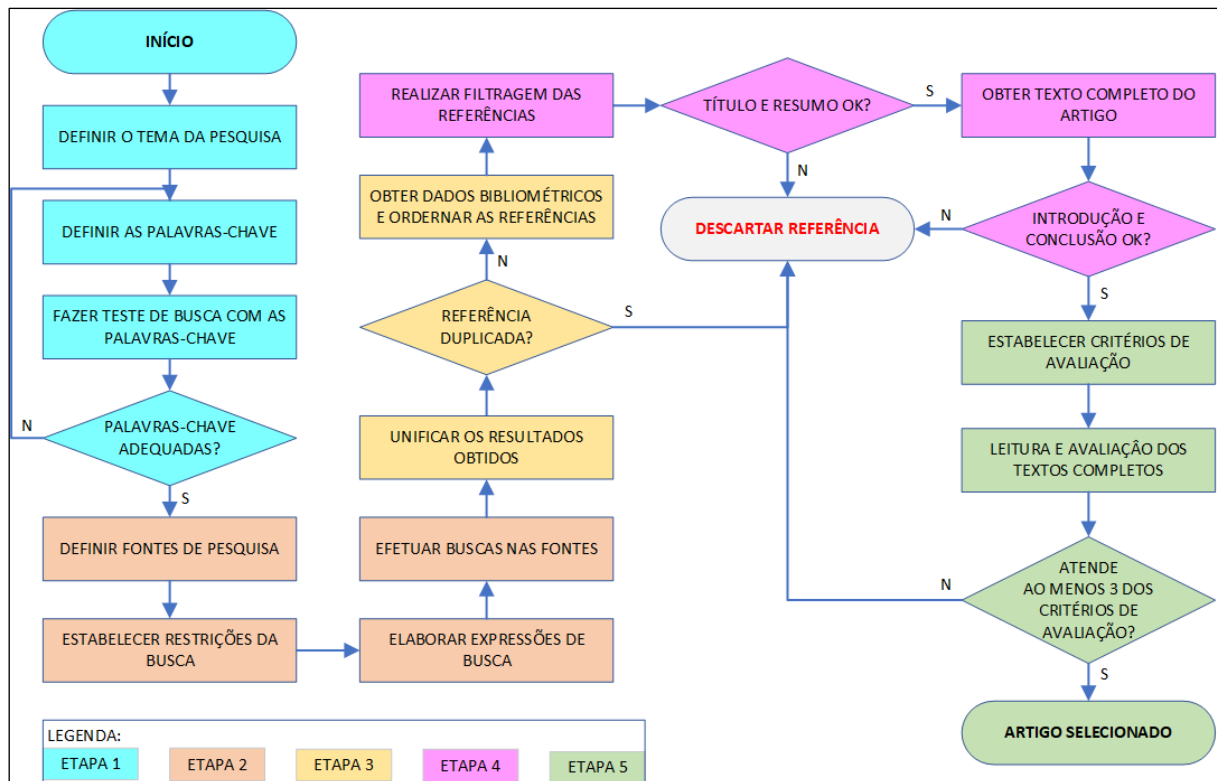
1.5.1 Metodologia da Revisão de Literatura

A revisão de literatura deste trabalho aplicou como procedimento metodológico específico o processo *ProKnow-C*, o qual, de acordo com Afonso et al. (2012), “se constitui em uma metodologia para construção do conhecimento”, composta de quatro etapas: “1) seleção do portfólio bibliográfico que proporcionará a revisão de literatura; 2) análise bibliométrica do portfólio bibliográfico; 3) análise sistêmica do portfólio bibliográfico; 4) elaboração dos objetivos de pesquisa.”. Cada etapa possui uma série de procedimentos sequenciais pré-estabelecidos, que vão desde a definição das fontes de busca até a seleção do portfólio relevante sobre o tema de pesquisa.

Dessa forma, os procedimentos executados nessa revisão de literatura foram agrupados, de forma sequencial, ao longo de cinco etapas, que são uma adaptação das etapas 1, 2 e 3 do processo *ProKnow-C* (AFONSO *et al.*, 2012). A descrição dos procedimentos e a ordem de execução dos mesmos, bem como o sequenciamento das etapas (diferenciadas por cores) estão representados no fluxograma apresentado na Figura 4.

Cada grupo de elementos do fluxograma dos procedimentos da revisão de literatura, apresentado na Figura 4, recebeu uma cor diferentes com a finalidade de identificar cada uma das etapas do processo ilustrado. Assim, a etapa 1 tem seus elementos na cor ciano, os elementos da etapa 2 receberam a cor bege, os elementos da etapa 3 receberam cor amarela, os elementos da etapa 4 receberam cor magenta e, por fim, os elementos da etapa 5 receberam cor verde.

Figura 4 – Fluxograma dos procedimentos da revisão de literatura.



Fonte: Adaptado de Afonso *et al.* (2012).

Nas subseções a seguir são descritos, de forma detalhada, os procedimentos executados em cada uma das etapas representadas na figura anterior.

1.5.1.1 Etapa 1: Definição do Tema e Palavras-Chave

Na primeira etapa dessa revisão de literatura (cor ciano no fluxograma), foram realizados os procedimentos de definição do tema de pesquisa e, em seguida, de escolha das palavras-chave que o representem adequadamente. Convém ressaltar que o procedimento de escolha de palavras-chave é o que requer maior atenção do pesquisador, uma vez que determinará quais artigos serão selecionados para as etapas posteriores. Caso as palavras-chave não sejam escolhidas adequadamente, artigos relevantes sobre o tema de pesquisa podem ficar de fora quando da realização do procedimento de busca nas bases de dados, o que irá impactar negativamente na qualidade da revisão de literatura e, conseqüentemente, na definição do estágio de maturidade do tema da pesquisa. O Quadro 2 apresenta o tema e as respectivas palavras-chave definidas para esta pesquisa.

Quadro 2 – Definição do tema e das palavras-chave que representam a pesquisa.

TEMA DA PESQUISA	– Publicação de metadados geoespaciais na <i>Web Semântica</i> .
PALAVRAS-CHAVE INICIAIS	– Metadados geoespaciais (<i>geospatial metadata</i>). – <i>Web Semântica</i> (<i>semantic Web</i>). – <i>Web</i> de dados (<i>Web of data</i>). – Dados Abertos Ligados (<i>linked open data</i>). – <i>Framework</i> para Descrição de Recursos (<i>resource description framework</i>). – Informação Geoespacial (<i>geospatial information</i>).

Fonte: Elaborado pelo autor.

Uma vez escolhidas as palavras-chave, as traduções destas para o inglês foram testadas, tanto isoladamente quanto de maneira combinada, nas seguintes plataformas de busca: *Scopus* e *Web of Science*. Levando em consideração os resultados obtidos nestes testes, a palavra-chave ‘*geospatial metadata*’ foi substituída por ‘*geo* metadata*’, onde o caractere ‘*’ representa ‘qualquer variação’, ou seja, abrangerá tanto o termo ‘*geospatial*’ como o termo ‘*geographic*’, já que ambos costumam ser usados indistintamente em artigos para se referir a este tipo de metadado. Da mesma forma, a palavra-chave ‘*linked data*’ foi utilizada no lugar de ‘*linked open data*’, diante da constatação de que vários trabalhos consideram que nem todo dado ligado (*linked data*) é aberto (*open*). Por fim, a palavra-chave ‘*resource description framework*’ foi substituída por seu acrônimo ‘*RDF*’, o qual é mais comumente encontrado nos artigos, e a palavra-chave ‘*geospatial information*’, por não trazer resultados relevantes, foi descartada.

1.5.1.2 Etapa 2: Buscas em Bases de Dados Indexadas

Após a definição do tema de pesquisa e a escolha e teste das palavras-chave que o representam, na segunda etapa (cor bege no fluxograma) foram definidas as fontes de busca de artigos, bem como as restrições da busca nessas fontes, tais como os tipos de documentos a serem encontrados, o período de abrangência e as respectivas expressões de busca, elaboradas a partir das palavras-chave definidas na etapa anterior. O Quadro 3 apresenta as fontes e as restrições de busca estipuladas para esta pesquisa.

Quadro 3 – Definição das fontes e das restrições das buscas a serem efetuadas.

FONTES DE BUSCA	– <i>Scopus</i> – <i>Web of Science</i> – <i>SciELO</i> – <i>Brapi</i>
TIPOS DE DOCUMENTOS	– Artigos em periódicos – Artigos em conferência

PERÍODO DE ABRANGÊNCIA	– Últimos 10 anos (2011 a 2020/21)
EXPRESSÕES DE BUSCA	– Scopus: TITLE-ABS-KEY("geo* metadata") AND TITLE-ABS-KEY("semantic Web" OR "Web of data" OR "linked data" OR "linked open data" OR "rdf") AND DOCTYPE(ar OR cp) AND PUBYEAR > 2010 – Web of Science: TS=("geo* metadata" AND ("semantic Web" OR "Web of data" OR "linked data" OR "linked open data" OR "rdf")) AND DT=(Article OR Proceedings Paper) AND PY=(2011-2021) – SciELO: ("geo* metadata") AND ("semantic Web" OR "linked data" OR "linked open data" OR "Web of data" OR "rdf") – Brapi: "geospatial metadata" AND "semantic Web"

Fonte: Elaborado pelo autor.

Definidas as fontes e as restrições, foram efetuadas as buscas. Os resultados obtidos, em cada uma das bases consultadas, foram exportados para arquivos no formato *BibTeX*, que foram salvos para serem organizados na etapa seguinte. Convém ressaltar que as buscas nas bases *SciELO* e *Brapi* não retornaram resultados para a respectiva expressão de busca.

1.5.1.3 Etapa 3: Organização dos Resultados das Buscas

Na terceira etapa (cor amarela no fluxograma), os resultados obtidos nas buscas da etapa anterior foram organizados para facilitar a execução das próximas etapas, de filtragem e de seleção de textos, respectivamente. Para isso, foram executados os seguintes passos: 1) unificar os metadados coletados nas bases de dados consultadas, descartando as referências duplicadas (por meio de um *script* em linguagem *R*); e 2) obter dados bibliométricos e ordenar as referências (equação *InOrdinatio* - Ano+Citações+Fator Impacto).

Como passo inicial desta etapa, os diferentes arquivos (em formato *BibTeX*) contendo os resultados das buscas efetuadas em cada uma das bases consultadas foram unificados, sendo removidas as referências duplicadas. Os dados unificados foram então salvos em um arquivo no formato CSV (texto separado por vírgulas, em português), o qual pode ser facilmente convertido em uma planilha para ser usada no próximo passo. Para auxiliar na execução deste passo, foi elaborado um *script* em linguagem *R*, o qual é apresentado no Quadro 4.

Quadro 4 – *Script* para unificação e remoção de duplicidades nos resultados das buscas.

```
library(bibliometrix)
x <- convert2df("path_to_file/scopus.bib", dbsource = "scopus", format = "bibtex")
y <- convert2df("path_to_file/wos.bib", dbsource = "isi", format = "bibtex")
z <- mergeDbSources(X, Y, remove.duplicated = TRUE)
w <- z[,c("AU", "TI", "SO", "AB", "DE", "ID", "DI", "LA", "DT", "TC", "PY")]
write.table(w, "path_to_file/artigos.csv", sep=";", row.names=FALSE)
```

Fonte: Elaborado pelo autor.

No *script* apresentado, é utilizada uma biblioteca (*library*) chamada *bibliometrix* (Aria; Cuccurullo, 2017). Por meio desta biblioteca é que são lidos os arquivos de resultados, em formato *BibTeX*, obtidos nas buscas nas bases *Scopus* (variável X) e *Web of Science* (variável Y), e efetuada a unificação deles, com a remoção das duplicidades (variável Z). Nas duas últimas linhas do *script* são selecionadas e organizadas as colunas autor (AU), título (TI), nome do periódico ou dos anais de evento (SO), resumo (AB), palavras-chave do autor (DE), palavras-chave da indexação (ID), identificador DOI (DI), idioma (LA), tipo de artigo (DT), total de citações (TC) e ano de publicação (PY), que foram gravadas em um arquivo no formato CSV (*write.table*), a ser utilizado no passo seguinte.

No segundo passo desta etapa, foram coletados dados bibliométricos das referências unificadas no passo anterior, mais especificamente os valores de Fator de Impacto (FI) de cada periódico. As fontes de consulta mais comumente utilizadas para se obter o FI são o *Scimago Journal Rank* (SJR), publicado pela *Scopus*, e o *Journal Citation Report* (JCR), publicado pela *Clarivate* (*Web of Science*). Os valores de FI obtidos para cada referência foram inseridos em uma nova coluna acrescentada às existentes no arquivo CSV gerado no passo anterior.

A obtenção dos valores de FI dos periódicos se fez necessária neste passo para possibilitar a ordenação dos resultados por meio da aplicação da equação *InOrdinatio*. A equação *InOrdinatio* é produto de um modelo multicritério de avaliação, utilizado para ranquear artigos científicos, baseado na combinação de três critérios: a) o fator de impacto do periódico em que o artigo foi publicado; b) o número de citações; e c) o ano de publicação (PAGANI; KOVALESKI; RESENDE, 2015). O Quadro 5 apresenta a equação e a descrição de suas variáveis.

Quadro 5 – Equação *InOrdinatio* para ranqueamento de artigos científicos.

***InOrdinatio* = (IF/1000) + α * [10 – (RY – PY)] + (Σ Ci), onde:**

IF = *Impact Factor*: fator de impacto do periódico no qual o artigo foi publicado;

α = Peso a ser dado ao critério “ano de publicação”, podendo variar de 1 a 10;

RY = *Research Year*: ano em que a pesquisa bibliográfica está sendo feita;

PY = *Publication Year*: ano de publicação do artigo analisado;

Σ Ci = *Citation*: número de citações do artigo analisado.

Fonte: Adaptado de Pagani, Kovaleski e Resende (2015).

Na equação *InOrdinatio*, o fator do impacto (IF) do periódico é dividido por 1.000, com o intuito de normalizar seu valor (troca do milhar por decimal) em relação aos demais critérios (ano de publicação e número de citações). Por conseguinte, o critério ano é calculado pela diferença entre o ano em que a pesquisa bibliográfica está sendo realizada (RY) e o ano de

publicação do artigo (PY), sendo o resultado dessa diferença (RY-PY) subtraído da constante 10, que representa o horizonte máximo da pesquisa bibliográfica realizada (últimos 10 anos), e o resultado obtido é multiplicado pelo peso (α , de 1 a 10) dado a este critério em contraponto aos demais. Por fim, o critério número de citações (ΣCi) é a reprodução do valor apresentado na coluna TC da planilha, derivada dos resultados das buscas efetuadas. O valor final da equação é a soma dos valores resultantes desses três critérios.

Assim, obtido o valor da equação *InOrdinatio* para cada referência da planilha, foi realizada a ordenação (ranqueamento) das referências, partindo do maior para o menor valor dessa coluna, o que auxiliou na realização da próxima etapa desta revisão de literatura, que foi a aplicação de filtros para seleção dos textos a serem analisados.

1.5.1.4 Etapa 4: Aplicação de Filtros para Seleção de Textos

A seguir, na quarta etapa (cor magenta no fluxograma), foi realizada a filtragem das referências obtidas nas buscas efetuadas na etapa 2, após os procedimentos de unificação e ordenação executados na etapa 3. Como resultado dessa etapa, serão obtidos os textos completos dos artigos a serem analisados na próxima etapa desta revisão de literatura. A aplicação de filtros foi realizada a partir de três passos: 1) analisar o título e o resumo; 2) obter o texto completo do artigo; e 3) analisar as seções Introdução e Conclusão.

Assim, o primeiro filtro aplicado nesta etapa foi a análise do título e do resumo de cada um dos artigos, na ordem em que foram organizados na planilha (valor da coluna *InOrdinatio* decrescente). Caso estes dois elementos do artigo estivessem alinhados ao tema de pesquisa, então prosseguia-se para o segundo passo, de obtenção do texto completo do artigo. Caso contrário, a referência era imediatamente descartada.

Para auxiliar na execução do segundo passo, de obtenção dos textos completos, foi utilizado o gerenciador de referências *Zotero*, que auxiliou na organização dos textos baixados junto às fontes indicadas nos registros selecionados nas bases de dados, bem como facilitar a posterior elaboração de referências bibliográficas.

Após a obtenção dos textos completos dos artigos selecionados foi realizada, no terceiro passo, a análise das seções de introdução e de conclusão de cada artigo. Da mesma forma que no filtro do passo 1, caso estas duas seções do artigo estivessem alinhadas ao tema de pesquisa, então o artigo era selecionado para a próxima etapa e, caso contrário, a referência e o respectivo texto completo do artigo eram descartados.

1.5.1.5 Etapa 5: Análise dos Textos Completos

Como última etapa (cor verde no fluxograma) da revisão de literatura, foi realizada a análise do texto completo de cada um dos artigos selecionados na etapa anterior. Para isso, foram executados os seguintes passos: 1) estabelecer os critérios de avaliação dos artigos; e 2) efetuar a leitura e avaliação dos textos completos selecionados.

No primeiro passo, foram estabelecidos os critérios para avaliação dos artigos selecionados na etapa anterior, que se dará por meio da leitura do texto completo. Estes critérios estão relacionados no Quadro 6.

Quadro 6 – Critérios para avaliação dos artigos selecionados para leitura do texto completo.

Nº.	CRITÉRIO
1	Propõe processo, método ou técnica para a publicação de metadados (geo)espaciais na <i>Web</i> ?
2	Aborda fundamentos/conceitos e/ou aplica tecnologias da <i>Web Semântica</i> (LOD/RDF/SparQL) a metadados (geo)espaciais?
3	Detalha ao menos um dos seguintes procedimentos para publicação de metadados (geo)espaciais na <i>Web</i> ? a) Conversão de formatos (ISO 19115/GML/XML) para RDF; b) Armazenamento de triplas RDF (<i>Triple Store</i>); c) Disponibilização e acesso na <i>Web</i> de Dados;
4	Aborda a aplicação de boas práticas para publicação de dados na <i>Web</i> (<i>Linked Data</i> e/ou <i>W3C/OGC</i>)?
5	Apresenta metodologia bem definida e resultados condizentes com os objetivos propostos?

Fonte: Elaborado pelo autor.

Uma vez estabelecidos os critérios para avaliação dos artigos, no passo seguinte, os respectivos textos completos foram lidos integralmente e avaliados quanto aos critérios estabelecidos. Ao final, os artigos que atenderam ao menos três dos cinco critérios foram selecionados para compor o portfólio de trabalhos relacionados. Os resultados obtidos nessa revisão de literatura, e os trabalhos relacionados são apresentados na Seção a seguir.

Concluindo a presente Seção de introdução, o restante do trabalho está estruturado da seguinte forma: a seguir, a Seção 2 apresenta o portfólio de trabalhos relacionados; a Seção 3 apresenta os conceitos de dados e metadados geoespaciais, bem como a arquitetura e os componentes de uma Infraestrutura de Dados Espaciais; na sequência, a Seção 4 faz uma revisão dos conceitos, tecnologias e ferramentas da *Web Semântica*, e das boas práticas de *Linked Data* e de publicação de dados espaciais na *Web*; a Seção 5 detalha o *framework* para publicação de metadados geoespaciais semânticos na *Web*, bem como sua validação e resultados obtidos; e, por fim, a Seção 6 apresenta as considerações finais.

2 REVISÃO DE LITERATURA E TRABALHOS RELACIONADOS

Nesta Seção são apresentados os resultados da revisão de literatura realizada com o intuito de compor um portfólio de trabalhos relacionados, bem como determinar o atual estágio de evolução do tema ‘publicação de metadados geoespaciais na Web Semântica’. Também será apresentado o posicionamento deste trabalho quando comparado aos trabalhos relacionados que foram selecionados para o portfólio.

2.1 REVISÃO DE LITERATURA

A partir do tema de pesquisa e das respectivas palavras-chave que o representam, definidos na etapa 1 (vide Quadro 2), e das fontes de pesquisa, das restrições e das expressões de busca, definidas na etapa 2 (vide Quadro 3), foram realizadas as buscas nas bases de dados *Scopus* e *Web of Science*. Como resultados, foram obtidas 21 referências na *Scopus* e 11 referências na *Web of Science*. No entanto, durante o procedimento de unificação dos resultados (passo 1 da etapa 3), foi possível notar que 9 das 11 referências encontradas na *Web of Science* também foram encontradas na *Scopus*, e que uma das referências encontradas nessa última estava duplicada. Com isso, foram obtidas, ao todo, 22 referências exclusivas.

Já no passo 2 da terceira etapa, foram coletados, por meio de consulta ao *Scimago Journal Rank* (SJR)², quando encontrados, os valores do Fator de Impacto dos periódicos ou dos anais de eventos, conforme o tipo de artigo de cada referência. O SJR foi escolhido como fonte de consulta para o FI dos artigos selecionados neste trabalho devido ao fato de que 20 das 22 referências encontradas nas buscas estavam indexadas na base de dados *Scopus*.

Uma vez coletado o valor do FI de cada referência, em seguida foi calculado o respectivo valor da equação *InOrdinatio*, o que possibilitou a ordenação (ranqueamento) das referências obtidas na etapa anterior. Na equação, foi atribuído o valor 1 para a variável α , o que significa que o ano de publicação do artigo tem o mesmo peso que os demais critérios (fator de impacto e total de citações). A Tabela 1 mostra, para cada artigo em análise, os dados das colunas ‘Autores’, ‘Ano de publicação’, ‘Tipo de documento’, ‘Total de citações’, ‘Fator de Impacto’ e ‘*InOrdinatio*’, ordenada por essa última coluna.

² Disponível em <https://www.scimagojr.com/journalrank.php>. Acesso via Portal de Periódicos CAPES.

Tabela 1 – Valores do Fator de Impacto e da equação *InOrdinatio* de cada artigo.

Autores (AU)	Ano de Publicação (PY)	Tipo de Documento (DT)	Total de Citações (TC)	Fator de Impacto (IF/1000)	<i>InOrdinatio</i> ($\alpha = 1$)
YUAN <i>et al.</i>	2013	Artigo em Periódico	15	2,616	19,62
KALANTARI <i>et al.</i>	2014	Artigo em Periódico	12	1,356	16,36
LI <i>et al.</i>	2020	Artigo em Conferência	0	4,027	13,03
CHEATHAM <i>et al.</i>	2018	Artigo em Periódico	6	-	13,00
YUE <i>et al.</i>	2012	Artigo em Periódico	11	0,491	12,49
FUGAZZA <i>et al.</i>	2018	Artigo em Periódico	3	0,655	10,66
FUGAZZA <i>et al.</i>	2016a	Artigo em Periódico	5	0,655	10,66
FUGAZZA <i>et al.</i>	2021	Artigo em Periódico	0	0,386	10,39
GRAUPNER e NÜST	2020	Artigo em Conferência	0	0,537	9,54
FUGAZZA <i>et al.</i>	2016b	Artigo em Conferência	4	0,175	9,18
MCKEE, G.	2019	Artigo em Periódico	0	0,771	8,77
FUGAZZA <i>et al.</i>	2017	Artigo em Periódico	2	0,655	8,66
SHVAIKO <i>et al.</i>	2012	Artigo em Conferência	7	-	8,00
ZHANG <i>et al.</i>	2013	Artigo em Conferência	2	-	4,00
MENEZES <i>et al.</i>	2014	Artigo em Periódico	0	0,772	3,77
RENTERIA-AGUALIMPIA <i>et al.</i>	2013	Artigo em Conferência	1	0,200	3,20
DO <i>et al.</i>	2014	Artigo em Conferência	0	0,177	3,18
FERNANDEZ-ORDONEZ e CAROLINA	2014	Artigo em Conferência	0	-	3,00
GIANNOPOULOS <i>et al.</i>	2014	Artigo em Conferência	0	-	3,00
RITSCHER <i>et al.</i>	2012	Artigo em Periódico	1	0,309	2,31
BAPTISTA <i>et al.</i>	2013	Artigo em Conferência	0	0,117	2,12
AKLI-ASTOUATI, GUEBAILI e MOKHTARI	2011	Artigo em Periódico	0	0,111	0,11

Fonte: Elaborado pelo autor.

A seguir, no primeiro passo da etapa 4, foram aplicados os filtros de seleção de textos, sendo o primeiro deles a leitura do título e do resumo de cada referência. A Tabela 2 mostra as referências selecionadas ao término deste passo.

Tabela 2 – Resultados da análise do título e do resumo de cada artigo.

Autores (AU)	Ano de Publicação (PY)	Tipo de Documento (DT)	Total de Citações (TC)	Fator de Impacto (IF/1000)	<i>InOrdinatio</i> ($\alpha = 1$)	Título ok?	Resumo ok?	Obter texto completo?
YUAN <i>et al.</i>	2013	Periódico	15	2,616	19,62	Sim	Sim	Sim
KALANTARI <i>et al.</i>	2014	Periódico	12	1,356	16,36	Sim	Não	Não
LI <i>et al.</i>	2020	Conferência	0	4,027	13,03	Sim	Sim	Sim
CHEATHAM <i>et al.</i>	2018	Periódico	6	-	13,00	Sim	Sim	Sim
YUE <i>et al.</i>	2012	Periódico	11	0,491	12,49	Sim	Sim	Sim
FUGAZZA <i>et al.</i>	2018	Periódico	3	0,655	10,66	Sim	Sim	Sim
FUGAZZA <i>et al.</i>	2016a	Periódico	5	0,655	10,66	Sim	Sim	Sim
FUGAZZA <i>et al.</i>	2021	Periódico	0	0,386	10,39	Sim	Sim	Sim

GRAUPNER e NÜST	2020	Conferência	0	0,537	9,54	Não	Não	Não
FUGAZZA <i>et al.</i>	2016b	Conferência	4	0,175	9,18	Sim	Sim	Sim
MCKEE, G.	2019	Periódico	0	0,771	8,77	Não	Não	Não
FUGAZZA <i>et al.</i>	2017	Periódico	2	0,655	8,66	Sim	Sim	Sim
SHVAIKO <i>et al.</i>	2012	Conferência	7	-	8,00	Sim	Sim	Sim
ZHANG <i>et al.</i>	2013	Conferência	2	-	4,00	Sim	Sim	Sim
MENEZES <i>et al.</i>	2014	Periódico	0	0,772	3,77	Não	Não	Não
RENTERIA-AGUALIMPIA <i>et al.</i>	2013	Conferência	1	0,200	3,20	Sim	Sim	Sim
DO <i>et al.</i>	2014	Conferência	0	0,177	3,18	Não	Não	Não
FERNANDEZ-ORDONEZ e CAROLINA	2014	Conferência	0	-	3,00	Sim	Sim	Sim
GIANNOPOULOS <i>et al.</i>	2014	Conferência	0	-	3,00	Não	Não	Não
RITSCHER <i>et al.</i>	2012	Periódico	1	0,309	2,31	Sim	Não	Não
BAPTISTA <i>et al.</i>	2013	Conferência	0	0,117	2,12	Não	Não	Não
AKLI-ASTOUATI, GUEBAILI e MOKHTARI	2011	Periódico	0	0,111	0,11	Sim	Sim	Sim

Fonte: Elaborado pelo autor.

Nesta tabela é possível verificar que, ao final deste primeiro passo, foram descartadas oito referências devido ao fato dos respectivos títulos e/ou resumos não estarem alinhados ao tema de pesquisa. As demais referências tiveram seus textos completos coletados para análise no próximo passo.

Por sua vez, no terceiro passo desta etapa, foi efetuada a leitura das seções de introdução e de conclusão de cada artigo selecionado no primeiro passo. A Tabela 3 apresenta os resultados da aplicação deste filtro de seleção.

Tabela 3 – Resultados da análise das seções de introdução e conclusão de cada artigo.

Autores (AU)	Ano de Publicação (PY)	Tipo de Documento (DT)	Total de Citações (TC)	Fator de Impacto (IF/1000)	InOrdinatio ($\alpha = 1$)	Introdução e conclusão ok?
YUAN <i>et al.</i>	2013	Artigo em Periódico	15	2,616	19,62	Sim
LI <i>et al.</i>	2020	Artigo em Conferência	0	4,027	13,03	Sim
CHEATHAM <i>et al.</i>	2018	Artigo em Periódico	6	-	13,00	Sim
YUE <i>et al.</i>	2012	Artigo em Periódico	11	0,491	12,49	Sim
FUGAZZA <i>et al.</i>	2018	Artigo em Periódico	3	0,655	10,66	Sim
FUGAZZA <i>et al.</i>	2016a	Artigo em Periódico	5	0,655	10,66	Sim
FUGAZZA <i>et al.</i>	2021	Artigo em Periódico	0	0,386	10,39	Sim
FUGAZZA <i>et al.</i>	2016b	Artigo em Conferência	4	0,175	9,18	Sim
FUGAZZA <i>et al.</i>	2017	Artigo em Periódico	2	0,655	8,66	Sim
SHVAIKO <i>et al.</i>	2012	Artigo em Conferência	7	-	8,00	Sim
ZHANG <i>et al.</i>	2013	Artigo em Conferência	2	-	4,00	Sim

RENTERIA-AGUALIMPIA <i>et al.</i>	2013	Artigo em Conferência	1	0,200	3,20	Sim
FERNANDEZ-ORDONEZ e CAROLINA	2014	Artigo em Conferência	0	-	3,00	Não
AKLI-ASTOUATI, GUEBAILI e MOKHTARI	2011	Artigo em Periódico	0	0,111	0,11	Sim

Fonte: Elaborado pelo autor.

Ao observar esta tabela, é possível notar que, como resultado da leitura das seções de introdução e de conclusão de cada artigo, foi descartada mais uma referência (FERNANDEZ-ORDONEZ e CAROLINA, 2014) por não estar alinhada ao tema de pesquisa.

Por fim, na Etapa 5, foi efetuada a leitura integral dos textos selecionados na etapa anterior. Para cada texto lido, foram respondidos os critérios de análise estabelecidos no primeiro passo desta etapa (conforme Quadro 6). A Tabela 4 apresenta os resultados da análise de cada texto completo, indicando os que foram selecionados, por atender ao menos três dos critérios de avaliação.

Tabela 4 – Resultados da leitura e análise dos textos completos.

Autores (AU)	Ano de Publicação (PY)	Critério 1?	Critério 2?	Critério 3?				Critério 4?	Critério 5?	Artigo Selecionado?
				questão a)	questão b)	questão c)	atende critério?			
YUAN <i>et al.</i>	2013	SIM	SIM	Sim	Não	Sim	SIM	NÃO	SIM	SIM
LI <i>et al.</i>	2020	NÃO	SIM	Não	Não	Não	NÃO	NÃO	SIM	NÃO
CHEATHAM <i>et al.</i>	2018	NÃO	SIM	Não	Não	Sim	SIM	NÃO	SIM	SIM
YUE <i>et al.</i>	2012	NÃO	SIM	Não	Não	Sim	SIM	NÃO	SIM	SIM
FUGAZZA <i>et al.</i>	2018	NÃO	SIM	Sim	Não	Sim	SIM	NÃO	SIM	SIM
FUGAZZA <i>et al.</i>	2016a	NÃO	SIM	Não	Não	Não	NÃO	NÃO	NÃO	NÃO
FUGAZZA <i>et al.</i>	2021	SIM	SIM	Sim	Não	Sim	SIM	NÃO	SIM	SIM
FUGAZZA <i>et al.</i>	2016b	NÃO	SIM	Sim	Não	Sim	SIM	NÃO	SIM	SIM
FUGAZZA <i>et al.</i>	2017	NÃO	SIM	Sim	Não	Sim	SIM	NÃO	SIM	SIM
SHVAIKO <i>et al.</i>	2012	SIM	SIM	Não	Não	Sim	SIM	NÃO	SIM	SIM
ZHANG <i>et al.</i>	2013	NÃO	SIM	Não	Não	Não	NÃO	NÃO	SIM	NÃO
RENTERIA-AGUALIMPIA <i>et al.</i>	2013	NÃO	NÃO	Não	Não	Não	NÃO	NÃO	SIM	NÃO
AKLI-ASTOUATI, GUEBAILI e MOKHTARI	2011	NÃO	SIM	Não	Não	Não	NÃO	NÃO	NÃO	NÃO

Fonte: Elaborado pelo autor.

A partir dos dados apresentados nesta tabela, é possível identificar que ao término desta etapa foram descartadas outras cinco referências, por não terem atendido a, pelo menos, três dos critérios estabelecidos para análise dos textos completos. Também é possível notar, quanto ao atendimento individual dos critérios e questões avaliados, que: a) somente três dos artigos

analisados atenderam ao critério 1 (propor processo, método ou técnica para a publicação de metadados (geo)espaciais na *Web*, tema da pesquisa); b) nenhum artigo atendeu ao quesito b (armazenamento de triplas RDF) do critério 3 (detalhar pelo menos um dos procedimentos (quesitos) para publicação de metadados (geo)espaciais na *Web*); e, c) nenhum artigo atendeu ao critério 4 (abordar a aplicação de boas práticas para publicação de dados na *Web*). Isso mostra que o tema de pesquisa escolhido ainda é pouco explorado na literatura, o que pode abrir várias frentes de estudo relacionadas ao tema nos próximos anos.

Por fim, como resultado desta revisão de literatura, é apresentada a Tabela 5, que relaciona os artigos escolhidos para compor o portfólio de trabalhos relacionados ao tema de pesquisa deste trabalho.

Tabela 5 – Artigos selecionados para compor o portfólio de trabalhos relacionados.

Autores (AU)	Ano de Publicação (PY)	Tipo de Documento (DT)	Total de Citações (TC)	Fator de Impacto (IF/1000)	<i>InOrdinatio</i> ($\alpha = 1$)
YUAN <i>et al.</i>	2013	Artigo em Periódico	15	2,616	19,62
CHEATHAM <i>et al.</i>	2018	Artigo em Periódico	6	-	13,00
YUE <i>et al.</i>	2012	Artigo em Periódico	11	0,491	12,49
FUGAZZA <i>et al.</i>	2018	Artigo em Periódico	3	0,655	10,66
FUGAZZA <i>et al.</i>	2021	Artigo em Periódico	0	0,386	10,39
FUGAZZA <i>et al.</i>	2016b	Artigo em Conferência	4	0,175	9,18
FUGAZZA <i>et al.</i>	2017	Artigo em Periódico	2	0,655	8,66
SHVAIKO <i>et al.</i>	2012	Artigo em Conferência	7	-	8,00

Fonte: Elaborado pelo autor.

Com relação à utilização da equação *InOrdinatio* (PAGANI; KOVALESKI; RESENDE, 2015) para ordenar (ranquear) as referências obtidas nas buscas efetuadas, ao visualizar a Tabela 5 é possível verificar que todos os artigos selecionados nesta revisão de literatura apresentaram valor calculado de *InOrdinatio* $\geq 8,00$, ou seja, todos os artigos escolhidos estavam na metade superior da listagem ordenada. Assim, caso fosse adotado este valor de *InOrdinatio* como ‘nota de corte’ (filtro) para descarte de artigos logo no início da Etapa 4, essa revisão de literatura atingiria o mesmo resultado, porém com uma quantidade menor de artigos a serem avaliados por meio da leitura do texto completo.

2.2 TRABALHOS RELACIONADOS

Uma vez concluída a revisão de literatura foi possível obter, como resultado, um portfólio de trabalhos relacionados, os quais serão comparados com a proposta do presente trabalho. Todos os trabalhos selecionados para compor o portfólio apresentam uma abordagem, técnica ou ferramenta para a disponibilização de metadados geoespaciais na *Web Semântica*. Nos parágrafos a seguir são apresentadas algumas considerações sobre cada um destes trabalhos.

Yue et al. (2012) apresentam uma abordagem para a geração automática de metadados semânticos da execução de serviços *Web* encadeados para a geração de produtos de geoprocessamento, de forma a possibilitar ao usuário obter informações sobre a proveniência dos dados e a consistência semântica dos serviços utilizados para gerar tal produto. A abordagem tem seu foco na especificação detalhada do mecanismo de geração e propagação dos metadados. Os metadados semânticos são gerados em RDF para que possam ser consultados por *queries* SPARQL antes da execução de novos serviços *Web* em que serão aplicados. No entanto, essa abordagem não tem como objetivo fazer a ligação dos metadados com outras fontes nem os publicar como conjuntos de dados RDF na *Web* de dados.

Seguindo essa mesma linha, Yuan et al. (2013) também apresentam uma abordagem para a publicação de metadados de proveniência, catalogados em um servidor de catálogo CSW, seguindo os princípios de *Linked Data*. Na abordagem proposta, todos os campos de metadados relacionados às informações de proveniência são coletados do catálogo CSW (em formato XML) e convertidos para triplas RDF com suporte de uma ontologia de proveniência de dados (*Provenir*). Em seguida estes campos de metadados RDF são enriquecidos por meio da inserção de URIs de fontes de dados relacionados a este domínio na nuvem LOD.

Por outro lado, Shvaiko et al. (2012) fazem um relato das experiências e lições aprendidas na publicação, de forma experimental, de alguns conjuntos de dados e metadados geoespaciais da Província Autônoma de Trento, na Itália. Para isto, propõem uma abordagem baseada no pipeline proposto por Maali, Cyganiak e Peristeras (2012), constituída das seguintes etapas: conversão e ligação, compartilhamento e, avaliação. Para a publicação dos metadados geoespaciais, que se encontram catalogados em um servidor de catálogo CSW, os registros em formato XML foram convertidos para RDF e, em seguida, foram inseridos manualmente URIs apontando para recursos do *Geonames* por meio de associações do tipo *owl:sameAs*. Os arquivos RDF enriquecidos foram, então, disponibilizados no servidor de geoportal da referida província para *download*.

Por sua vez, Cheatham et al. (2018) apresentam o *GeoLink*, um grafo de conhecimento que converte em RDF os metadados de diversos repositórios de dados de geociências dos Estados Unidos, com o objetivo de integrá-los e oferecer aos usuários uma fonte única para consulta, raciocínio e visualização de tais metadados. O conjunto de dados RDF que forma o grafo de conhecimento *GeoLink* é publicado na nuvem LOD seguindo os princípios de *Linked Data* sendo que, em meados de 2015, este grafo já continha cerca de 48 milhões de triplas. Para possibilitar tal integração de fontes diversas, o projeto elaborou uma ontologia de base, a partir da qual cada provedor de metadados a serem coletados personalizava sua própria ontologia e, a partir dela, realizada o despejo (*dump*) de seus conjuntos de metadados já em RDF. Para o enriquecimento semântico das triplas geradas, foi disponibilizado aos provedores um algoritmo para inserir URIs para os dados existentes no *GeoLink*. No entanto os autores relatam que essa ligação de dados automatizada pelo algoritmo não é precisa, sendo necessária intervenção manual em alguns casos.

Foram ainda selecionados quatro trabalhos de autoria de Fugazza et al. (2016b; 2017; 2018; 2021), cada qual apresentando um método ou ferramenta que, juntos, compõem uma abordagem para o enriquecimento semântico de metadados geoespaciais catalogados em IDEs.

Os autores chamam essa abordagem de ‘gestão descentralizada (ou delegação) de metadados’, uma vez que os referidos elementos de metadados podem receber como valor um URI para uma fonte de dados na *Web* que forneça maiores detalhes e permita a descoberta de novas informações a respeito, de forma a garantir que tais elementos estejam sempre atualizados e preenchidos de maneira padronizada e consistente. No entanto, a abordagem não tem o intuito de publicar as triplas RDF geradas, elas são mantidas em um servidor de aplicação e, quando um usuário consulta um registro de metadados pelo serviço CSW, as triplas são resgatadas e consultadas, sendo os valores obtidos (literais e URIs) então inseridos no arquivo XML a ser entregue ao usuário.

Neste contexto, Fugazza et al. (2016b) implementaram um fluxo de atividades para a criação de metadados geoespaciais semânticos, estruturados sobre vocabulários controlados e informações de contexto dentro do projeto italiano de pesquisa marinha (RITMARE), com o intuito de promover a integração destes metadados dentro da Infraestrutura de Dados Espaciais da União Europeia (plataforma INSPIRE).

As etapas do fluxo implementado consistem na utilização de uma ferramenta, desenvolvida pelo projeto, para a edição de metadados que produz descrições de acordo com perfis genéricos de metadados baseados em XML, a partir de uma linguagem de modelo, bem como permite a associação de URIs como valores de elementos destes metadados, gerando uma

versão em RDF do registro. Para isso, são associados os seguintes vocabulários: FOAF (para pontos de contato), SKOS (base de conhecimento do projeto RITMARE), GEONAMES (características geográficas) e DCAT (descrição de recursos geoespaciais). A seguir, os metadados em RDF são armazenados e, quando um usuário solicita um conjunto de metadados contendo campos na forma de URI, o aplicativo recupera as triplas URIs e as resolve (consultas SPARQL) para obter o valor atualizado para campo, o qual é atualizado no arquivo XML antes de ser entregue ao usuário.

Em Fugazza et al. (2017), os autores apresentam um método, implementado por meio de um aplicativo chamado *Liftboy*, e de uma plataforma de edição de metadados chamada EDI, para a normalização e enriquecimento semântico de campos de metadados geoespaciais. Tais ferramentas são aplicadas durante as fases de conversão de padrões e perfis e de edição, para composição do registro de metadados XML descritivos de um determinado recurso geoespacial, antes de ser catalogado em um servidor CSW.

O método emprega uma técnica chamada ‘aumento semântico’, na qual o mantenedor dos metadados, utilizando uma interface gráfica, especifica uma URI ou seleciona termos de um tesauro específico como valor para um determinado campo do registro de metadados que está sendo criado. As URIs especificadas são incluídas como anotações no documento XML que será catalogado no servidor CSW, permitindo que os valores dos campos anotados sejam atualizados quando forem acessados pelos usuários. Uma versão em RDF do registro de metadados é mantida em um *triple store* (armazenamento de triplas RDF) no servidor EDI para possibilitar as consultas semânticas. Com esse método pretende resolver problemas relacionados ao multilinguismo e à heterogeneidade semântica.

Em continuidade a essa abordagem, Fugazza et al. (2018) detalham a concepção e arquitetura da ferramenta EDI. Para isso, descrevem os modelos que expressam esquemas de metadados personalizados, os quais são utilizados para orientar a criação de interfaces de autoria baseadas na *Web* com o intuito de ajudar os usuários na edição de metadados geoespaciais. São ilustrados os três principais recursos oferecidos pela ferramenta, para o que chamam de ‘edição assistida de metadados’ a partir do enriquecimento semântico de itens de metadados ligando-os a fontes de dados externas: funcionalidades de autocompletar; compilação automática condicional de campos; e preenchimento dinâmico de listas suspensas.

Com o intuito de solucionar os problemas de inconsistência, interoperabilidade e heterogeneidade semântica em metadados geoespaciais, Fugazza et al. (2021) propõem uma técnica chamada "delegação de autoridade" onde alguns campos dos metadados (contato, palavras-chaves, extensão espacial e extensão temporal) podem receber, no lugar de valores

puramente textuais, URIs para outras fontes de dados externas tais como vocabulários, diretórios de autoridades, dicionários geográficos etc. Dessa forma seria possível implementar consultas multilíngues e expandidas de forma a solucionar problemas de heterogeneidade.

A Implementação foi feita em dois cenários: o primeiro em metadados formatados em XML com enriquecimento por meio de microdados, sendo o documento final de metadados gerado por um processador XSLT quando da solicitação pelo usuário; o segundo na conversão dos metadados para RDF com o enriquecimento por meio de URIs para fontes de dados externas, onde foi desenvolvido um software para efetuar as consultas SPARQL para obtenção dos valores atualizados dos campos de metadados enriquecidos quando o documento de metadados for requisitado pelo usuário. No entanto, essa resolução das URIs conforme demanda pode levar a problemas de desempenho dependendo da latência das fontes de dados externas utilizadas.

Por fim, foi elaborada a Tabela 6 que apresenta uma comparação entre os trabalhos relacionados, apresentados nesta seção, e a proposta do presente trabalho. Para isso, foram estabelecidos alguns critérios que permitirão identificar quais aspectos o presente trabalho propõe acrescentar ao estágio de maturidade do tema ‘publicação de metadados geoespaciais na Web Semântica’. Os critérios estabelecidos foram os seguintes:

- CRITÉRIO 1: O trabalho tem como principal finalidade a publicação de metadados geoespaciais semânticos na *Web*?
- CRITÉRIO 2: O trabalho tem como contexto de aplicação metadados geoespaciais catalogados, por meio do serviço CSW, em Infraestruturas de Dados Espaciais?
- CRITÉRIO 3: O trabalho descreve uma sequência de métodos, técnicas ou ferramentas para implementar as seguintes etapas visando a publicação de metadados semânticos na *Web*: (a) conversão de formatos; (b) enriquecimento semântico; (c) armazenamento e disponibilização?
- CRITÉRIO 4: O trabalho propõe o enriquecimento semântico dos metadados geoespaciais por meio do mapeamento de vocabulários e da ligação (via URIs) a outras fontes de dados na *Web*, seguindo as boas práticas de *Linked Data*?
- CRITÉRIO 5: O trabalho aplica as boas práticas para Dados Espaciais na *Web*, propostas pelo *World Wide Web Consortium* e *Open Geospatial Consortium* (2017)?
- CRITÉRIO 6: O trabalho detalha o uso de infraestrutura de *triplestore* para armazenamento e disponibilização na *Web* dos metadados semânticos?

Tabela 6 – Comparação entre trabalhos relacionados.

TRABALHO	CRITÉRIO 1	CRITÉRIO 2	CRITÉRIO 3	CRITÉRIO 4	CRITÉRIO 5	CRITÉRIO 6
YUE <i>et al.</i> , 2012	NÃO	SIM	Somente (a) e (c)	NÃO	NÃO	NÃO
SHVAIKO <i>et al.</i> , 2012	SIM	SIM	SIM	SIM	NÃO	NÃO
YUAN <i>et al.</i> , 2013	SIM	SIM	SIM	SIM	NÃO	Apenas menciona
FUGAZZA <i>et al.</i> , 2016b	NÃO	SIM	Somente (a) e (b)	SIM	NÃO	NÃO
FUGAZZA <i>et al.</i> , 2017	NÃO	SIM	Somente (a) e (b)	SIM	NÃO	NÃO
CHEATHAM <i>et al.</i> , 2018	SIM	NÃO	Somente (b) e (c)	SIM	NÃO	NÃO
FUGAZZA <i>et al.</i> , 2018	NÃO	SIM	Somente (a) e (b)	SIM	NÃO	Apenas menciona
FUGAZZA <i>et al.</i> , 2021	NÃO	SIM	Somente (a) e (b)	SIM	NÃO	NÃO
ESTE TRABALHO	SIM	SIM	SIM	SIM	SIM	SIM

Fonte: Elaborado pelo autor.

A partir da análise da tabela anterior é possível verificar que este trabalho atende a todos os critérios estabelecidos para o comparativo realizado. Assim, como principais diferenciais deste trabalho, em relação aos demais trabalhos relacionados, convém destacar, na etapa de armazenamento e disponibilização, a especificação de uma plataforma de armazenamento de triplas RDF (*triplestore*), o que foi tratado de forma superficial em somente dois dos trabalhos relacionados. Destaca-se também neste trabalho a aplicação das ‘boas práticas para dados espaciais na Web’, recomendadas pelo World Wide Web Consortium e Open Geospatial Consortium (2017), em cada uma das etapas do *framework* proposto, o que não foi abordado por nenhum dos trabalhos relacionados que foram publicados após o ano de 2017.

Convém ainda destacar que somente dois dos trabalhos relacionados (SCHVAIKO *et al.*, 2012; YUAN *et al.*, 2013) também propõe a publicação de metadados geoespaciais como forma de integrar as IDEs com a Web Semântica, porém diferencia-se deste trabalho por não apresentar os diferenciais mencionados no parágrafo anterior.

Como conclusão, apesar do reduzido número de trabalhos encontrados na literatura, o que indica que o tema ainda é pouco explorado, foi possível determinar o atual estágio de evolução do tema, bem como o posicionamento deste trabalho quando comparado aos trabalhos relacionados selecionados para o portfólio.

A seguir, na próxima Seção, são apresentados os conceitos de metadados e de dados e metadados geoespaciais, essenciais para o entendimento do contexto de aplicação deste trabalho.

3 DADOS E METADADOS GEOESPACIAIS

Nesta Seção são apresentados os conceitos de metadados e de dados e metadados geoespaciais, seus formatos e perfis, bem como os padrões e vocabulários utilizados para sua construção. Além disso, são apresentados os componentes e arquitetura das Infraestruturas de Dados Espaciais (IDEs), com foco no serviço de catálogo de metadados, contexto de aplicação deste trabalho. Com isso, pretende-se cumprir o objetivo específico I proposto na Seção de Introdução.

3.1 METADADOS

Diante do contínuo desenvolvimento e aprimoramento das Tecnologias da Informação e Comunicação (TICs), bem como a evolução da *Web* notada nas últimas décadas, a produção e disseminação de recursos informacionais digitais vem crescendo de forma expressiva a cada ano, consolidando o fenômeno que atualmente chamamos de “sociedade da informação”.

Diante disso, a imensa quantidade de recursos informacionais sendo disponibilizados a cada dia na *Web* traz à tona a necessidade de melhorar seus processos de busca e recuperação pelos usuários que, antes de acessarem um recurso, como por exemplo um *dataset* contendo alguns milhares de registros, precisam constatar sua proveniência e confiabilidade, bem como verificar se tal *dataset* realmente atende suas necessidades informacionais.

É nesse contexto que surgem os metadados, como elementos descritores de recursos informacionais, quer sejam físicos ou digitais, com o propósito de identificar e descrever detalhadamente suas características, de forma que possam ser mais fácil e rapidamente encontrados e recuperados pelos usuários, sejam eles humanos ou máquinas.

Para Feitosa (2006, p. 50), o termo ‘metadados’ tem origem no “[...] prefixo grego *metá*, que significa mudança, posteridade, além de transcendência e reflexão crítica sobre alguma coisa.”, que junto com a palavra ‘dados’, leva a “[...] uma noção de transcendência sobre a informação, ou de uma possível descrição sobre a informação que os dados representam.”.

Assim, a descrição eficiente dos recursos informacionais e a representação adequada dos conjuntos de dados podem contribuir para a obtenção de resultados relevantes no processo de recuperação da informação. Neste sentido, ainda de acordo os mesmos autores, a Ciência da Informação, no que tange os metadados, “[...] tem um importante papel na medida em que se torna fundamental proporcionar a representação adequada do conteúdo informacional para

garantir resultados relevantes no processo de recuperação dos conjuntos de dados.” (MOREIRA et al., 2017, p. 162).

Os metadados estão cada vez mais presentes em nosso cotidiano, tornando-se indispensáveis para a identificação e descrição não somente de dados armazenados na *Web*, mas também fora dela, como em arquivos, sistemas e bases de dados fechadas (não publicadas na *Web*). Constituem ainda elementos indispensáveis para a consolidação de uma *Web* de dados, no lugar da atual *Web* de documentos, como previsto por Berners-Lee (2006, não paginado, tradução nossa), “[...] ao possibilitarem que conjuntos de dados possam ser mais facilmente encontrados, acessados, interoperáveis e reutilizáveis pelos usuários.”.

Para Riley (2017, p. 1, tradução nossa), metadados são “[...] a informação que nós criamos, armazenamos e compartilhamos para descrever coisas, permitindo-nos interagir com essas coisas para obter o conhecimento que necessitamos.”. E vai além dessa definição abrangente, ao afirmar que os metadados “são a chave para a funcionalidade de sistemas que detém conteúdo, habilitando os usuários a encontrarem itens de interesse, gravar informações essenciais sobre eles, e compartilhar a informação com outros.”.

De acordo com Gilliland (2016), até a década de 1990, o termo metadados era empregado principalmente pelas comunidades de dados geoespaciais e de sistemas de informação, no contexto das quais se referiam a um conjunto de padrões setoriais ou disciplinares utilizados para estabelecer os elementos descritores necessários para a identificação, representação, interoperabilidade, gerenciamento e uso dos dados contidos em um sistema de informação. No entanto é possível constatar que tal entendimento ainda se mantém atual, quando Riley (2017, p. 2, tradução nossa) afirma que “as principais funcionalidades da maioria dos pacotes de software que usamos todos os dias são orientadas por metadados.”.

Neste contexto, os metadados podem ser definidos, de acordo com Alves (2010), como:

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda 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.

Além de identificar, representar e descrever um recurso informacional, os metadados também documentam como esse recurso se comporta, qual sua função e uso, quais relacionamentos possui com outros recursos, e como é gerenciado ao longo do tempo (GILLILAND, 2016).

Na perspectiva da Ciência da Informação, diversos autores que têm discutido definições para o termo ‘metadados’ acabam por convergir para um ponto em comum, no qual o conceito de metadados “[...] está atribuído a uma informação estruturada para as ações de identificação, descoberta, seleção, uso, acesso e gerenciamento” (ARAKAKI; ARAKAKI, 2019, p. 4).

Ademais, os metadados fornecem informações adicionais que ajudam os consumidores de dados a entenderem melhor o significado dos dados, sua estrutura e a esclarecer outras questões, como direitos e termos de licença, a organização que gerou os dados, qualidade dos dados, métodos de acesso aos dados e o cronograma de atualização dos conjuntos de dados (W3C, 2017). Tudo isso possibilita ao usuário discernir se aquele conjunto de dados é confiável, útil e de que forma poderá ser reutilizado. Na Figura 5 são apresentados dois exemplos de conjuntos de metadados para a descrição de diferentes recursos de informação.

Figura 5 – Exemplos de conjuntos de metadados.

Metadados de um Documento Bibliográfico		Metadados de um Arquivo de Imagem	
Tipo de Referência	Tese ou Dissertação (Thesis)	Propriedade	Valor
Site	mtc-m2ld.sid.inpe.br	Imagem	
Código do Detentor	isadg [BR SPINFE] ibi 8JMKD3MGPCW/3DT298S	Identificação de Imagem	
Identificador	8JMKD3MGPCW347/447SK2L	Dimensões	2592 x 1944
Repositório	sid.inpe.br/mtc-m2ld/2021/06.24.14.27	Largura	2592 pixels
Última Atualização	2021:09.16.14.57.57 (UTC) simone	Altura	1944 pixels
Metadados	sid.inpe.br/mtc-m2ld/2021/06.24.14.27.36	Resolução horizontal	300 dpi
Atualização dos Metadados	2021:10.04.19.58.28 (UTC) simone	Resolução vertical	300 dpi
Chave Secundária	INPE-18480-TDI/3128	Intensidade de bits	24
Chave de Citação	Freitas:2021:ExInFl	Compactação	
Título	Exposição a incêndios florestais em unidades de conservação c	Unidade de resolução	2
Título Alternativo	Exposure to forest fires in sustainable use conservation unit	Representação de cores	sRGB
Curso	SER-SRE-DIPGR-INPE-MCTI-GOV-BR	Bits/pixel compactados	
Ano	2021	Câmera	
Data	2021-05-28	Fabricante da câmera	Nokia
Data de Acesso	13 out. 2021	Modelo da câmera	N85
Tipo da Tese	Dissertação (Mestrado em Sensoriamento Remoto)	Escala de número f	f/2.8
Tipo Secundário	TDI	Tempo de exposição	1/1000 s
Número de Páginas	134	Velocidade ISO	ISO-100
		Ajuste de exposição	

Fonte: Elaborado pelo Autor.

Para Gilliland (2016), os metadados de um recurso informacional devem ser capazes de: certificar a autenticidade e o grau de integridade de seu conteúdo; estabelecer e documentar o contexto desse conteúdo; identificar e explorar as relações existentes dentro e entre objetos de informação; relacionar uma gama de pontos de acesso intelectual; e, apresentar algumas das informações que um profissional da informação tenha fornecido como referência.

Quanto aos diferentes tipos, ou categorias, de metadados existentes para a descrição de recursos informacionais, estes são usualmente divididos nas categorias apresentadas no Quadro 7 a seguir.

Quadro 7 – Categorias de metadados.

CATEGORIA DE METADADOS	FINALIDADE	EXEMPLOS DE CAMPOS
Direitos (Rights)	Informar os direitos de propriedade intelectual e de uso do recurso.	– <i>Copyright status</i> – <i>License terms</i> – <i>Rights holder</i>
Técnicos (Technical)	Fornecer detalhes técnicos do recurso, tais como formato de arquivo e instruções de acesso e uso.	– <i>File type</i> – <i>File size</i> – <i>Creation date/time</i>
Preservação (Preservation)	Registro da série de intervenções em um recurso visando garantir sua longevidade e integridade.	– <i>Checksum</i> – <i>Preservation event</i>
Descritivos (Descriptive)	Descrever o conteúdo e identificar as características do recurso.	– <i>Title</i> – <i>Author</i> – <i>Subject</i> – <i>Publication date</i>
Estruturais (Structural)	Descrever como as peças de um único objeto se encaixam e como um objeto se relaciona com outros.	– <i>Sequence</i> – <i>Place in hierarchy</i>

Fonte: Adaptado de Jaffe (2017).

Disso resulta que os metadados são estruturas complexas que podem demandar um alto custo e esforço para serem criadas e mantidas (GILLILAND, 2016). Por isso, a escolha de esquemas de estruturação de metadados para padronizar a descrição e representação de um recurso informacional não deve ser feita de forma aleatória, sendo que o ideal é que seja escolhido um padrão de metadados já estabelecido e mantido pela comunidade no qual aquele recurso de informação se insere, como será visto a seguir.

3.1.1 Padrões e vocabulários para a estruturação de metadados

Os padrões de metadados tem como principal finalidade assegurar que diferentes tipos de metadados, utilizados por diferentes comunidades para a descrição de diferentes tipos de recursos de informação, sejam capazes de interoperar entre si.

Neste sentido, a utilização de padrões de metadados, de acordo com Moreira et al. (2017, p. 167), “[...] pode ajudar os sistemas de busca na *Web* a responder consultas que exigem a

extração e combinação de dados de múltiplas fontes, permitindo o acesso integrado a estes serviços”. Já para Formenton et al. (2017, p. 83), a adoção de padrões de metadados é considerada uma estratégia para “[...] apoiar a gestão e a preservação de objetos digitais em ambientes informacionais, tais como os repositórios institucionais.”.

Um dos componentes básicos de um conjunto de elementos de metadados é o conteúdo, que trata das declarações ou instruções de quais valores devem ser atribuídos a cada elemento. Neste caso, os padrões de metadados servem para prover regras para os conteúdos a serem inseridos, bem como para representação deste conteúdo e quais valores são permitidos (por exemplo, a partir de um vocabulário controlado específico). Além disso, um padrão de metadados pode ainda determinar o formato de codificação (XML ou RDF, por exemplo) deste conjunto de elementos (CHAN; ZENG, 2006).

Atualmente existe uma grande variedade de padrões de metadados altamente detalhados, que foram desenvolvidos por comunidades individuais, como as de bibliotecas, museus, arquivos, geoinformação, governo eletrônico, entre outras. Por outro lado, existem também os padrões de metadados extremamente simples e genéricos para serem usados por qualquer comunidade, especialista ou não, para descrever uma ampla variedade de recursos e promover a interoperabilidade entre eles (GILLILAND, 2016). Um exemplo de padrão de metadados altamente detalhado é o estabelecido pela norma ISO (*International Organization for Standardization*) 19115 para dados geoespaciais. Já o *Dublin Core Metadata Element Set* (DCMES) é um exemplo de padrão de metadados enxuto e genérico. Para ilustrar, o Quadro 8 mostra um exemplo de metadados, em formato XML, estruturados no padrão *Dublin Core*.

Quadro 8 – Exemplo de metadados, em formato XML, estruturados no padrão Dublin Core.

```
<?xml version="1.0" encoding="UTF-8"?>
<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/http://www.openarchives.org/OAI/2.0/oai_dc/oai_dc.xsd">
  <dc:title>Análise do desflorestamento no distrito florestal sustentável da BR-163, PA</dc:title>
  <dc:creator>OBT/INPE</dc:creator>
  <dc:subject>Distrito Florestal</dc:subject>
  <dc:subject>desmatamento</dc:subject>
  <dc:subject>degradação florestal</dc:subject>
  <dc:subject>corte seletivo</dc:subject>
  <dc:subject>Unidade de Conservação</dc:subject>
  <dc:subject>Terra Indígena</dc:subject>
  <dc:description>Esse trabalho teve como objetivo analisar relações entre as diferentes formas de desflorestamento no DFS da BR-163 para o ano de 2008 por corte raso (PRODES), extração seletiva de madeira (DETEX) ou por degradação
```

```

florestal (DEGRAD), e as distintas categorias de Unidades de Conservação (UC),
Terras Indígenas (TI) e Projetos de Assentamento do INCRA. </dc:description>
<dc:date>2008</dc:date>
<dc:type>text</dc:type>
<dc:identifier>http://urlib.net/sid.inpe.br/mtc-m19@80/2010/02.09.19.52
</dc:identifier>
<dc:identifier>http://urlib.net/8JMKD3MGP7W/36TNMBS</dc:identifier>
<dc:language>pt</dc:language>
<dc:rights>Copyright information available at source archive</dc:rights>
</oai_dc:dc>

```

Fonte: Elaborado pelo autor.

Além da utilização dos padrões de metadados para padronizar a sintaxe e formatação dos elementos, os vocabulários controlados podem ser usados para padronizar e controlar os valores a serem atribuídos aos elementos de um conjunto de metadados. Um vocabulário controlado pode ser definido como uma lista de termos predefinidos para um determinado assunto ou tema, na qual cada conceito é identificado por uma palavra ou frase, podendo também apresentar mapeamentos para outros termos e definir relacionamentos entre eles. Normalmente são concebidos em um único idioma, porém algumas comunidades trabalham na conexão entre vocabulários de diferentes idiomas ou na construção de novos vocabulários que juntam termos de diferentes idiomas (RILEY, 2017).

Assim, são exemplos de vocabulários controlados os tesouros, as listas de cabeçalho de assunto e as listas de autoridades, dentre os quais podem ser mencionados: *Library of Congress Subject Headings* (LCSH); *The Getty Institute Thesaurus of Geographical Names* (TGN); *Library of Congress Name Authority File* (LCNAF); *Virtual International Authority File* (VIAF); e, como exemplo brasileiro, o Vocabulário Controlado de Governo Eletrônico (VCGE). Diante disso, o Quadro 9 apresenta a relação dos 15 elementos centrais (*core elements*) que compõem o padrão DCMES, acompanhados de possíveis vocabulários controlados que podem ser empregados para padronizar seus respectivos valores.

Quadro 9 – Elementos do padrão de metadados *Dublin Core Metadata Element Set*.

ELEMENTO	DESCRIÇÃO/USO	POSSÍVEL VOCABULÁRIO PARA O VALOR
<i>Title</i>	Nome dado ao recurso.	
<i>Subject</i>	O tema/assunto a que se refere o recurso.	<i>Library of Congress Subject Headings (LCSH)</i>
<i>Description</i>	Descrição do recurso.	
<i>Creator</i>	Entidade responsável pela criação/geração do recurso.	<i>Virtual International Authority File (VIAF)</i>
<i>Publisher</i>	Entidade responsável por tornar o recurso disponível.	

Contributor	Entidade responsável por contribuir para revisão/adaptação/modificação do recurso.	<i>Library of Congress Name Authority File (LCNAF)</i>
Date	Um instante ou período de tempo associado a um evento no ciclo de vida do recurso.	<i>W3CDTF</i>
Type	Natureza ou gênero do recurso.	<i>DCMI Type Vocabulary</i>
Format	Formato de arquivo, meio físico, ou dimensões do recurso.	<i>Internet Media Types (MIME)</i>
Identifier	Identificador único para referenciar o recurso em um dado contexto.	
Source	Um recurso relacionado, a partir do qual o recurso descrito é derivado.	
Language	Linguagem do recurso.	<i>ISO 639</i>
Relation	Um recurso relacionado.	
Coverage	O tópico espacial ou temporal do recurso, a aplicabilidade espacial do recurso, ou a jurisdição sob a qual o recurso é relevante.	<i>Thesaurus of Geographic Names (TGN)</i>
Rights	Informação sobre direitos de propriedade intelectual e de uso do recurso.	

Fonte: Adaptado de Jaffe (2017).

Mesmo com o passar do tempo, os padrões de metadados e suas estruturas estão em um estado de contínua acomodação de novos modelos de compartilhamento de informações (WOODLEY, 2016). Com o surgimento da *Web Semântica* e a constante evolução de suas tecnologias e ferramentas, como o *Resource Description Framework* (RDF), além das boas práticas dos Dados Ligados (*Linked Data*), abre-se caminho para promover a automação e interoperabilidade semântica dos metadados, que poderão ser consumidos não somente por humanos, mas também por máquinas e agentes de software na *Web*. Além disso, poderão ser enriquecidos por meio da conexão com outros conjuntos de dados na chamada ‘nuvem de dados abertos ligados’ (*LOD Cloud*, em inglês), como será visto na próxima Seção deste trabalho.

3.2 DADOS GEOESPACIAIS

Os dados geoespaciais (também chamados de dados geográficos ou georreferenciados) são caracterizados por registrarem uma localização espacial (ou geográfica) onde um determinado fenômeno ocorre, assim como os atributos que descrevem esse fenômeno (VINHAS, 2020). A localização geográfica neste tipo de dado tem como sistema de referência a superfície da Terra, sendo geralmente empregados sistemas de coordenadas para descrever seu posicionamento (W3C; OGC, 2017).

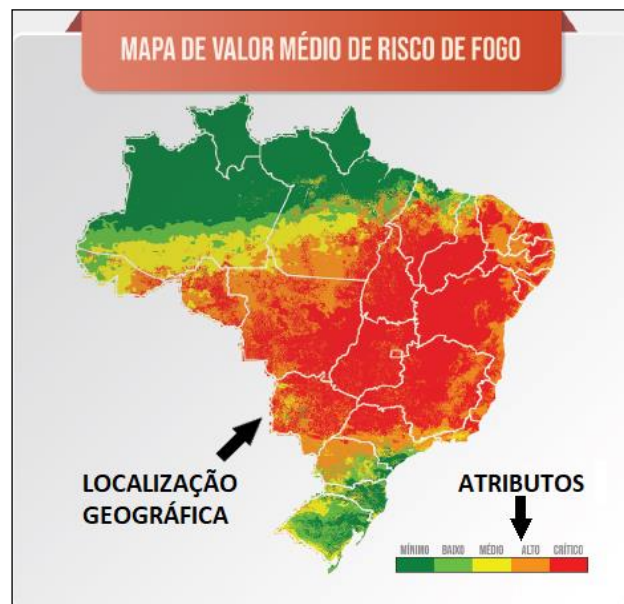
Em Brasil (2008, não paginado), o termo ‘dado geoespacial’ é definido como:

[...] aquele que se distingue por sua componente espacial, que associa a cada entidade ou fenômeno uma localização na Terra, traduzida por sistema geodésico de referência, podendo ser derivada das tecnologias de levantamento associadas a sistemas globais de posicionamento apoiados por satélites, bem como de mapeamento ou de sensoriamento remoto.

Por sua vez, Lisboa Filho e Iochpe (1996) definem este tipo de dado como “[...] objetos com atributos contendo informações sobre sua localização geográfica em relação a um sistema de coordenadas.”. Como exemplos de fenômeno geográfico, podem ser mencionados aqueles relacionados à agricultura, climatologia, uso da terra, vegetação, economia, educação, política, entre outros (ROSA, 2013).

Para que possam ser visualizados e manipulados em um sistema de informação geográfica (SIG), os dados geoespaciais são representados por pontos, linhas e polígonos, aos quais são associados atributos, ou seja, características das feições a serem representadas. Para exemplificar, um ponto pode representar locais em processo de desertificação, uma linha pode representar estradas ou rios, bem como um polígono pode representar tipos de vegetação por área, cobertura de uso da terra, entre outras feições (ROSA, 2013). Na Figura 6 é apresentada uma imagem que permite identificar as características descritas até o momento sobre os dados geoespaciais.

Figura 6 – Exemplo de dados geoespaciais.



Fonte: Brasil (2021).

Além das definições apresentadas anteriormente, uma conceituação de dados geoespaciais mais alinhada à Ciência da Informação é encontrada em Rosa (2013, p. 76):

[...] os dados geográficos são armazenados por meio de entidades gráficas, que representam os elementos do mundo real que se deseja analisar e os dados descritivos, que nada mais são do que tabelas que contém informações alfanuméricas que descrevem as características das entidades gráficas. Essa representação que se constrói do mundo real é fruto de um processo de interpretação dos elementos que compõem o mundo real e que nos interessam para uma determinada finalidade. Nessa representação utilizamos nossos processos cognitivos que envolvem habilidades de seleção, generalização, simulação e síntese, para expressar nossa percepção do mundo real no computador).

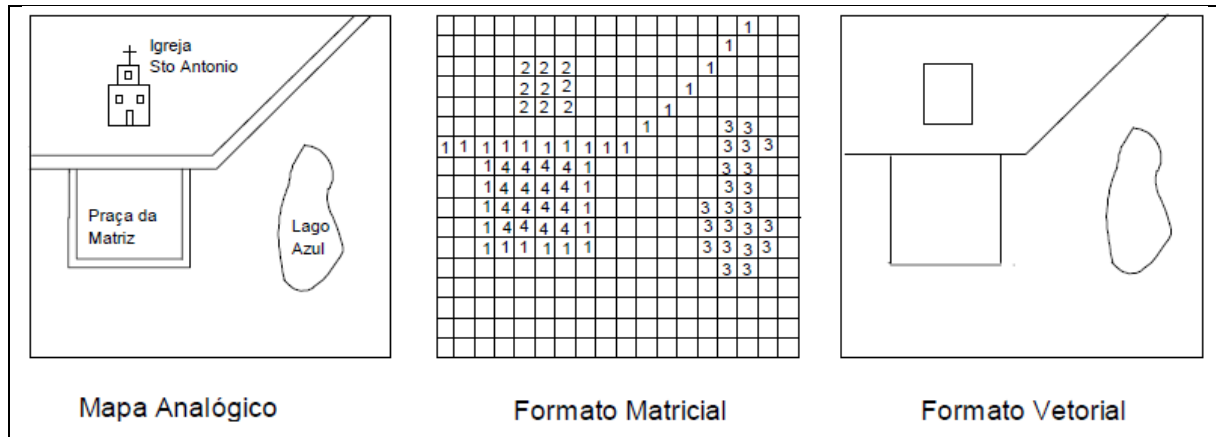
As características mais relevantes dos dados geoespaciais são o grande volume de dados gerados em um curto espaço de tempo e a grande variação dos atributos ao longo de uma região ou ao longo do tempo (variação espaço-temporal), o que traz maiores dificuldades para seu armazenamento e processamento. Daí surgem os formatos de representação de dados geoespaciais, como será visto a seguir.

3.2.1 Formatos de representação de dados geoespaciais

Devido à enorme complexidade da variação geográfica do mundo real, os dados geoespaciais precisam ser reduzidos a uma quantidade finita e gerenciável para que possam ser armazenados e processados de forma digital. Isso se dá por meio da utilização de formatos de representação de dados geoespaciais, os quais fornecem um conjunto de regras para que as variações geográficas do mundo real sejam adequadamente representadas na forma digital. Os dois formatos mais comumente utilizados para essa finalidade são o matricial (ou *raster*) e o vetorial (LISBOA FILHO; IOCHPE, 1996). A Figura 7 apresenta um exemplo comparativo entre estes dois formatos de representação de dados geoespaciais.

No formato matricial, a representação de cada entidade associada a um dado geoespacial é feita célula a célula sobre uma malha quadriculada regular, onde cada célula é individualmente integrada ao plano por meio de suas coordenadas (x,y). Assim, ao considerar uma matriz $p(x,y)$, composta de x linhas e y colunas, haverá para cada célula um número de linha, um número de coluna e um valor correspondente ao atributo sendo representado (espaços vazios são preenchidos com zeros). Dessa forma, é possível representar um ponto por uma única célula, uma linha por um conjunto de células vizinhas numa determinada direção, ou uma área por um aglomerado de células. Uma vez que este formato de representação usa um plano bidimensional, somente um atributo (tipo de solo, hidrologia, relevo, etc.) pode ser representado por plano, de forma que para a representação de vários atributos simultaneamente um conjunto de planos superpostos (camadas) deve ser utilizado (ROSA, 2013).

Figura 7 – Exemplo dos formatos de representação matricial³ e vetorial.



Fonte: Lisboa Filho e Iochpe (1996).

Já no formato vetorial, a representação das entidades associadas a um dado geoespacial se dá por meio de pontos (com coordenadas x,y), linhas e polígonos, sendo estes dois últimos formados por séries de pontos. Neste formato assume-se que as coordenadas dos pontos são matematicamente exatas, com a finalidade de tentar reproduzir cada entidade da forma mais precisa possível (ROSA,2013). Assim, os polígonos (cadeias de linhas) servem para representar os limites de entidades do tipo área, enquanto as linhas, além de formarem polígonos, também representam entidades do tipo rios, estradas, redes elétricas, entre outras (LISBOA FILHO; IOCHPE, 1996).

No Quadro 10 são apresentadas as vantagens e desvantagens da utilização de ambos os formatos de representação de dados geoespaciais.

Quadro 10 – Comparação entre os formatos de representação matricial e vetorial.

	FORMATO MATRICIAL	FORMATO VETORIAL
VANTAGENS	• Utiliza estrutura mais simples. • Representação eficiente de áreas com alta variação espacial. • Tratamento adequado de imagens produzidas por satélites.	• Utiliza uma estrutura de dados mais compacta que o formato matricial. • Permite a realização de cálculos de medidas espaciais mais precisos. • Gera saídas gráficas de melhor qualidade.
DESVANTAGENS	• Estruturas de dados são menos compactas. Armazenam uma maior quantidade de dados, independente da utilização de métodos de compactação. • Saída gráfica com menor resolução.	• As estruturas de dados são mais complexas que no formato matricial. • A representação de áreas com alta variação espacial é ineficiente. • É inadequado para manipulação de imagens geradas por satélites.

Fonte: Adaptado de Lisboa Filho e Iochpe (1996).

³ Para simplificar a figura, no formato matricial os autores deixaram em branco as células com valor zero.

Uma vez conhecidas as definições, características e formatos de representação dos dados geoespaciais, a seção a seguir apresenta como os conjuntos de dados (*datasets*) geoespaciais são descritos por meio dos metadados geoespaciais, com o intuito de facilitar sua localização e recuperação junto aos respectivos provedores e serviços de distribuição.

3.3 METADADOS GEOESPACIAIS

Os metadados geoespaciais podem ser considerados uma especialização de metadados utilizados para a descrição de recursos de informação contendo dados geoespaciais. Dada as características específicas deste tipo de dado, apresentadas na seção anterior, os conjuntos de dados geoespaciais dependem quase que inteiramente de metadados para sua descoberta e recuperação, uma vez que este tipo de dado se encontra tipicamente em formato não textual.

Para Goodchild et al. (1997 apud BRASIL, 2010, p. 22), metadados geoespaciais são definidos como uma “descrição de alto nível, disponibilizando informações sobre referenciamento espacial, qualidade, proveniência, periodicidade, acesso e distribuição dos dados geoespaciais”.

No Brasil, o art. 2º, inciso II, do Decreto nº 6.666, de 27/11/2008 (BRASIL, 2008) define metadados geoespaciais como sendo um “conjunto de informações descritivas sobre os dados, incluindo as características do seu levantamento, produção, qualidade e estrutura de armazenamento, essenciais para promover a sua documentação, integração e disponibilização, bem como possibilitar a sua busca e exploração.”.

Nos EUA, o instrumento legal mais claro em relação aos metadados geoespaciais foi a Circular nº A-16 (*US Office of Management and Budget*, 2002), emitida pelo Gabinete Federal de Gestão e Orçamento. Nessa Circular, os metadados geoespaciais são definidos como:

[...] informação sobre dados e/ou serviços geoespaciais, tais como conteúdo, fonte, coleta, escala espacial, precisão, projeção, responsável, telefone de contato, método de coleção, e outras descrições. Os metadados são fundamentais para documentar, preservar e proteger os recursos de dados espaciais das agências. Metadados confiáveis, estruturados de forma padronizada, são essenciais para assegurar que os dados geoespaciais sejam utilizados adequadamente, e que qualquer análise resultante seja crível. Os metadados também podem ser utilizados para facilitar a pesquisa e o acesso aos conjuntos de dados ou serviços geoespaciais dentro de uma plataforma ou biblioteca de dados.

A expansão dos sistemas de informação geográfica (SIGs), das bases de dados geoespaciais, e da geração de imagens da Terra, associada ao desejo de se trocar conjuntos de dados entre usuários e organizações, é que motivou o surgimento dos metadados geoespaciais padronizados. Dessa forma, os metadados para um recurso geoespacial devem permitir a um

potencial usuário determinar se são adequados para serem utilizados em uma aplicação específica (COWEN, 2017).

Os metadados geoespaciais são uma preocupação de longa data na área da geoinformação. De acordo com Fernández-Ordóñez et al. (2014, p. 118, tradução nossa), os metadados geoespaciais “[...] emergem como tema de pesquisa e desenvolvimento diante da crescente produção mundial de dados geoespaciais e dos desafios para sua disponibilidade através da *Web*.”.

Dentre as principais funções dos metadados, na área da geoinformação, estão a descrição detalhada e a transmissão de informações fundamentais sobre conjuntos de dados geoespaciais disponibilizados pelos provedores. Assim, os metadados geoespaciais possibilitam a localização e recuperação dos conjuntos de dados geoespaciais pelos usuários, além de auxiliá-los na escolha daqueles *datasets* que realmente atendam suas necessidades informacionais, evitando assim o *download* desnecessário de grandes quantidades de dados.

Para que isso seja possível, os metadados geoespaciais devem especificar quais serviços ou softwares podem consumir os dados geoespaciais e onde eles são oferecidos para permitir a automatização dos processos. Devem ainda fornecer informações a respeito da proveniência e qualidade, ou seja, descrever a forma como os dados geoespaciais foram gerados, bem como informar quais os erros potenciais e a incerteza ou variabilidade nos cálculos e medições afetos (POUCHARD et al., 2013).

No entanto, para garantir que todas essas informações sejam transmitidas pelos metadados de forma completa e acurada, se faz necessário o envolvimento de todo o corpo técnico responsável pela produção dos dados geoespaciais na atividade de geração e catalogação dos metadados. Conforme Brasil (2010, p. 88),

[...] o ideal é que a confecção dos metadados seja feita pelo responsável pela elaboração do respectivo produto que está sendo documentado e em paralelo ao processo de geração dos dados, ou seja, na linha de produção dos mesmos. Na impossibilidade de isso acontecer, sugere-se minimamente uma supervisão dos metadados criados, feita pelo produtor dos dados geoespaciais.

Além disso, um instrumento importante para identificar os recursos necessários e os diversos papéis na atividade de confecção de metadados geoespaciais é a implantação de uma ‘política de metadados’ na organização, a qual deve contar com o apoio da alta direção (BRASIL, 2010).

Assim, para os provedores de dados geoespaciais, a tarefa de geração e catalogação de metadados deve ter o mais alto nível de prioridade, uma vez que os dados que possuem metadados tornam-se mais informativos e de melhor qualidade (DIBS, 2013). Além disso,

quando metadados são criados e catalogados, contribui-se para o acréscimo de semântica ao dado geoespacial descrito, uma vez por meio dos metadados fica possível identificar o significado real do dado descrito. Para ajudar nessa tarefa, padrões e perfis (nacionais e internacionais) para metadados geoespaciais têm sido produzidas e continuam a evoluir ao longo do tempo, como será visto a seguir.

3.3.1 Padrões e perfis para metadados geoespaciais

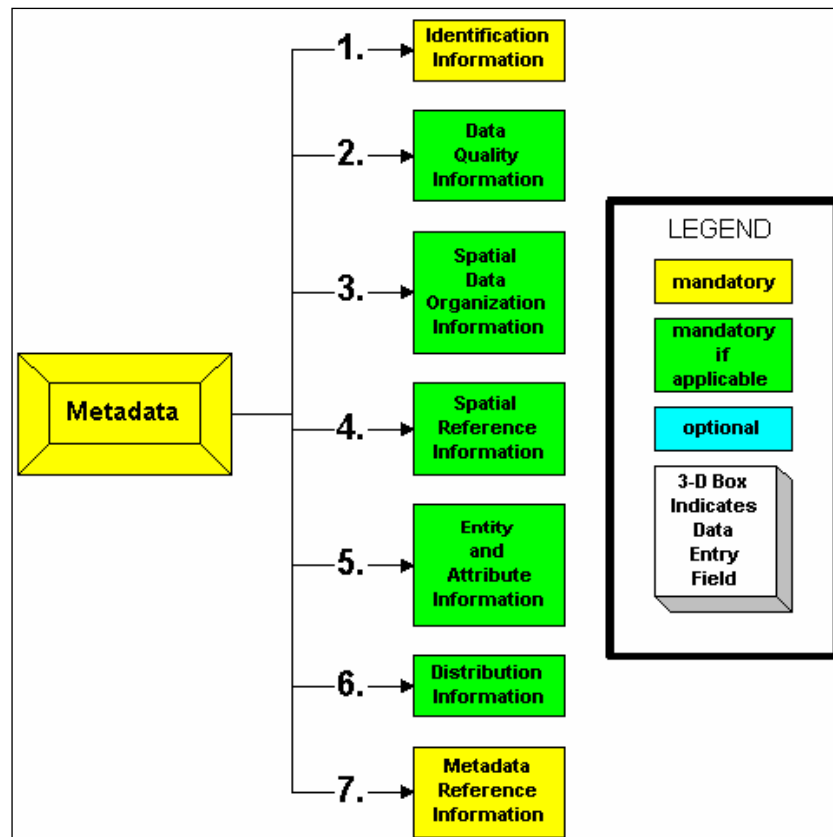
Para que seja possível representar as características inerentes aos conjuntos de dados geoespaciais, os metadados que os descrevem necessitam de elementos que permitam incluir sua extensão espacial (isto é, a caixa de delimitação), qual o sistema de referência de coordenadas (CRS – *Coordinate Reference System*) utilizado, as informações de qualidade e proveniência, dentre outros detalhes espaciais (W3C; OGC, 2017). Isso requer um padrão de metadados, a fim de que os conjuntos de dados geoespaciais a serem descritos sejam corretamente recuperados, codificados e processados pelas ferramentas de propósito específico do domínio (FUGAZZA et al., 2016).

Com o surgimento e a popularização das Infraestruturas de Dados Espaciais (IDEs), surgiu também a necessidade de se padronizar as diferentes formas de criar e gerenciar metadados sobre os recursos de geoinformação disponibilizados nessas plataformas (BRASIL, 2021). Nesse sentido, um padrão de metadados geoespaciais deve estabelecer um conjunto comum de terminologia e definições para a documentação do dado geográfico, incluindo elementos para sua identificação, bem como para informar os usuários sobre a qualidade do dado, a organização espacial, o sistema de referência de coordenadas aplicado, entre outros (FGDC, 1997).

Em nível local, nos EUA, o FGDC (*Federal Geographic Data Committee*), que é um comitê entre agências para promover a coordenação do uso, troca e disseminação de dados espaciais propôs o padrão CSDGM - *Content Standard for Digital Geospatial Metadata* (FGDC, 1998), que especifica os elementos necessários para suportar os principais usos de metadados geoespaciais, de forma a prover informação sobre o domínio de dados de uma organização, e sobre como processar e interpretar dados transferidos de uma fonte externa.

No padrão CSDGM, os metadados geoespaciais devem indicar o 'quem', 'o quê', 'quando', 'onde', 'porque', e 'como' de um conjunto de dados geoespacial (CASANOVA; BRAUNER; CÂMARA, 2005). A Figura 8 ilustra as categorias de metadados deste padrão.

Figura 8 – Categorias de metadados do padrão CSDGM.



Fonte: FGDC (1998).

As informações gerais contidas no padrão CSDGM foram distribuídas nas seguintes categorias, mostradas na Figura 8: identificação, qualidade do dado, organização do dado espacial, referência espacial, informação de entidade e atributo, distribuição e referências, sendo que a primeira e a última categorias são mandatórias em quaisquer casos, e as demais somente se aplicável.

De acordo com Brasil (2010), o CSDGM foi um padrão pioneiro para metadados geoespaciais, que acabou por ser adotado internacionalmente. No entanto, sua versão mais atual foi publicada em 1998, tendo em vista que o FGDC vem, desde 2003, encorajando a adoção e transferindo seus metadados para o padrão ISO 19115, assim como vem sendo feito por outros organismos internacionais que possuem padrões locais de metadados geoespaciais.

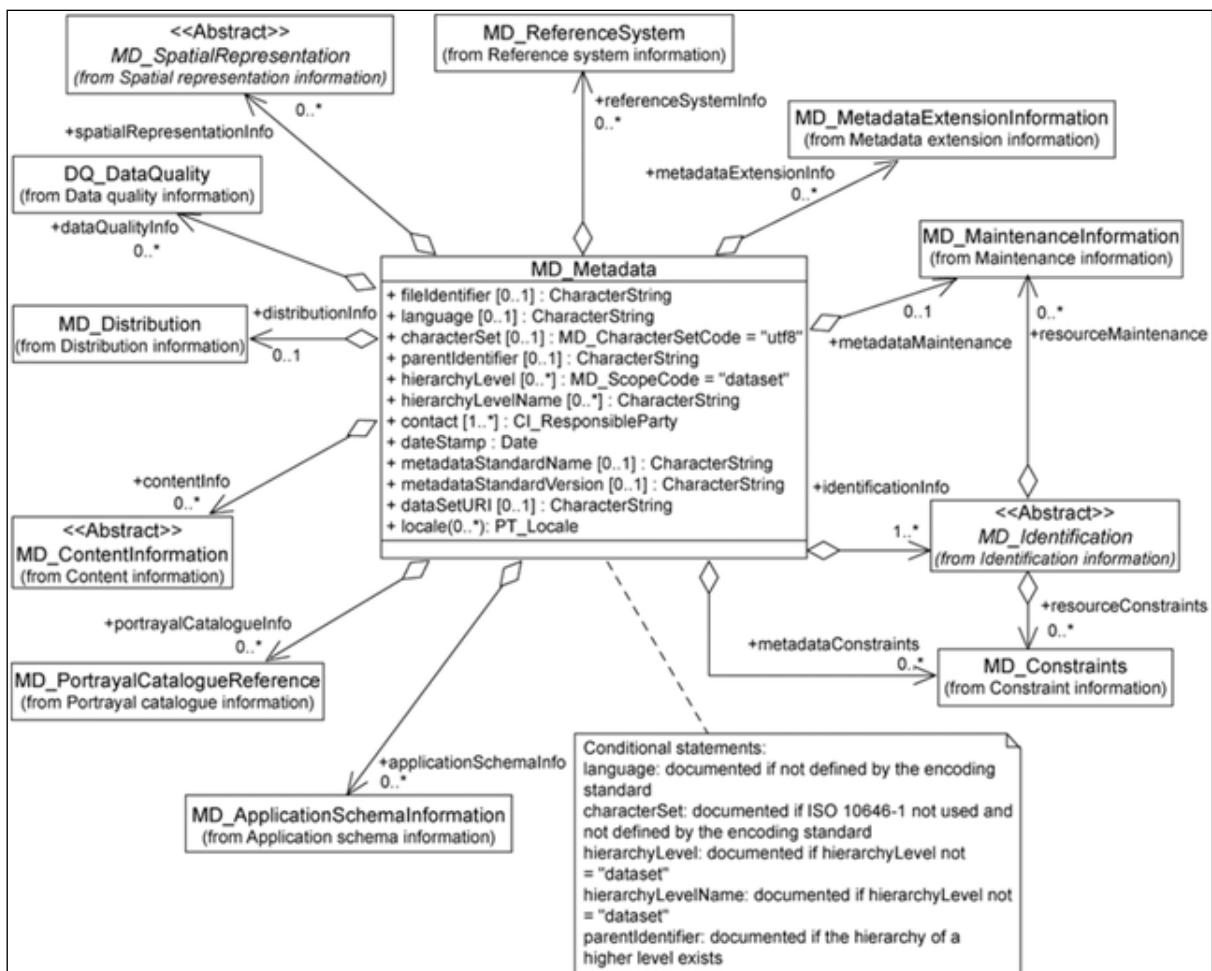
Em nível internacional, o ISO 19115 *Geographic Information – Metadata* (ISO, 2003) é um padrão de metadados de informação geográfica cuja finalidade é definir um esquema de metadados para a descrição de dados geoespaciais, incluindo a identificação, qualidade, extensão espacial, horizonte temporal, conteúdo, sistema de referência espacial, distribuição,

entre outras informações específicas. É atualmente o padrão mais aplicado internacionalmente para intercâmbio, integração e recuperação de dados geoespaciais (FAN, J. et al., 2017).

Para Cowen (2017, p. 4, tradução nossa), “o surgimento e a adoção de um padrão internacional de metadados foi um grande passo em direção à maturidade das disciplinas geoespaciais”. O principal benefício de se utilizar um padrão internacional para metadados é a promoção da interoperabilidade entre Sistemas de Informação Geográfica (SIGs) de diferentes instituições, empresas, comunidades e países (BRASIL, 2009).

A ISO 19115 define elementos de metadados, suas propriedades e os relacionamentos entre os elementos, além de estabelecer um conjunto comum de terminologia, definições e procedimentos de extensão de metadados (ISO, 2003). Faz parte de uma família de várias normas para informação geográfica e suporta o referenciamento espacial. Utiliza a modelagem UML para representar suas seções, entidades e elementos de metadados, conforme ilustrado na Figura 9.

Figura 9 – Esquema UML do padrão de metadados ISO 19115.



Além disso, este padrão combina aspectos de muitos outros padrões de metadados. Apesar de sua estrutura ser similar à do padrão CSDGM, com elementos obrigatórios e opcionais, a partir do qual podem ser derivados perfis, o padrão ISO 19115 apresenta diferenças marcantes em sua implementação (BRASIL, 2010).

Conforme mostrado na Figura 9, por ser aderente ao paradigma orientado ao objeto, os metadados do padrão ISO 19115 foram organizados em seções (componente superior), classes/subclasses (componente secundário), atributos, domínios e listas de códigos (FERNÁNDEZ-ORDÓÑEZ et al., 2014). As seções estão descritas no Quadro 11 a seguir.

Quadro 11 – Seções de metadados do padrão ISO 19115.

SEÇÃO DE METADADOS	DESCRIÇÃO
MD_Metadata	Informações do Conjunto de Entidades de Metadados: define metadados de um produto e estabelece hierarquia
MD_Identification	Informações de Identificação: informação básica requerida para identificar univocamente um produto
MD_Constraints	Informações de Restrições: restrições legais e de segurança no acesso e no uso dos dados
DQ_DataQuality	Informações de Qualidade dos Dados: descreve sua linhagem (fontes e processos de produção) e qualidade/ teste dos dados
MD_MaintenanceInformation	Informações de Manutenção dos Dados: descreve práticas de manutenção e atualização
MD_SpatialRepresentation	Informações de Representação Espacial: descreve mecanismo usado para representar os dados geoespaciais (matricial ou vetorial)
MD_ReferenceSystem	Informações do Sistema de Referência: descreve sistema de referência espacial e temporal usado
MD_ContentInformation	Informações de Conteúdo: descreve conteúdo do(s) catálogo(s) de abrangência e de feições usado(s) para definir feições de dados geoespaciais
MD_Distribution	Informações do Distribuidor: informações do distribuidor e métodos de acesso
MD_MetadataExtensionInformation	Descreve as extensões (novos elementos para atender aplicações específicas) e as informações sobre cada elemento de extensão criado.
MD_PortrayalCatalogueReference	Informações que identificam o catálogo de representação cartográfica usado
MD_ApplicationInformationSchema	Informações sobre o esquema de aplicativo usado para construir o conjunto de dados

Fonte: Adaptado de Brasil (2009).

Entre as seções do padrão ISO 19115 apresentadas no Quadro 11, estão distribuídos os cerca de 300 elementos de metadados de seu esquema, o que permite a definição de perfis e extensões para aplicações específicas, servindo de base para o estabelecimento de perfis de metadados geoespaciais em diversos países, uma vez que leva em consideração diferenças culturais e linguísticas. No entanto, dentre essa grande quantidade de elementos, oito devem ser

de uso obrigatório em qualquer perfil que seja baseado nesse padrão, de forma a garantir a interoperabilidade entre as diversas implementações (BRASIL, 2009).

No ano de 2014 essa norma passou por um processo de atualização pelo Comitê Técnico 211 da ISO, dando origem à versão ISO 19115-1:2014, com o intuito de adaptar-se às características e evolução da informação geoespacial. Nessa atualização⁴ foram acrescentados 194 elementos e excluídos 66 de seus elementos originais.

Por sua vez, um perfil de metadados geoespaciais reúne um conjunto de elementos necessários, baseados em um ou mais padrões de metadados, para representar as características dos recursos geoespaciais de uma determinada comunidade, ou de um produto geoespacial, de forma a garantir sua identificação, avaliação e utilização consistente. Esse conjunto de elementos é proposto como o núcleo básico e comum a vários tipos de recursos geoespaciais, de forma a organizar a informação de uma instituição, permitindo sua recuperação, além de descrever recursos informacionais capazes de auxiliar os usuários na identificação, autenticação e contextualização desses recursos (PINTO et al., 2017).

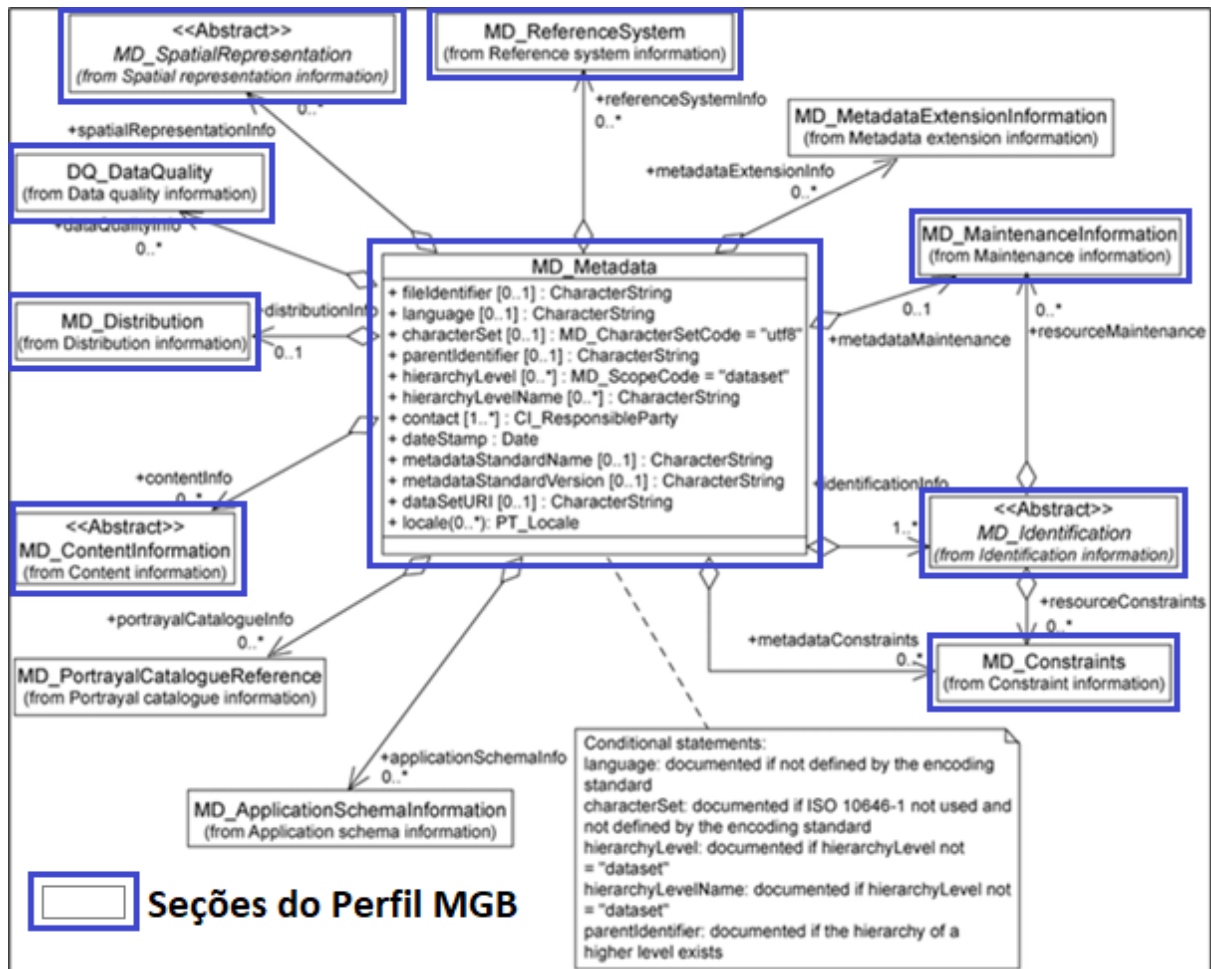
No Brasil foi estabelecido o Perfil de Metadados Geoespaciais Brasileiro (Perfil MGB), o qual é apresentado em detalhes a seguir.

3.3.2.1 Perfil de Metadados Geoespaciais Brasileiro

O Perfil de Metadados Geoespaciais Brasileiro (Perfil MGB) foi estabelecido, em 2009, pela então Comissão Nacional de Cartografia (CONCAR), a partir da análise do padrão para metadados geoespaciais especificado pela norma técnica ISO 19115 (*Geographic Information – Metadata*), bem como de outros perfis internacionais também baseados nessa norma. Conforme visto anteriormente, dentre a grande quantidade de elementos que compõem as diversas seções da norma ISO 19115, oito são de uso obrigatório em qualquer perfil que seja baseado nesse padrão, de forma a garantir a interoperabilidade entre as diversas implementações. Assim, o perfil MGB contempla as principais seções dessa norma técnica, além de empregar os oito elementos de uso obrigatório (BRASIL, 2009). A Figura 10 mostra as seções da ISO 19115 que foram adotadas pelo Perfil MGB.

⁴ Um breve estudo comparativo das alterações ocorridas no padrão ISO 19115-1 pode ser encontrado em Loti et al. (2019).

Figura 10 – Seções da ISO 19115 adotadas pelo Perfil MGB.



Fonte: Adaptado de Brasil (2009).

Conforme mostrado na Figura 10, o Perfil MGB, em sua concepção, adotou as seguintes seções do padrão ISO 19115 (vide descrições no Quadro 11): *MD_Metadata*, *MD_Identification*, *MD_Constraints*, *MD_SpatialRepresentation*, *MD_ReferenceSystem*, *MD_ContentInformation*, *MD_Distribution*, *MD_MaintenanceInformation*, e *DQ_DataQuality*.

Devido à atualização ocorrida em 2014 no padrão ISO 19115, que deu origem à versão ISO 19115-1, foi iniciado, em 2019, um projeto de atualização do Perfil MGB que resultou, em 2021, na publicação do Perfil MGB versão 2.0⁵. O Quadro 12 apresenta os pacotes de elementos que compõem este perfil, bem como as respectivas definições, tipos e obrigatoriedade de utilização de cada um deles.

⁵ Disponível em: <https://www.ibge.gov.br/geociencias/metodos-e-outros-documentos-de-referencia/normas/30717-perfil-de-metadados-geoespaciais-do-brasil.html?=&t=publicacoes>

Quadro 12 – Pacotes de elementos do Perfil MGB 2.0.

NOME PACOTE	NOME ISO 19115	DEFINIÇÃO	OBRIGATORIEDADE
Identificador do metadado	<i>metadataIdentifier</i>	Identificador único para a instância de metadado	Obrigatório
Localidade padrão	<i>defaultLocale</i>	Idioma e código de caracteres usados para documentar o metadado	Obrigatório
Metadado de nível acima	<i>parentMetadata</i>	Citação para um metadado que agregue esse registro, se for o caso	Opcional
Escopo do metadado	<i>metadataScope</i>	Escopo ou tipo do recurso que este metadado descreve	Obrigatório
Responsável pelo metadado	<i>contact</i>	Entidade ou pessoa responsável pelas informações descritas no metadado	Obrigatório
Data do metadado	<i>dateInfo</i>	Datas associadas ao metadado. Incluir pelo menos a data de criação	Obrigatório
Padrão de metadados	<i>metadataStandard</i>	Citação para o padrão de metadados usado	Opcional
Perfil de metadados	<i>metadataProfile</i>	Citação para o perfil de metadados utilizado para documentar essa instância	Obrigatório
Link do metadado	<i>metadataLink</i>	Endereço Web onde o metadado pode ser encontrado	Opcional
Representação espacial	<i>spatialRepresentationInfo</i>	Informação sobre a representação espacial do recurso descrito	Opcional
Sistema de referência	<i>referenceSystemInfo</i>	Informação sobre o sistema de referência espacial e temporal do recurso descrito	Opcional
Identificação	<i>identificationInfo</i>	Informação básica sobre o recurso que o metadado descreve	Obrigatório
Conteúdo	<i>contentInfo</i>	Informações técnicas sobre os dados vetoriais ou matriciais descritos	Opcional
Distribuição	<i>distributionInfo</i>	Informação sobre o distribuidor e as formas de se obter o recurso	Opcional
Qualidade	<i>dataQualityInfo</i>	Informação sobre a qualidade do recurso	Opcional
Linhagem do recurso	<i>resourceLineage</i>	Informação sobre os insumos e processos de produção aplicados no recurso	Opcional

Catálogo de representação	<i>portrayalCatalogueInfo</i>	Informação sobre o catálogo de representações cartográficas do recurso	Opcional
Restrições do metadado	<i>metadataConstraints</i>	Restrições ao acesso e uso do metadado	Opcional

Fonte: Brasil (2009).

A utilização deste perfil de metadados pelos provedores de dados tem por objetivo: "a) identificar o produtor e a responsabilidade técnica; b) padronizar a terminologia utilizada; c) assegurar a transferência de dados; d) viabilizar a integração de informações; e) identificar a qualidade da informação geográfica; e, f) garantir os requisitos mínimos de uso dos dados geoespaciais." (BRASIL, 2009, p. 10). É esperado também que a adoção deste perfil de metadados pelos provedores de dados geoespaciais promova a interoperabilidade entre diversos sistemas de informação geográfica, de forma a facilitar o compartilhamento de dados entre diferentes instituições e organizações.

3.3.2 Codificação de metadados geoespaciais

Os esforços da comunidade de geoinformação para tornar a criação de metadados uma prática comum, levou ao desenvolvimento de ferramentas padronizadas que permitiram a criação de metadados na forma de documentos em formato XML (*eXtensible Markup Language*) usando apenas um editor de texto simples. O XML é um padrão industrial aberto para a publicação e distribuição de informação através da Internet, que pode ser utilizado em qualquer serviço de catálogo de metadados. Neste sentido, a indústria de softwares de informação geoespacial também desenvolveu ferramentas baseadas em formulários para criar metadados em formato XML (COWEN, 2017).

Os metadados geoespaciais, serializados em formato XML, utilizam o padrão de codificação GML (*Geography Markup Language*). O GML é uma gramática XML, publicada na forma do padrão ISO 19136:2007, que é utilizada para expressar características geoespaciais (coordenadas, pontos, linhas e polígonos), além de servir como um formato de intercâmbio aberto (*OpenGIS*) para transações geoespaciais na Internet (OGC, 2007). O Quadro 13 apresenta uma amostra de metadados geoespaciais codificados neste padrão.

Quadro 13 – Amostra de metadados geoespaciais codificados em GML/XML.

```

<?xml version="1.0" encoding="UTF-8"?>
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd"
  xmlns:gml="http://www.opengis.net/gml"
  xmlns:gco="http://www.isotc211.org/2005/gco"
  xsi:schemaLocation="http://www.isotc211.org/2005/gmd
http://www.isotc211.org/2005/gmd/gmd.xsd http://www.isotc211.org/2005/gmx
http://www.isotc211.org/2005/gmx/gmx.xsd http://www.isotc211.org/2005/srv
http://schemas.opengis.net/iso/19139/20060504/srv/srv.xsd">
  <gmd:fileIdentifier>
    <gco:CharacterString>a5220c18-f7fa-4e3e-b39b-feeb3ccc4830</gco:CharacterString>
  </gmd:fileIdentifier>
  <gmd:language>
    <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/"
      codeListValue="por"/>
  </gmd:language>
  ...
  <gmd:referenceSystemInfo>
    <gmd:MD_ReferenceSystem>
      <gmd:referenceSystemIdentifier>
        <gmd:RS_Identifier>
          <gmd:code>
            <gco:CharacterString>EPSG:4674</gco:CharacterString>
          </gmd:code>
        </gmd:RS_Identifier>
      </gmd:referenceSystemIdentifier>
    </gmd:MD_ReferenceSystem>
  </gmd:referenceSystemInfo>
  ...
  <gmd:extent>
    <gmd:EX_Extent>
      <gmd:temporalElement>
        <gmd:EX_TemporalExtent>
          <gmd:extent>
            <gml:TimePeriod gml:id="d18857e395a1051934">
              <gml:beginPosition>2008-04-01</gml:beginPosition>
              <gml:endPosition>2018-09-24</gml:endPosition>
            </gml:TimePeriod>
          </gmd:extent>
        </gmd:EX_TemporalExtent>
      </gmd:temporalElement>
    </gmd:EX_Extent>
  </gmd:extent>
  <gmd:extent>
    <gmd:EX_Extent>
      <gmd:geographicElement>
        <gmd:EX_GeographicBoundingBox>
          <gmd:westBoundLongitude>
            <gco:Decimal>-74</gco:Decimal>
          </gmd:westBoundLongitude>
          <gmd:eastBoundLongitude>
            <gco:Decimal>-44</gco:Decimal>
          </gmd:eastBoundLongitude>
          <gmd:southBoundLatitude>
            <gco:Decimal>-17.5</gco:Decimal>
          </gmd:southBoundLatitude>
          <gmd:northBoundLatitude>
            <gco:Decimal>6</gco:Decimal>
          </gmd:northBoundLatitude>
        </gmd:EX_GeographicBoundingBox>
      </gmd:geographicElement>
    </gmd:EX_Extent>
  </gmd:extent>
  ...
</gmd:MD_Metadata>

```

Fonte: Elaborado pelo autor.

No Quadro 13 são apresentadas algumas seções de um registro de metadados, que fornecem informações sobre o identificador único associado, idioma dos metadados, sistema de referência de coordenadas utilizado, e as coordenadas do polígono que demarca a área geográfica a que se referem os metadados. Os conceitos-chave do GML são a *feature*, uma abstração de um fenômeno do mundo real, a *geographic feature*, uma característica que está associada a uma localização relativa à Terra, e a *feature collection*, uma coleção de *features* que pode ser considerada como uma única *feature* e dá uma representação digital do mundo real. Diferentes das *features*, as *geographic features* têm propriedades cujo valor pode ser uma geometria (BRINK et al., 2014).

3.4 INFRAESTRUTURA DE DADOS ESPACIAIS (IDE)

Vários produtos, sistemas e organizações dependem do acesso imediato a informações geoespaciais precisas e oficiais (PATROUMPAS et al., 2015). Quando a implementação dos Sistemas de Informações Geográficas (SIGs) migrou dos computadores de grande porte (*mainframe*) para os computadores pessoais, os novos utilizadores destes sistemas desejavam muitas vezes incorporar os dados geoespaciais existentes. Essas demandas encorajaram o estabelecimento de parcerias entre provedores, baseadas em abordagens partilhadas para a criação e distribuição deste tipo de dado, o que levou à concepção das atuais Infraestruturas de Dados Espaciais (COWEN, 2017).

Nos EUA, o *Federal Geographic Data Committee* define uma IDE como um "guarda-chuva de políticas, padrões e procedimentos sob os quais as organizações e tecnologias interagem para promover o uso, gerenciamento e produção mais eficientes de dados geoespaciais" (FGDC, 1996). Já no Brasil, as IDEs podem ser definidas como um "meio de empregar tecnologias, políticas, padrões e partes interessadas para auxiliar no acesso aos dados espaciais e possibilitar sua disseminação" (ASSIS et al., 2019, p. 3). Neste sentido, para Rajabifard, Feeney e Williamson (2002, p. 13, tradução nossa) é "[...] a interação entre usuários e provedores de dados, além de qualquer agente agregador de valor entre eles, que deve direcionar o desenvolvimento de qualquer IDE.”.

Diante disso, a concepção de qualquer IDE deve se basear em um modelo que leve em conta as dimensões humana, organizacional e técnica. O Quadro 14 apresenta alguns dos elementos a serem observados em cada uma dessas dimensões.

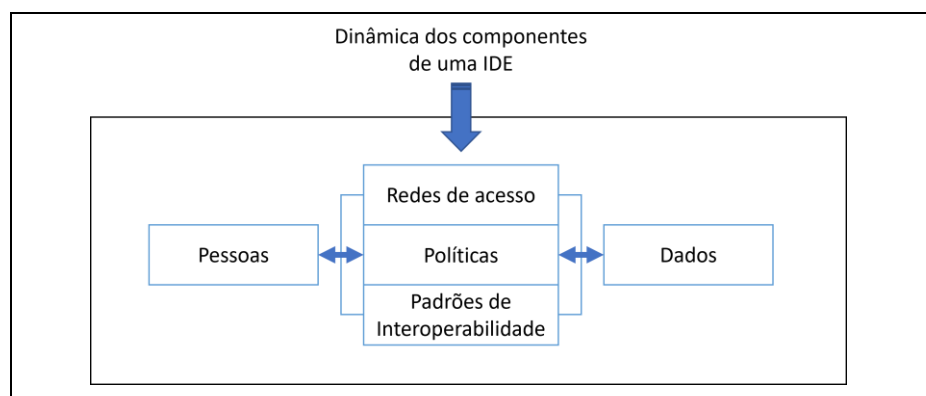
Quadro 14 - Dimensões e elementos de um modelo para concepção de uma IDE.

DIMENSÕES	ELEMENTOS
Humana	• Produtores e Usuários • Recursos Humanos
Organizacional	• Marco Legal • Organização e Gestão • Fortalecimento Institucional
Técnica	• Dados e Metadados • Normas e Especificações • Tecnologia

Fonte: Adaptado de Brasil (2010).

Inicialmente, o objetivo das IDEs era reduzir a duplicação de esforços entre agências, melhorar a qualidade e reduzir os custos relacionados com a informação geográfica, tornar os dados geográficos mais acessíveis ao público, aumentar os benefícios da utilização dos dados disponíveis, e estabelecer parcerias chave com os setores público e privado para aumentar a disponibilidade de dados (COWEN, 2017). Atualmente, o objetivo principal das IDEs é oferecer serviços de acesso à informação geográfica com base em grandes catálogos de metadados, não importando para o usuário o local, meio ou estrutura física onde os dados são armazenados (DAVIS JÚNIOR; SOUZA; BORGES, 2005).

Para isso, as IDEs possuem como componentes centrais pessoas, redes de acesso, políticas, padrões e dados, ou seja, a dinâmica de uma IDE não trata somente da disponibilização de dados geoespaciais aos usuários finais por meio de serviços de valor agregado, mas envolve questões relacionadas à padrões de interoperabilidade, políticas e redes (RAJABIFARD et al., 2006). A Figura 11 ilustra o arranjo dessa dinâmica entre os componentes de uma IDE.

Figura 11 – Dinâmica dos componentes de uma IDE.

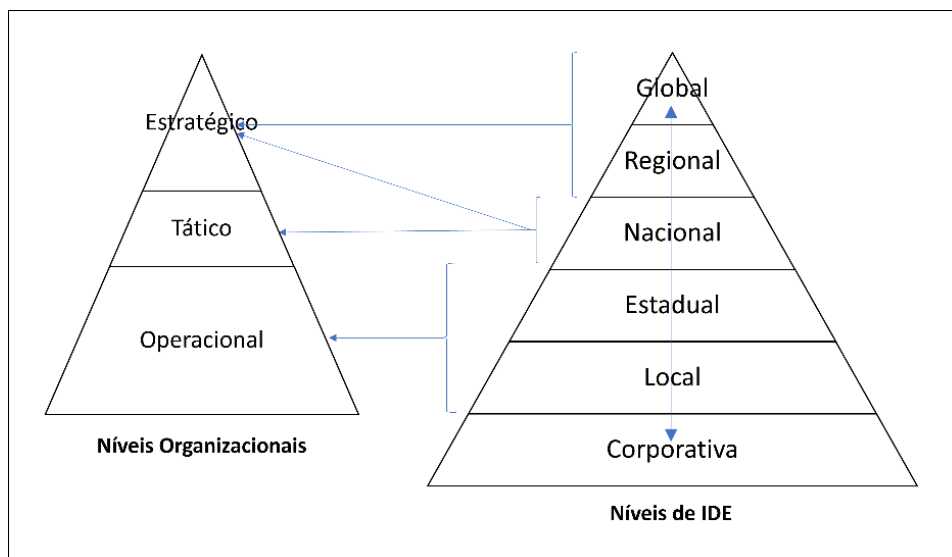
Fonte: Adaptado de Rajabifard et al. (2006).

Na dinâmica apresentada na Figura 11, as pessoas são uma componente chave para o processamento das transações e para a tomada de decisões. No entanto todas as decisões requerem dados e, à medida que os dados se tornam mais mutáveis, questões humanas de compartilhamento de dados, segurança, precisão e acesso incentivam a necessidade de relações mais definidas entre as pessoas e os dados (RAJABIFARD et al., 2006).

Ao longo do tempo, as IDEs vem sendo implementadas em diferentes níveis político-administrativo: global, regional, nacional, estadual, local ou organizacional. Estes níveis podem ser representados de forma hierárquica por meio de uma pirâmide, onde, baseado nos objetivos e estratégias de diferentes iniciativas, podem ser relacionados a uma pirâmide de gestão organizacional (estratégico, tático e operacional), conforme ilustra a Figura 12.

Como mostra a figura 12, os níveis multinacionais (global e regional) podem ser considerados similares ao nível estratégico de uma estrutura organizacional. O nível nacional pode ser considerando similar ao nível gerencial ou ao nível estratégico, a depender do sistema político do país. Já os níveis estadual e local podem ser similares ao nível operacional de uma organização, uma vez que esses níveis produzem, coletam e gerenciam dados em larga escala de forma a contribuir com os níveis superiores.

Figura 12 – Relação entre níveis de IDEs e níveis organizacionais.



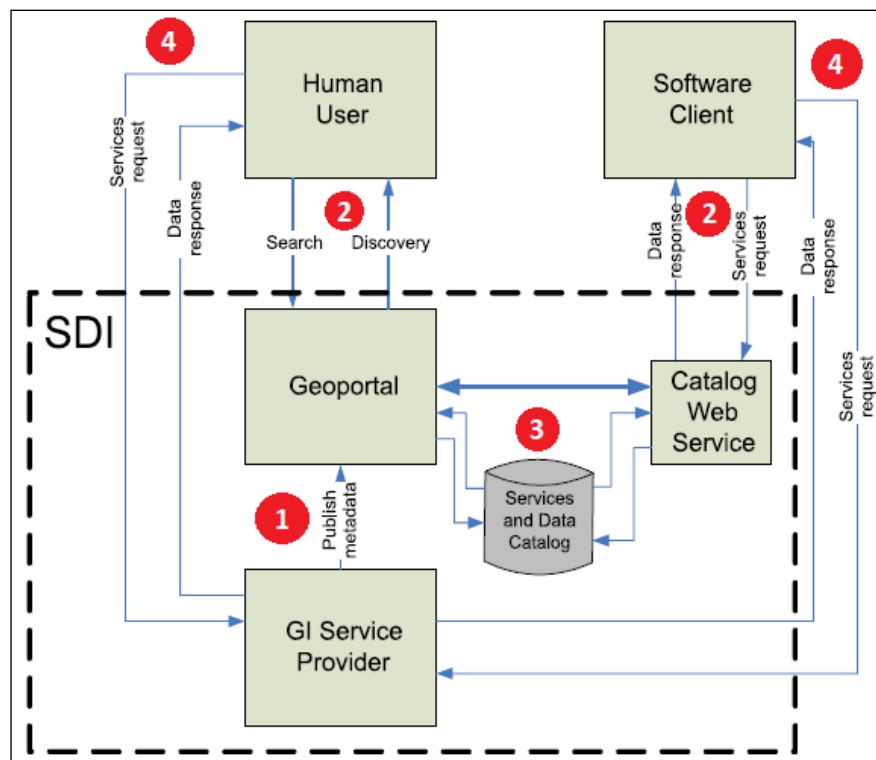
Fonte: Adaptado de Rajabifard et al. (2006).

Hoje em dia as IDEs são sistemas maduros, desenvolvidos após décadas de pesquisa e desenvolvimento, baseados em padrões liderados por organismos internacionais como ISO, OGC e W3C (PATROUMPAS et al., 2015). Dessa forma, elas não devem impor a adoção de produtos específicos aos seus participantes, mas sim prover uma visão de sua arquitetura e

determinar um conjunto mínimo de padrões, de ampla aceitação, para que efetivamente haja interoperabilidade dos dados. De acordo com Brasil (2010, p. 32), "dentre os fatores críticos de sucesso na construção da arquitetura informacional de uma IDE devem ser considerados os aspectos referentes aos dados, à sua normalização e à tecnologia necessária para a sua geração, disponibilização, acesso e manuseio (incluindo análises)". A Figura 13 permite visualizar a arquitetura conceitual de uma IDE.

Na arquitetura apresentada na Figura 13, o acesso aos dados é realizado apenas por meio de serviços *Web*, sendo que cada provedor se encarrega de registrar (1), junto a um serviço de catalogação, quais dados possui, onde estão, como estão organizados, e onde estão os metadados. Assim, basta que o usuário acione um serviço (2) para determinar se os dados que procura estão disponíveis, um segundo serviço (3) para avaliar detalhes sobre a fonte e o processo de captura e, se os dados atenderem suas necessidades informacionais, acione um terceiro serviço (4) para recuperá-los (DAVIS JÚNIOR; SOUZA; BORGES, 2005).

Figura 13 – Arquitetura conceitual de uma IDE.



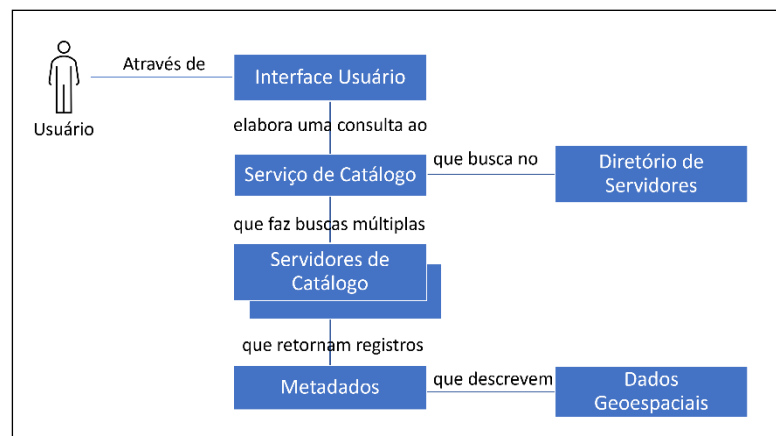
Fonte: Adaptado de Davis Júnior e Alves (2006).

Uma vez que o propósito de uma IDE é facilitar o acesso e troca de informações geoespaciais, torna-se fundamental a existência de ferramentas que possibilitem ao usuário interagir com as diversas bases de provedores de dados geoespaciais envolvidos, na busca pelos

conjuntos de dados que atendam às suas necessidades informacionais, usando para isso critérios espaciais, temporais ou outras características associadas às informações. Estas interações com as IDEs em busca de informação se dividem em duas frentes: a de descoberta e a de recuperação de informação geoespacial (CAMBOIM; SLUTER, 2013).

Nas interações de descoberta, considerando que provedores e usuários estão distribuídos através da rede, as IDEs implementam catálogos de metadados onde os provedores publicam as características dos conjuntos de dados geoespaciais disponibilizados. Tais catálogos de metadados possuem ferramentas de busca para que os usuários possam encontrar os conjuntos de dados que atendam a determinados critérios, como área de abrangência, forma de execução, data de atualização, nível de detalhamento, entre outros (CAMBOIM; SLUTER, 2013). A Figura 14 ilustra essas interações de descoberta em IDEs. Em geral, este catálogo de metadados é operacionalizado pelo Serviço de Catálogo de Metadados para a Web (CSW), que será visto mais adiante em detalhes.

Figura 14 – Interações de descoberta de metadados em IDEs.



Fonte: Adaptado de GSDI (2004, p.41).

Conforma ilustrado na Figura 14, nas interações de descoberta de metadados em IDEs, os usuários, por meio de uma interface no geoportal da IDE, envia uma consulta ao serviço de catálogo local, que além de buscar em suas entradas e diretórios de servidores, também pesquisa em outros servidores de catálogo em IDEs interligadas (como por exemplo a INDE brasileira), que retornam os registros de metadados que satisfaçam aos parâmetros da consulta efetuada.

Já nas interações de recuperação, uma vez que os usuários tenham encontrado os conjuntos de dados geoespaciais que satisfaçam suas necessidades informacionais, o próximo passo é a obtenção dos mesmos junto aos respectivos provedores acessando-os diretamente ou por meio de serviços Web (CAMBOIM; SLUTER, 2013).

O OGC já realizou estudos (REED; REICHARDT, 2008) sobre os desafios para as próximas gerações de IDEs tendo concluído que um fator crítico para o sucesso de longo prazo destas plataformas é a interoperabilidade semântica, o que está diretamente relacionado à proposta deste trabalho.

A seguir será detalhado um dos componentes centrais de uma IDE, o Catálogo de Metadados, operacionalizado pelo Serviço de Catálogo de Metadados para a Web (CSW), que constitui a fonte primária de metadados a serem publicados de forma semântica na Web por meio do *framework* proposto neste trabalho.

3.4.1 Serviço de Catálogo de Metadados Geoespaciais para a Web (CSW)

Um catálogo pode ser definido como um sistema para a publicação, consulta e recuperação de itens de informação de forma sistemática (LOPEZ-PELLICER et al., 2010). Nesse sentido, os registros de metadados de recursos geoespaciais fornecem a base para a construção de catálogos de metadados geoespaciais. Embora a criação de catálogos de metadados tenha, inicialmente, o propósito de suprir as necessidades informacionais de um único usuário ou organização, seu real valor é percebido quando tais catálogos são compartilhados entre organizações e com outros usuários na Web (COWEN, 2017).

O projeto de geoportais em IDEs é influenciado principalmente pelos padrões da OGC e pelo conceito de Arquiteturas Orientadas a Serviços (SOA). O centro desse conceito arquitetônico é baseado no paradigma ‘publique, encontre, consuma’, no qual um provedor de serviços registra ou publica um recurso em um catálogo e os usuários posteriormente podem pesquisar e encontrar esse recurso, obtendo todas as informações necessárias para consumi-lo. Nesse paradigma, a capacidade de descoberta de recursos desempenha um papel central, pois somente quando os usuários podem descobrir um recurso é que eles podem usá-lo. Neste sentido, os catálogos de uma IDE não contêm essencialmente os dados em si, mas sim as informações de metadados descritivas sobre os conjuntos de dados ou serviços, bem como um link para os dados ou serviços reais (VOCKNER; MITTBOCK, 2014).

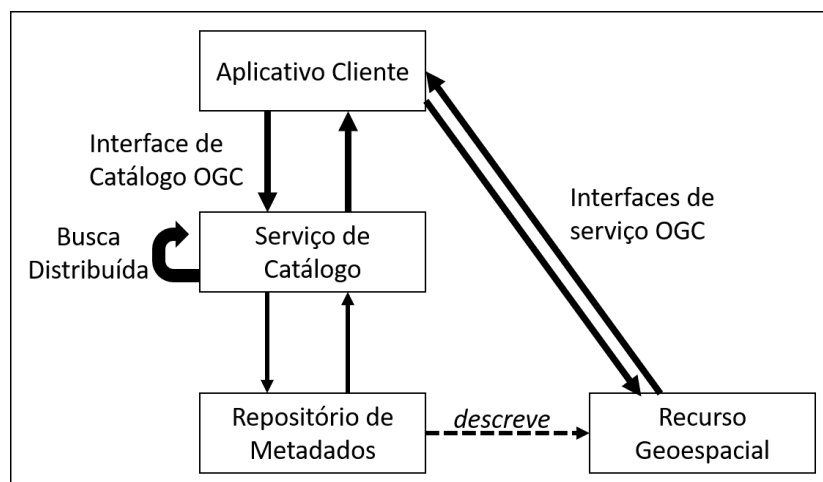
Os catálogos de metadados geoespaciais disponibilizam descrições de *datasets* armazenados em diversas fontes. Além disso, oferecem serviços de localização, consulta e gerência dos metadados, assim como serviços de solicitação dos *datasets*, os quais são redirecionados às respectivas fontes (CASANOVA; BRAUNER; CÂMARA, 2005). Para isso, a especificação *OpenGIS Catalogue Services* (CS) detalha interfaces de descoberta, acesso, manutenção e organização de catálogos de metadados de informação geoespacial e de recursos

relacionados em sistemas distribuídos (OGC, 2007b). Essa especificação define uma ligação de protocolo HTTP chamada de *Catalog Service for the Web* (CSW).

O CSW, ou Serviço de Catálogo para a Web, é um padrão da OGC que descreve perfis de aplicativos para publicação, acesso e pesquisa sobre coletas de metadados em dados geoespaciais, serviços e recursos relacionados. Nele, os metadados devem ser codificados em XML, sendo as respostas às solicitações dos usuários devolvidas no mesmo formato (PATROUMPAS et al., 2015). Sua arquitetura é apresentada na Figura 15.

Na Figura 15, os clientes utilizam o CSW para localizar os metadados dos recursos geoespaciais de que necessitam. Em geral, os catálogos possuem ferramentas de busca distribuída nos metadados para selecionar somente os *datasets* que atendam determinados critérios descritos pelos metadados (área de abrangência, data de atualização, nível de detalhamento, entre outros). Para descrição do conteúdo do repositório de metadados são usadas palavras-chave, o que pode gerar problemas na busca caso seja usada na descrição do conteúdo uma terminologia diferente da empregada pelo usuário (CAMBOIM; SLUTER, 2013).

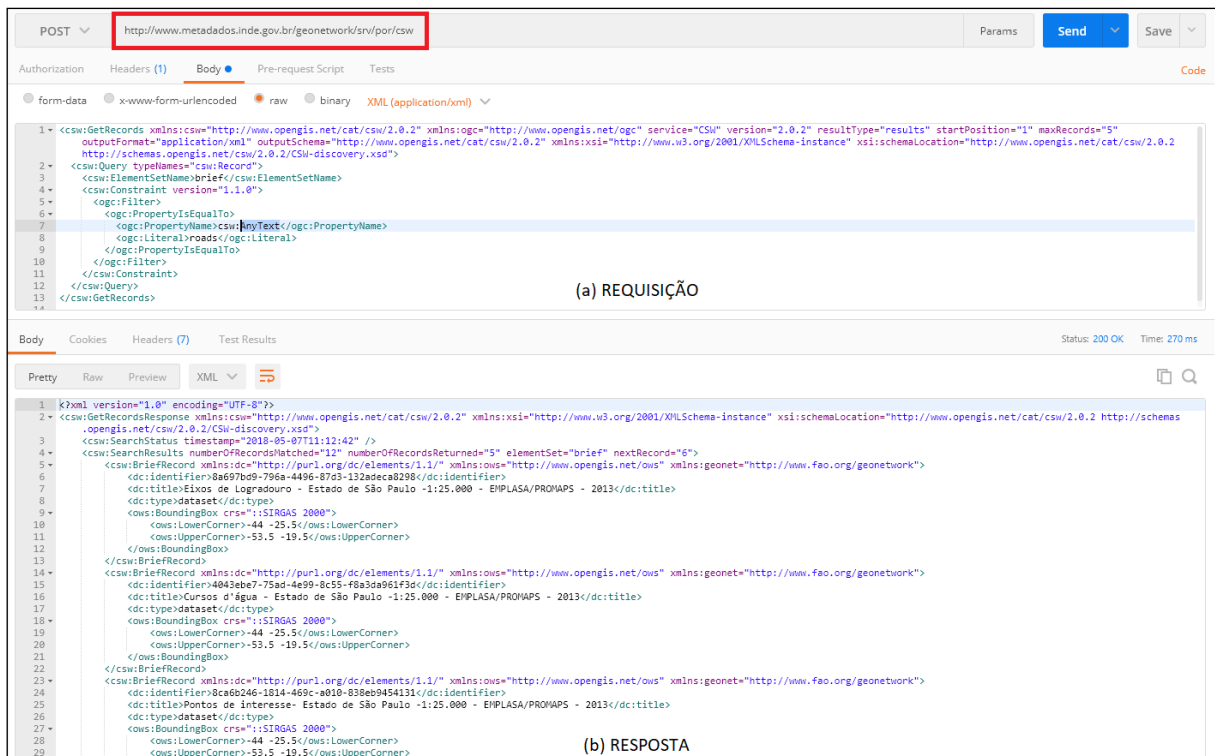
Figura 15 – Arquitetura do Serviço de Catálogo para a Web (CSW).



Fonte: Adaptado de OGC (2007b).

Por estes aspectos, a operação do serviço CSW é bastante complexa. Como exemplo, a operação *GetRecordById*, ilustrada na Figura 16, busca representações de registros de metadados utilizando o identificador dos metadados no repositório local.

Figura 16 – Exemplo de requisição e resposta do Serviço de Catálogo para a Web (CSW).



Fonte: Brasil (2022).

Na consulta CSW apresentada na Figura 16, é efetuado em (a) uma requisição do tipo `<csw:GetRecords>` ao servidor de catálogo de metadados da INDE (<http://www.metadados.inde.gov.br/geonetwork/srv/por/csw>), solicitando uma relação breve (`csw:ElementSetName = brief`) dos metadados, contendo qualquer texto (`csw:AnyText`), que descrevam *datasets* relacionados a rodovias (*roads*). Como resposta, em (b) são retornados metadados de identificação e de extensão espacial de três *datasets* encontrados no servidor de catálogo que atenderam aos parâmetros da consulta: ‘Eixos de Logradouro – Estado de São Paulo...’, ‘Cursos d’Água – Estado de São Paulo’ e ‘Pontos de Interesse – Estado de São Paulo’.

A implementação de um servidor CSW pode ser feita por meio de diversos softwares livres atualmente disponíveis, dentre os quais: *ArcIMS*, da empresa *ESRI*; *GeoConnect*, da empresa *Intergraph*; *GeoNetwork*, da FAO⁶; entre outros. No Brasil, para as instituições integrantes da INDE, a recomendação da Comissão Nacional de Cartografia (CONCAR) é pelo uso do software *GeoNetwork*, que tem entre suas principais características o suporte nativo a padrões de metadados geoespaciais conhecidos (FGDC, ISO 19115, ...), mecanismos de busca

⁶ Food and Agriculture Organization, ou Organização para Alimentação e Agricultura, da Organização das Nações Unidas.

avançados, sincronização entre catálogos distribuídos, além de se ser livre e de código aberto (BRASIL, 2010).

Com isso, cumpre-se o objetivo específico I inicialmente estabelecido, de apresentar os conceitos de dados e metadados geoespaciais, bem como a arquitetura e os componentes de uma Infraestrutura de Dados Espaciais, com foco no serviço de catálogo de metadados, que são essenciais para a compreensão de alguns dos componentes do *framework* que será proposto neste trabalho.

4 WEB SEMÂNTICA E *LINKED DATA*

Nesta Seção será abordado outro elemento chave para a construção do *framework* proposto neste trabalho, que é a *Web Semântica*, de forma a cumprir com o objetivo específico II inicialmente proposto. A *Web Semântica* é o meio que possibilitará a publicação de metadados geoespaciais, enriquecidos semanticamente e ligados a outras fontes de dados, com o intuito de proporcionar uma recuperação mais precisa dos metadados pelos usuários de uma IDE, diretamente na *Web* e sem a necessidade de recorrer a consultas sintáticas em servidores de ‘Serviço de Catálogo para a *Web*’ (CSW).

A primeira subseção discorrerá sobre como surgiu e como se deu a evolução da *Web Semântica* até os dias atuais. Na segunda subseção são apresentadas as tecnologias e ferramentas concebidas ao longo da evolução da *Web Semântica*. Na terceira subseção são tratados os princípios dos dados ligados (*Linked Data*) e o paradigma dos dados abertos ligados (*linked open data*). Além disso, nessa Seção são abordadas a aplicação de *Linked Data* aos dados geoespaciais, também conhecida como *linked geodata*, e as boas práticas para publicação de dados espaciais na *Web*. Por fim, a quarta subseção mostra de que forma as boas práticas de *Linked Data* e de dados espaciais na *Web* podem ajudar a promover a integração das IDEs com a *Web Semântica* por meio da publicação dos metadados catalogados em tais plataformas.

4.1 EVOLUÇÃO DA WEB SEMÂNTICA

A *Web*, tal como a conhecemos atualmente, ainda pode ser considerada uma ‘*Web* de documentos’, ou seja, totalmente voltada para a leitura, extração e relacionamento de conteúdos por humanos, que foi formada a partir da criação crescente de *links* entre documentos de hipertexto, escritos em linguagem HTML, e da disponibilização de arquivos nos mais diferentes formatos.

Quando começou a desenvolver a *Web*, no início da década de 1990, Tim Berners-Lee imaginou uma *Web* de leitura e gravação, porém o que surgiu foi uma *Web* predominantemente de leitura, chamada de *Web 1.0*, onde a interação com os usuários se limitava à busca e leitura de conteúdo. Tal falta de interação entre usuários e a *Web 1.0* levou ao surgimento, em 1999, da chamada *Web 2.0*, na qual usuários comuns podiam escrever e compartilhar informações entre si, o que culminou no surgimento dos *blogs*, das redes sociais e das plataformas de *streaming* de vídeo (Po et. al., 2020).

Com o passar do tempo, os usuários começaram a fazer *upload* de conteúdo textual e de multimídia a taxas cada vez mais altas, o que somado ao enorme volume de páginas na *Web* e o maior número de requisições, acabaram por exigir dos aplicativos novas maneiras de lidar com documentos, de forma que as máquinas pudessem entender com quais dados estavam lidando. Isso levou ao projeto de uma *Web* 3.0, chamada de ‘*Web Semântica*’, cuja ideia principal era fornecer um contexto aos documentos na *Web* em um formato legível por máquina (Po et. al., 2020). Nela, o conteúdo da *Web* é gerado por aplicações baseadas em modelos, com foco na interoperabilidade de dados, sensíveis ao contexto. Com isso, a *Web* pode vir a ser explorada por agentes inteligentes e motores semânticos, uma forma de aproximá-la do campo da Inteligência Artificial (IA).

Para essa aproximação da *Web* com a IA, de acordo com Feitosa (2006, p. 64), é necessário “adicionar-se lógica à *Web*, por meio da utilização de uma linguagem que seja capaz de expressar dados e regras para inferências sobre tais dados, fornecendo também as regras para a exportação para a *Web* de qualquer sistema de representação do conhecimento existente.”. Diante disso,

A concepção da *Web Semântica*, que almeja tais objetivos, é uma estrutura de representação de dados na *World Wide Web*, por meio de um esforço colaborativo liderado pelo W3C, com a participação de pesquisadores e de empresas de desenvolvimento de todo o mundo. A ideia da *Web Semântica* é uma evolução do grupo de trabalho *Metadata Activity*, vinculado ao W3C, cujos maiores objetivos eram o estudo e o desenvolvimento de um *framework* para representação da informação compreensível por máquinas na *Web* (FEITOSA, 2006, p. 64).

O nome ‘*Web Semântica*’ foi escolhido pelo W3C para este projeto da *Web* 3.0, uma vez que inicialmente se pretendia, de acordo com Souza e Alvarenga (2004, p. 2), “embutir inteligência e contexto nos códigos XML utilizados para confecção de páginas *Web*, de modo a melhorar a forma com que programas podem interagir com estas páginas e possibilitar um uso mais intuitivo por parte dos usuários”.

Com a *Web Semântica* passou-se a estimular a publicação de conteúdo junto com os metadados que os descrevem, de uma forma compreensível por máquina. Para Po et al. (2020), essas descrições legíveis por máquina permitem aos gerenciadores adicionarem significado ao conteúdo, de forma que uma máquina pode processar o próprio conhecimento, com o intuito de obter resultados mais significativos e ajudar os computadores a realizarem a coleta e pesquisa automatizada de informações. Com isso, torna-se evidente o potencial da *Web Semântica*, também conhecida como ‘*Web inteligente*’, que é capaz de gerar avanços em diversos campos científicos, como a busca e recuperação da informação (CONEGLIAN et al., 2017). Nesse sentido, Ramalho, Vidotti e Fujita (2008, p. 2) observam que:

[...] comparando com as abordagens tradicionalmente desenvolvidas, o projeto *Web Semântica* constitui-se como uma tentativa inversa de solução que tem como objetivo desenvolver meios para que as máquinas possam servir aos humanos de maneira mais eficiente, mas para isso torna-se necessário construir instrumentos que forneçam sentido lógico e semântico aos computadores.

De acordo com Berners-Lee (2006), a *Web Semântica* não serve apenas para colocar dados na *Web*, mas sim criar *links* entre eles, permitindo que pessoas e máquinas possam explorar uma *Web* de Dados. Diante disso, para Berners-Lee, Hendler e Lassila (2001, não paginado, tradução nossa):

A *Web Semântica* trará estrutura para o conteúdo significativo das páginas da *Web*, criando um ambiente onde os agentes de software ao se deslocarem de página em página podem prontamente realizar tarefas sofisticadas para os usuários. [...] A *Web Semântica* não é uma *Web* separada, mas uma extensão da atual, onde a informação possui significados bem definidos, possibilitando que computadores e pessoas trabalhem em cooperação.

Assim, um dos principais benefícios proporcionados pela *Web Semântica* é a possibilidade de interligar dados de diferentes fontes, não importando o formato dos dados ou a plataforma de *software* e de sistema operacional, dentre outros aspectos técnicos. Para isso, de acordo com Ankolekar (2008), a *Web Semântica* é construída sobre uma infraestrutura descentralizada e aberta, o que facilita a distribuição, o reuso e a interoperabilidade de dados.

Tais benefícios representam, para Marcondes (2012, p. 176), “[...] uma evolução da *Web*, em direção a mais amplas possibilidades de interação e de recuperação semântica da informação”. Neste sentido, de acordo com Gill (2016, não paginado, tradução nossa), a visão de Berners-Lee é que “[...] a *Web* evolua para uma rede contínua de dados interoperáveis, significativamente interligados, que possam ser compartilhados e reutilizados através dos limites do software, da empresa e da comunidade.”.

Ademais, de acordo com Souza e Alvarenga (2004, p. 138):

Podemos esperar que a *Web* tenha grande melhoria dos índices de revocação e precisão no atendimento das necessidades de informação, porque a semântica embutida nos documentos permitirá aos dispositivos de recuperação evitar os problemas comuns de polissemia e sinonímia, além de considerar as informações em seus contextos de significado.

Isso quer dizer que, na *Web Semântica*, os significados dos termos são levados em consideração na formulação das consultas, o que possibilita a procura de informações considerando aspectos semânticos dos dados, com o intuito de obter resultados capazes de satisfazer as necessidades de informação do usuário (PABÓN; GONZÁLEZ, 2014).

4.2 TECNOLOGIAS E FERRAMENTAS DA WEB SEMÂNTICA

A implementação e consolidação da *Web Semântica* se dão por meio do contínuo desenvolvimento de padrões tecnológicos e ferramentas, de forma a sustentar uma infraestrutura aberta e distribuída, com o propósito de possibilitar o compartilhamento, a interoperabilidade e o reuso de dados, tanto por pessoas quanto por agentes de *software*.

Para isso, de acordo com Souza e Alvarenga (2004, p. 134), “[...] é necessária uma padronização de tecnologias, de linguagens e de metadados descritivos, de forma que todos os usuários da *Web* obedeçam a determinadas regras comuns e compartilhadas sobre como armazenar dados e descrever a informação armazenada”.

Nas subseções a seguir são apresentadas algumas das principais tecnologias e ferramentas que sustentam a infraestrutura da *Web Semântica*.

4.2.1 Identificador Uniforme de Recursos (URI - *Uniform Resource Identification*)

A infraestrutura da *Web* atual suporta uma rede distribuída de páginas que podem se referir umas às outras por meio de ligações globais chamadas de ‘Localizadores de Recursos Uniformes (URL – *Uniform Resource Locator*)’. Nesse sentido, uma ideia da *Web Semântica* é suportar uma *Web* distribuída no nível dos dados, ou seja, em vez de ter uma página apontando para outra, um dado aponta para outro usando referências globais chamadas ‘Identificadores de Recursos Uniformes (URI – *Uniform Resource Identifier*)’. Breitman et al. (2007, p. 59, tradução nossa) definem um URI como “[...] uma cadeia de caracteres que identifica um recurso abstrato ou físico na *Web*”.

O uso de URIs sob o protocolo HTTP para identificar recursos na *Web* é recomendado, uma vez que eles são considerados uma maneira simples de criar nomes exclusivos de maneira descentralizada na *Web*. Além de fornecer um nome, os URIs também podem servir como meio de acesso à informação que descreve o recurso identificado (HEATH; BIZER, 2011).

Além disso, as URIs HTTP devem ser desreferenciáveis, ou seja, os usuários na *Web* podem pesquisar o URI usando o protocolo HTTP e recuperar uma descrição do recurso identificado pelo URI na forma de um documento da *Web*. As descrições legíveis por humanos geralmente são representadas como HTML, e as descrições legíveis por máquina são representadas como dados RDF, conforme será visto adiante (HEATH; BIZER, 2011).

O Quadro 16 apresenta exemplos de URIs para identificar um recurso (Amazônia Legal) e o respectivo documento (legível por humano ou legível por máquina) que descreve esse recurso.

Quadro 16 – Exemplos de URIs para identificação e descrição de um recurso.

http://terrabrasilis.dpi.inpe.br/resource/Amazônia_Legal (URI que identifica a região da Amazônia Legal)
http://terrabrasilis.dpi.inpe.br/resource/Amazônia_Legal.rdf (URI identificando o documento RDF/XML que descreve a região da Amazônia Legal)
http://terrabrasilis.dpi.inpe.br/resource/Amazônia_Legal.html (URI identificando o documento HTML que descreve a região da Amazônia Legal)

Fonte: Elaborado pelo autor.

Ademais, um URI deve fornecer uma identificação global para um recurso que seja comum em toda a *Web*, de forma que se dois agentes na *Web* quiserem se referir ao mesmo recurso, a prática recomendada é que seja definido, em consenso, um URI comum para o recurso. Isso vale não somente para a *Web Semântica*, mas para a *Web* em geral (ALLEMANG; HENDLER, 2008). Ao utilizar URIs para identificar recursos na *Web*, é desejável que estes sejam definidos e designados de forma clara e inequívoca, e que sejam estáveis e persistentes (REID et al., 2012).

4.2.1.1 Namespaces

Um *namespace* XML, ou simplesmente *namespace*, é um conjunto de conceitos, identificado por um URI, com um prefixo associado. Uma definição, no campo da Ciência da Informação, para o conceito de *namespace* é dada por Souza e Alvarenga (2004, p. 136):

Um *namespace* (NS) define um vocabulário controlado que identifica um conjunto de conceitos de forma única para que não haja ambiguidade na sua interpretação. Os *namespaces* XML são conjuntos de tipos de elementos e atributos possíveis para cada tipo. [...] Os *namespaces* podem ser referenciados por meio de uma URI, que se constitui em um repositório compartilhado, e não-ambíguo, onde usuários e programas de validação de código XML podem consultar a sintaxe e propriedades semânticas dos conceitos cobertos.

Para declarar um *namespace* em uma página XML deve ser usada uma família de atributos reservados, cujo nome deve ser ‘*xmlns*’ ou ter ‘*xmlns:*’ como prefixo, e cujo valor é um URI que identifica o *namespace*. A seguir, o Quadro 17 apresenta um fragmento de código que mostra como são feitas as declarações de *namespace* em uma página *Web*.

Quadro 17 – Declaração de *namespaces* em uma página Web.

```

1. <rdf:RDF
2.     xmlns="http://www.cat.com/docs"
3.     xml:base="http://www.cat.com/docs"
4.     xmlns:cs="http://www.cat.com/schema/"
5.     xmlns:dc="http://purl.org/dc/elements/1.1/"
6.     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
7.     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
8.     xmlns:owl="http://www.w3.org/2002/07/owl#"
9.     xmlns:xsd="http://www.w3.org/2001/XMLSchema#">

```

Fonte: Adaptado de Breitman et al. (2007)

No Quadro 17, a linha 2 indica que o *namespace* padrão do conceito apresentado é o URI *http://www.cat.com/docs*. Todos os termos não qualificados utilizados no documento XML que definem o conceito pertencerão a este *namespace*. A linha 3 identifica o URI base para o conceito, a linha 4 atribui o prefixo ‘cs’ ao URI ‘*http://www.cat.com/schema/*’, e a linha 5 atribui o prefixo ‘dc’ ao URI ‘*http://purl.org/dc/elements/1.1/*’, que identifica o *namespace* do esquema de metadados do padrão *Dublin Core*, a partir do qual serão importados outros conceitos. Por fim, as linhas 6 a 9 definem os prefixos para os *namespaces* dos modelos RDF, RDF Schema, OWL, e XML Schema, respectivamente (BREITMAN et al., 2007).

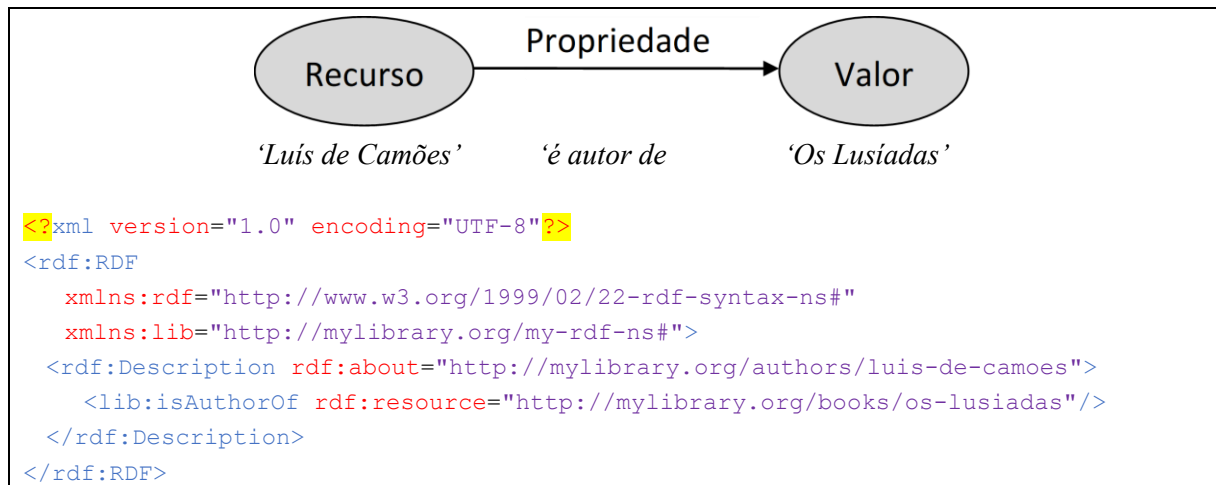
4.2.2 Modelo de dados RDF (*Resource Description Framework*)

Para possibilitar que máquinas e agentes de *software* processem conteúdo da *Web*, é necessário o compartilhamento de uma estrutura de dados comum, o que foi possibilitado pelo surgimento do padrão *Resource Description Framework* (RDF).

O padrão RDF foi desenvolvido pelo W3C com o objetivo de codificar descrições de recursos em formato legível por máquina, para que computadores possam compartilhar e processar informações de forma significativa e útil. Para atingir tal objetivo, o modelo de dados RDF representa as informações como triplas na forma de grafos, compostos por arcos e nós, o que possibilita a representação integrada de informações provenientes de múltiplas fontes. Além disso, o modelo de dados é estruturado de forma heterogênea e pode ser representado por diferentes esquemas (HEATH; BIZER, 2011).

As triplas RDF são representadas tal como exibido na Figura 16, em que um recurso (sujeito) é representado por um nó, um valor (objeto) por outro nó, e uma propriedade (predicado) como um arco rotulado. Dessa forma, uma tripla pode ser lida como uma sentença composta por sujeito, predicado e objeto.

Figura 16 – Representação de uma tripla RDF na forma de grafo e seu código XML.



Fonte: Elaborado pelo autor.

Em suma, o RDF fornece uma maneira flexível de descrever as coisas no mundo - como pessoas, locais ou conceitos abstratos - e como eles se relacionam com outras coisas. Estas declarações de relações entre as coisas são, em essência, vínculos que conectam as coisas no mundo (GILL, 2016). Para Heath e Bizer (2011), o RDF fornece um modelo de dados extremamente simples, mas estritamente adaptado à arquitetura *Web*, podendo ser empregado como uma ‘língua franca’, capaz de moderar entre outros modelos de dados usados na *Web*.

4.2.3 Sintaxes para serialização de dados RDF

Para que seja possível publicar dados na forma de grafos RDF na *Web*, é necessário serializá-los usando uma sintaxe RDF, isto é, pegar as triplas que compõem um grafo RDF e gravá-las em um arquivo usando uma sintaxe específica. Neste sentido, diversas sintaxes para serialização de dados RDF se tornaram ‘recomendação W3C’ nos últimos anos (AUER et al., 2011). A seguir são apresentadas as características de algumas dessas sintaxes.

A sintaxe RDF/XML⁷ é amplamente utilizada para publicar dados na *Web*. Por ser uma sintaxe XML para as triplas RDF, é vista como de difícil leitura e escrita por humanos. Assim, deve-se considerar o uso de outras serializações caso humanos sejam os consumidores finais dos dados a serem publicados (HEATH; BIZER, 2011).

Outra sintaxe ‘recomendação W3C’ é a RDFa⁸, que embute triplas RDF em páginas escritas em *eXtensible HyperText Markup Language* (XHTML). No entanto, como o RDFa está

⁷ <http://www.w3.org/TR/rdf-syntax-grammar/>

⁸ <http://www.w3.org/TR/xhtml-rdfa-primer/>

incorporado em XHTML, a sobrecarga é maior em comparação com outras sintaxes de serialização, além de ser pouco legível para consumo por humanos (AUER et al., 2011).

Já a *Turtle*⁹ é uma sintaxe de texto simples para serializar dados RDF. Por oferecer suporte para prefixos de *namespace* e outras abreviações, a *Turtle* é normalmente a sintaxe de serialização escolhida para leitura e escrita humana de triplas RDF (HEATH; BIZER, 2011).

A *JSON-LD*¹⁰ é uma das mais recentes sintaxes de serialização, sendo bastante intuitiva para programadores já familiarizados com o formato JSON (ISOTANI; BITTENCOURT, 2015). Diversas linguagens de programação fornecem suporte nativo ao formato JSON, dentre as quais *JavaScript* e PHP. Dessa forma, os desenvolvedores da *Web* podem manipular triplas RDF sem a necessidade de instalar bibliotecas de *software* adicionais (HEATH; BIZER, 2011).

Por fim, a *N-Triples*¹¹ é considerada a mais simples e intuitiva dentre as sintaxes apresentadas. Como o modelo RDF é estruturado em triplas na forma ‘recurso-propriedade-valor’, a sintaxe *N-Triples* simplesmente estrutura o código dessa forma, ou seja, cada linha dessa serialização equivale a uma tripla RDF (ISOTANI; BITTENCOURT, 2015).

Para exemplificar a serialização de dados RDF, o Quadro 18 mostra algumas triplas contendo afirmações sobre a cidade paulista de Marília, serializadas nas sintaxes até aqui apresentadas.

Quadro 18 – Exemplos de serialização de dados RDF em diferentes sintaxes.

RDF/XML

```
<rdf:Description rdf:about="http://dbpedia.org/resource/Marília">
  <rdf:type rdf:resource="http://www.wikidata.org/entity/Q486972" />
  <rdfs:label xml:lang="pt">Marília</rdfs:label>
  <rdfs:comment xml:lang="pt">Marília é um município brasileiro do estado de
    São Paulo. Situa-se na região Centro-Oeste Paulista. Fica distante da
    capital do estado 443 quilômetros por rodovia; 529 quilômetros por ferrovia
    e 376 quilômetros em linha reta.</rdfs:comment>
  <dbp:leaderName xml:lang="en">Daniel Alonso</dbp:leaderName>
  <geo:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-22.2139
  </geo:lat>
  <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-49.9458
  </geo:long>
  <foaf:homepage rdf:resource="http://www.marilia.sp.gov.br" />
  <dbo:areaTotal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
    1.17e+09</dbo:areaTotal>
  <dbo:elevation rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
    660.0</dbo:elevation>
  <dbo:populationTotal rdf:datatype="http://www.w3.org/2001/XMLSchema#
    nonNegativeInteger">240590</dbo:populationTotal>
</rdf:Description>
```

⁹ <http://www.w3.org/TR/turtle/>

¹⁰ <http://www.w3.org/TR/json-ld/>

¹¹ <http://www.w3.org/TR/n-triples/>

RDFa

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
    "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html version="XHTML+RDFa 1.0" xml:lang="en"
    xmlns="http://www.w3.org/1999/xhtml"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:dbo="http://dbpedia.org/ontology/"
    xmlns:dbp="http://dbpedia.org/property/"
    xmlns:dbt="http://dbpedia.org/resource/"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
  <head>
    <meta charset="utf-8" />
    <title>HTML Microdata document</title>
  </head>
  <body><dl itemscope itemid="http://dbpedia.org/resource/Marília">
    <dt>Subject Item</dt><dd>dbt:Marília</dd>
    <dt>rdf:type</dt><dd>
      <a itemprop="http://www.w3.org/1999/02/22-rdf-syntax-ns#type"
        href="http://www.wikidata.org/entity/Q486972">wikidata:Q486972</a></dd>
    <dt>rdfs:label</dt><dd>
      <span itemprop="http://www.w3.org/2000/01/rdf-schema#label" xml:lang="pt">
        Marília</span></dd>
    <dt>dbo:abstract</dt><dd>
      <span itemprop="http://dbpedia.org/ontology/abstract" xml:lang="pt">
        Marília é um município brasileiro do estado de São Paulo. Situa-se na
        região Centro-Oeste Paulista. Fica distante da capital do estado 443
        quilômetros por rodovia; 529 quilômetros por ferrovia e 376 quilômetros
        em linha reta.</span></dd>
    <dt>dbp:leaderName</dt><dd>
      <span itemprop="http://dbpedia.org/property/leaderName"
        xml:lang="en">Daniel Alonso</span></dd>
    <dt>geo:lat</dt><dd>
      <span itemprop="http://www.w3.org/2003/01/geo/wgs84_pos#lat">-22.2139
        </span></dd>
    <dt>geo:long</dt><dd>
      <span itemprop="http://www.w3.org/2003/01/geo/wgs84_pos#long">-49.9458
        </span></dd>
    <dt>dbo:areaTotal</dt><dd>
      <span itemprop="http://dbpedia.org/ontology/areaTotal">1.17e+09</span></dd>
    </dd><dt>foaf:homepage</dt><dd>
      <a itemprop="http://xmlns.com/foaf/0.1/homepage"
        href="http://www.marilia.sp.gov.br">n9:www.marilia.sp.gov.br</a></dd>
    <dt>dbo:elevation</dt><dd>
      <span itemprop="http://dbpedia.org/ontology/elevation">660.0</span></dd>
    <dt>dbo:populationTotal</dt><dd>
      <span itemprop="http://dbpedia.org/ontology/populationTotal">240590
        </span></dd></dl>
  </body>
</html>

```

N-Triples

```

<http://dbpedia.org/resource/Marília> <http://www.w3.org/1999/02/22-rdf-syntax-
ns#type> <http://www.wikidata.org/entity/Q486972> .
<http://dbpedia.org/resource/Marília> <http://www.w3.org/2000/01/rdf-schema#
label> "Marília"@pt .
<http://dbpedia.org/resource/Marília> <http://www.w3.org/2000/01/rdf-schema#
comment> "Marília é um município brasileiro do estado de São Paulo. Situa-
se na região Centro-Oeste Paulista. Fica distante da capital do estado 443
quilômetros por rodovia; 529 quilômetros por ferrovia e 376 quilômetros em
linha reta."@pt .
<http://dbpedia.org/resource/Marília> <http://dbpedia.org/property/leaderName>
"Daniel Alonso"@en .
<http://dbpedia.org/resource/Marília> <http://www.w3.org/2003/01/geo/wgs84_pos#
lat> "-22.2139"^^<http://www.w3.org/2001/XMLSchema#float> .

```

```

<http://dbpedia.org/resource/Marília> <http://www.w3.org/2003/01/geo/wgs84_pos#
long> "-49.9458"^^<http://www.w3.org/2001/XMLSchema#float> .
<http://dbpedia.org/resource/Marília> <http://xmlns.com/foaf/0.1/homepage>
<http://www.marilia.sp.gov.br> .
<http://dbpedia.org/resource/Marília> <http://dbpedia.org/ontology/areaTotal>
"1.17e+09"^^<http://www.w3.org/2001/XMLSchema#double> .
<http://dbpedia.org/resource/Marília> <http://dbpedia.org/ontology/elevation>
"660.0"^^<http://www.w3.org/2001/XMLSchema#double> .
<http://dbpedia.org/resource/Marília> <http://dbpedia.org/property/population
Total> "240590"^^<http://www.w3.org/2001/XMLSchema#integer> .

```

Turtle

```

@prefix wikidata: <http://www.wikidata.org/entity/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix dbp: <http://dbpedia.org/property/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
<http://dbpedia.org/resource/Marília>
rdf:type      wikidata:Q486972 ;
rdfs:label    "Marília"@pt ;
rdfs:comment  "Marília é um município brasileiro do estado de São
Paulo. Situa-se na região Centro-Oeste Paulista. Fica distante da
capital do estado 443 quilômetros por rodovia; 529 quilômetros por
ferrovia e 376 quilômetros em linha reta."@pt ;
geo:lat      "-22.213888168334960938"^^xsd:float ;
geo:long     "-49.945835113525390625"^^xsd:float ;
foaf:homepage <http://www.marilia.sp.gov.br> ;
dbp:leaderName "Daniel Alonso"@en ;
dbo:areaTotal 1170000000.0 ;
dbo:elevation 660.0 ;
dbo:populationTotal "240590"^^xsd:nonNegativeInteger .

```

RDF/JSON

```

{
  "http://dbpedia.org/resource/Marília" : {
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" : [ { "type" : "uri",
"value" : "http://www.wikidata.org/entity/Q486972" } ],
    "http://www.w3.org/2000/01/rdf-schema#label" : [ { "type" : "literal",
"value" : "Marília" , "lang" : "pt" } ],
    "http://www.w3.org/2000/01/rdf-schema#comment" : [ { "type" : "literal",
"value" : " Marília é um município brasileiro do estado de São Paulo. Situa-
se na região Centro-Oeste Paulista. Fica distante da capital do estado 443
quilômetros por rodovia; 529 quilômetros por ferrovia e 376 quilômetros em
linha reta." , "lang" : "pt" } ],
    "http://xmlns.com/foaf/0.1/homepage" : [ { "type" : "uri", "value" :
"http://www.marilia.sp.gov.br" } ] ,
    "http://www.w3.org/2003/01/geo/wgs84_pos#lat" : [ { "type" : "literal",
"value" : -22.21388816833496 , "datatype" :
"http://www.w3.org/2001/XMLSchema#float" } ] ,
    "http://www.w3.org/2003/01/geo/wgs84_pos#long" : [ { "type" : "literal",
"value" : -49.94583511352539 , "datatype" :
"http://www.w3.org/2001/XMLSchema#float" } ] ,
    "http://dbpedia.org/property/leaderName" : [ { "type" : "literal", "value" :
"Daniel Alonso" , "lang" : "en" } ] ,
    "http://dbpedia.org/ontology/areaTotal" : [ { "type" : "literal", "value" :
1170000000 , "datatype" : "http://www.w3.org/2001/XMLSchema#double" } ] ,
    "http://dbpedia.org/ontology/elevation" : [ { "type" : "literal", "value" :
660 , "datatype" : "http://www.w3.org/2001/XMLSchema#double" } ] ,
    "http://dbpedia.org/ontology/populationTotal" : [ { "type" : "literal",
"value" : "240590" , "datatype" :
"http://www.w3.org/2001/XMLSchema#nonNegativeInteger" } ]
  }
}

```

No Quadro 18 é possível perceber as diferenças entre as sintaxes de serialização de dados no modelo RDF, principalmente no que se refere à legibilidade por humanos, para a qual a sintaxe *N-Triples* é a que mais se aproxima de uma representação direta de uma tripla em seu formato nativo ‘recurso-propriedade-valor’.

4.2.4 Armazenamento e consulta de dados no modelo RDF

Da mesma forma que os dados estruturados podem ser armazenados em Sistemas Gerenciadores de Bancos de Dados (SGBDs) e consultados por meio da linguagem *Structured Query Language* (SQL), os dados modelados como RDF podem ser armazenados em *triplestores* (bancos de triplas) e consultados por meio do *Sparql Protocol And Rdf Query Language* (SPARQL), como será visto a seguir.

4.2.4.1 Triplestores

Os *triplestores* são repositórios semânticos que funcionam como SGBDs que podem interpretar os dados neles armazenados utilizando motores de inferência, de forma a apoiar a manipulação eficiente de dados da *Web Semântica*. Com base na semântica dos esquemas, esses repositórios podem inferir fatos implícitos e considerá-los no processo de avaliação da consulta. Além disso, os *triplestores* assumem o papel de servidores *Web*, uma vez que fornecem acesso à *Web* de dados (KIRYAKOV; DAMOVA, 2011).

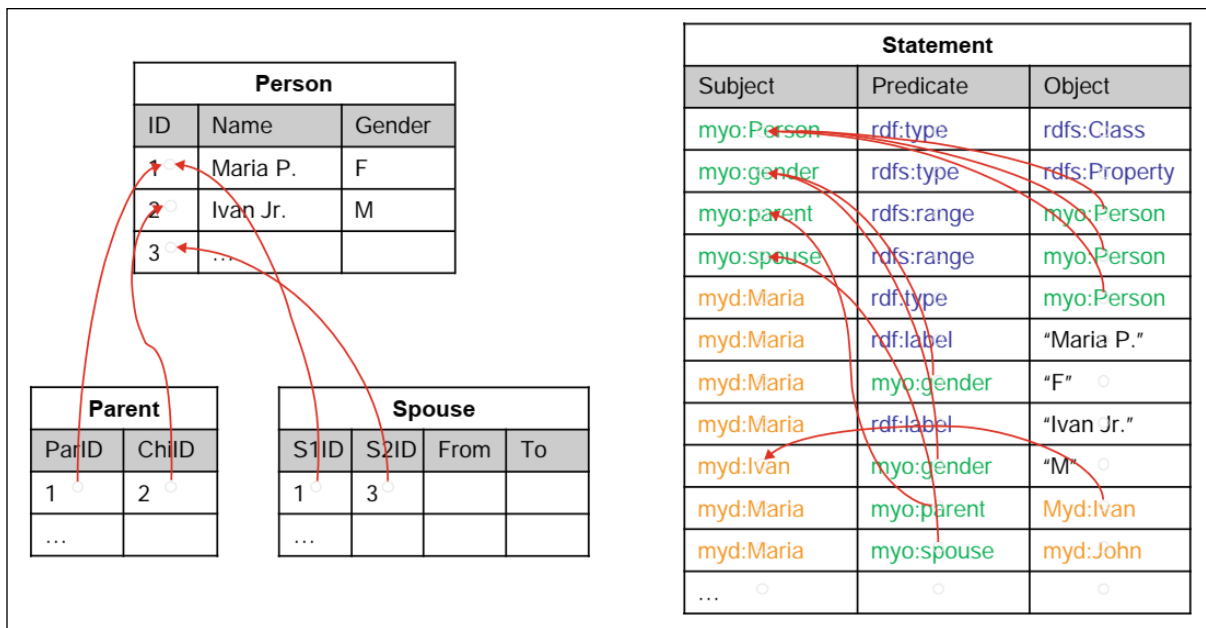
Algumas das principais características dos repositórios semânticos, de acordo com Kiryakov e Damova (2011) são: fácil integração de múltiplas fontes de dados, desde que os esquemas destas fontes estejam semanticamente alinhados; fácil consulta a esquemas de dados enriquecidos semanticamente, independentemente do vocabulário e dos padrões de modelagem usados para codificar os dados; grande poder analítico, uma vez que a semântica será aplicada de forma completa mesmo quando isso exigir inferências recursivas em múltiplas etapas; e, interoperabilidade eficiente dos dados, uma vez que a importação de dados RDF de uma *triplestore* para outra é simples, com base no uso de URIs.

Tais características e vantagens podem oferecer, na visão de Kiriakov e Damova (2011, p. 237, tradução nossa),

[...] uma nova gama de cenários de acesso à informação, onde a pessoa que formula a consulta não está ciente dos detalhes dos esquemas que foram usados para a codificação dos dados. Isto só é possível em um SGBD onde o motor compreende (até certo ponto) a semântica dos dados e a semântica da consulta e pode interpretá-los a fim de encontrar correspondências apesar das variações de sintaxe.

Quanto à arquitetura, os *triplestores* utilizam um modelo físico genérico, no qual os dados são representados internamente em estruturas de dados na forma de grafo, onde cada tripla é armazenada em uma linha composta por três colunas, sendo uma para o recurso, uma para a propriedade e uma para o valor. A Figura 17 ilustra essa estrutura, comparando-a com a de um SGBD relacional.

Figura 17 – Comparação entre as estruturas de um *triplestore* e de um SGBD relacional.



Fonte: Kiryakov e Damova (2011).

Na Figura 17 é apresentada uma comparação entre a representação dos dados em um SGBD relacional (à esquerda) e em um *triplestore* (à direita). Nesse tipo de estrutura, uma mudança no esquema (a definição de uma nova propriedade, por exemplo) não requer nenhuma mudança na representação dos fatos afirmados no *triplestore*, uma vez que não impõe uma mudança estrutural na representação física. Em contraste, no SGBD relacional, a representação física do esquema de dados é dependente do esquema (KIRYAKOV; DAMOVA, 2011).

De acordo com Laufer (2015), os *triplestores* mais utilizados em aplicações da Web Semântica são o *Virtuoso*¹² (banco de dados relacional com inúmeras adaptações orientadas a RDF), o *Sesame*¹³ (um *framework Java* com um banco de triplas nativo), o *Jena TDB*¹⁴ (componente do *framework Apache Jena* para armazenamento e consulta de dados RDF), o

¹² <http://vos.openlinksw.com/owiki/wiki/VOS>

¹³ <http://www.openrdf.org/>

¹⁴ <https://jena.apache.org/documentation/tdb/>

*GraphDB*¹⁵ (antigo *OWLIN*, é um dos poucos que podem realizar inferências semânticas em escala), o *4store*¹⁶ (tem como pontos fortes o desempenho, a escalabilidade e a estabilidade, porém não oferece mecanismos de inferências), e o *Blazegraph*¹⁷ (escrito em *Java*, tem alta performance e escalabilidade, porém oferece um conjunto restrito de inferências).

Para efetuar consultas a dados RDF armazenados em *triplestores*, deve ser utilizada uma linguagem de consulta semântica, sendo a *Sparql Protocol and RDF Query Language* (SPARQL¹⁸) um dos padrões do W3C para a *Web Semântica*, que será apresentada a seguir.

4.2.4.2 Linguagem de consulta semântica SPARQL

A SPARQL é uma linguagem para consulta de grafos RDF por meio da combinação de padrões em triplas. Apesar de possuir semelhança sintática com a linguagem SQL, é uma linguagem de consulta muito mais poderosa do que essa, uma vez que foi projetada para a *Web* aberta, descentralizada e fluida (VALLE; CERI, 2011). Além disso, a SPARQL é também um protocolo para o envio de consultas e recebimento de resultados de pontos de acesso (*endpoints*) remotos (AUER et al., 2011).

Todavia, a linguagem SPARQL não contém cláusulas da *Data Definition Language* (DDL), uma vez que os esquemas são representados como grafos RDF padrão, o que não exige linguagem específica para lidar com eles. Além disso, a SPARQL não é uma *Data Modification Language* (DML), uma vez que não é possível inserir, modificar ou excluir grafos RDF por meio dela (VALLE; CERI, 2011).

Por conta disso, a linguagem SPARQL suporta apenas quatro tipos de consultas: *SELECT* (retorna ‘n’ tuplas de resultados, tal como em SQL), *DESCRIBE* (retorna um grafo RDF que descreve os recursos que combinam com as restrições da consulta), *ASK* (retorna uma resposta positiva ou negativa, que indica se um padrão de consulta pode ser satisfeito ou não) e *CONSTRUCT* (retorna um grafo RDF construído a partir de um resultado de pesquisa).

De acordo com Laufer (2015), os padrões de triplas da cláusula *SELECT* têm a mesma forma de triplas normais, exceto que qualquer uma de suas partes (recurso, propriedade ou valor) pode ser substituída por uma variável (que começa com um caractere ‘?’). Além disso,

¹⁵ <https://graphdb.ontotext.com/documentation/free/about-graphdb.html>

¹⁶ <https://github.com/4store/4store>

¹⁷ <http://blazegraph.com>

¹⁸ <http://www.w3.org/TR/rdf-sparql-protocol/>

essa cláusula aceita que seja especificado o grafo RDF a ser consultado (em geral um URI), por meio do termo FROM. O Quadro 19 apresenta um exemplo de consulta SPARQL.

Quadro 19 – Exemplo de consulta SPARQL.

- a. Considere o seguinte conjunto de triplas RDF serializadas em *Turtle*:

```
base <http://example.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

<#gabriela-cravo-canela>
  dc:identifier "9788535912388" ;
  foaf:name "Gabriela, Cravo e Canela" ;
  dc:creator "Jorge Amado" ;
  dc:publisher <#companhia-das-letras> .

<#vidas-secas>
  dc:identifier "9788501067340" ;
  foaf:name "Vidas Secas" ;
  dc:creator "Graciliano Ramos" ;
  dc:publisher <#grupo-record> .

<#antologia-poetica>
  dc:identifier "9788535921199" ;
  foaf:name "Antologia Poética" ;
  dc:creator "Carlos Drummond de Andrade" ;
  dc:publisher <#companhia-das-letras> .

<#companhia-das-letras>
  foaf:name "Companhia das Letras".

<#grupo-record>
  foaf:name "Grupo Editorial Record" .
```

- b. Ao realizar a seguinte consulta em linguagem SPARQL sobre este conjunto de triplas:

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX ex: <http://example.org/>
SELECT ?name ?creator
WHERE
{
  ?book dc:publisher ex:companhia-das-letras .
  ?book foaf:name ?name .
  ?book dc:creator ?creator .
}
```

- c. É retornado como resultado:

name	creator
“Gabriela, Cravo e Canela”	“Jorge Amado”
“Antologia Poética”	“Carlos Drummond de Andrade”

Fonte: Adaptado de Laufer (2015).

No Quadro 19, a consulta apresentada em (b) primeiramente define os *namespaces* por meio da cláusula *PREFIX*. Em seguida, solicita a exibição em (c) do nome da obra (*?name*) e do nome do autor (*?creator*) para todos os livros publicados pela editora ‘Companhia das Letras’ (*?book dc:publisher ex:companhia-das-letras*) encontrados nas triplas em (a). No entanto, as consultas SPARQL não retornam conjuntos de triplas, e sim conjuntos de tuplas (como linhas de uma tabela), conforme apresentado no passo (c) da figura (LAUFER, 2015).

Recentemente foi padronizada uma extensão da linguagem SPARQL para consulta de dados geoespaciais semânticos, a GeoSPARQL, composta de um vocabulário leve para representar e consultar dados geoespaciais. (HUANG et al., 2019). A seguir são apresentados alguns *triplestores* compatíveis com essa extensão.

4.2.4.3 *Triplestores* com suporte a consultas geoespaciais

No campo da geoinformação, para o armazenamento de dados geoespaciais modelados como RDF é necessário que o *triplestore* a ser utilizado possua capacidade para manipular tais tipos de dados e, além disso, ofereça um ponto de acesso para consultas geoespaciais, o que pode ser alcançado por meio da conformidade com a extensão GeoSPARQL¹⁹. O Quadro 20 apresenta alguns *triplestores* que oferecem compatibilidade com essa extensão.

Quadro 20 – *Triplestores* compatíveis com a extensão GeoSPARQL.

	RDF4J	GeoSPARQL -Jena	Virtuoso	Stardog	GraphDB
Solução de armazenamento de triplas RDF	Nativa	Nativa	Baseada em SGBD	Nativa	Nativa
Sintaxe de serialização de geometria	WKT	WKT, GML	WKT	WKT	WKT, GML
Conformidade com GeoSPARQL	Total	Total	Total	Parcial	Total
Sistema de Referência Espacial (SRS) suportado	WGS84	DIVERSOS	WGS84	WGS84	WGS84

Fonte: Adaptado de Huang et al., (2019)

No Quadro 20, as soluções de armazenamento adotadas pelos *triplestores* podem ser de dois tipos: nativas (projetadas do zero), ou baseadas em um SGBD relacional existente. Apenas o *Virtuoso* se baseia nesse último. Quanto à sintaxe de serialização, todos os *triplestores* suportam operações espaciais para geometrias serializadas em *Well-Know Text* (WKT), sendo que *GraphDB* e *GeoSPARQL-Jena* também suportam serialização em formato GML. Quanto à conformidade com a extensão GeoSPARQL, todos apresentam conformidade total ou parcial. Por fim, quanto ao sistema de referência espacial (SRS – *Spatial Reference System*), todos oferecem suporte ao WGS84, que é o sistema de referência geodésico global. Como o

¹⁹ <http://www.opengeospatial.org/standards/geosparql>

GeoSPARQL-Jena foi desenvolvido muito recentemente, ele também suporta a transformação entre diferentes SRS (HUANG et al., 2019).

4.2.5 Vocabulários para descrição de dados em RDF

A integração de dados no modelo RDF requer uma ponte entre os esquemas que são usados por diferentes fontes para publicar seus dados. O termo ‘esquema’ deve ser aqui entendido como uma mistura de termos distintos, oriundos de diferentes vocabulários (em RDF), que são usados por uma fonte para publicar dados na *Web* (HEATH; BIZER, 2011).

Neste contexto, um vocabulário pode ser definido tecnicamente como um conjunto de URIs únicas, cada qual identificando, de forma não ambígua, um conceito abrangido por um *namespace*.

Outra definição é dada por Nacuoungue, Rozsa e Dutra (2018, p. 24):

Um vocabulário reúne as definições de um conjunto de classes e propriedades para descrever tipos específicos de conceitos em um determinado domínio. Os termos de vocabulários também fornecem os *links* em dados ligados. As definições de termos fornecidos pelos vocabulários contêm uma semântica que permite que os dados se tornem significativos.

No entanto, diferentes comunidades podem ter diferentes preferências sobre os vocabulários a serem utilizados para publicar dados na *Web* a respeito de um determinado recurso. Isso significa que a *Web* de dados está aberta à utilização, em paralelo, de vocabulários arbitrários. Mas, apesar dessa abertura geral, considera-se uma boa prática reutilizar termos de vocabulários RDF conhecidos e amplamente adotados, a fim de facilitar o processamento de dados interligados pelos aplicativos dos clientes. Somente se tais vocabulários não fornecerem os termos necessários para descrição de um recurso, é que se deve definir uma nova terminologia específica para a fonte de dados (HEATH; BIZER, 2011b).

Nas subseções a seguir são apresentados, de forma breve, alguns dos vocabulários RDF mais utilizados na *Web* de dados.

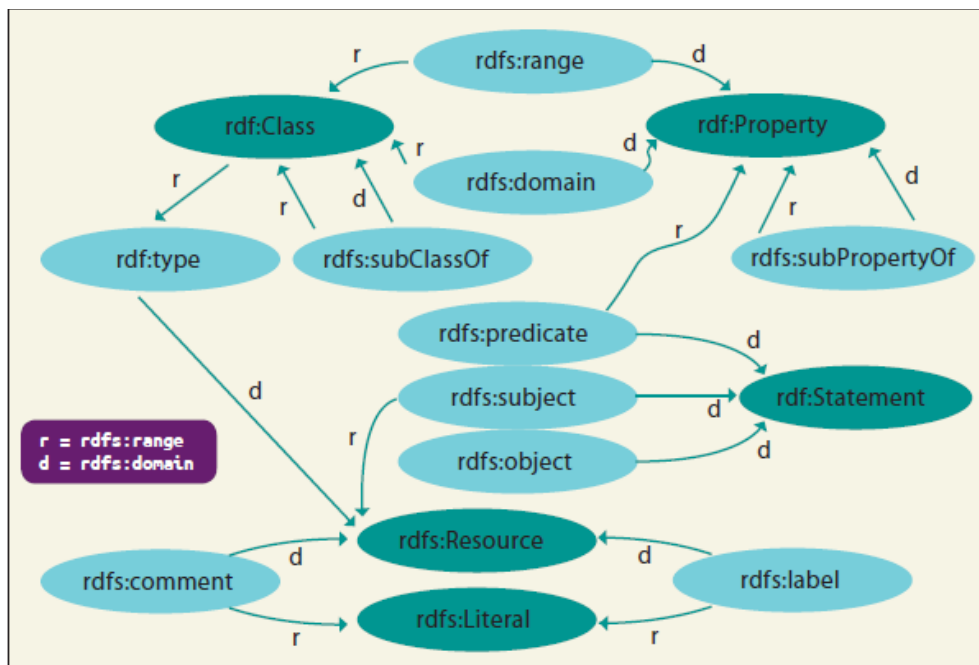
4.2.5.1 RDF Schema Vocabulary (RDFS)

O modelo de dados RDF não possui termos específicos para a descrição de classes e propriedades de coisas no mundo e seus relacionamentos por meio de triplas. Para essa função

foi criado um vocabulário, que amplia a expressividade do modelo RDF, chamado *RDF Schema Vocabulary* (RDFS²⁰).

O vocabulário RDFS serve para descrever modelos conceituais (ou taxonomias) em termos de classes e propriedades, bem como nos casos em que relações de subsunção entre termos devem ser representadas (por exemplo, todos os atletas ‘também são’ pessoas). Assim, quando combinado com um mecanismo de raciocínio, o RDFS permite que raciocínios implícitos sejam inferidos para os dados (HEATH; BIZER, 2011). A Figura 18 mostra um diagrama com os principais elementos deste vocabulário.

Figura 18 – Diagrama de elementos do vocabulário RDFS.



Fonte: Isotani e Bittencout (2015).

Conforme mostrado na Figura 18, os dois principais conceitos a serem observados são a ‘Classe’ (*rdf:Class*), utilizada para descrever recursos (*rdfs:Resource* e *rdfs:Literal*), e a ‘Propriedade’ (*rdf:Property*), utilizada para definir relações entre sujeito e objeto. Dentre os tipos de propriedade mais usados estão o *rdf:type* (indica que ‘é uma instância de’), *rdfs:domain* (indica o sujeito de uma relação) e *rdfs:range* (indica o objeto da relação), sendo os dois últimos representados no diagrama por meio dos arcos ‘d’ e ‘r’ respectivamente (ISOTANI; BITTENCOURT, 2015). Há ainda outros dois elementos que são utilizados para a anotação de

²⁰ <https://www.w3.org/TR/rdf-schema/>

recursos: o *rdfs:label*, utilizado para prover um nome, e o *rdfs:comment*, utilizado para prover uma descrição, ambos legíveis por humanos, para um recurso (HEATH; BIZER, 2011).

4.2.5.2 Simple Knowledge Organization System (SKOS)

O Sistema Simples de Organização do Conhecimento (em inglês, SKOS²¹) é uma recomendação do W3C, amplamente usada para representar taxonomias, tesauros, sistemas de cabeçalho de assunto e vocabulários controlados. Conforme afirmam Allemang e Hendler (2008), o SKOS tem como base a Web Semântica, o que o diferencia de outros padrões de tesauros. Essa característica, de acordo com Laufer (2015), permite que conceitos possam ser identificados por meio de URIs, e que “Os diversos conceitos podem ser relacionados semanticamente entre si, em hierarquias e redes de associação informais, e agrupados em esquemas conceituais” (LAUFER, 2015, p. 96).

O SKOS permite estruturar termos de acordo com a especificidade ou relação, agrupá-los em coleções, além de especificar um recipiente geral para termos e coleções. Permite ainda que os desenvolvedores de tesauros associem identificadores neutros de idioma com termos e os relacionem com múltiplas representações de texto em outros idiomas, permitindo assim consultas multilíngues (FUGAZZA, 2011). O Quadro 21 apresenta as descrições e exemplos dos principais elementos do SKOS.

Quadro 21 – Principais elementos do SKOS.

ELEMENTO	DEFINIÇÃO	EXEMPLO
<i>Concept</i>	Classe que define que um determinado recurso é um conceito	@prefix ex: <http://example.com> . ex:animals rdf:type skos:Concept.
<i>prefLabel</i>	Rótulo preferido a ser exibido. Um rótulo permite fazer referência a conceitos em linguagem natural.	ex:animals rdf:type skos:Concept ; skos:prefLabel "animals" ; skos:altLabel "creatures" .
<i>altLabel</i>	Rótulo alternativo, utilizado para indicar um sinônimo. Um rótulo permite fazer referência a conceitos em linguagem natural.	
<i>broader</i>	Ao fazer ligação com outros conceitos do vocabulário, indica que um conceito é mais abrangente que um outro.	ex:mammals rdf:type skos:Concept ; skos:prefLabel "mammals" ; skos:broader ex:animals .
<i>narrower</i>	Inverso de <i>broader</i> .	ex:animals rdf:type skos:Concept ; skos:prefLabel "animals" ; skos:narrower ex:mammals .

²¹ <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>

<i>related</i>	Indica uma relação associativa entre dois conceitos.	ex:ornithology rdf:type skos:Concept ; skos:prefLabel "ornithology" . ex:birds rdf:type skos:Concept ; skos:prefLabel "birds" ; skos:related ex:ornithology .
<i>note</i>	Indica uma observação a respeito do conceito. Variações: “ <i>scopeNote</i> “, “ <i>historyNote</i> “, “ <i>editorialNote</i> “ e “ <i>changeNote</i> ”	ex:microwaveFrequencies skos:scopeNote "Used for frequencies between 1GHz to 300Ghz"@en . ex:tomato skos:changeNote "Moved from 'fruits' to 'vegetables'" .
<i>definition</i>	Fornece uma definição do conceito.	ex:documentation rdf:type skos:Concept ; skos:definition "the process of storing and retrieving information in all fields of knowledge" .
<i>ConceptScheme</i>	Representa tesouros ou esquemas nos quais os conceitos podem ser criados e usados como entidades independentes.	ex:animalThesaurus rdf:type skos:ConceptScheme ; dct:title "Simple animal thesaurus" ; dct:creator ex:antoineIsaac . ex:mammals rdf:type skos:Concept ; skos:inScheme ex:animalThesaurus . ex:cows rdf:type skos:Concept; skos:broader ex:mammals ; skos:inScheme ex:animalThesaurus . ex:fish rdf:type skos:Concept ; skos:inScheme ex:animalThesaurus .

Fonte: Adaptado de Laufer (2015).

Dentre os primeiros adeptos do SKOS, como formato nativo ou alternativo para representar tesouros, podem ser mencionados o AGROVOC²², um recurso multilíngue com mais de 28.000 termos relacionados a todos os campos temáticos na agricultura, alimentação e domínios relacionados; e o GEMET²³, que fornece um conjunto mais restrito de termos (aproximadamente 5.800), porém cobre 31 idiomas diferentes (Fugazza, 2011).

4.2.5.3 Dublin Core Metadata Initiative (DCMI)

O vocabulário dos ‘Termos de Metadados da *Dublin Core Metadata Initiative* (DCMI²⁴)’ define atributos gerais de metadados, tais como título, criador, data e assunto

²² <https://agrovoc.fao.org/browse/agrovoc/en/>

²³ <https://www.eionet.europa.eu/gemet/en/about/>

²⁴ <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/2012-06-14/>

(HEATH; BIZER, 2011). Nele estão incluídos os quinze termos do *Dublin Core Metadata Element Set* (também conhecido como *Dublin Core - DC*), além de várias dezenas de propriedades, classes, tipos de dados e esquemas de codificação de vocabulário. O DC mais esses vocabulários de extensão são chamados coletivamente de *DCMI Metadata Terms* (ou *DCTerms*). Esses termos devem ser usados em combinação com termos de metadados de outros vocabulários compatíveis no contexto de perfis de aplicativos (DCMI, 2020).

Os termos de metadados DCMI são expressos em vocabulários RDF, sendo que cada termo é identificado com um URI para *namespaces* DCMI, geralmente declarados na forma de prefixos para tornar os dados, consultas e esquemas mais concisos e legíveis (DCMI, 2020).

De acordo com Laufer (2015, p. 93), a DCMI

[...] é uma das primeiras iniciativas a pensar na definição de um vocabulário para a descrição de metadados, tendo como premissas que as descrições tenham independência em relação à sintaxe e tenham uma semântica bem definida. É um vocabulário adequado para a descrição de recursos (documentos), e começou como um conjunto básico de 15 propriedades, que tinham uma ideia análoga a dos elementos de catalogação de bibliotecas.

A seguir, o Quadro 22 descreve os 15 termos do *Dublin Core Metadata Element Set*.

Quadro 22 – Termos do Dublin Core Metadata Element Set.

TERMO	DESCRIÇÃO
<i>title</i>	Nome do recurso
<i>creator</i>	Nome do criador do recurso
<i>subject</i>	Tópico do recurso
<i>description</i>	Descrição do recurso (resumo, sumário etc.)
<i>publisher</i>	Entidade responsável por tornar o recurso disponível
<i>contributor</i>	Nome dos contribuidores do recurso
<i>date</i>	Data associada ao recurso
<i>type</i>	Tipo do recurso
<i>format</i>	Formato de arquivo, meio físico de armazenamento ou dimensões do recurso
<i>identifier</i>	Referência única ao recurso dentro de um determinado contexto
<i>source</i>	Fonte que originou o recurso
<i>language</i>	Linguagem do recurso
<i>relation</i>	Relação entre dois recursos
<i>coverage</i>	Cobertura espacial ou temporal do recurso
<i>rights</i>	Direitos associados ao recurso

Fonte: Adaptado de Laufer (2015)

Para a criação do DCTerms foram acrescentadas aos quinze elementos apresentados no Quadro 22, propriedades como *abstract*, *audience*, *conformsTo*, *hasVersion*, *replaces*, *requires* etc., além de classes como *FileFormat*, *Jurisdiction*, *Location*, *ProvenanceStatement*, *MediaType*, entre outras (LAUFER, 2015).

4.2.5.4 Friend Of A Friend Vocabulary (FOAF)

O vocabulário *Friend of a Friend* (FOAF²⁵) define termos para descrever pessoas, suas atividades e suas relações com outras pessoas e objetos (HEATH; BIZER, 2011). Para esse propósito, o FOAF foi concebido sob o princípio de que ‘Qualquer pessoa pode dizer qualquer coisa sobre qualquer assunto’ (ALLEMANG; HENDLER, 2008).

O FOAF aproveita a natureza distribuída do RDF para fornecer uma representação distribuída da informação (ALLEMANG; HENDLER, 2008). O Quadro 23 apresenta as principais classes (primeira letra maiúscula) e propriedades (primeira letra minúscula) deste vocabulário.

Quadro 23 – Classes e propriedades do vocabulário FOAF.

ELEMENTO	DESCRIÇÃO
<i>Agent</i>	coisas que realizam algo, que podem ser pessoas, organizações, robôs etc. Tem como subclasses “ <i>Person</i> ”, “ <i>Organization</i> ” e “ <i>Group</i> ”.
<i>Person</i>	entidade central do vocabulário: representa as pessoas.
<i>name</i>	cadeia de caracteres com um nome.
<i>title</i>	forma de tratamento, como, por exemplo, “Sr.”, “Sra.”, etc.
<i>img</i>	uma imagem que representa uma pessoa.
<i>depiction</i>	relaciona alguma coisa a uma imagem.
<i>familyName</i>	descreve parte do nome de uma pessoa (sobrenome).
<i>givenName</i>	descreve parte do nome de uma pessoa (primeiro nome).
<i>knows</i>	relaciona duas pessoas.
<i>based_near</i>	relação espacial entre duas coisas.
<i>age</i>	a idade da pessoa.
<i>made</i>	alguma coisa feita por alguém.
<i>primaryTopic</i>	principal tópico de um documento.
<i>Project</i>	um projeto.
<i>Organization</i>	uma organização.

²⁵ <http://www.foaf-project.org/>

<i>Group</i>	um grupo.
<i>Member</i>	um membro de um grupo.
<i>Document</i>	um documento.
<i>Image</i>	uma imagem.

Fonte: Adaptado de Laufer (2015).

De acordo com Laufer (2015, p. 95), “Uma das ideias criadas com o vocabulário FOAF foi a de que cada pessoa poderia definir um arquivo, chamado arquivo FOAF, com suas informações pessoais [...]”.

Por utilizar o modelo de dados RDF, o FOAF oferece uma estrutura que é extensível. Dessa forma, como qualquer pessoa pode dizer qualquer coisa sobre qualquer tópico, o FOAF permite que qualquer pessoa faça declarações inéditas sobre pessoas, projetos etc., e relacionar tais declarações a outras já feitas (ALLEMANG; HENDLER, 2008).

4.2.5.5 *Data Catalog Vocabulary (DCAT)*

O vocabulário DCAT²⁶, padrão W3C atualmente na versão 2 (DCAT-2), fornece classes e propriedades para que conjuntos de dados (*datasets*) possam ser descritos e incluídos em catálogos de metadados, facilitando o consumo e a agregação de metadados de diversos catálogos (W3C, 2020). Para Laufer (2015, p. 102), “O uso de uma forma padrão para a descrição dos catálogos aumenta a capacidade de descoberta e permite que as aplicações possam encontrar metadados distribuídos em diferentes catálogos.”.

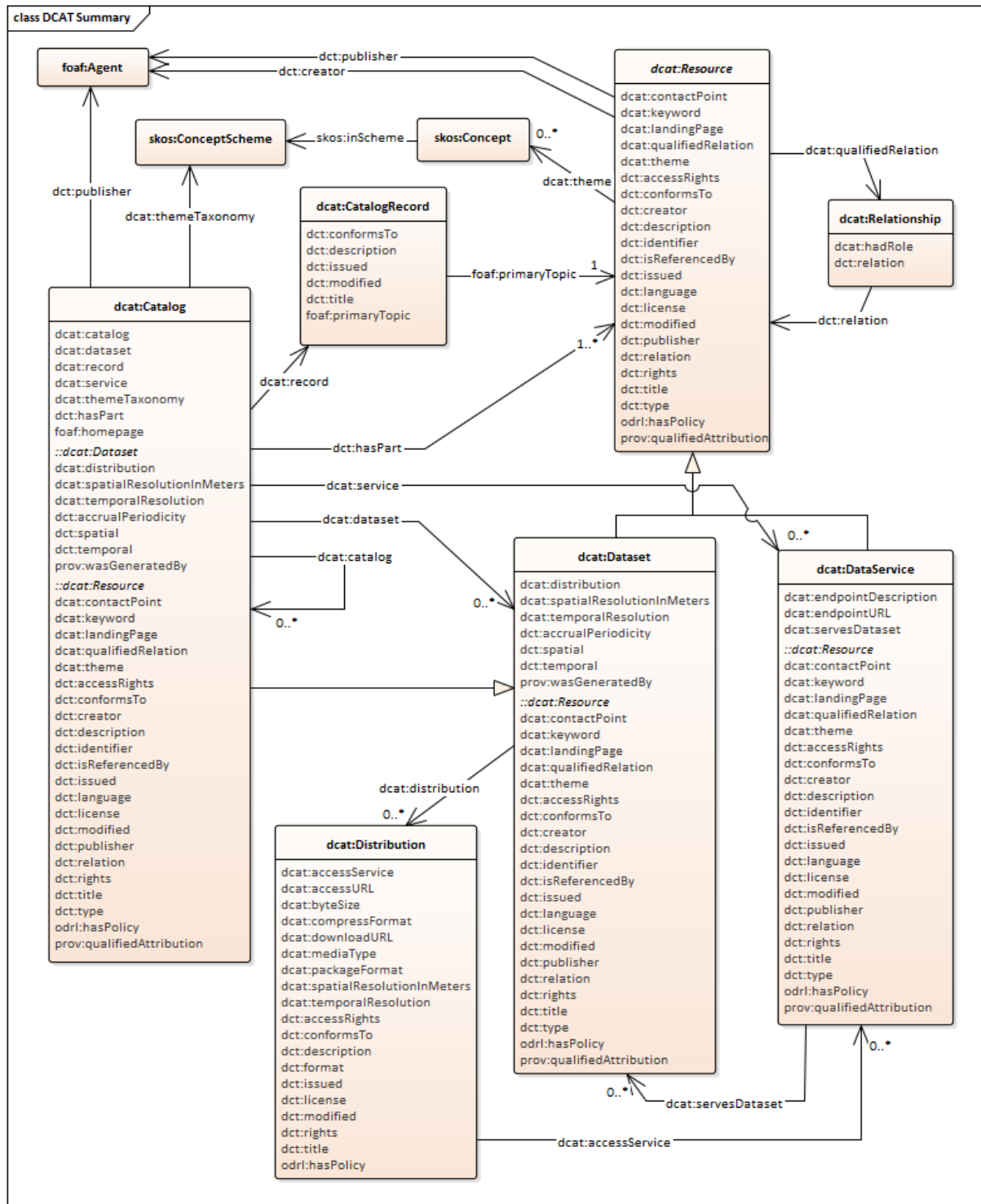
No esquema do DCAT-2, um conjunto de dados é definido como uma “coleção de dados, publicada ou com curadoria por um único agente, e disponível para acesso ou download em uma ou mais serializações ou formatos” (W3C, 2020, não paginado, tradução nossa), e que pode ser representado por uma ou mais distribuições, as quais podem ser disponibilizadas por meio de serviços *Web*.

Neste vocabulário, os elementos são definidos em um *namespace* próprio (<http://www.w3.org/ns/dcat#>), bem como podem ser incorporados elementos de outros vocabulários padronizados, tais como o FOAF, o DCTerms e o VOID²⁷ (*Vocabulary Of Interlinked Datasets*). A Figura 19 mostra uma visão geral do esquema do vocabulário DCAT-2, mostrando as classes de recursos e os relacionamentos entre elas.

²⁶ <https://www.w3.org/TR/vocab-dcat-2/>

²⁷ <https://www.w3.org/TR/void/>

Figura 19 – Visão geral do esquema do vocabulário DCAT-2.



Fonte: W3C (2020)

Conforme mostrado na Figura 19, as principais classes de recursos deste vocabulário são: *dcat:Catalog*, que representa um catálogo, no qual seus registros são coleções de metadados sobre *conjuntos de dados* ou *serviços de dados*; *dcat:CatalogRecord*, que representa um registro de um catálogo; *dcat:Dataset*, que é simplesmente um pseudônimo para *dc:Dataset*,

e que representa um conjunto de dados em um catálogo; *dc*at:DataService, que representa um serviço de dados, acessível através de uma API; e *dc*at:Distribution, que representa uma forma de acesso ao conjunto de dados catalogado (W3C, 2020).

O DCAT-2 é comumente utilizado em portais de dados abertos baseados no *software* de código aberto CKAN²⁸ (*Comprehensive Knowledge Archive Network*), com o intuito de facilitar a integração entre diferentes portais.

4.2.5.6 Outros Vocabulários RDF

Além dos vocabulários apresentados nas subseções anteriores, existem diversos outros vocabulários que podem ser utilizados para a descrição de dados em RDF para diferentes propósitos em diferentes domínios. O Quadro 24 relaciona alguns destes vocabulários.

Quadro 24 – Outros vocabulários para descrição de dados em RDF.

NOME	PREFIXO	URI	PROPÓSITO
<i>Bio</i>	bio:	http://purl.org/vocab/bio/0.1/	Informação bibliográfica
<i>Creative Commons Rights</i>	cc:	http://creativecommons.org/ns#	Licenças
<i>DOAP</i>	doap:	http://usefulinc.com/ne/doap#	Projetos
<i>Geo</i>	pos:	http://www.w3.org/2003/01/geo/wgs84_pos#	Posições
<i>GeoNames</i>	gn:	http://www.geonames.org/ontology#	Localizações
<i>Good Relations</i>	gr:	http://purl.org/goodrelations/v1#	Produtos
<i>Object Reuse and Exchange</i>	ore:	http://www.openarchives.org/ore/terms	Recursos de mapa
<i>SIOC</i>	sioc:	http://rdfs.org/sioc/ns#	Comunidades online
<i>vCard</i>	vcad:	http://www.w3.org/2006/vcard/ns#	Cartão de visita
<i>VoID</i>	void:	http://rdfs.org/ns/void#	Interligação de <i>datasets</i>
<i>WordNet</i>	wn:	http://xmlns.com/wordnet/1.6/	Palavras em inglês
<i>PROV-O</i>		http://www.w3.org/ns/prov#	Proveniência de dados
<i>ORG</i>	org:	http://www.w3.org/ns/org#	Estruturas organizacionais
<i>QB</i>	qb:	http://purl.org/linked-data/cube#	Dados multidimensionais
<i>SCOVO</i>	scv:	http://purl.org/NET/scovo	Dados estatísticos
<i>Hydra Core Vocabulary</i>	hydra:	http://www.w3.org/ns/hydra/core#	APIs
<i>Internationalization Tag Set</i>	its:	http://www.w3.org/2005/11/its	Internacionalização

Fonte: Adaptado de Isotani e Bittencourt (2015).

²⁸ <http://ckan.org/>

Os conceitos e as ferramentas e tecnologias da *Web Semântica* vistos nesta seção são fundamentais para a compreensão do conteúdo da próxima seção, que trata dos dados ligados (*Linked Data*) e suas boas práticas para a publicação de dados na *Web*.

4.3 DADOS LIGADOS (*LINKED DATA*)

Para que a *Web Semântica* seja uma realidade em nosso cotidiano, é necessária a criação de instrumentos que possam estimular a efetiva publicação de grandes volumes de dados na *Web*, de forma consistente, seguindo padrões, a fim de que possam ser facilmente descobertos, consumidos, compartilhados e reutilizados de forma livre na *Web*. De acordo com Berners-Lee (2006), a *Web Semântica* não serve apenas para colocar dados na *Web*, mas também criar *links* entre eles, permitindo que pessoas ou máquinas possam explorar uma *Web* de Dados.

Assim, um instrumento criado com este propósito consiste em um conjunto de boas práticas para a exposição, compartilhamento e conexão de dados estruturados na *Web*. Essas boas práticas são conhecidas como o paradigma dos ‘dados ligados’ (em inglês, *Linked Data*).

Nhancuougue, Rozsa e Dutra (2019, p. 24) definem *Linked Data* como “um conjunto de dados publicados na *Web*, em formatos que sejam legíveis por máquinas, com estruturas semânticas bem definidas, que estejam ligados a outros dados externos.”.

O principal objetivo do paradigma do *Linked Data* é possibilitar a transição da atual *Web*, orientada a documentos, para uma *Web* de dados interligados no ambiente da *Web Semântica*. Para Auer et al. (2011, p. 1, tradução nossa),

Um dos maiores desafios na área de gerenciamento inteligente da informação é a exploração da *Web* como plataforma de integração de dados e informações, bem como de busca e consulta. Da mesma forma que publicamos informações textuais não estruturadas na *Web* como páginas HTML e pesquisamos tais informações usando motores de busca baseados em palavras-chave, logo seremos capazes de publicar facilmente as informações estruturadas, interligando-as de forma confiável com outros dados publicados na *Web* e pesquisar o espaço de dados resultante usando consultas mais expressivas, em vez das buscas simples por palavras-chave.

Para Laufer (2015, p. 70), o que se pretende alcançar com o *Linked Data* é “[...] um banco de dados global, onde um conjunto crescente de informações possa ser acessado por um conjunto diversificado de aplicações com os mais diferentes propósitos.”.

Assim, as aplicações que desejam consumir dados ligados devem simplesmente compreender os vocabulários ou esquemas utilizados para descrever esses dados. O próprio processo de desenvolvimento de vocabulários é completamente aberto, no qual qualquer pessoa

pode desenvolver e publicar um vocabulário RDF em seu próprio *namespace*, para que seja utilizado por terceiros.

4.3.1 Princípios de *Linked Data*

Com o intuito de orientar o processo de publicação de dados ligados na *Web*, Berners-Lee (2006) propôs quatro regras a serem seguidas para que tal publicação seja bem-sucedida e possa, assim, agregar valor aos dados:

- 1) Use URIs como nome para coisas. Isso levará à criação de um sistema onde todos os recursos sejam identificados de forma unívoca (PO et al., 2020);
- 2) Use URIs HTTP para que as pessoas possam pesquisar estes nomes.
- 3) Quando alguém procurar uma URI, forneça informações úteis usando os padrões do W3C (RDF, SPARQL);
- 4) Inclua *links* para outros URIs a fim de que eles possam descobrir mais coisas. Essa regra viabilizará a formação de uma rede global de dados ligados (PO et al., 2020).

A partir da aplicação dessas regras, os dados ligados fornecerão acesso a um espaço de dados global e sem limites. Isto melhora a capacidade de descoberta de dados relevantes, uma vez que as fontes de dados podem ser mais facilmente rastreadas pelos mecanismos de busca e acessadas usando navegadores de dados genéricos. Além disso, as aplicações da *Web* poderão descobrir e acessar novas fontes de dados sem a necessidade de alterar seu código. (HEATH; BIZER, 2011).

Para Nhacuoungue, Rozsa e Dutra (2019), as ligações de dados no Linked Data:

[...] são construídas, basicamente, criando-se triplas RDF nas quais o sujeito é um recurso no dataset local e o objeto se refere a um recurso em um dataset externo, unidos por algum predicado conveniente para a ligação. Uma vez acessado o recurso externo, podem-se descobrir ligações para outros recursos e assim por diante.

Dessa forma, as tecnologias de *Linked Data* (ou dados ligados) podem contribuir para conectar os diferentes silos de dados que existem atualmente na *Web*, formando um único espaço de informação global. Dados relacionados sobre a mesma entidade, porém de diferentes fontes, podem ser agregados, fundidos e consultados, tal como se um banco de dados local fosse consultado (HEATH, BIZER, 2011).

4.3.2 Dados Abertos Ligados (LOD)

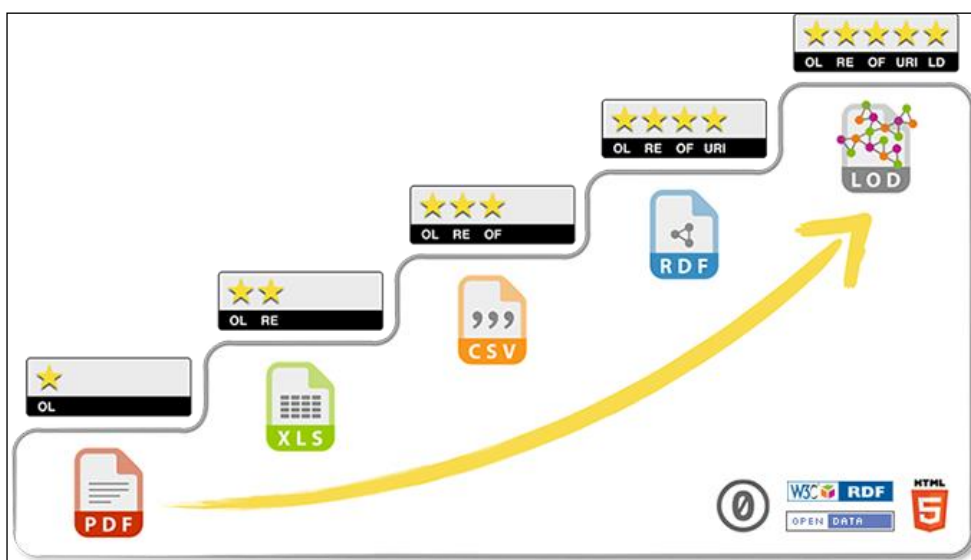
A transição da atual *Web* de documentos para uma *Web* de dados tem estimulado os usuários, principalmente governos, a promoverem a abertura e publicação de dados *online*, aproveitando as vantagens que tanto os dados abertos quanto os dados ligados oferecem, o que levou à fusão destas duas práticas, por meio do projeto ‘dados abertos ligados’ (*Linked Open Data* - LOD).

De acordo com Gill (2016, não paginado, tradução nossa), o *Linked Open Data*:

[...] oferece grande promessa para o uso semanticamente rico, mais fácil e mais difundido, reutilização e compartilhamento tanto de registros de metadados quanto dos vocabulários controlados que são usados para povoar esses registros e fornecer conexões significativas entre os recursos. O LOD tem o potencial de revolucionar a forma como os dados podem ser divulgados e integrados de forma a melhorar significativamente o processo de busca e utilização de informações e recursos.

O principal objetivo do projeto LOD era iniciar a *Web* de dados, identificando conjuntos de dados disponibilizados sob licenças abertas, os quais eram convertidos em RDF, de acordo com os princípios de *Linked Data*, e então publicados na *Web* (HEATH; BIZER, 2011). Este objetivo fica evidente ao observar o ‘esquema de classificação por estrelas’ formulado por Berners-Lee (2006), para incentivar a publicação de dados na *Web*. Nesse esquema, a quantidade de estrelas atribuídas aumenta à medida que o conjunto de dados se torna mais fácil de ser encontrado, compartilhado e reutilizado pelos usuários. A Figura 20 ilustra os cinco níveis desse esquema de classificação.

Figura 20 – Esquema de classificação de 5 estrelas para dados na *Web*.



Fonte: Berners-Lee (2006).

Como pode ser observado na Figura 20, para receber a classificação inicial de uma estrela é necessário apenas que os dados sejam disponibilizados na *Web* em qualquer formato (independentemente de ser proprietário ou não, aberto ou não), desde que sob uma licença aberta (sem restrições quanto a perfis de usuários e finalidade de uso). Um arquivo em formato PDF (*Portable Document Format*) é um exemplo.

Para evoluir para o segundo nível (duas estrelas), além dos dados serem publicados sob uma licença aberta, devem ser disponibilizados em formato estruturado, de forma que possam ser facilmente manipulados e exportados para outros formatos estruturados. Um exemplo seria uma planilha em formato XLS, que apesar de ser originada em um aplicativo proprietário (*Microsoft Excel*), é possível manipular e exportar os dados com facilidade.

Já o terceiro nível (três estrelas) pode ser alcançado ao adicionar a exigência de que os dados sejam disponibilizados em um formato não proprietário. Assim, no exemplo anterior, em vez de disponibilizar a planilha em formato XLS, uma opção seria disponibilizá-la no formato CSV (*Comma Separated Values*), que também é um formato estruturado, porém não depende de aplicativos proprietários para que seja manipulado.

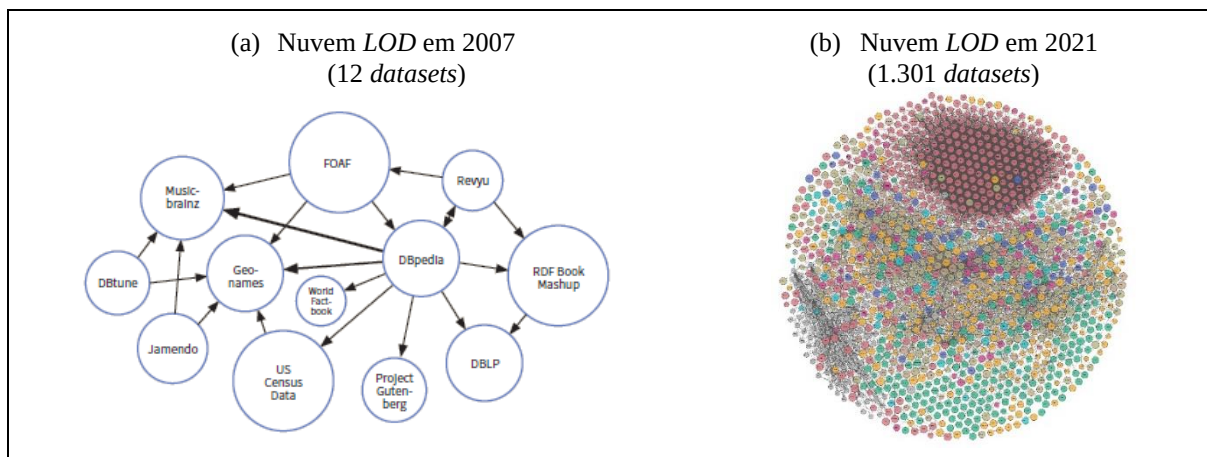
Por fim, para se alcançar a quarta e quinta estrelas, além de disponibilizar os dados em formato aberto, estruturado e não proprietário, é necessário seguir os padrões de dados abertos do W3C. Assim, a quarta estrela é conquistada ao adotar o modelo de dados RDF, de modo que qualquer pessoa ou máquina possa identificar e acessar os dados de forma padronizada na *Web*. E a quinta estrela é conquistada quando os dados RDF forem ligados aos dados de outras bases publicadas na *Web* sob o mesmo padrão, seguindo as boas práticas de *Linked Open Data*.

Diante disso, para Santarem Segundo (2015, p. 225) o projeto *Linked Data*:

[...] atualmente apresenta-se como a melhor forma de materialização dos conceitos e tecnologias da *Web Semântica*, é um projeto, com um conjunto de normas a serem seguidas, que usa os mesmos princípios de ligação semântica da *Web de Dados*. Entretanto tem particularidades específicas, indicando um grau de exigência maior na constituição de sua rede de interligações.

Neste sentido, com a finalidade de representar os conjuntos de dados abertos ligados disponíveis na *Web*, foi criado um diagrama chamado ‘nuvem LOD (LOD Cloud)’. O diagrama é mantido e atualizado regularmente pelo *Insight Center for Data Analytics*, um dos maiores centros de pesquisa em Ciência de Dados da Europa. A Figura 21 exhibe em (a) o estado inicial, em 2007, e em (b) o estado atual da nuvem.

Figura 21 – Estados inicial e atual da nuvem de dados ligados (*LOD Cloud*).



Fonte: McCrae (2021).

Como é possível perceber na Figura 21, a partir dos esforços do projeto LOD, centenas de conjuntos de dados já foram ligados uns aos outros. Na maioria dos casos essa ligação é feita utilizando uma propriedade do tipo ‘owl:sameAs’, a qual liga a descrição de um recurso em um domínio a uma outra descrição do mesmo recurso em um outro domínio.

O nó central da nuvem LOD é o *dataset DBpedia*, que disponibiliza em formato RDF dados estruturados extraídos da *Wikipedia*, tais como imagens, coordenadas geográficas, categorizações e *links* para páginas da *Web*, possibilitando que sejam conectados a diferentes conjuntos de dados na *Web*. Cada entidade da *DBpedia* (denotado por uma URI na forma *http:dbpedia.org/resource/{nome}*) está conectada diretamente a um artigo da *Wikipedia* (LAUFER, 2015).

4.3.3 Dados Geoespaciais Ligados (*Linked GeoData*)

Um elemento comum em vários dos conjuntos de dados disponibilizados na *Web* é a localização geográfica. Isso torna os dados geoespaciais especialmente úteis à *Web* de Dados (W3C; OGC, 2017).

A geografia é um fator que muitas vezes pode conectar informações dos mais variados domínios. Isso é aparente na *Web* de dados, onde o dicionário de nomes geográficos *Geonames* frequentemente serve como um *hub* para outros conjuntos de dados que possuem algum componente geográfico, como visto anteriormente na Figura 21(a).

No contexto geoespacial, a transformação e publicação de dados geográficos na forma de *Linked Data* teve como iniciativas pioneiras os projetos *GeoNames*²⁹, *OpenStreetMap*³⁰, e *Ordnance Survey*³¹. Atualmente, o vocabulário *Basic Geo*³² (WGS84 lat/long) é um dos principais vocabulários usados para a descrição de conjuntos de dados geoespaciais ligados. O *Basic Geo* é um vocabulário básico que fornece um *namespace* para representar lat(itude), long(itude) e outras informações sobre coisas localizadas espacialmente, usando o sistema de coordenadas WGS84 como um dado de referência (VILCHES-BLÁSQUEZ; SAADREVA, 2019). O Quadro 25 mostra um breve exemplo de aplicação do vocabulário WGS84.

Quadro 25 – Exemplo de aplicação do vocabulário WGS84.

```
@prefix geo: <https://www.w3.org/2003/01/geo/wgs84_pos#> .
a geo:Point ;
  geo:lat "55.701" ;
  geo:long "12.552" .
```

Fonte: Vilches-Blásquez e Saadreja (2019).

Dessa forma, ao aplicar os princípios de *Linked Data* a localizações geográficas, torna-se possível relacionar locais entre si e interligar dados sobre locais a diferentes fontes de dados, bem como para o conteúdo adicional da *Web*. Por exemplo, uma região poderia ser ligada a suas localidades abrangentes em uma hierarquia administrativa, bem como a dados sobre pessoas que nasceram, morreram ou trabalharam naquela região. Os clientes da *Web Semântica* podem então navegar através de *links* explícitos para recuperar dados descrevendo as entidades interligadas (BECKER; BIZER, 2009).

Na *DBpedia*, os locais são frequentemente interligados com outros recursos usando propriedades como ‘*dbpedia-owl:localização*’, ‘*dbpedia-owl:cidade*’ e ‘*dbpedia-owl:país*’, que relacionam um recurso a um local, cidade e país, respectivamente, ou ainda a propriedade ‘*dbpedia-owl:capital*’, que relaciona um país ou região à sua capital (BECKER; BIZER, 2009).

Para aplicações geoespaciais, o principal *hub* de dados ligados na nuvem LOD é o serviço *GeoNames*, que fornece URIs persistentes e metadados sobre uma variedade de localizações geográficas. Uma das principais características do *GeoNames* é que cada entidade descrita em seu banco de dados inclui uma serialização RDF, permitindo *links* diretos de outros serviços vinculados. Assim, o *GeoNames* se posicionou como uma fonte de dados essencial

²⁹ <https://www.geonames.org/ontology/documentation.html>

³⁰ <https://www.openstreetmap.org/>

³¹ <https://www.ordnancesurvey.co.uk/>

³² <https://www.w3.org/2003/01/geo/>

para qualquer desenvolvedor que trabalhe em uma aplicação de dados ligados com um componente geoespacial. (BIDNEY; CLAIR, 2014)

Assim, com a incorporação flexível da semântica geoespacial na *Web*, a recuperação de informações geoespaciais se tornará precisa até o nível em que os resultados das consultas dos usuários serão imediatamente úteis, sem eliminar os resultados irrelevantes. (EGENHOFER, 2002)

Diante disso, é possível constatar que o paradigma de *Linked Data* oferece grandes oportunidades no domínio das geociências, uma vez que dados geoespaciais convencionais isolados e heterogêneos podem ser expostos como dados ligados na *Web*, promovendo assim o amplo compartilhamento e integração das informações geoespaciais (YUAN, 2013).

4.3.4 Boas Práticas para Publicação de Dados como *Linked Data*

A publicação de dados na forma de dados ligados requer a adoção dos princípios de *Linked Data* mencionados anteriormente. A conformidade do processo de publicação com os padrões e boas práticas que sustentam tais princípios é o que permite que o paradigma de *Linked Data* simplifique a interoperabilidade e a reutilização de dados na *Web* (HEATH; BIZER, 2011).

Neste sentido, o ‘Grupo de Trabalho do W3C para Dados Abertos Governamentais’³³ publicou um conjunto de dez ‘boas práticas para a publicação de *Linked Data*’ (HYLAND; ATEMEZING; VILLAZÓN-TERRAZAS, 2013), as quais são resumidas no Quadro 26.

Quadro 26 – Boas práticas para a publicação de *Linked Data*.

#	PRÁTICA	DESCRIÇÃO
1	Preparar os <i>stakeholders</i>	Preparar as partes interessadas explicando o processo de criação e manutenção de <i>Linked Data</i> .
2	Selecionar o conjunto de dados	Selecionar o conjunto (ou fonte) de dados a ser publicado e conectado a outros conjuntos de dados.
3	Modelar os dados	Representar os objetos de dados e como eles são relacionados de forma independente da aplicação.
4	Especificar a licença	Especificar uma licença de dados aberta apropriada. A reutilização de dados é mais provável de ocorrer quando há uma declaração clara sobre a origem, propriedade e termos relacionados ao uso dos dados publicados.
5	Nomear bons <i>URIs</i>	Criar uma estratégia de nomenclatura de <i>URIs</i> e um plano de implementação baseado em <i>URIs HTTP</i> . Considerar nesse plano como

³³ <http://www.w3.org/2011/gld/>

		lidar com a nomeação de objetos, o suporte multilíngue, a mudança de dados ao longo do tempo, bem como uma estratégia de persistência.
6	Usar vocabulários conhecidos	Descrever, sempre que possível, os objetos com vocabulários já existentes. Ampliar os vocabulários existentes sempre que necessário e criar vocabulários somente quando necessário e seguindo as boas práticas.
7	Converter os dados	Converter os dados para um formato de representação de <i>Linked Data</i> . Isso normalmente é feito por script ou outros processos automatizados
8	Prover acesso aos dados por máquinas	Oferecer várias formas de busca e outros processos automatizados para acessar os dados por meio de mecanismos padrão da <i>Web</i> .
9	Anunciar novos conjuntos de dados ligados	Anunciar os novos conjuntos de dados publicados em um domínio de autoridade.
10	Reconhecer a função social dos conjuntos de dados	Reconhecer a responsabilidade de manutenção dos dados publicados. Certificar-se que os conjuntos de dados permaneçam disponíveis ao longo do tempo.

Fonte: Adaptado de Hyland, Atezing e Villazón-Terrazas (2013).

Além de seguir estas boas práticas, Heath e Bizer (2011) recomendam que os conjuntos de dados a serem publicados devem permitir que os agentes de software na *Web* descubram todas as metainformações relevantes necessárias para integrar dados de diferentes fontes. Para isso, elaboraram uma lista de verificação, apresentada no Quadro 27, que pode ser usada pelos provedores para verificar se os dados atendem aos principais requisitos para publicação como *Linked Data*.

Quadro 27 – Lista de verificação para publicação de *Linked Data*.

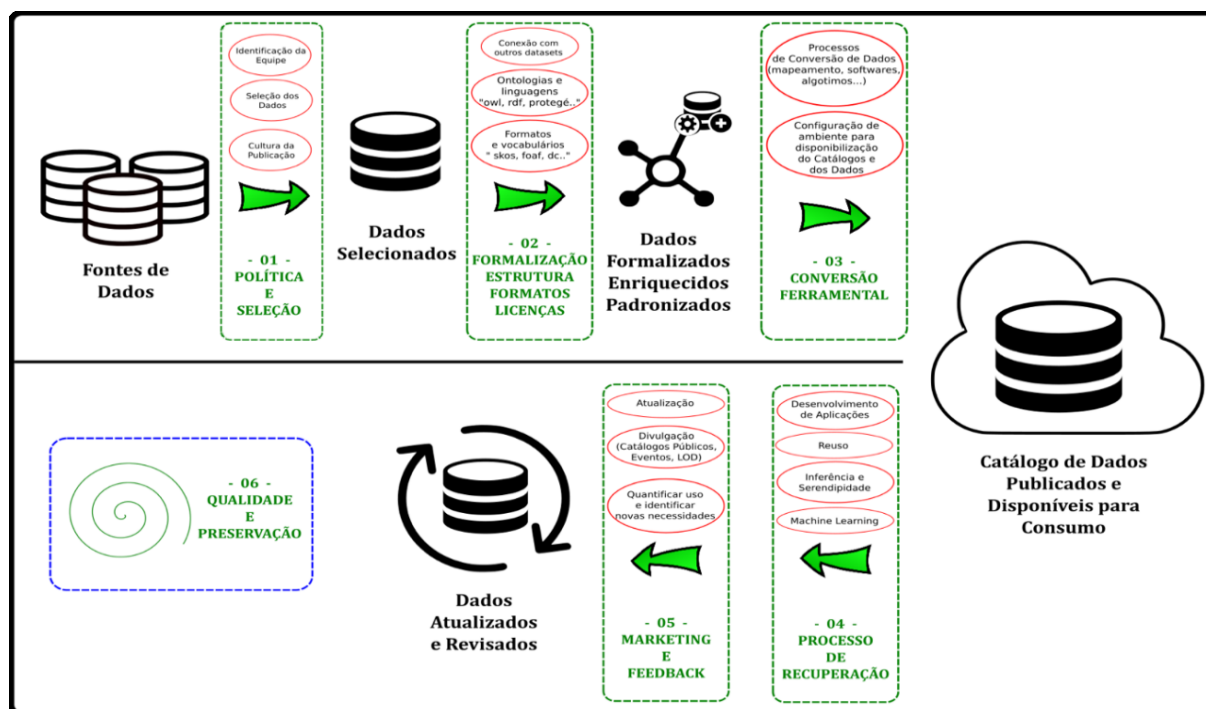
#	ITEM DE VERIFICAÇÃO	DESCRIÇÃO
1	O conjunto de dados é ligado a outros conjuntos de dados?	Os <i>links RDF</i> conectam dados de diferentes fontes em um grafo <i>RDF</i> global e permitem que navegadores e rastreadores de <i>Linked Data</i> naveguem entre as fontes de dados.
2	São fornecidos metadados de proveniência?	Para permitir que os aplicativos tenham certeza sobre a origem dos dados, bem como para permitir que eles avaliem a qualidade dos dados, a fonte de dados deve publicar metadados de proveniência juntamente com os dados primários.
3	São fornecidos metadados de licenciamento?	Os dados da <i>Web</i> devem ser auto descritivos em relação a quaisquer restrições que se apliquem ao seu uso. Uma maneira comum de expressar tais restrições é anexar uma licença de dados aos dados publicados. Fazer isso é essencial para permitir que os aplicativos usem dados da <i>Web</i> sob uma base legal segura.
4	São usados termos de vocabulários amplamente implantados?	Para facilitar a compreensão dos dados ligados pelos aplicativos, os provedores de dados devem usar termos de vocabulários amplamente implantados para representar os dados.
5	Os <i>URIs</i> de termos de vocabulários proprietários são desreferenciáveis?	Os provedores de dados geralmente definem termos proprietários que são usados além dos termos de vocabulários amplamente implantados. Para permitir que os aplicativos recuperem automaticamente a definição de termos de vocabulário da <i>Web</i> , os

		URIs que identificam termos de vocabulário proprietários devem ser desreferenciáveis.
6	Os termos de vocabulário proprietários são mapeados para outros vocabulários?	Os termos do vocabulário proprietário devem ser relacionados aos termos correspondentes em outros vocabulários para permitir que os aplicativos entendam o máximo de dados possível.
7	São fornecidos metadados no nível do conjunto de dados?	Além de tornar os dados de instância auto descritivos, também é desejável que os editores de dados forneçam metadados que descrevam características de conjuntos de dados completos, por exemplo, o tópico de um conjunto de dados e estatísticas mais detalhadas.
8	São oferecidos métodos de acesso adicionais?	A principal maneira de publicar <i>Linked Data</i> na <i>Web</i> é tornar os URIs que identificam itens de dados desreferenciáveis em descrições <i>RDF</i> . Além disso, vários provedores de dados <i>LOD</i> optaram por oferecer dois meios alternativos de acesso a seus dados por meio de <i>endpoints SPARQL</i> ou fornecer <i>dumps RDF</i> do conjunto de dados completo.

Fonte: Adaptado de Heath e Bizer (2011).

Em outra frente, Santarem Segundo (2018) propõe um ‘fluxo organizacional para publicação de dados’, que tem por finalidade identificar e orientar as diversas fases que um projeto de publicação de dados na *Web* deve ter. A Figura 22 ilustra o fluxo organizacional proposto.

Figura 22 – Fluxo Organizacional para publicação de Dados.



Fonte: Santarem Segundo (2018).

Na Figura 22 é apresentada a proposta de divisão das várias tarefas, organizadas em etapas, que compõem o fluxo organizacional proposto. As etapas, suas tarefas, entradas e saídas são as apresentadas no Quadro 28.

Quadro 28 – Etapas e tarefas que compõem o fluxo organizacional para publicação de dados.

ETAPAS	TAREFAS	ENTRADAS	SAÍDAS
1. Política e seleção	• Identificar a equipe; • Selecionar os dados; • Criar cultura de publicação.	Fontes de Dados	Dados selecionados
2. Formalização, estrutura, formatos e licenças	• Conectar com outros datasets; • Definir Ontologias e linguagens; • Escolher formatos e vocabulários.	Dados selecionados	Dados formalizados, enriquecidos e padronizados
3. Conversão ferramental	• Definir processos de conversão de dados; • Configurar ambiente para disponibilização do catálogo de dados.	Dados formalizados, enriquecidos e padronizados	Catálogo de dados publicados e disponíveis para consumo.
4. Recuperação	• Desenvolver aplicações para consumo e reuso de dados; • Criar axiomas para inferência e serendipidade; • Usar técnicas de <i>Machine Learning</i>.	Catálogo de dados publicados e disponíveis para consumo.	
5. Marketing e feedback	• Manter os dados atualizados; • Divulgação dos dados publicados; • Quantificar o uso e identificar novas necessidades.		Dados atualizados e revisados.

Fonte: Adaptado de Santarem Segundo (2018)

Uma vez entendido o que são os dados ligados e como são estruturados, conectados e publicados na *Web* de dados, formando uma nuvem de dados abertos ligados, na seção a seguir é discutido como integrar as IDEs atuais e futuras com o universo da *Web Semântica* por meio da publicação de seus catálogos de metadados seguindo as boas práticas de *Linked Data*.

4.4 INTEGRAÇÃO DE INFRAESTRUTURAS DE DADOS ESPACIAIS COM A WEB SEMÂNTICA

A aplicação das tecnologias da *Web Semântica* na comunidade geoespacial tem se desenvolvido consideravelmente nos últimos anos. Embora as IDEs e a *Web Semântica* sejam bastante distintas em relação a suas arquiteturas, fluxos de trabalho e padrões aplicados, ainda assim podem ser consideradas complementares (WIEMANN; BERNARD, 2016).

No entanto, a derivação de informações significativas a partir de fontes de dados espaciais distribuídas na *Web* continua sendo um grande desafio. Neste aspecto, de acordo com Wiemann e Bernard (2016), as discussões giram em torno de duas questões principais: como os dados espaciais na *Web* podem ser relacionados de forma eficiente? E, como essas relações podem ser utilizadas de forma significativa e produtiva?

Outra questão significativa nas IDEs é a heterogeneidade semântica, que é um impedimento para integrar dados geoespaciais de múltiplas fontes e fundir dados geoespaciais com outros tipos de dados, já que a semântica dos metadados, esquemas e conteúdo de dados não são normalmente harmonizados (HUANG et al., 2019).

Diante disso, os dados ligados são um meio promissor para resolver as limitações relativas à integração de dados e heterogeneidade semântica das atuais IDEs, de forma que eles têm sido vistos como um dos principais elementos na concepção da próxima geração destas plataformas. As boas práticas de *Linked Data* para estruturação, publicação, descoberta, acesso e integração de dados na *Web* tem o potencial de promover o compartilhamento efetivo e ampliar a descoberta de recursos geoespaciais em IDEs (HUANG et al., 2019).

Nas subseções a seguir, serão abordadas as boas práticas de *Linked Data* e do grupo de trabalho formado por membros da OGC e do W3C para a publicação de dados e metadados geoespaciais na *Web*. Será também abordada a publicação de metadados geoespaciais catalogados em IDEs, que é o objeto do *framework* proposto neste trabalho como forma de promover a integração das IDEs com a *Web Semântica*.

4.4.1 Boas práticas da OGC e W3C para Dados Espaciais na *Web*

Tendo em vista que governos e diversos setores econômicos, tais como energia, transporte, manufatura e saúde estão se tornando integrados ao utilizarem a *Web* como plataforma comum de compartilhamento de dados, torna-se necessário estabelecer um conjunto de recomendações aplicáveis à publicação de dados na *Web*. Essas recomendações devem abranger aspectos relacionados à formatos e identificadores de dados, metadados, licenciamento, proveniência e acesso a dados, dentre outros (W3C, 2017).

Assim, para orientar, padronizar e estimular a publicação de dados geoespaciais na *Web*, a OGC elaborou, em conjunto com o W3C, um documento definindo as ‘boas práticas para dados espaciais na *Web*’ (em inglês, *Spatial Data on the Web Best Practices*³⁴). Anteriormente,

³⁴ <https://www.w3.org/TR/sdw-bp/>

o W3C já havia publicado, em janeiro de 2017, um documento (*Data on the Web Best Practices*³⁵) descrevendo um conjunto de boas práticas para a publicação de quaisquer tipos de dados na *Web*. Neste documento, são descritas 35 práticas recomendadas, as quais são listadas no Quadro 29.

Quadro 29 – Relação das boas práticas do W3C para publicação de dados na *Web*.

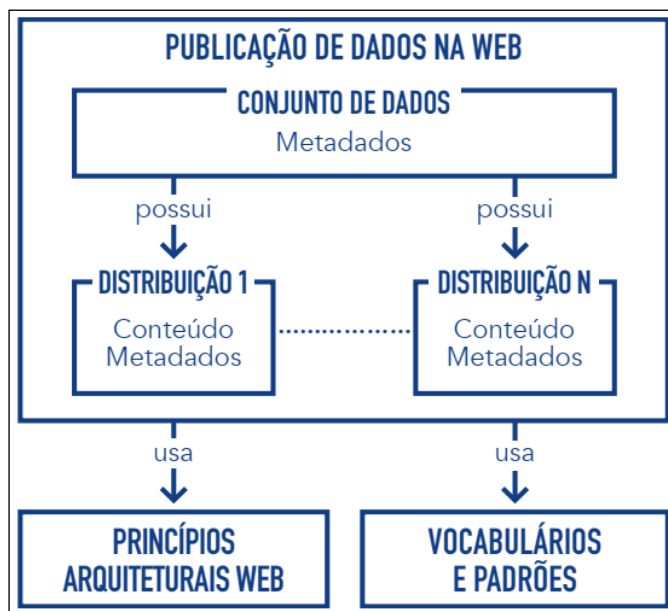
BOAS PRÁTICAS PARA PUBLICAÇÃO DE DADOS NA WEB			
1	Forneça metadados	19	Use negociação de conteúdo para servir dados disponíveis em vários formatos
2	Forneça metadados descritivos	20	Forneça acesso em tempo real
3	Forneça metadados estruturais	21	Forneça dados atualizados
4	Forneça informações de licença de dados	22	Forneça uma explicação para os dados que não estão disponíveis
5	Forneça informações de proveniência de dados	23	Disponibilize dados por meio de uma API
6	Forneça informações de qualidade de dados	24	Use os padrões da <i>Web</i> como a base das APIs
7	Forneça um indicador de versão	25	Forneça documentação completa para sua API
8	Forneça histórico de versão	26	Evite interromper alterações em sua API
9	Use URIs persistentes como identificadores de conjuntos de dados	27	Preserve identificadores
10	Use URIs persistentes como identificadores dentro dos conjuntos de dados	28	Avalie a cobertura do conjunto de dados
11	Atribua URIs a versões e séries de conjuntos de dados	29	Colete feedback dos consumidores de dados
12	Use formatos de dados padronizados legíveis por máquina	30	Disponibilize feedback
13	Use representações de dados neutras em termos de localidade	31	Enriqueça os dados gerando novos dados
14	Forneça dados em vários formatos	32	Forneça Apresentações Complementares
15	Reutilize vocabulários, de preferência padronizados	33	Forneça feedback ao editor original
16	Escolha o nível de formalização correto	34	Siga os termos de licenciamento
17	Forneça download em massa	35	Cite a publicação original
18	Forneça subconjuntos para grandes conjuntos de dados		

Fonte: Traduzido de W3C (2017).

³⁵ <https://www.w3.org/TR/dwbp/>

As boas práticas propostas nesse documento seguem os princípios da arquitetura da *Web* e usam vocabulários e padrões para promover a interoperabilidade entre conjuntos de dados, conforme ilustra a Figura 23.

Figura 23 – Esquema para publicação de conjuntos de dados na *Web*.



Fonte: Lóscio *et al.* (2017).

Na Figura 23, um conjunto de dados (*dataset*), descrito por um conjunto de metadados, possui uma ou mais distribuições, cada qual com seus conteúdos e metadados descritivos. De acordo com Lóscio *et al.* (2017), as distribuições “[...] são uma forma física específica de um conjunto de dados. Essas distribuições facilitam o compartilhamento de dados em larga escala [...]”. A publicação desses conjuntos de dados e metadados na *Web* devem seguir os princípios da arquitetura da *Web*, além de usar vocabulários e padrões.

Contudo, o documento publicado pelo W3C não previa em suas práticas como lidar com a publicação de dados que contenham uma extensão espacial (tamanho, forma ou posição), os chamados dados (geo)espaciais. Diante disso, e considerando o grande volume de dados espaciais sendo usados por aplicações na *Web* e disponibilizados por meio de IDEs, é que foi formado um grupo de trabalho constituído por especialistas do W3C e da OGC, para elaborar um documento de boas práticas para a publicação de dados espaciais na *Web* (W3C; OGC, 2017).

Assim, em setembro de 2017, é que o documento ‘Boas práticas para dados espaciais na *Web*’ foi publicado pela W3C. O documento reúne 14 práticas recomendadas, relacionadas no Quadro 30, com o intuito de resolver problemas relacionados à descoberta, acessibilidade e

interoperabilidade de dados geoespaciais, de forma que sejam integrados mais amplamente à *Web* de Dados por meio de padrões e de soluções que possibilitem sua descoberta e reuso por usuários não especialistas (W3C; OGC, 2017).

Quadro 30 - Boas práticas para publicação de dados espaciais na *Web*.

BOAS PRÁTICAS PARA PUBLICAÇÃO DE DADOS ESPACIAIS NA WEB			
1	Use URIs HTTP persistentes globalmente exclusivos para coisas espaciais	8	Declare como os valores das coordenadas são codificados
2	Torne seus dados espaciais indexáveis por mecanismos de pesquisa	9	Descreva o posicionamento relativo
3	Conecte recursos para criar a <i>Web</i> de Dados	10	Use tipos de relação apropriados para vincular coisas espaciais
4	Use codificações de dados espaciais que correspondam ao seu público-alvo	11	Forneça informações sobre a natureza mutável das coisas espaciais
5	Forneça geometrias na <i>Web</i> de uma forma utilizável	12	Exponha dados espaciais por meio de 'APIs de conveniência'
6	Forneça geometrias no nível certo de exatidão, precisão e tamanho	13	Inclua metadados espaciais em metadados de conjunto de dados
7	Escolha sistemas de referência de coordenadas para atender às aplicações do seu usuário	14	Descreva a precisão posicional dos dados espaciais

Fonte: W3C e OGC (2017).

Ao observar os conjuntos de boas práticas apresentados nos dois quadros anteriores, é possível perceber que algumas das práticas são voltadas para a publicação de metadados, sejam metadados em geral (práticas 1, 2 e 3 no Quadro 29) ou metadados geoespaciais (práticas 13 e 14 no Quadro 30). Tais práticas visam apoiar a descoberta e reutilização de dados, bem como orientar a inclusão da extensão espacial, do sistema de referência de coordenadas e de outros detalhes espaciais dos conjuntos de dados a serem descritos.

A seguir são apresentadas algumas técnicas e abordagens encontradas na literatura para orientar a publicação de catálogos de metadados das IDEs como *Linked Data*.

4.4.2 Publicação de Catálogos de Metadados Geoespaciais como *Linked Data*

Antes de acessar os conjuntos de dados geoespaciais disponibilizados por uma IDE, eles devem ser descobertos. Conforme visto anteriormente, nas IDEs, as coleções de descrições estruturadas (metadados) sobre o conteúdo de conjuntos de dados, informações de procedência e formatos de dados são expostas na *Web* por meio de serviços de catálogo. Por conseguinte,

os geoportais disponibilizam tais catálogos por meio de um serviço *Web* (CSW), que permite a descoberta e exploração dos conjuntos de dados disponibilizados pela IDE.

No entanto, para que os metadados catalogados em uma IDE possam ser indexados por um mecanismo de pesquisa da *Web*, uma solução possível seria publicá-los utilizando as ferramentas e tecnologias da *Web Semântica*. Para isso, os metadados geoespaciais devem ser preparados e organizados usando formatos de dados que possam ser lidos e processados por máquinas e agentes de software.

Além disso, os metadados geoespaciais nas IDEs são vagamente acoplados aos conjuntos de dados, e têm seus próprios padrões, como o ISO 19115, de forma que ainda é uma questão desafiadora como metadados geoespaciais podem ser representados nos formatos da *Web* de dados (YUE et al., 2016).

A fim de superar esses desafios e ter os metadados geoespaciais de uma IDE publicados na *Web* de forma indexável e legível por máquina, a adoção das boas práticas de *Linked Data* e de dados espaciais na *Web* pode ser um fator preponderante de sucesso. E o primeiro passo nesse sentido é realizar o mapeamento do esquema dos metadados catalogados na IDE para o modelo de dados RDF, ou seja, transformar metadados ISO 19115/XML em RDF/XML.

Para isso, Lopez-Pellicer et al. (2012) propõe uma abordagem para coletar os registros de um servidor de catálogo em uma IDE e converter os respectivos campos de metadados para o modelo de dados RDF na forma de *Linked Data*. A abordagem consiste em três etapas distintas descritas nos parágrafos a seguir.

Na primeira etapa, a coleta de registros de metadados no servidor catálogo da IDE é feito da seguinte maneira: por meio do serviço CSW são enviadas solicitações do tipo *GetRecords* ao servidor de catálogo, que retorna o esquema e os conjuntos de metadados requisitados em formato XML.

Na segunda etapa, para a conversão de padrão dos metadados coletados, uma vez que a especificação do serviço *Catalogue Service* (CS) define um conjunto mínimo de metadados ISO 19115 que podem ser representados usando o padrão *Dublin Core* (DC), é feito um mapeamento cruzado (*crosswalk*) do primeiro padrão para o segundo, ambos serializados em formato XML.

De acordo com o *Dublin Core Metadata Glossary*, um *crosswalk* “[...] é uma tabela que mapeia as relações e equivalências entre dois ou mais formatos de metadados. Os *crosswalks* suportam a capacidade dos mecanismos de busca de pesquisar efetivamente em bancos de dados heterogêneos, ou seja, ajudam a promover a interoperabilidade” (NOGUERAS-ISO, 2005, p.

91, tradução nossa). O Quadro 31 apresenta a correspondência entre elementos utilizada no *crosswalk* do padrão ISO 19115 para o padrão DC.

Quadro 31 – Correspondência entre elementos no *crosswalk* ISO 19115 para *Dublin Core*.

Elemento ISO 19115 (M) Mandatório, (O) Opcional, (C) Condicional	Elemento <i>Dublin Core</i>
Título do <i>dataset</i> (M) (MD_Metadata.identificationInfo.citation.title)	TITLE
Criador do <i>dataset</i> (O) (MD_Metadata.identificationInfo.pointOfContact.role="originator")	CREATOR
Categoria/assunto do <i>dataset</i> (M) (MD_Metadata.identificationInfo.topicCategory)	SUBJECT
Descrição resumida do <i>dataset</i> (M) (MD_Metadata.identificationInfo.abstract)	DESCRIPTION
Publicador do <i>dataset</i> (O) (MD_Metadata.identificationInfo.pointOfContact.role="publisher") Ponto de contato (M) (MD_Metadata.contact)	PUBLISHER
Data de referência do <i>dataset</i> (M) (MD_Metadata.identificationInfo.citation.date) Carimbo de data dos metadados (M) (MD_Metadata.dateStamp)	DATE
Tipo de representação espacial (O) (MD_Metadata.identificationInfo.spatialRepresentationType)	TYPE
Formato de distribuição (O) (MD_Metadata.distributionInfo.distributionFormat)	FORMAT
Recurso <i>online</i> (O) (MD_Metadata.distributionInfo.transferOptions.onLine.linkage)	IDENTIFIER
Linhagem (O) (MD_Metadata.dataQualityInfo.lineage.source.description)	SOURCE
Idioma (M) (MD_Metadata.identificationInfo.defaultLocale.language)	LANGUAGE
Localização geográfica (por quatro coordenadas ou por identificador geográfico) (C) (MD_Metadata.identificationInfo.extent.geographicElement) Informação de extensão adicional (vertical e temporal) (O) (MD_Metadata.identificationInfo.extent.temporalElement.extent)	COVERAGE (refinement <i>spatial</i>) (refinement <i>temporal</i>)
Licença (O) (MD_Metadata.identificationInfo.resourceConstraints.useConstraints="license")	LICENSE
Periodicidade de atualização (O) (MD_Metadata.identificationInfo.resourceMaintenance.maintenanceAndUpdateFrequency)	ACCRUALPERIODICITY

Fonte: Adaptado de Noguera-Iso (2005).

Os conjuntos de metadados obtidos após o *crosswalk*, já representados no padrão DC, ainda podem ser complementados com outros campos personalizados conforme a necessidade dos provedores e usuários, tais como metadados de proveniência.

Por fim, na terceira etapa, os metadados representados em DC são convertidos para o modelo de dados RDF (triplificação), no formato RDF/XML, seguindo as recomendações da

DCMI para expressar o DC usando RDF (Nilsson et al., 2007). Durante esse processo de conversão para triplas RDF, são inseridos os URIs de identificação de recursos e os URIs de ligação a outros *datasets* na *Web* de dados, seguindo as boas práticas de *Linked Data*.

Essa abordagem, no entanto, deixa algumas questões em aberto. Conforme enfatizam Reid et al. (2012), o padrão de metadados DC não contém nenhum elemento para a descrição de conjuntos de dados abstratos, tal como um recurso para *download*. Também não contém nenhuma propriedade para falar sobre pessoas e organizações que possam ser autoras ou mantenedoras de conjuntos de dados.

Como solução, para a descrição de catálogos de dados existe um vocabulário específico, recomendação W3C, que foi projetado para aumentar o DC, o *Data Catalog Vocabulary* (DCAT). Assim, na segunda etapa da abordagem apresentada, as propriedades *dcatalog:Catalog*, *dcatalog:CatalogRecord* e *dcatalog:Distribution* podem ser adicionadas aos metadados obtidos após o *crosswalk*. Já para se referir a pessoas, sejam naturais ou corporativas, a prática comum é usar o vocabulário *Friend-of-a-Friend* (FOAF), por meio das propriedades: *foaf:Agent*, quando a natureza precisa da entidade em questão não é clara; *foaf:Person*, quando fizer referência a uma pessoa natural; e, *foaf:Organization*, quando fizer referência a uma organização (REID et al., 2012).

Por fim, Ronzhin et al. (2019), em sua abordagem para a publicação de *datasets* geoespaciais da plataforma *Inspire* (IDE oficial da União Europeia) na *Web*, fazem duas recomendações que podem ser aplicadas na publicação de metadados geoespaciais como *Linked Data*: usar o vocabulário Geo (WGS84) para representar as geometrias e, em relação à ligação com a *Web* de dados, incluir URIs para recursos da *DBpedia*.

Diante do exposto, cumpre-se com o objetivo específico II inicialmente proposto, de ‘conceituar e descrever as tecnologias e ferramentas da *Web Semântica* e as boas práticas de *Linked Data* e de dados espaciais na *Web*’. Com isso, tem-se todo o referencial teórico e a descrição das práticas que nortearão a construção do *framework* proposto neste trabalho, visando a publicação de metadados geoespaciais catalogados em IDEs seguindo as boas práticas de *Linked Data*, o qual será apresentado em detalhes na Seção a seguir.

5 **FRAMEWORK PARA PUBLICAÇÃO DE METADADOS GEOESPACIAIS SEMÂNTICOS NA WEB**

Nesta Seção será apresentado o *framework* proposto no presente trabalho, que tem por finalidade promover a publicação de metadados geoespaciais semânticos na *Web* seguindo as boas práticas de *Linked Data* e de dados espaciais na *Web*. Cada uma das camadas do *framework*, e seus respectivos procedimentos, são descritos de forma detalhada, especificando as interações (de entrada e de saída) e o fluxo de execução. Também será mostrado como as tecnologias e ferramentas da *Web Semântica*, apresentadas na seção anterior, foram empregadas em sua construção.

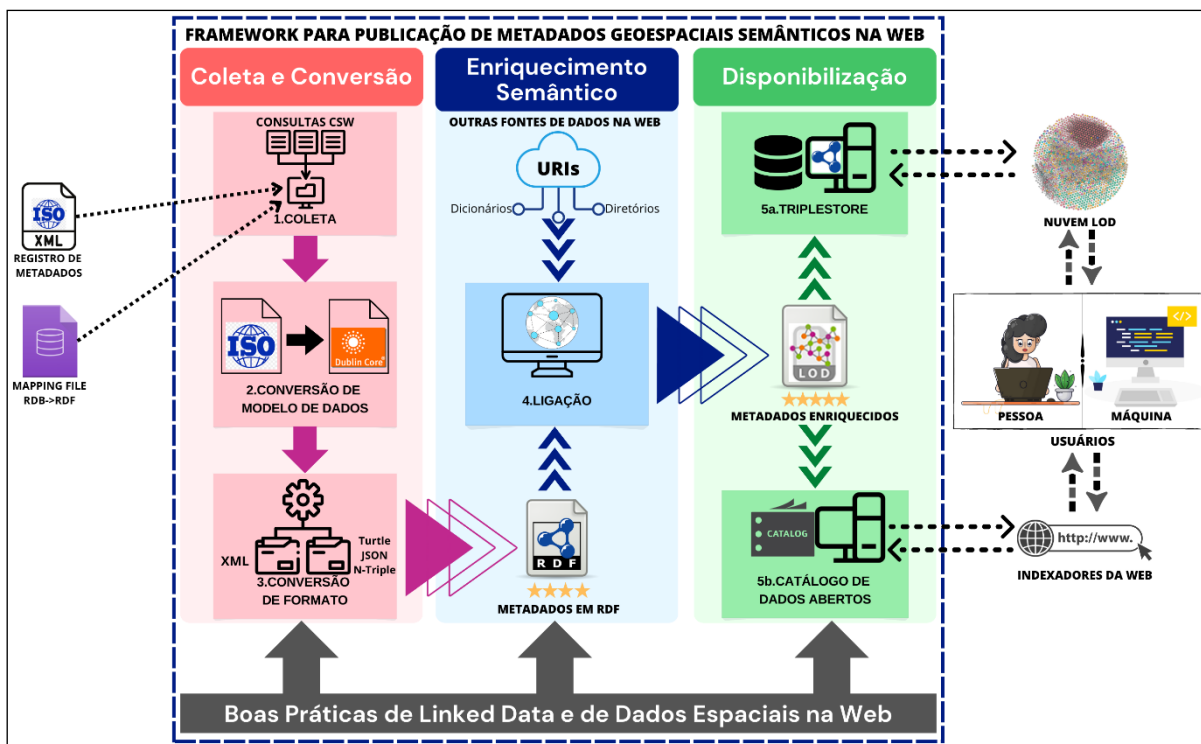
Assim, a primeira subseção trará uma visão geral do *framework*, com suas etapas, procedimentos e interações, e de que forma é feito seu acoplamento a uma IDE. Já na segunda subseção é apresentado o detalhamento de cada uma das camadas (‘Coleta e Conversão’, ‘Enriquecimento Semântico’, e ‘Disponibilização’), cada qual com seus procedimentos, entradas e saídas, além das boas práticas aplicáveis. Por fim, a terceira subseção descreve a prova de conceito realizada para validação do *framework*, bem como os resultados obtidos, que serão analisados frente às questões de pesquisa a serem respondidas por este trabalho.

Com isso, espera-se cumprir com os objetivos específicos: III - Construir um *framework* conceitual para a publicação de metadados geoespaciais na *Web*, descrevendo as etapas e respectivos procedimentos e atores envolvidos; e IV - Validar o *framework* por meio da realização de uma prova de conceito na qual serão publicados os metadados catalogados na IDE de um provedor de dados geoespaciais.

5.1 VISÃO GERAL DO *FRAMEWORK*

O *framework* proposto visa estabelecer um fluxo organizacional para orientar a execução das etapas de um processo de publicação de metadados geoespaciais semânticos na *Web*. Para isso, sua construção teve como base a fundamentação teórica, as tecnologias e ferramentas da *Web Semântica*, as melhores práticas para publicação de dados na *Web*, e as técnicas para publicação de catálogos de metadados na forma de *Linked Data*, todas apresentadas anteriormente neste trabalho. A seguir, a Figura 24 apresenta uma visão geral do *framework*.

Figura 24 – Visão geral do *framework* proposto neste trabalho.



Fonte: Elaborado pelo autor.

Na Figura 24 é possível observar que o *framework* é formado pelo acoplamento de três camadas, nomeadamente ‘Coleta e Conversão’, ‘Enriquecimento Semântico’ e ‘Disponibilização’. Tais camadas representam, respectivamente, as etapas de um fluxo orientado (representado pelas setas vermelhas) para a publicação de metadados geoespaciais semânticos na *Web*, cada qual com entradas e saídas (representadas pelas setas pontilhadas), bem como os procedimentos, que devem observar em sua execução as boas práticas de *Linked Data* e de Dados Espaciais na *Web* (representadas pelas setas de cor grafite).

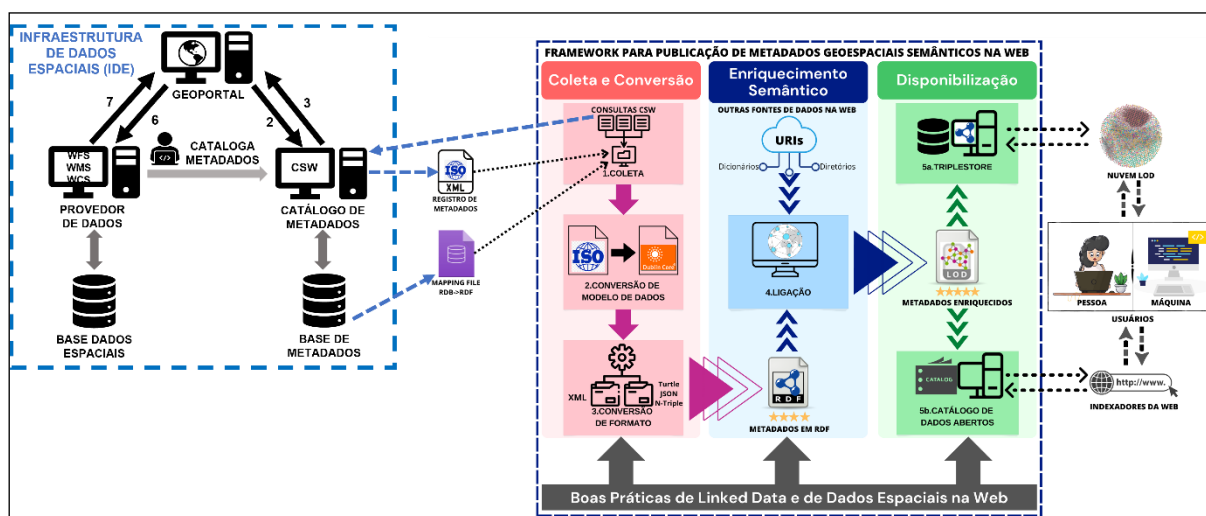
Assim, a primeira camada (Coleta e Conversão) é a responsável pela coleta dos registros de metadados sintáticos no servidor de catálogo de uma IDE e pelas respectivas conversões de modelo de dados e de formato de serialização, resultando em registros de metadados semânticos no formato RDF.

Em seguida, na segunda camada (Enriquecimento Semântico), os metadados em RDF passam por um processo de ligação onde recebem URIs de outras fontes de dados na *Web* (tais como dicionários ou diretórios de nomes geográficos, tesouros, *DBpedia*, etc.) como valores para algumas de suas propriedades. Ao final são entregues metadados enriquecidos na forma de *Linked Data*.

Por fim, a terceira camada (Disponibilização) é a responsável pela publicação na *Web* dos metadados enriquecidos semanticamente, o que pode ser feito por meio de uma plataforma de *triplestore* (para disponibilização na LOD Cloud) e/ou de uma plataforma de catálogo de dados abertos (para indexação na *Web*).

Com isso, uma vez que os metadados geoespaciais tenham sido coletados do catálogo de uma IDE e publicados de forma semântica na *Web* utilizando o *framework* proposto, a expectativa é de que tais metadados (e os respectivos recursos por eles descritos) possam ser descobertos e acessados diretamente na *Web* pelos usuários (humanos e máquinas) por meio de mecanismos de busca ou de uma URI HTTP, ou ainda por meio de consultas semânticas (via console SPARQL) nas quais seja feita a desambiguação de palavras com vários sentidos e a incorporação de todos os significados e sinônimos dos termos fornecidos pelo usuário, de forma a expandir a consulta formulada. A Figura 25 ilustra a integração do *framework* com a arquitetura de uma IDE.

Figura 25 – Integração do *framework* a uma IDE.



Fonte: Elaborado pelo autor.

É possível observar na Figura 25 que a interface entre o *framework* e a plataforma de IDE se dá pela coleta dos registros de metadados que atendem aos critérios estabelecidos nas consultas CSW formuladas ao servidor de catálogo da IDE. Também é possível coletar os registros de metadados por meio de conexão direta (*dump*) à base de dados do servidor de catálogo. Considerando que os registros de metadados dos catálogos das IDEs são predominantemente fornecidos no padrão ISO 19115 e serializados em XML, tal como ocorre

nas IDEs integradas à INDE no Brasil, não há a necessidade de prever ferramentas de interoperabilidade de padrões de metadados nessa interface.

Após essa visão geral, na subseção a seguir são apresentadas, de forma detalhada, cada uma das camadas deste *framework*.

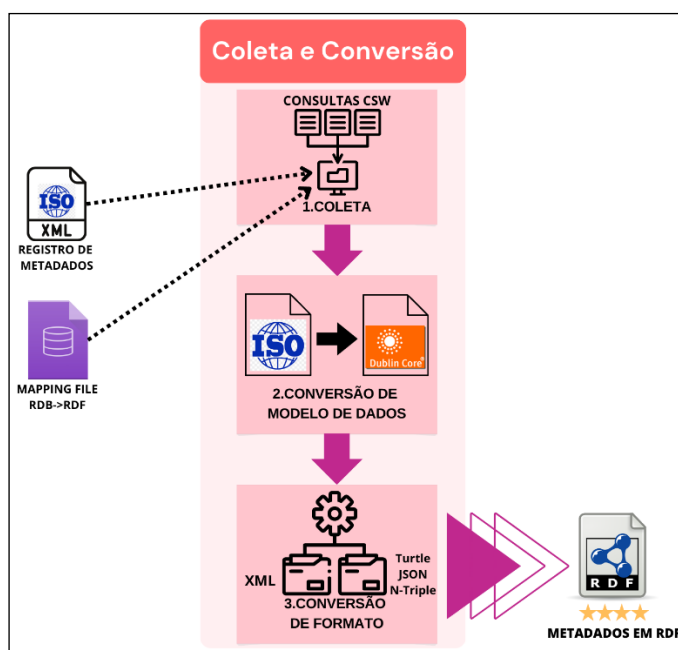
5.2 DETALHAMENTO DO *FRAMEWORK*

Nesta subseção serão descritas cada uma das camadas (vide Figura 24) que compõem o *framework* proposto neste trabalho. É importante ressaltar que todas as camadas foram construídas seguindo as boas práticas já apresentadas, a saber: ‘Boas práticas para publicação de *Linked Data*’; ‘Boas práticas do W3C para publicação de dados na *Web*’; e, ‘Boas práticas da OGC e W3C para dados espaciais na *Web*’. Ao final da descrição de cada camada são relacionadas quais dessas boas práticas foram seguidas.

5.2.1 Camada de Coleta e Conversão

A camada de ‘Coleta e Conversão’, mostrada na Figura 26, é a responsável pela coleta das entradas do *framework*, ou seja, dos registros de metadados geoespaciais sintáticos, em formato XML (três estrelas na classificação de Berners-Lee, 2006).

Figura 26 – Visão da camada de ‘Coleta e Conversão’ do *framework*.



Fonte: Elaborado pelo autor.

Uma vez coletados os registros de metadados, nos procedimentos seguintes dessa camada, ilustradas na Figura 26, são efetuadas a conversão de modelo de dados (de ISO 19115 para RDF) e a conversão de formato de serialização de seus respectivos elementos. Ao final, como resultado (saída), são entregues registros de metadados geoespaciais semânticos em formato RDF (quatro estrelas no sistema de classificação de Berners-Lee, 2006).

A seguir são descritos, em detalhe, os três procedimentos que compõem essa camada.

5.2.1.1 Procedimento de Coleta

O primeiro procedimento dessa camada é o de coleta, a qual pode ser feita das seguintes maneiras: 1) pelo carregamento de um registro individual de metadados, diretamente pelo usuário, ou; 2) a partir de consultas CSW formuladas ao servidor de catálogo da IDE, na quais são obtidos registros individuais ou conjuntos de registros que satisfaçam as restrições das consultas, ou; 3) por *dump* (transferência de todos os registros de uma só vez) da base de dados relacional (RDB – *Relational DataBase*) do servidor de catálogo da IDE.

Assim, a coleta individual de registros de metadados geoespaciais, quando fornecido pelo usuário, se dá através da inserção de arquivos (formato XML) contendo os metadados formatados no modelo de dados ISO 19115.

Já quando os registros de metadados estiverem catalogados numa plataforma de IDE, a coleta é realizada por meio da formulação de consultas para o servidor de catálogo da IDE utilizando o serviço CSW (padrão OGC). Por meio de um *Webservice* são enviadas requisições no formato especificado pelo protocolo do serviço CSW, que são respondidas pelo servidor de catálogo na forma de arquivos XML contendo os registros de metadados que atendam aos critérios da consulta formulada em cada requisição. Um exemplo de requisição e resposta do serviço CSW pode ser visualizado no Quadro 14 apresentado anteriormente.

Por sua vez, quando a coleta de metadados se der por meio de *dump* da base de dados do servidor de catálogo de uma IDE, é necessário fazer um mapeamento do esquema da base de dados relacional (RDB) para o modelo de dados RDF. Tal mapeamento é feito por meio da edição de um ‘arquivo de mapeamento’ (*mapping file*) utilizando a linguagem R2RML³⁶ (*RDB to RDF Mapping Language*), uma recomendação W3C que emprega a sintaxe *Turtle*. Nesse arquivo de mapeamento é feita a associação dos nomes de colunas e dos valores das linhas de cada tabela RDB para, respectivamente, predicados e objetos no modelo de dados RDF. A

³⁶ <https://www.w3.org/TR/r2rml/>

identificação dos sujeitos RDF se dá pela coluna que contém a chave primária da tabela. O Quadro 32 mostra um exemplo de arquivo de mapeamento RDB para RDF utilizando a linguagem R2RML.

Quadro 32 – Exemplo de arquivo de mapeamento RDB para RDF na linguagem R2RML.

```
@prefix rr: <http://www.w3.org/ns/r2rml#>.
@prefix gld: <http://geodados.com.br/ns#>.

<#TriplesMap1>
  rr:logicalTable [ rr:tableName "FUNCIONARIOS" ];
  rr:subjectMap [
    rr:template "http://data.geodados.com.br/funcionarios/{EMPNO}";
    rr:class gld:Funcionario;
  ];
  rr:predicateObjectMap [
    rr:predicate gld:Nome;
    rr:objectMap [ rr:column "NOME" ];
  ];
  rr:predicateObjectMap [
    rr:predicate gld:Cargo;
    rr:objectMap [ rr:column "CARGO" ];
  ];
  rr:predicateObjectMap [
    rr:predicate gld:Depto;
    rr:objectMap [ rr:column "DEPTO" ];
  ];
```

Fonte: Elaborado pelo autor.

No Quadro 32, considere uma base de dados relacional contendo uma tabela chamada ‘FUNCIONARIOS’, que contém três colunas, chamadas ‘EMPNO’, ‘NOME’, ‘CARGO’ e ‘DEPTO’. Assim, para que sejam geradas triplas RDF a partir das linhas dessa tabela, o arquivo de mapeamento deve ser editado da seguinte forma: na seção <#TriplesMap1> a tabela ‘FUNCIONARIOS’ (rr:tableName) tem mapeado como sujeitos das triplas (rr:subjectMap) os registros da coluna ‘EMPNO’ (chave primária da tabela) sob a forma de URI (http://data.geodados.com.br/funcionarios/<valores_de_EMPNO>); já as colunas ‘NOME’, ‘CARGO’ e ‘DEPTO’ tem seus nomes mapeados como predicados (rr:predicate) na forma ‘gld:Nome’, ‘gld:Cargo’ e ‘gld:Depto’ respectivamente, sendo seus registros (linhas) mapeados como objetos (rr:objectMap) das triplas. É possível notar que o prefixo ‘gld:’ nos predicados foi declarado no início do arquivo como *namespace* para <http://geodados.com.br/ns#>.

A edição do arquivo de mapeamento, bem como a configuração da conexão à base de dados relacional e a execução do *dump* para o *framework* pode ser feita através de plataformas disponíveis na Web para tal finalidade, como a D2RQ (Bizer et al., 2007), *Virtuoso* (Blakeley, 2007), entre outras. A execução do *dump* cria um arquivo contendo as triplas RDF geradas a

partir de todos os registros da base de dados mapeada. O Quadro 33 apresenta uma amostra, serializada em *Turtle*, das triplas a serem geradas quando da execução do *dump* da RDB mapeada no Quadro 32.

Quadro 33 – Amostra das triplas RDF geradas no *dump* da RDB mapeada no Quadro 32.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix gld: <http://geodados.com.br/ns#> .

< http://data.geodados.com.br/funcionarios>
a rdfs:Resource ;
    gld:Funcionario <http://data.geodados.com.br/funcionarios/201> ;
    gld:Funcionario <http://data.geodados.com.br/funcionarios/202> ;
    gld:Funcionario <http://data.geodados.com.br/funcionarios/203> .

<http://data.geodados.com.br/funcionarios/201>
a gld:funcionario ;
    gld:Nome "Carlos" ;
    gld:Depto "Exploracao" ;
    gld:Cargo "Geólogo" .

<http://data.geodados.com.br/funcionarios/202>
a gld:funcionario ;
    gld:Nome "Alice" ;
    gld:Depto "Automacao" ;
    gld:Cargo "Engenheira" .

<http://data.geodados.com.br/funcionarios/203>
a gld:funcionario ;
    gld:Nome "Paulo" ;
    gld:Depto "Juridico" ;
    gld:Cargo "Advogado" .
```

Fonte: Elaborado pelo autor.

Cada tripla apresentada no Quadro 33 tem como sujeito (recurso) o código (chave primária), na forma de URI, de um determinado empregado (201, 202 ou 203); como predicado, um termo de vocabulário (*gld:*) criado a partir do nome da respectiva coluna RDB mapeada (*gld:Nome*, *gld:Cargo* ou *gld:Depto*); e, como objeto, o valor de um determinado registro da coluna mapeada.

Assim, uma vez coletados os metadados sintáticos, o próximo procedimento é o de conversão do modelo de dados, descrito a seguir.

5.2.1.2 Procedimento de conversão do modelo de dados

No segundo procedimento desta camada, os metadados coletados (ISO 19115/XML, três estrelas no sistema de classificação de Berners-Lee, 2006) são convertidos para o modelo

de dados RDF. Para isso é realizado um mapeamento cruzado (*crosswalk*) entre esses dois modelos de dados.

No *crosswalk*, elementos do registro de metadados ISO 19115 são selecionados e convertidos para triplas RDF por meio da aplicação do vocabulário DCAT-2. Com isso são produzidos conjuntos de metadados geoespaciais semânticos em RDF (quatro estrelas no sistema de classificação de Berners-Lee, 2006), mantendo a serialização na sintaxe XML. Caso desejado, esses conjuntos de metadados podem ser convertidos, no procedimento seguinte, para quaisquer das sintaxes de serialização apresentadas anteriormente (vide Quadro 18).

É importante ressaltar que nem todos os elementos do registro de metadados ISO 19115 coletado encontrarão elementos correspondentes no vocabulário DCAT-2, porém, os elementos mapeados no *crosswalk* cobrem a maior parte dos elementos comumente encontrados em registros de metadados geoespaciais.

Assim, efetuada a conversão do modelo de dados, os registros de metadados em RDF/XML seguem para o terceiro procedimento desta camada, que é o de conversão do formato de serialização, apresentado a seguir.

5.2.1.3 Procedimento de conversão do formato de serialização

Neste procedimento os conjuntos de metadados já convertidos para o modelo de dados RDF, originalmente serializados em XML, podem ter o formato de serialização convertido para qualquer outro dos formatos apresentados anteriormente (vide Quadro 18). Para implementar este procedimento pode ser utilizada alguma das ferramentas de conversão RDF disponíveis na Web, tais como *EasyRDF*³⁷ e *Datalift*³⁸.

Mais adiante, na subseção de ‘validação do *framework*’, será possível visualizar os resultados das conversões de modelo de dados e de formato de serialização efetuadas em um registro de metadados geoespaciais real, coletado do catálogo de uma IDE atualmente em operação.

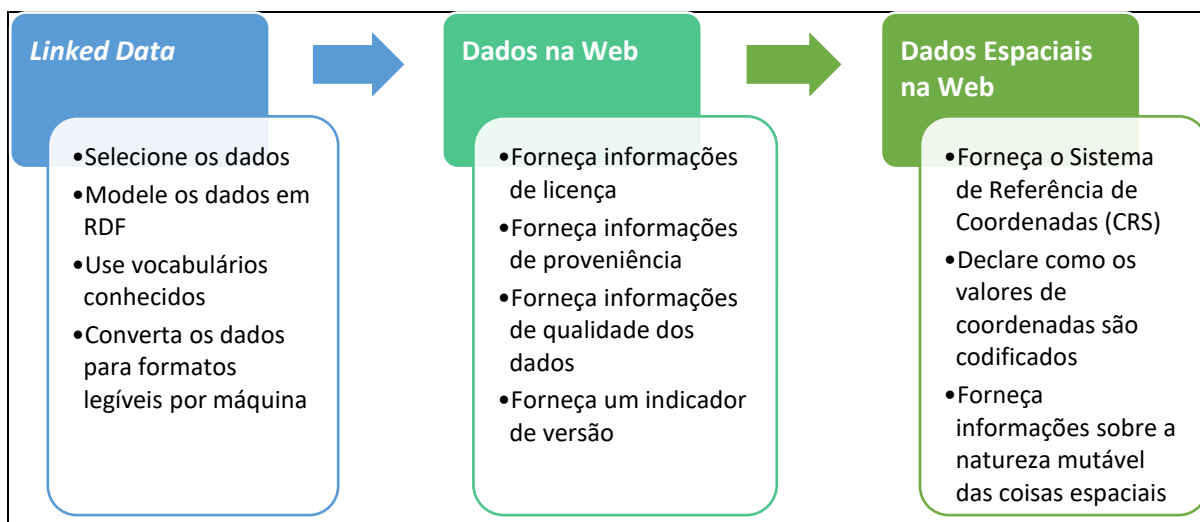
5.2.1.4 Boas práticas seguidas na primeira camada do *framework*

As boas práticas seguidas nesta camada de ‘coleta e conversão’ do *framework* são apresentadas na Figura 27.

³⁷ <https://www.easyrdf.org/>

³⁸ <http://www.datalift.org/>

Figura 27 – Boas práticas seguidas na camada de ‘Coleta e Conversão’ do *framework*.



Fonte: Elaborado pelo autor.

Dentre as boas práticas para publicação de *Linked Data*, listadas na Figura 27, estão as que tratam da seleção, modelagem e conversão dos conjuntos de dados, neste caso os registros de metadados geoespaciais. A partir dessas boas práticas, é recomendado que a escolha dos registros de metadados geoespaciais a serem publicados de forma semântica na *Web* seja feita levando-se em consideração as necessidades e expectativas dos usuários, da forma mais completa possível, usando vocabulários conhecidos da comunidade e convertendo-os para o modelo de dados RDF, o que permite que se tornem legíveis por máquinas.

Quanto às boas práticas para publicação de dados na *Web*, estão aquelas relacionadas às informações mínimas essenciais que os registros de metadados, a serem convertidos para o modelo de dados RDF, devem conter: informações sobre a licença, proveniência e qualidade dos conjuntos de dados que descrevem. É importante também informar a versão do conjunto de dados descrito por esses metadados.

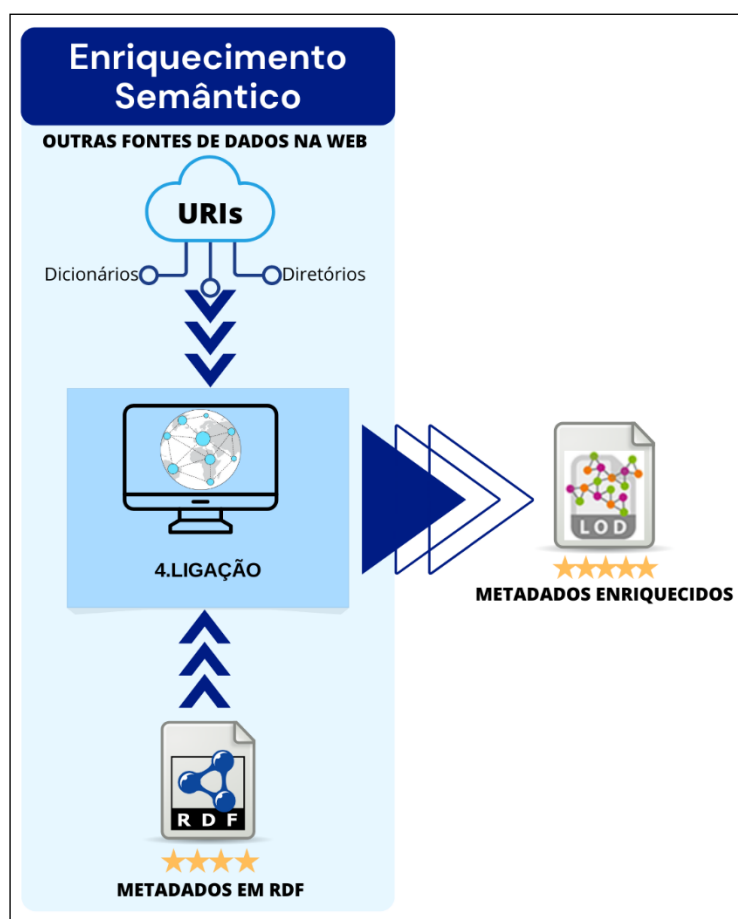
Por fim, dentre as boas práticas para dados espaciais na *Web* estão as que se referem à informação da extensão espacial dos registros de metadados a serem publicados, tais como informar o Sistema de Referência de Coordenadas e qual o formato de valores de coordenadas utilizados nos dados. Além disso, é recomendado que sejam incluídas nos metadados a serem publicados informações sobre a natureza mutável das coisas espaciais, tais como o período temporal ou momento específico a que se referem os conjuntos de dados que estão sendo descritos por esses metadados.

Com isso, o conjunto de metadados geoespaciais em RDF obtido nessa camada podem seguir para a camada seguinte do *framework*, que será descrita a seguir.

5.2.2 Camada de Enriquecimento Semântico

Na segunda camada do *framework*, os conjuntos de metadados geoespaciais em RDF recebem um enriquecimento semântico por meio da ligação de alguns de seus elementos a outras fontes de dados na *Web* (publicadas na nuvem LOD), com a finalidade de levar o usuário à descoberta de outros dados relacionados ou complementares não previstos em sua busca inicial, bem como encontrar informações sobre o mesmo recurso em outras fontes e idiomas. A seguir, a Figura 28 apresenta uma visão dessa camada.

Figura 28 – Visão da camada de ‘Enriquecimento Semântico’.



Fonte: Elaborado pelo autor.

Conforme mostra a Figura 28, os metadados em RDF gerados na camada anterior passam por um procedimento de ligação, onde alguns de seus recursos (ou sujeitos) e valores (ou objetos) recebem URIs para outras fontes na *Web* de dados (nuvem LOD). Ao final deste procedimento são produzidos conjuntos de metadados enriquecidos semanticamente e conectados à nuvem LOD (cinco estrelas no sistema de classificação de Berners-Lee, 2006).

A seguir, é descrito de forma detalhada o único procedimento dessa camada, chamado de ‘Ligação’.

5.2.2.1 Procedimento de ‘Ligação’

No procedimento de ‘Ligação’, o conjunto de metadados em RDF, recebido da camada anterior, tem algumas das propriedades de seus recursos selecionadas para receberem valores na forma de URIs desreferenciáveis para outras fontes de dados na *Web*. Dentre as possíveis fontes a serem conectadas estão os dicionários de nomes geográficos (*Geonames*, TGN – *Getty Thesaurus of Geographic Names*, etc.), os tesouros (*Agrovoc*, *Gemet*, etc.) e os diretórios de autoridades (*VIAF*, *LCNS*, etc.).

A seguir, o Quadro 34 relaciona os elementos de metadados geoespaciais, dentre os provenientes do *crosswalk* efetuado na camada anterior, que podem receber URIs como valores, bem como as indicações de fontes na *Web* para efetivar tais ligações.

Quadro 34 – Sugestões de elementos de metadados que podem ser ligados à *Web* de dados.

ELEMENTO DE METADADOS	FONTES INDICADAS PARA LIGAÇÃO NA WEB DE DADOS	ELEMENTOS A SEREM CONECTADOS NA FONTE
dct:subject	<i>DBpedia</i> <i>Geonames</i> <i>Wikidata</i> <i>Agrovoc</i>	<i>dct:subject / dbo:wikiPageWikiLink</i> <i>gn:Feature</i> <i>wikibase:Item</i> <i>skos:Concept</i>
dct:publisher	<i>VIAF</i> <i>LCNAF</i>	<i>skos:Concept / foaf:primaryTopic</i> <i>madsrdf:CorporateName</i>
adms:representationTechnique	<i>inspire.eu</i> (Diretiva da União Europeia para IDEs)	<i>dct:identifier / dct:title</i>
dct:language	<i>Library of Congress - LOC</i>	<i>madsrdf:Language</i>
dct:spatial	<i>Geonames</i> <i>Getty Thesaurus of Geographic Names</i> (TGN)	<i>gn:Feature</i> <i>gvp:PhysPlaceConcept</i>
dct:license	<i>Creative Commons</i>	<i>cc:attributionName</i> <i>cc:attributionURL</i>
dcat:accrualperiodicity	<i>Statistical Data and Metadata eXchange Community - SDMX</i>	<i>sdmx:Concept</i> <i>sdmx-code:Freq</i>

Fonte: Elaborado pelo autor.

Como pode ser observado no Quadro 34, as principais fontes de dados na *Web* às quais os elementos de metadados *dct:subject*, *dct:publisher*, *dct:language* e *dct:spatial* podem ser ligados são a *DBpedia* (nó central da nuvem LOD), o TGN e *Geonames* (dicionários de nomes

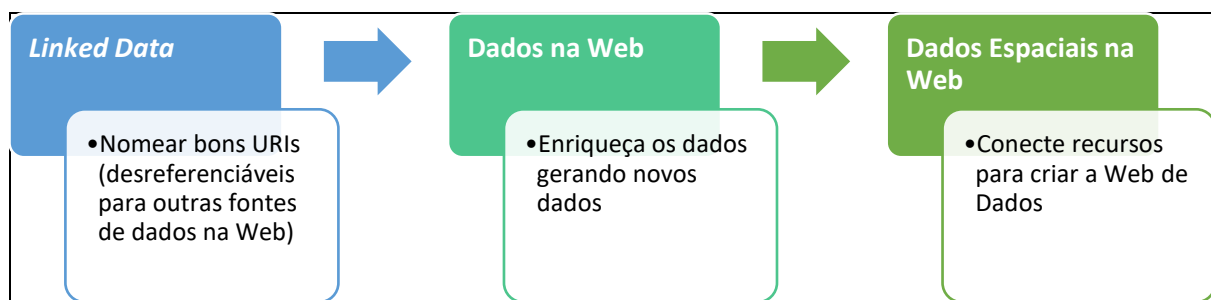
geográficos), e o VIAF (diretório de instituições). Para os demais elementos existem fontes específicas, tais como: *Creative Commons*, para o elemento *dct:license*; *Statistical Data and Metadata eXchange (SDMX) Community*, para o elemento *dcat:accrualperiodicity*; e a Diretiva da União Europeia para IDEs (*inspire.eu*), para o elemento *adms:representationTechnique*.

É importante ressaltar que as listas de elementos de metadados para *crosswalk* apresentados no Quadro 31 e no Quadro 34 não são exaustivas, sendo que adaptações e novas correspondências podem ser formuladas dependendo das necessidades de cada comunidade de usuários. Além disso, resalta-se que não foram criados novos elementos de metadados nesse procedimento, apenas foram estabelecidas correspondências entre elementos já existentes.

5.2.2.2 Boas práticas seguidas na segunda camada do *framework*

As boas práticas para *Linked Data*, dados na *Web* e dados espaciais na *Web*, seguidas nesta camada do *framework* são as indicadas na Figura 29.

Figura 29 – Boas práticas seguidas na camada de ‘Ligação’ do *framework*.



Fonte: Elaborado pelo autor.

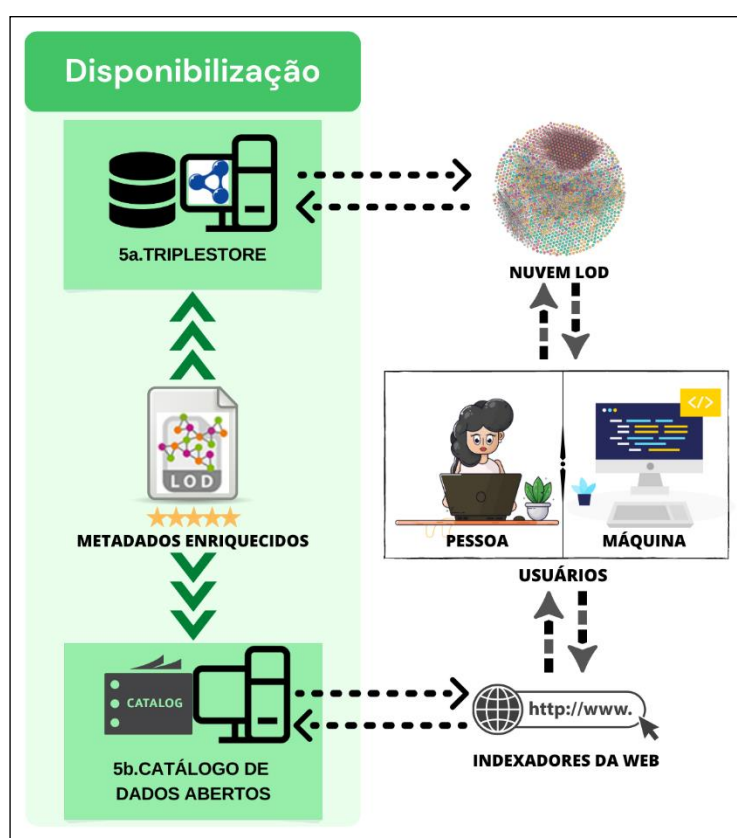
De acordo com a Figura 29, as três boas práticas seguidas nessa camada remetem à recomendação de ligar os elementos de metadados a outras fontes de dados na *Web* a fim de criar novos recursos (expandindo a *Web* de dados) e possibilitar a expansão de consultas a partir desses metadados.

Assim, uma vez obtidos os metadados semânticos (em RDF), ligados à *Web* de dados por meio de URIs, na próxima camada será tratada a disponibilização desses metadados.

5.2.3 Camada de Disponibilização

Na terceira camada do *framework*, os conjuntos de metadados geoespaciais enriquecidos, produzidos na etapa anterior, são armazenados e disponibilizados para indexação, consulta e acesso pelos usuários (pessoas ou máquinas) na *Web*. É a camada que entrega o resultado final (metadados semânticos publicados na *Web*), de forma que o *framework* cumpra com sua finalidade. A Figura 30 apresenta uma visão dessa camada.

Figura 30 – Visão da camada de ‘Disponibilização’ do *framework*.



Fonte: Elaborado pelo autor.

Conforme mostra a Figura 30, a disponibilização dos conjuntos de metadados semânticos pode ser feita a partir de dois componentes: de uma plataforma de *triplestore*, a qual implementa um banco de triplas RDF e uma console (*endpoint*) de consulta no protocolo SPARQL, disponibilizando assim as triplas de metadados na *Web* de dados (nuvem LOD); e de uma plataforma de catálogo de dados abertos (CKAN, *Socrata*, entre outros), permitindo assim que os metadados semânticos sejam mais facilmente indexados pelos motores de busca na *Web*. Esses dois componentes são apresentados, em detalhe, a seguir.

5.2.3.1 Disponibilização via plataforma de *triplestore*

Para que as triplas de metadados geoespaciais semânticos, produzidas por este *framework*, possam ser devidamente publicadas e conectadas na nuvem LOD, é proposta a implantação de uma plataforma de *triplestore* que possua capacidade para lidar com dados e consultas espaciais (compatibilidade com a extensão GeoSPARQL), conforme Quadro 20 apresentado anteriormente.

Dessa forma, a partir do procedimento de ‘Ligação’, executado na camada anterior, os conjuntos de triplas RDF, enriquecidas por meio de URIs para outras fontes na *Web* de dados, podem ser armazenados no banco de triplas e disponibilizados para consultas na *Web* através da console SPARQL, nativos da plataforma escolhida. A partir da plataforma (servidor *Web*) também deve ser possível a publicação desses arquivos, ou efetuar o *dump* do banco de triplas, diretamente na nuvem LOD.

5.2.3.2 Disponibilização via plataforma de catálogo de dados abertos

Com a finalidade de possibilitar que os conjuntos de metadados semânticos sejam indexados na *Web* e, dessa forma, sejam encontrados, consultados e consumidos pelos usuários, é proposta a implantação de uma plataforma de catálogo de dados abertos (CKAN, *Socrata*, entre outras) que possua capacidade (nativa ou por meio de extensão/complemento) para coletar (*harvesting*) registros de metadados no formato RDF.

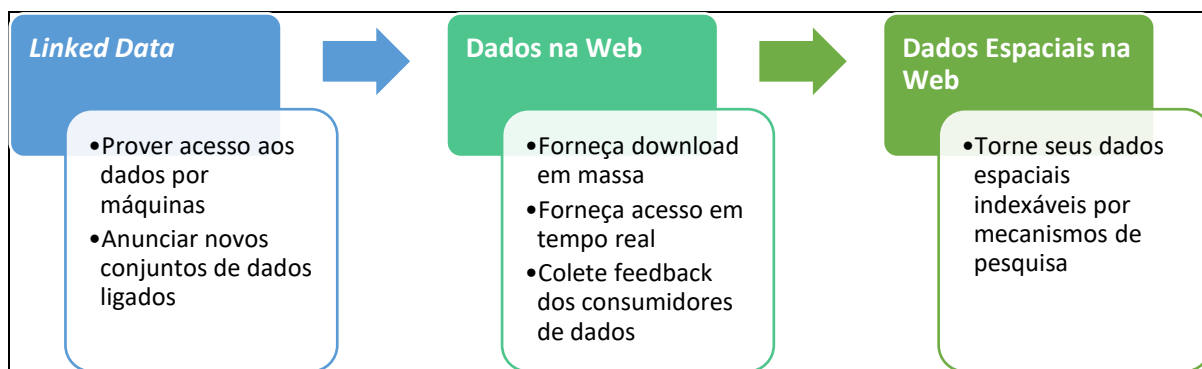
Por meio deste tipo de disponibilização, o procedimento de ‘Ligação’ da camada anterior disponibiliza os metadados em RDF (em qualquer serialização) para coleta pela plataforma de catálogo, a fim de que sejam disponibilizados e indexados na *Web*.

Com isso, pretende-se que qualquer usuário, utilizando qualquer navegador e mecanismo de busca na *Web*, possa encontrar e consultar os metadados catalogados e, a partir deles, explorar outros dados relacionados, ou recursos em outros idiomas, em outras fontes. A partir dessa plataforma também será possível coletar o *feedback* dos usuários, de forma a possibilitar a melhoria contínua do *framework*.

5.2.3.3 Boas práticas seguidas na terceira camada do *framework*

Na camada de ‘Disponibilização’ do *framework* foram seguidas as boas práticas que estão listadas na Figura 31.

Figura 31 – Boas práticas seguidas na camada de ‘Disponibilização’ do *framework*.



Fonte: Elaborado pelo autor.

As ‘boas práticas para publicação de *Linked Data*’, mostradas na Figura 31, se referem às seguintes recomendações: a) prover acesso aos dados por máquinas, uma vez que os metadados publicados estão no modelo de dados RDF com atribuição de URIs, o que possibilita seu processamento por máquinas e agentes de software na *Web*; e b) anunciar novos conjuntos de dados ligados, o que será efetivado a partir da catalogação dos metadados semânticos na plataforma de catálogo de dados abertos e da publicação na nuvem LOD.

Já dentre as ‘boas práticas para dados na *Web*’ se referem às seguintes recomendações: a) forneça *download* em massa, o que será possibilitado pela opção de efetuar *dump* do banco de triplas da plataforma de *triplestore*; b) forneça acesso em tempo real, uma vez que as plataformas de *triplestore* e de catálogo de dados abertos são baseadas em servidores *Web*, bem como os dados ligados a outras fontes na *Web*; e c) colete *feedback* dos consumidores de metadados, o que será possível através da plataforma de catálogo de dados abertos.

Por fim, a recomendação oriunda das ‘boas práticas para dados espaciais na *Web*’ seguida nessa camada foi ‘torne seus dados espaciais indexáveis por mecanismos de pesquisa’, o que será efetivado a partir da disponibilização dos metadados semânticos na plataforma de catálogo de dados abertos.

Concluída a descrição das camadas que compõem o *framework* proposto, na subseção a seguir será realizada sua validação por meio da realização de uma prova de conceito.

5.3 VALIDAÇÃO DO *FRAMEWORK*

Nessa subseção será descrita a prova de conceito realizada com o intuito de validar o *framework* proposto, e que teve como entrada uma amostra de metadados geoespaciais coletado de uma IDE atualmente em operação. Será mostrado o passo a passo de execução dos

procedimentos em cada camada, apresentando como resultado (saída) metadados geoespaciais semânticos publicados na *Web*. Por fim, os resultados obtidos são analisados frente aos objetivos e questões desta pesquisa.

5.3.1 Descrição da amostra

A amostra utilizada para a prova de conceito, com o intuito de validar o *framework* proposto, constitui-se de um registro de metadados que descreve um conjunto de dados (*dataset*) geoespaciais referente aos ‘Incrementos de Desmatamento para o Bioma Amazônia a partir do ano de 2008’.

Este registro de metadados foi coletado da plataforma TerraBrasilis³⁹, uma IDE mantida pelo INPE/MCTI como parte do Projeto de Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite (PRODES). O referido registro encontra-se originalmente formatado no modelo de dados ISO 19115, serializado em XML. Alguns trechos desse registro de metadados são apresentados na Figura 32. O registro completo pode ser consultado no Anexo A deste trabalho.

Figura 32 – Trechos do registro de metadados geoespaciais coletado para a validação.

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gml="http://www.opengis.net/gml"
3    xmlns:gco="http://www.isotc211.org/2005/gco"
4    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
5    xsi:schemaLocation="http://www.isotc211.org/2005/gmd http://www.isotc211.org/2005/gmd/gmd.xsd http://www.isotc211.org/2005/gmx
      http://www.isotc211.org/2005/gmx/gmx.xsd http://www.isotc211.org/2005/srv http://schemas.opengis.net/iso/19139/20060504/srv/srv.xsd"
6  >
7    <gmd:fileIdentifier>
8      <gco:CharacterString>b75b83db-8026-43f9-9537-ee1dfa308158</gco:CharacterString>
9    </gmd:fileIdentifier>
10   <gmd:language>
11     <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="por"/>
12   </gmd:language>
13   <gmd:characterSet>
14     <gmd:MD_CharacterSetCode codeListValue="utf8"
15       codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#MD_CharacterSetCode"/>
16   </gmd:characterSet>
17   <gmd:contact>
18     <gmd:CI_ResponsibleParty>
19       <gmd:individualName>
20         <gco:CharacterString>André Fernando Araújo de Carvalho</gco:CharacterString>
21       </gmd:individualName>
22       <gmd:organisationName>
23         <gco:CharacterString>Divisão de Processamento de Imagens - DPI</gco:CharacterString>
24       </gmd:organisationName>
25       <gmd:positionName gco:nilReason="missing">
26         <gco:CharacterString/>
27       </gmd:positionName>
28     </gmd:CI_ResponsibleParty>
29   </gmd:contact>
30   ...

```

³⁹ <http://terrabrasilis.dpi.inpe.br>

```

63      <gmd:role>
64      <gmd:CI_RoleCode codeListValue="pointOfContact"
65      codeList=
        "http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#CI_RoleCode"/>
66    </gmd:role>
67  </gmd:CI_ResponsibleParty>
68  </gmd:contact>
69  <gmd:dateStamp>
70    <gco:DateTime>2022-06-15T00:14:49</gco:DateTime>
71  </gmd:dateStamp>
72  <gmd:metadataStandardName>
73    <gco:CharacterString>ISO 19115:2003/19139</gco:CharacterString>
74  </gmd:metadataStandardName>
75  <gmd:metadataStandardVersion>
76    <gco:CharacterString>1.0</gco:CharacterString>
77  </gmd:metadataStandardVersion>
78  <gmd:referenceSystemInfo>
79    <gmd:MD_ReferenceSystem>
80      <gmd:referenceSystemIdentifier>
81        <gmd:RS_Identifier>
82          <gmd:code>
83            <gco:CharacterString>EPSG:4674</gco:CharacterString>
84          </gmd:code>
85        </gmd:RS_Identifier>
86      </gmd:referenceSystemIdentifier>
87    </gmd:MD_ReferenceSystem>
88  </gmd:referenceSystemInfo>
89  <gmd:identificationInfo>
90    <gmd:MD_DataIdentification>
91      <gmd:citation>
92        <gmd:CI_Citation>
93          <gmd:title>
94            <gco:CharacterString>Incremento de desmatamento do Bioma Amazônia à partir de 2008</gco:CharacterString>
95          </gmd:title>
96          <gmd:date>
97            <gmd:CI_Date>
98              <gmd:date>
99                <gco:DateTime>2022-06-13T12:00:00</gco:DateTime>
100              </gmd:date>
101              <gmd:dateType>
102                <gmd:CI_DateTypeCode codeListValue="publication"
103                codeList=
                  "http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#CI_DateTypeCode"/>
104              </gmd:dateType>
105            </gmd:CI_Date>
106          </gmd:date>
107        </gmd:CI_Citation>
108      </gmd:citation>
109      <gmd:abstract>
110
111    ...
112
113    <gmd:purpose>
114      <gco:CharacterString>O dado apresenta os incrementos de desmatamento para o Bioma Amazônia à partir de 2008.</gco:CharacterString>
115    </gmd:purpose>
116    <gmd:status>
117      <gmd:MD_ProgressCode codeListValue="onGoing"
118      codeList=
        "http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#MD_ProgressCode"/>
119    </gmd:status>
120    <gmd:pointOfContact>
121      <gmd:CI_ResponsibleParty>
122        <gmd:individualName>
123          <gco:CharacterString>André Fernando Araújo de Carvalho</gco:CharacterString>
124        </gmd:individualName>
125        <gmd:organisationName>
126          <gco:CharacterString>Divisão de Processamento de Imagens - DPI/OBT/INPE</gco:CharacterString>
127        </gmd:organisationName>
128        <gmd:positionName gco:nilReason="missing">
129          <gco:CharacterString/>
130        </gmd:positionName>
131        <gmd:contactInfo>
132
133      ...
134
135      <gmd:useConstraints>
136        <gmd:MD_RestrictionCode codeListValue="license"
137        codeList=
          "http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#MD_RestrictionCode"/>
138      </gmd:useConstraints>
139      <gmd:otherConstraints>
140        <gco:CharacterString>Creative Commons - Atribuição-CompartilhaIgual 4.0 Internacional</gco:CharacterString>
141      </gmd:otherConstraints>
142    </gmd:MD_LegalConstraints>
143    </gmd:resourceConstraints>
144    <gmd:spatialRepresentationType>
145      <gmd:MD_SpatialRepresentationTypeCode codeListValue="vector"
146      codeList=
        "http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#MD_SpatialRepresentationTypeCode"/>
147    </gmd:spatialRepresentationType>
148    <gmd:spatialResolution>
149      <gmd:MD_Resolution>
150        <gmd:equivalentScale>
151          <gmd:MD_RepresentativeFraction>
152            <gmd:denominator>
153              <gco:Integer>250000</gco:Integer>
154            </gmd:denominator>
155          </gmd:MD_RepresentativeFraction>
156        </gmd:equivalentScale>
157      </gmd:MD_Resolution>
158    </gmd:spatialResolution>
159    <gmd:language>
160      <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="por"/>
161    </gmd:language>
162
163    ...

```

```

323 <gmd:extent>
324   <gmd:EX_Extent>
325     <gmd:temporalElement>
326       <gmd:EX_TemporalExtent>
327         <gmd:extent>
328           <gml:TimePeriod gml:id="d18857e395a1051934">
329             <gml:beginPosition>2008-04-01</gml:beginPosition>
330             <gml:endPosition>2021-07-31</gml:endPosition>
331           </gml:TimePeriod>
332         </gmd:extent>
333       </gmd:EX_TemporalExtent>
334     </gmd:temporalElement>
335   </gmd:EX_Extent>
336 </gmd:extent>
337 <gmd:extent>
338   <gmd:EX_Extent>
339     <gmd:geographicElement>
340       <gmd:EX_GeographicBoundingBox>
341         <gmd:westBoundLongitude>
342           <gco:Decimal>-74</gco:Decimal>
343         </gmd:westBoundLongitude>
344         <gmd:eastBoundLongitude>
345           <gco:Decimal>-44</gco:Decimal>
346         </gmd:eastBoundLongitude>
347         <gmd:southBoundLatitude>
348           <gco:Decimal>-17.5</gco:Decimal>
349         </gmd:southBoundLatitude>
350         <gmd:northBoundLatitude>
351           <gco:Decimal>6</gco:Decimal>
352         </gmd:northBoundLatitude>
353       </gmd:EX_GeographicBoundingBox>
354     </gmd:geographicElement>
355   </gmd:EX_Extent>
356 </gmd:extent>
...
465 <gmd:dataQualityInfo>
466   <gmd:DQ_DataQuality>
467     <gmd:scope>
468       <gmd:DQ_Scope>
469         <gmd:level>
470           <gmd:MD_ScopeCode codeListValue="dataset"
471             codeList=
472               "http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#MD_ScopeCode"/>
473         </gmd:level>
474       </gmd:DQ_Scope>
475     </gmd:scope>
476     <gmd:lineage>
477       <gmd:LI_Lineage>
478         <gmd:statement>
479           <gco:CharacterString>O dado foi criado em 1988 e vem sendo mantido desde então. A série publicada com frequência anual inicia em
480             2008 e vem sendo mantida desde então.</gco:CharacterString>
481         </gmd:statement>
482       </gmd:LI_Lineage>
483     </gmd:lineage>
484   </gmd:DQ_DataQuality>
485 </gmd:dataQualityInfo>
486 </gmd:MD_Metadata>

```

Fonte: Elaborado pelo autor.

Os trechos do registro de metadados apresentados na Figura 32 são referentes aos seguintes grupos de elementos descritores: identificação, publicador, sistema de referência de coordenadas, restrições (licença), resolução e extensão espacial e qualidade dos dados.

5.3.2 Execução da prova de conceito

O primeiro procedimento executado no *framework* objetivando a publicação na *Web* da amostra de registro de metadados geoespaciais obtido do catálogo de uma IDE, é o ‘Procedimento de Coleta’ apresentado a seguir.

5.3.2.1 Procedimento de Coleta

Para esta prova de conceito, a amostra de registro de metadados foi carregada manualmente, na forma de um arquivo XML, conforme exibido na Figura 32. No entanto, caso o *framework* estivesse implementado e acoplado à plataforma TerraBrasilis, o mesmo registro

de metadados poderia ter sido coletado de maneira automatizada por meio da formulação de uma consulta ao servidor da referida plataforma utilizando o serviço CSW. No Quadro 35 é apresentado um modelo de consulta para essa forma de coleta.

Quadro 35 – Consulta CSW para coleta automatizada da amostra de registro de metadados.

```
GET/csw/servlet/cswservlet?request=GetRecordbyId&id=b75b83db-8026-43f9-9537-ee1dfa308158&elementSetName=full&outputSchema=http://www.opengis.net/cat/csw/2.0.2 HTTP/1.1
Host: terrabrasilis.dpi.inpe.br/geonetwork/srv/eng/catalog.search#
```

Fonte: Elaborado pelo autor.

A consulta apresentada no Quadro 35 faz uma requisição de um registro de metadados específico, identificado pelo ID (*GetRecordbyId*) ‘b75b83db-8026-43f9-9537-ee1dfa308158’, ao servidor de catálogo da IDE TerraBrasilis (*Host: terrabrasilis.dpi.inpe.br*), o qual deve ser retornado de acordo com o esquema de metadados CSW 2.0.2 (*outputSchema=http://www.opengis.net/cat/csw/2.0.2*) sob o protocolo HTTP/1.1. Nesse esquema, o conjunto de metadados é fornecido serializado em XML, tal como o registro de metadados fornecido manualmente.

Efetuada a coleta do registro de metadados, o próximo procedimento a ser executado é o de ‘Conversão de Modelo de Dados’, conforme apresentado a seguir.

5.3.2.2 Procedimento de Conversão do Modelo de Dados

Nesse procedimento é efetuado um *crosswalk* do modelo de dados ISO 19115 para o modelo de dados RDF utilizando o vocabulário DCAT-2 (recomendação do W3C para a descrição de conjuntos de dados, conforme visto anteriormente). O Quadro 36 apresenta o *crosswalk* dos elementos do registro de metadados ISO 19115 coletado no procedimento anterior.

Quadro 36 – *Crosswalk* entre elementos dos modelos de dados ISO 19115 e RDF utilizando o vocabulário DCAT-2.

ELEMENTO ISO 19115 (<i>MD_Metadata</i> .)	ELEMENTO RDF (vocabulário DCAT-2)
dateStamp	dct:issued
fileIdentifier	dct:identifier
identificationInfo.citation.title	dct:title

identificationInfo.abstract	dct:description
identificationInfo.citation.date	dct:issued
identificationInfo.pointOfContact (role="publisher")	dct:publisher [foaf:Organization]
identificationInfo.pointOfContact (role="author")	dct:creator [foaf:Organization]
identificationInfo.pointOfContact.organisationName	foaf:name
identificationInfo.pointOfContact.individualName	foaf:name
identificationInfo.pointOfContact.contactInfo.address.eletronicMail Address	foaf:mbox
identificationInfo.descriptiveKeywords.keyword (type="theme")	dcat:keyword
identificationInfo.descriptiveKeywords.keyword (type="place")	dct:location
identificationInfo.resourceConstraints (useConstraints="license")	dct:license
identificationInfo.language	dct:language
identificationInfo.topicCategory	dcat:theme [skos:Concept]
identificationInfo.graphicOverview.fileName	foaf:thumbnail
identificationInfo.graphicOverview.fileDescription	foaf:depiction
identificationInfo.resourceMaintenance.maintenanceAndUpdateFrequency	dct:accrualPeriodicity
identificationInfo.spatialRepresentationType	adms:representationTechnique
identificationInfo.spatialResolution	dcat:spatialResolutionInMeters
identificationInfo.extent.temporalElement	dct:temporal [dct:PeriodOfTime]
identificationInfo.extent.graphicElement	dct:spatial [dct:Location]
distributionInfo.transferOptions.onLine.name	dct:title
distributionInfo.transferOptions.onLine.description	dct:description
distributionInfo.transferOptions.onLine.linkage	dcat:accessURL
distributionInfo.distributionFormat	dct:format [dct:IMT]

Fonte: Elaborado pelo autor.

É possível notar, no Quadro 36, que cada elemento de metadado ISO 19115 previamente selecionado foi mapeado para um elemento correspondente do vocabulário DCAT-2, de acordo com o esquema apresentado anteriormente na Figura 19.

Ao final do *crosswalk* ainda é necessário acrescentar a classe *dcat:Catalog* no registro de metadados obtido (DCAT-2), a qual contém as propriedades que descrevem o catálogo de metadados em si, normalmente não encontradas no registro de metadados de origem. As propriedades a serem incluídas na classe *dcat:Catalog* estão relacionadas no Quadro 37.

Quadro 37 – Propriedades da classe *dcat:Catalog* a serem acrescentadas ao registro de metadados obtido do *crosswalk*.

PROPRIEDADE	DESCRIÇÃO
dct:title	Título do catálogo de metadados.
dcat:landingPage	URL da página Web do catálogo de metadados.
dct:publisher	Nome do provedor do catálogo de metadados.
dct:language	Idioma do catálogo de metadados (pode ser diferente do idioma dos registros de metadados)
dcat:catalog	URI do recurso 'Catalog'
dcat:record	URI do recurso 'CatalogRecord'
dcat:dataset	URI do recurso 'Dataset'
dcat:themeTaxonomy	URI do vocabulário temático do catálogo (skos:ConceptSchema)

Fonte: Elaborado pelo autor.

Assim, como resultado deste procedimento é entregue um registro de metadados estruturado no modelo de dados RDF, serializado em XML, com classes e propriedades provenientes do vocabulário DCAT-2, conforme exibido no Quadro 38.

Quadro 38 – Registro de metadados em RDF/XML entregue pelo procedimento de conversão de modelo de dados.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:adms="https://www.w3.org/ns/adms#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:locn="http://www.w3.org/ns/locn#">
  <dcat:Catalog rdf:about="http://terrabrasilis.dpi.inpe.br/catalogs/c0b7783e-
db46-45b6-aef5-d6268dfdef8f">
    <dct:title>TerraBrasilis catalogue</dct:title>
    <dcat:landingPage>http://terrabrasilis.dpi.inpe.br/geonetwork/srv/eng/
catalog.search#/home</dcat:landingPage>
    <dct:publisher rdf:resource="http://terrabrasilis.dpi.inpe.br/organizations/
INPE"/>
    <dct:language>eng</dct:language>
    <dcat:catalog rdf:resource="http://terrabrasilis.dpi.inpe.br/catalogs/
c0b7783e-db46-45b6-aef5-d6268dfdef8f"/>
    <dcat:record rdf:resource="http://terrabrasilis.dpi.inpe.br/records/
b75b83db-8026-43f9-9537-ee1dfa308158"/>
    <dcat:dataset rdf:resource="http://terrabrasilis.dpi.inpe.br/datasets/
b75b83db-8026-43f9-9537-ee1dfa308158"/>
    <dcat:themeTaxonomy rdf:resource="http://terrabrasilis.dpi.inpe.br/
registries/vocabularies/external.place.regions"/>
  </dcat:Catalog>
  <dcat:CatalogRecord rdf:about=
"http://terrabrasilis.dpi.inpe.br/records/b75b83db-8026-43f9-9537-ee1dfa308158">
    <dct:title>Incremento de desmatamento do Bioma Amazônia a partir de
2008</dct:title>
```

```

< dct: description >Área desmatada a partir de 2008 discretizadas por ano e
recortadas pelos limites do bioma Amazônia.</ dct: description >
< dct: issued rdf: datatype="http://www.w3.org/2001/XMLSchema#datetime">
2022-06-15T00:14:49</ dct: issued >
< foaf: primaryTopic rdf: resource="http://terrabrasilis.dpi.inpe.br/datasets/
b75b83db-8026-43f9-9537-ee1dfa308158"/>
</ dcat: CatalogRecord >
< dcat: Dataset rdf: about="http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-
8026-43f9-9537-ee1dfa308158">
  < dct: identifier >b75b83db-8026-43f9-9537-ee1dfa308158</ dct: identifier >
  < dct: title >Incremento de desmatamento do Bioma Amazônia a partir de
2008</ dct: title >
  < dct: issued rdf: datatype="http://www.w3.org/2001/XMLSchema#datetime">
2022-06-13T12:00:00</ dct: issued >
  < dct: description >Área desmatada a partir de 2008 discretizadas por ano e
recortadas pelos limites do bioma Amazônia.</ dct: description >
  < dct: publisher rdf: resource="http://terrabrasilis.dpi.inpe.br/organizations/
INPE"/>
  < dct: creator rdf: resource="http://terrabrasilis.dpi.inpe.br/organizations/
INPE"/>
  < dcat: accrualPeriodicity >annually</ dcat: accrualPeriodicity >
  < foaf: thumbnail >http://terrabrasilis.dpi.inpe.br/geonetwork/srv/api/records
/b75b83db-8026-43f9-9537-ee1dfa308158/attachments/Captura%20de%20tela%20de
%202021-05-21%2011-40-36.png</ foaf: thumbnail >
  < foaf: depiction >Print of yearly deforestation layer.</ foaf: depiction >
  < dcat: keyword >prodes</ dcat: keyword >
  < dcat: keyword >desmatamento</ dcat: keyword >
  < dcat: keyword >incremento</ dcat: keyword >
  < dcat: keyword >Bioma Amazônia</ dcat: keyword >
  < dcat: keyword >features</ dcat: keyword >
  < dct: location >Brazil</ dct: location >
  < dct: location >Amazônia Legal</ dct: location >
  < dct: license >Creative Commons - Atribuição-CompartilhaIgual 4.0
Internacional</ dct: license >
  < adms: representationTechnique >vector</ adms: representationTechnique >
  < dcat: spatialResolutionInMeters rdf: datatype="http://www.w3.org/2001/
/XMLSchema#decimal">250000.0</ dcat: spatialResolutionInMeters >
  < dct: language >por</ dct: language >
  < dcat: theme rdf: resource="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/isoTopicCategory/concepts/imageryBaseMapsEarthCover"/>
  < dcat: theme rdf: resource="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/isoTopicCategory/concepts/environment"/>
  < dcat: theme rdf: resource="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/isoTopicCategory/concepts/geoscientificInformation"/>
  < dct: temporal >
    < dct: PeriodOfTime >
      < dcat: startDate rdf: datatype=
"http://www.w3.org/2001/XMLSchema#date">2008-04-01</ dcat: startDate >
      < dcat: endDate rdf: datatype=
"http://www.w3.org/2001/XMLSchema#date">2021-07-31</ dcat: endDate >
    </ dct: PeriodOfTime >
  </ dct: temporal >
  < dct: spatial >
    < dct: Location >
      < locn: geometry rdf: datatype="http://www.opengis.net/ont/geosparql#
asWKT">POLYGON ((-74 -17.5, -74 6, -44 6, -44 -17.5, -74 -17.5))
      </ locn: geometry >
    </ dct: Location >
  </ dct: spatial >
  < dcat: Distribution rdf: resource="http://terrabrasilis.dpi.inpe.br/
distributions/map"/>
  < dcat: Distribution rdf: resource="http://terrabrasilis.dpi.inpe.br/
distributions/wms"/>
  < dcat: Distribution rdf: resource="http://terrabrasilis.dpi.inpe.br/
distributions/download"/>
</ dcat: Dataset >
< dcat: Distribution rdf: about="http://terrabrasilis.dpi.inpe.br/distributions/
map">

```

```

<dcats:accessURL>http://terrabrasilis.dpi.inpe.br/app/map/deforestation
</dcats:accessURL>
<dcats:title>Portal TerraBrasilis</dcats:title>
<dcats:description>Mapa para acesso online aos dados de desmatamento na
Amazônia e demais biomas</dcats:description>
<dcats:format>
  <dcats:IMT>
    <rdfs:label>WWW:LINK-1.0-http--link</rdfs:label>
  </dcats:IMT>
</dcats:format>
</dcats:Distribution>
<dcats:Distribution rdf:about="http://terrabrasilis.dpi.inpe.br/distributions/
wms">
  <dcats:accessURL>http://terrabrasilis.dpi.inpe.br/geoserver/prodes-amazon-
nb/wms</dcats:accessURL>
  <dcats:title>yearly_deforestation_biome</dcats:title>
  <dcats:description>Yearly deforestation biome since 2008</dcats:description>
  <dcats:format>
    <dcats:IMT>
      <rdfs:label>OGC:WMS-1.3.0-http-get-capabilities</rdfs:label>
    </dcats:IMT>
  </dcats:format>
</dcats:Distribution>
<dcats:Distribution rdf:about="http://terrabrasilis.dpi.inpe.br/distributions/
download">
  <dcats:accessURL>http://terrabrasilis.dpi.inpe.br/downloads/#cat_5
</dcats:accessURL>
  <dcats:title>Desmatamento - Bioma Amazônia</dcats:title>
  <dcats:description>Lista de arquivos Shapefile compactados em formato ZIP para
os temas relacionados ao desmatamento. Acesse para ver os detalhes.
</dcats:description>
  <dcats:format>
    <dcats:IMT>
      <rdfs:label>WWW:DOWNLOAD-1.0-http--download</rdfs:label>
    </dcats:IMT>
  </dcats:format>
</dcats:Distribution>
<skos:ConceptScheme rdf:about="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/external.place.regions">
  <skos:prefLabel xml:lang="pt">Continentes, países e regiões marítimas do
mundo.</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Continents, countries, sea regions of the
world.</skos:prefLabel>
</skos:ConceptScheme>
<skos:Concept rdf:about="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/isoTopicCategory/concepts/imageryBaseMapsEarthCover">
  <skos:prefLabel xml:lang="pt">Mapas com base em imagens de cobertura da
Terra</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Imagery base maps Earth cover</skos:prefLabel>
  <skos:inScheme rdf:resource="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/external.place.regions"/>
</skos:Concept>
<skos:Concept rdf:about="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/isoTopicCategory/concepts/environment">
  <skos:prefLabel xml:lang="pt">Meio Ambiente</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Environment</skos:prefLabel>
  <skos:inScheme rdf:resource="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/external.place.regions"/>
</skos:Concept>
<skos:Concept rdf:about="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/isoTopicCategory/concepts/geoscientificInformation">
  <skos:prefLabel xml:lang="pt">Informação geocientífica</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Geoscientific information</skos:prefLabel>
  <skos:inScheme rdf:resource="http://terrabrasilis.dpi.inpe.br/registries/
vocabularies/external.place.regions"/>
</skos:Concept>
<foaf:Organization rdf:about="http://terrabrasilis.dpi.inpe.br/organizations/
DPI-INPE">

```

```

<foaf:name>Divisão de Processamento de Imagens - DPI/OBT/INPE</foaf:name>
<foaf:homepage>https://www.dpi.inpe.br/DPI</foaf:homepage>
<foaf:member rdf:resource="http://terrabilis.dpi.inpe.br/persons/
prodes"/>
</foaf:Organization>
<foaf:Organization rdf:about="http://terrabilis.dpi.inpe.br/organizations/
INPE">
  <foaf:name>Instituto Nacional de Pesquisas Espaciais - INPE</foaf:name>
  <foaf:homepage>https://www.gov.br/inpe</foaf:homepage>
  <foaf:member rdf:resource="http://terrabilis.dpi.inpe.br/persons/
andre.carvalho"/>
</foaf:Organization>
<foaf:Agent rdf:about="http://terrabilis.dpi.inpe.br/persons/prodes">
  <foaf:name>Projeto de Monitoramento do Desmatamento da Floresta Amazônica
  Brasileira por Satélite</foaf:name>
  <foaf:mailbox>mailto:prodes@inpe.br</foaf:mailbox>
</foaf:Agent>
<foaf:Agent rdf:about="http://terrabilis.dpi.inpe.br/persons/andre.carvalho">
  <foaf:name>André Fernando Araújo de Carvalho</foaf:name>
  <foaf:mailbox>mailto:andre.carvalho@inpe.br</foaf:mailbox>
</foaf:Agent>
</rdf:RDF>

```

Fonte: Elaborado pelo autor.

Efetuada a conversão do modelo de dados, segue-se para o próximo procedimento da primeira camada do *framework*, que efetua a conversão desse registro de metadados para outro formato de serialização RDF como será visto a seguir.

5.3.2.3 Procedimento de Conversão de Formato de Serialização

Para validar este procedimento de conversão de formato de serialização RDF, o conjunto de metadados em RDF/XML, obtido no procedimento anterior, terá sua sintaxe de serialização convertida para *Turtle*, escolhida por ser uma sintaxe que proporciona melhor entendimento por humanos, como se pretende neste trabalho. No entanto, qualquer outra sintaxe de serialização dentre as apresentadas anteriormente (vide Quadro 18) poderia ser aqui utilizada. Diante disso, o Quadro 39 exhibe o conjunto de metadados em RDF serializado na sintaxe *Turtle*.

Quadro 39 – Conjunto de metadados RDF convertido para a sintaxe de serialização *Turtle*.

```

@prefix adms: <https://www.w3.org/ns/adms#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix geosparql: <http://www.opengis.net/ont/geosparql#> .
@prefix locn: <http://www.w3.org/ns/locn#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://terrabilis.dpi.inpe.br/catalogs/c0b7783e-db46-45b6-aef5-d6268dfdef8f>
a dcat:Catalog ;

```

```

dct:title "TerraBrasilis catalogue" ;
rdfs:label "TerraBrasilis catalogue (INPE)"@en , "Catálogo da plataforma TerraBrasilis (INPE)"@pt ;
dcat:landingPage <http://terrabrasilis.dpi.inpe.br/geonetwork/srv/eng/catalog.search#/home> ;
dct:publisher <http://terrabrasilis.dpi.inpe.br/organizations/INPE> ;
dct:language "eng" ;
dcat:catalog <http://terrabrasilis.dpi.inpe.br/catalogs/c0b7783e-db46-45b6-aef5-d6268dfdef8f> ;
dcat:record <http://terrabrasilis.dpi.inpe.br/records/b75b83db-8026-43f9-9537-ee1dfa308158> ;
dcat:dataset <http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-8026-43f9-9537-ee1dfa308158> ;
dcat:themeTaxonomy <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions> .

<http://terrabrasilis.dpi.inpe.br/records/b75b83db-8026-43f9-9537-ee1dfa308158>
a dcat:CatalogRecord ;
dct:title "Incremento de desmatamento do Bioma Amazônia à partir de 2008" ;
dct:description "Área desmatada à partir de 2008 discretizadas por ano e recortadas pelos limites do bioma Amazônia." ;
dct:issued "2022-06-15T00:14:49"^^xsd:datetime ;
foaf:primaryTopic <http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-8026-43f9-9537-ee1dfa308158> .

<http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-8026-43f9-9537-ee1dfa308158>
a dcat:Dataset ;
dct:identifier "b75b83db-8026-43f9-9537-ee1dfa308158" ;
dct:title "Incremento de desmatamento do Bioma Amazônia à partir de 2008" ;
dct:issued "2022-06-13T12:00:00"^^xsd:datetime ;
dct:description "Área desmatada à partir de 2008 discretizadas por ano e recortadas pelos limites do bioma Amazônia." ;
dct:publisher <http://terrabrasilis.dpi.inpe.br/organizations/INPE> ;
dct:creator <http://terrabrasilis.dpi.inpe.br/organizations/DPI-INPE> ;
foaf:thumbnail <http://terrabrasilis.dpi.inpe.br/geonetwork/srv/api/records/b75b83db-8026-43f9-9537-ee1dfa308158/
attachments/Captura%20de%20tela%20de%202021-05-21%2011-40-36.png> ;
foaf:depiction "Print of yearly deforestation layer." ;
dcat:keyword "prodes" ;
dcat:keyword "desmatamento" ;
dcat:keyword "incremento" ;
dcat:keyword "Bioma Amazônia" ;
dcat:keyword "features" ;
dct:location "Brazil" ;
dct:location "Amazônia Legal" ;
dct:license "Creative Commons - Atribuição-CompartilhaIgual 4.0 Internacional" ;
dct:language "por" ;
adms:representationTechnique "vector" ;
dcat:spatialResolutionInMeters "250000"^^xsd:decimal ;
dcat:accrualPeriodicity "annually" ;
dcat:theme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concepts/imageryBaseMapsEarth
Cover> ;
dcat:theme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concepts/environment> ;
dcat:theme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concepts/geoscientificInformation> ;
dct:spatial [
  a dct:Location ;
  locn:geometry ""POLYGON ((-74 -17.5, -74 6, -44 6, -44 -17.5, -74 -17.5))""^^geosparql:asWKT
] ;
dct:temporal [
  a dct:PeriodOfTime ;
  dcat:startDate "2008-04-01"^^xsd:date ;
  dcat:endDate "2021-07-31"^^xsd:date
] ;
dcat:distribution <http://terrabrasilis.dpi.inpe.br/distributions/download> ;
dcat:distribution <http://terrabrasilis.dpi.inpe.br/distributions/wms> ;
dcat:distribution <http://terrabrasilis.dpi.inpe.br/distributions/map> .

<http://terrabrasilis.dpi.inpe.br/distributions/download>
a dcat:Distribution ;
dcat:accessURL <http://terrabrasilis.dpi.inpe.br/downloads/#cat_5> ;
dct:title "Desmatamento - Bioma Amazônia" ;
dct:description "Lista de arquivos Shapefile compactados em formato ZIP para os temas relacionados ao desmatamento.
Acesse para ver os detalhes." ;
dct:format [
  a dct:IMT ;
  rdfs:label "WWW:DOWNLOAD-1.0-http--download"
] .

```

```

<http://terrabrasilis.dpi.inpe.br/distributions/wms>
a dcat:Distribution ;
dcat:accessURL <http://terrabrasilis.dpi.inpe.br/geoserver/prodes-amazon-nb/wms> ;
dct:title "yearly_deforestation_biome" ;
dct:description "Yearly deforestation biome since 2008" ;
dct:format [
  a dct:IMT ;
  rdfs:label "OGC:WMS-1.3.0-http-get-capabilities"
] .

<http://terrabrasilis.dpi.inpe.br/distributions/map>
a dcat:Distribution ;
dcat:accessURL <http://terrabrasilis.dpi.inpe.br/app/map/deforestation> ;
dct:title "Portal TerraBrasilis" ;
dct:description "Mapa para acesso online aos dados de desmatamento na Amazônia e demais biomas" ;
dct:format [
  a dct:IMT ;
  rdfs:label "WWW:LINK-1.0-http--link"
] .

<http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions>
a skos:ConceptScheme ;
skos:prefLabel "Continentes, países e regiões marítimas do mundo."@pt , "Continents, countries, sea regions of the world."@en .

<http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concepts/imageryBaseMapsEarthCover>
a skos:Concept ;
skos:prefLabel "Mapas com base em imagens de cobertura da Terra"@pt , "Imagery base maps Earth cover"@en ;
skos:inScheme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions> .

<http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concepts/environment>
a skos:Concept ;
skos:prefLabel "Meio Ambiente"@pt , "Environment"@en ;
skos:inScheme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions> .

<http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concepts/geoscientificInformation>
a skos:Concept ;
skos:prefLabel "Informação geocientífica"@pt , "Geoscientific information"@en ;
skos:inScheme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions> .

<http://terrabrasilis.dpi.inpe.br/organizations/INPE>
a foaf:Organization ;
foaf:name "Instituto Nacional de Pesquisas Espaciais - INPE" ;
foaf:homepage <https://www.gov.br/inpe> ;
foaf:member <http://terrabrasilis.dpi.inpe.br/persons/andre.carvalho> .

<http://terrabrasilis.dpi.inpe.br/organizations/DPI-INPE>
a foaf:Organization ;
foaf:name "Divisão de Processamento de Imagens - DPI/OBT/INPE" ;
foaf:homepage <https://www.dpi.inpe.br/DPI> ;
foaf:member <http://terrabrasilis.dpi.inpe.br/persons/prodes> .

<http://terrabrasilis.dpi.inpe.br/persons/andre.carvalho>
a foaf:Agent ;
foaf:name "André Fernando Araújo de Carvalho" ;
foaf:mbox <mailto:andre.carvalho@inpe.br> .

<http://terrabrasilis.dpi.inpe.br/persons/prodes>
a foaf:Agent ;
foaf:name "Projeto de Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite" ;
foaf:mbox <mailto:prodes@inpe.br> .

```

Fonte: Elaborado pelo autor.

Com isso, conclui-se a execução da primeira camada do *framework*, na qual foi produzido um conjunto de metadados em RDF (quatro estrelas na classificação de Berners-Lee, 2006), serializado na sintaxe *Turtle*, que será a entrada do próximo procedimento a ser executado, já na segunda camada, conforme descrito a seguir.

5.3.2.4 Procedimento de ‘Ligação’

Na segunda camada do *framework*, o conjunto de metadados em RDF, recebido da camada anterior, passou por um processo de ‘enriquecimento semântico’, no qual os valores (objetos) de algumas de suas triplas foram substituídos por URIs que conectam a outros conjuntos de dados semânticos publicados na *Web* de Dados (nuvem LOD). O Quadro 40 mostra quais elementos de metadados foram selecionados nessa prova de conceito, assim como as fontes de dados na *Web* ao qual foram conectados por meio de URIs.

Quadro 40 - Elementos de metadados RDF a serem conectados à *Web* de dados.

ELEMENTO	VALOR	FONTE DE DADOS NA NUVEM LOD
dct:publisher	“INPE”	VIAF (http://viaf.org/viaf/168786964)
dct:language	“por” “eng”	LOC (http://id.loc.gov/vocabulary/iso639-2/por) LOC (http://id.loc.gov/vocabulary/iso639-2/eng)
dcat:accrualPeriodicity	“Annually”	SDMX (http://purl.org/linked-data/sdmx/2009/code#freq-A)
dcat:keyword	“incremento” “desmatamento” “Bioma Amazônia” “features”	WIKIDATA (https://www.wikidata.org/wiki/Q17281278) DBpedia (http://dbpedia.org/resource/Deforestation) DBpedia (http://pt.dbpedia.org/resource/Amaz%C3%B4nia) WIKIDATA (https://www.wikidata.org/wiki/Q1921825)
dct:location	“Brazil” “Amazônia Legal”	Geonames (https://sws.geonames.org/3469034) Geonames (https://sws.geonames.org/12217728)
dct:license	“Creative Commons...”	CC (http://creativecommons.org/licenses/by-sa/4.0/)
adms:representationTechnique	“vector”	INSPIRE-EU (http://inspire.ec.europa.eu/metadata-codelist/SpatialRepresentationType/vector)
foaf:member	< http://terrabrasilis.dpi.inpe.br/persons/andre.carvalho >	ORCID (https://orcid.org/0000-0002-7893-6672)

Fonte: Elaborado pelo autor.

A partir das fontes de dados na nuvem LOD, mapeadas no Quadro 39, foi possível fazer, de forma manual, o enriquecimento de triplas do conjunto de metadados em RDF (serializado em *Turtle* no Quadro 40) recebido da camada anterior, cujo resultado é apresentado no Quadro

41. No entanto, o ideal é que, quando da implementação deste *framework*, seja construído um script com a finalidade de automatizar este procedimento.

Quadro 41 - Conjunto de metadados RDF enriquecido com ligações à nuvem LOD.

```
@prefix adms: <https://www.w3.org/ns/adms#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix geosparql: <http://www.opengis.net/ont/geosparql#> .
@prefix locn: <http://www.w3.org/ns/locn#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://terrabrasilis.dpi.inpe.br/catalogs/c0b7783e-db46-45b6-aef5-d6268dfdef8f>
a dcat:Catalog ;
dct:title "TerraBrasilis catalogue" ;
rdfs:label "TerraBrasilis catalogue (INPE)"@en ;
dcat:landingPage <http://terrabrasilis.dpi.inpe.br/geonetwork/srv/eng/
catalog.search#/home> ;
dct:publisher <http://viaf.org/viaf/168786964> ;
dct:language <http://id.loc.gov/vocabulary/iso639-2/eng> ;
dcat:catalog <http://terrabrasilis.dpi.inpe.br/catalogs/c0b7783e-db46-45b6-
aef5-d6268dfdef8f> ;
dcat:record <http://terrabrasilis.dpi.inpe.br/records/b75b83db-8026-43f9-9537-
eeldfa308158> ;
dcat:dataset <http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-8026-43f9-
9537-eeldfa308158> ;
dcat:themeTaxonomy <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/
external.place.regions> .

<http://terrabrasilis.dpi.inpe.br/records/b75b83db-8026-43f9-9537-eeldfa308158>
a dcat:CatalogRecord ;
dct:title "Incremento de desmatamento do Bioma Amazônia à partir de 2008" ;
dct:description "Área desmatada à partir de 2008 discretizadas por ano e
recortadas pelos limites do bioma Amazônia." ;
dct:issued "2022-06-15T00:14:49"^^xsd:datetime ;
foaf:primaryTopic <http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-8026-43f9-
9537-eeldfa308158> .

<http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-8026-43f9-9537-eeldfa308158>
a dcat:Dataset ;
dct:identifier "b75b83db-8026-43f9-9537-eeldfa308158" ;
dct:title "Incremento de desmatamento do Bioma Amazônia à partir de 2008" ;
dct:issued "2022-06-13T12:00:00"^^xsd:datetime ;
dct:description "Área desmatada à partir de 2008 discretizadas por ano e
recortadas pelos limites do bioma Amazônia." ;
dct:publisher <http://viaf.org/viaf/168786964> ;
dct:creator <http://terrabrasilis.dpi.inpe.br/organizations/DPI-INPE> ;
foaf:thumbnail <http://terrabrasilis.dpi.inpe.br:/geonetwork/srv/api/records/
b75b83db-8026-43f9-9537-eeldfa308158/attachments/Captura%20de%20tela%20de%20
2021-05-21%2011-40-36.png> ;
foaf:depiction "Print of yearly deforestation layer." ;
dcat:keyword "prodes" ;
dcat:keyword <http://dbpedia.org/resource/Deforestation> ;
dcat:keyword <https://www.wikidata.org/wiki/Q17281278> ;
dcat:keyword <http://pt.dbpedia.org/resource/Amaz%C3%B4nia> ;
dcat:keyword <https://www.wikidata.org/wiki/Q1921825> ;
dct:location <https://sws.geonames.org/3469034> ;
dct:location <https://sws.geonames.org/12217728> ;
dct:license <http://creativecommons.org/licenses/by-sa/4.0/> ;
dct:language <http://id.loc.gov/vocabulary/iso639-2/por> ;
```

```

adms:representationTechnique <http://inspire.ec.europa.eu/metadata-
codelist/SpatialRepresentationType/vector> ;
dcat:spatialResolutionInMeters "250000"^^xsd:decimal ;
dcat:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
dcat:theme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/
isoTopicCategory/concepts/imageryBaseMapsEarthCover> ;
dcat:theme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/
isoTopicCategory/concepts/environment> ;
dcat:theme
<http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concep
ts/geoscientificInformation> ;
  dct:spatial [
    a dct:Location ;
    locn:geometry ""POLYGON ((-74 -17.5, -74 6, -44 6, -44 -17.5, -74 -
17.5))""^^geosparql:asWKT
  ] ;
  dct:temporal [
    a dct:PeriodOfTime ;
    dcat:startDate "2008-04-01"^^xsd:date ;
    dcat:endDate "2021-07-31"^^xsd:date
  ] ;
  dcat:distribution <http://terrabrasilis.dpi.inpe.br/distributions/download> ;
  dcat:distribution <http://terrabrasilis.dpi.inpe.br/distributions/wms> ;
  dcat:distribution <http://terrabrasilis.dpi.inpe.br/distributions/map> .

<http://terrabrasilis.dpi.inpe.br/distributions/download>
a dcat:Distribution ;
dcat:accessURL <http://terrabrasilis.dpi.inpe.br/downloads/#cat_5> ;
dct:title "Desmatamento - Bioma Amazônia" ;
dct:description "Lista de arquivos Shapefile compactados em formato ZIP para os
temas relacionados ao desmatamento. Acesse para ver os detalhes." ;
dct:format [
  a dct:IMT ;
  rdfs:label "WWW:DOWNLOAD-1.0-http--download"
] .

<http://terrabrasilis.dpi.inpe.br/distributions/wms>
a dcat:Distribution ;
dcat:accessURL <http://terrabrasilis.dpi.inpe.br/geoserver/prodes-amazon-
nb/wms> ;
dct:title "yearly_deforestation_biome" ;
dct:description "Yearly deforestation biome since 2008" ;
dct:format [
  a dct:IMT ;
  rdfs:label "OGC:WMS-1.3.0-http-get-capabilities"
] .

<http://terrabrasilis.dpi.inpe.br/distributions/map>
a dcat:Distribution ;
dcat:accessURL <http://terrabrasilis.dpi.inpe.br/app/map/deforestation> ;
dct:title "Portal TerraBrasilis" ;
dct:description "Mapa para acesso online aos dados de desmatamento na Amazônia
e demais biomas" ;
dct:format [
  a dct:IMT ;
  rdfs:label "WWW:LINK-1.0-http--link"
] .

<http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions>
a skos:ConceptScheme ;
skos:prefLabel "Continentes, países e regiões marítimas do mundo."@pt ,
"Continents, countries, sea regions of the world."@en .

<http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concep
ts/imageryBaseMapsEarthCover>
a skos:Concept ;
skos:prefLabel "Mapas com base em imagens de cobertura da Terra"@pt , "Imagery
base maps Earth cover"@en ;

```

```

    skos:inScheme
    <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions>
    .

    <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concepts/environment>
    a skos:Concept ;
    skos:prefLabel "Meio Ambiente"@pt , "Environment"@en ;
    skos:inScheme
    <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions>
    .

    <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/isoTopicCategory/concepts/geoscientificInformation>
    a skos:Concept ;
    skos:prefLabel "Informação geocientífica"@pt , "Geoscientific information"@en ;
    skos:inScheme <http://terrabrasilis.dpi.inpe.br/registries/vocabularies/external.place.regions> .

    <http://viaf.org/viaf/168786964>
    a foaf:Organization ;
    foaf:name "Instituto Nacional de Pesquisas Espaciais - INPE" ;
    foaf:homepage <https://www.gov.br/inpe> ;
    foaf:member <https://orcid.org/0000-0002-7893-6672> .

    <http://terrabrasilis.dpi.inpe.br/organizations/DPI-INPE>
    a foaf:Organization ;
    foaf:name "Divisão de Processamento de Imagens - DPI/OBT/INPE" ;
    foaf:homepage <https://www.dpi.inpe.br/DPI> ;
    foaf:member <http://terrabrasilis.dpi.inpe.br/persons/prodes> .

    <https://orcid.org/0000-0002-7893-6672>
    a foaf:Agent ;
    foaf:name "André Fernando Araújo de Carvalho" ;
    foaf:mbox <mailto:andre.carvalho@inpe.br> .

    <http://terrabrasilis.dpi.inpe.br/persons/prodes>
    a foaf:Agent ;
    foaf:name "Projeto de Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite" ;
    foaf:mbox <mailto:prodes@inpe.br> .

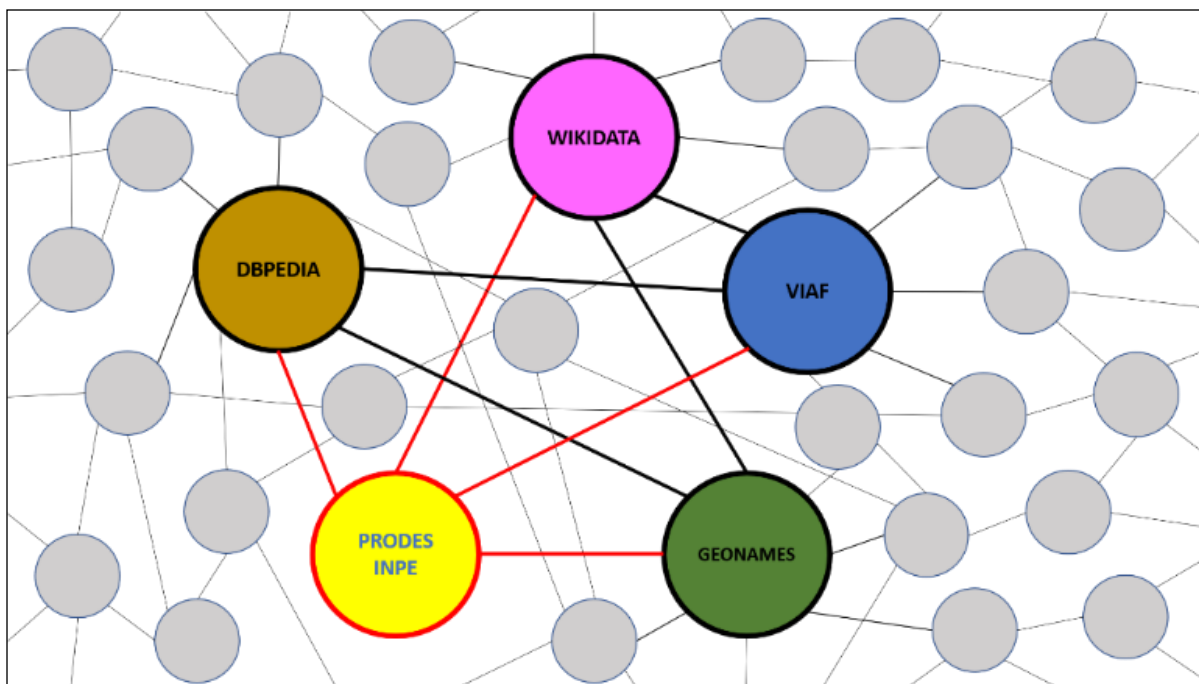
```

Fonte: Elaborado pelo autor.

No Quadro 41 é possível visualizar os elementos do conjunto de metadados desta prova de conceito que foram conectados à Web de dados (destaque em verde), de acordo com o mapeamento efetuado no Quadro 40. Com isso, foi obtido um conjunto de metadados em RDF, conectado à nuvem LOD, conforme ilustra a Figura 33, de forma a alcançar a quinta estrela (conceito máximo) na classificação de Berners-Lee (2006).

Na Figura 33 é possível visualizar que a ligação da amostra de metadados (PRODES INPE) à nuvem LOD se deu por meio de URIs para os seguintes nós: *DBpedia*, *Geonames*, *VIAF* e *Wikidata*. A partir deles é possível chegar a diversas outras fontes de dados ligadas a esses nós, o que levará o usuário a explorar novas fontes de informação, em diversos idiomas, relacionados ao campo de metadados de origem.

Figura 33 – Conexão do conjunto de metadados do *framework* à nuvem LOD.



Fonte: Elaborado pelo autor.

Assim, conclui-se a execução da segunda camada do *framework*, que entregou como resultado um conjunto de metadados em RDF ligados à nuvem LOD, que seguirá para a próxima camada, responsável por publicá-lo na *Web*, como será visto a seguir.

5.3.2.5 Publicação dos metadados semânticos

Na terceira camada do *framework* foi efetuada a publicação na *Web* do conjunto de metadados em RDF (ligados à nuvem LOD) que foram produzidos na camada anterior. Tal publicação foi feita pelas duas formas apresentadas anteriormente: por meio de uma plataforma de *triplestore*, e de uma plataforma de catálogo de dados abertos, conforme a seguir.

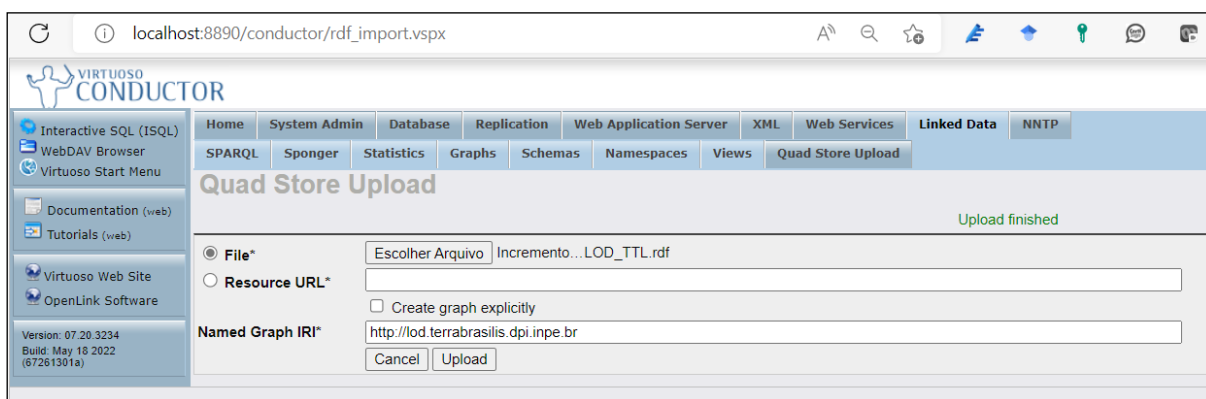
5.3.2.5.1 Disponibilização por meio de plataforma de *triplestore*

Para possibilitar o armazenamento, a disponibilização na *Web* de dados e a consulta semântica (via console SPARQL) do conjunto de triplas RDF, gerado a partir da amostra de registro de metadados processada nas camadas anteriores desse *framework*, foi usada a plataforma de *triplestore* ‘*Virtuoso Open Source (VOS)*’. A plataforma VOS foi escolhida para essa prova de conceito devido a interface de sua ferramenta de administração, via console *Web*,

e por seu banco de triplas ser compatível com consultas GeoSPARQL, uma vez que se trata de triplas contendo extensão espacial.

A seguir, na Figura 34, é exibida a console da ferramenta de gerenciamento da plataforma VOS (*Virtuoso Conductor*), na guia onde o arquivo contendo as triplas RDF, conectadas à LOD e serializadas em sintaxe *Turtle*, recebido da camada anterior, é carregado no banco de triplas (*Quad Store Upload*).

Figura 34 – Carregamento do conjunto de triplas RDF na plataforma de *triplestore*.



Fonte: Elaborado pelo autor.

Por meio da interface *Web* de administração (*Virtuoso Conductor*) da plataforma VOS, mostrada na Figura 34, na aba '*Linked Data*', ao selecionar a opção '*Quad Store Upload*' é aberto um formulário para selecionar o arquivo, ou indicar a URL do recurso, que contém as triplas RDF a serem carregadas para o *triplestore*.

Assim, uma vez que o arquivo contendo o conjunto de triplas RDF esteja carregado na plataforma de *triplestore*, passa a ser possível consultá-los por meio da console SPARQL, conforme exemplo ilustrado na Figura 35.

Na consulta SPARQL apresentada na Figura 35, foi solicitada a exibição de todas as triplas (cláusula *select **) do conjunto de triplas carregado no *triplestore* (cláusula *from*) que tenham como recurso (sujeito) a URI `<http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-8026-43f9-9537-ee1dfa308158>` (cláusula *where*). Como resultado, a console retorna as propriedades (predicados) e respectivos valores (objetos) das triplas de metadados que descrevem o referido recurso. Nesses resultados é possível observar que alguns dos valores (objetos) estão na forma de URI (em azul) desreferenciável para outro recurso na *Web* de dados.

Figura 35 – Consulta SPARQL sobre o conjunto de triplas carregado no *triplestore*.

The screenshot shows the Virtuoso SPARQL Execution interface. On the left is a sidebar with navigation links like 'Interactive SQL (ISQL)', 'WebDAV Browser', 'Virtuoso Start Menu', 'Documentation (web)', 'Tutorials (web)', 'Virtuoso Web Site', and 'OpenLink Software'. The main area has a top navigation bar with tabs: Home, System Admin, Database, Replication, Web Application Server, XML, Web Services, Linked Data, and NNTP. Below this is the 'SPARQL Execution' section. It includes a 'Query' tab and a 'Saved Queries' tab. The 'Default Graph IRI' is set to 'http://lod.terrabrasilis.dpi.inpe.br'. The query text is:


```
select * from <http://lod.terrabrasilis.dpi.inpe.br>
where {<http://terrabrasilis.dpi.inpe.br/datasets/b75b83db-8026-43f9-9537-ee1dfa308158> ?p ?o}
```

 Below the query are buttons for 'Execute', 'Save', 'Load', and 'Clear'. The results are displayed in an 'HTML5 table' with two columns: 'p' and 'o'. The table contains 14 rows of data, each representing a triple from the dataset.

p	o
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/ns/dcat#Dataset
http://www.w3.org/ns/dcat#spatialResolutionInMeters	250000
https://www.w3.org/ns/adms#representationTechnique	http://inspire.ec.europa.eu/metadata-codelist/SpatialRepresentationType/vector
http://purl.org/dc/terms/creator	http://terrabrasilis.dpi.inpe.br/organizations/DPI-INPE
http://purl.org/dc/terms/description	"Área desmatada à partir de 2008 discretizadas por ano e recortadas pelos limites do bioma Amazônia."
http://purl.org/dc/terms/identifier	"b75b83db-8026-43f9-9537-ee1dfa308158"
http://purl.org/dc/terms/issued	"2022-06-13T12:00:00"^^<http://www.w3.org/2001/XMLSchema#datetime>
http://purl.org/dc/terms/language	http://id.loc.gov/vocabulary/iso639-2/por
http://purl.org/dc/terms/license	http://creativecommons.org/licenses/by-sa/4.0/
http://purl.org/dc/terms/location	https://sws.geonames.org/3469034
http://purl.org/dc/terms/location	https://sws.geonames.org/12217728
http://purl.org/dc/terms/publisher	http://viaf.org/viaf/168786964

Fonte: Elaborado pelo autor.

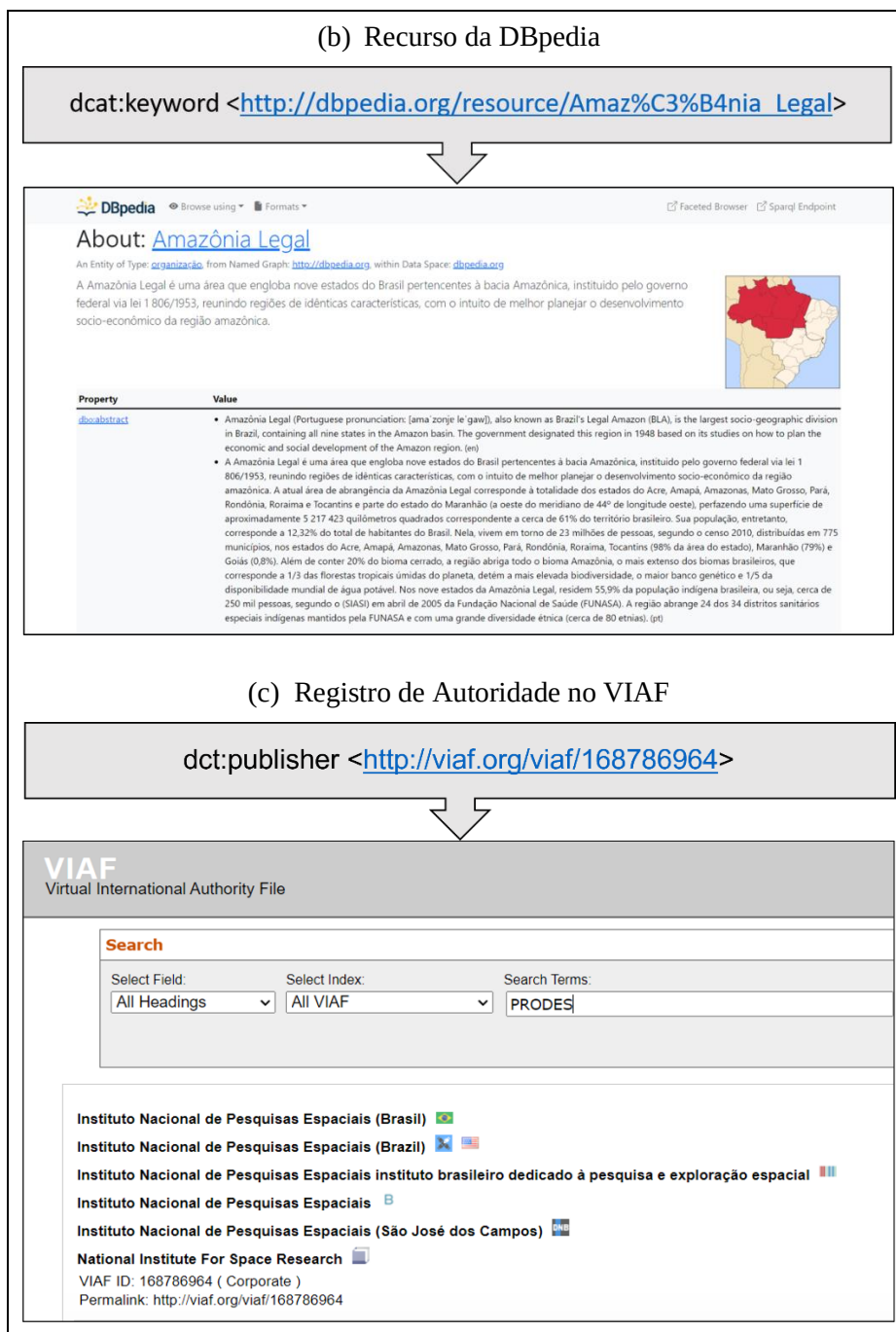
Na Figura 36 são ilustrados algumas das fontes de dados na *Web* a que levam tais URIs, de forma a demonstrar que os metadados foram disponibilizados na *Web* sob o paradigma dos dados ligados (*Linked Data*), cumprindo, dessa forma, o propósito do *framework* apresentado neste trabalho.

Figura 36 – Fontes de dados alcançáveis a partir dos resultados da consulta SPARQL.

(a) Localidade no Geonames

dct:location <<https://sws.geonames.org/12217728>>

The screenshot shows the GeoNames website interface. At the top, there's a navigation bar with links like 'About', 'Browse', 'Download', 'API', and 'Help'. The main content area displays a map of South America, specifically focusing on Brazil. A location marker is placed over the Amazon region, with a pop-up window showing details for 'Amazonia Legal'. The pop-up includes the location name, coordinates (5° 7' 45.07" S, 63° 02' 17" W), and a link to the location's page on GeoNames. The map is interactive, with various controls for zooming and navigation.



Fonte: Elaborado pelo autor.

A partir dos valores (objetos) das triplas na forma de URI desreferenciável, apresentadas nos resultados da consulta Sparql executada, caso o usuário deseje saber mais sobre a que localidade se refere o *dataset* descrito pelos metadados apresentados, basta clicar no valor correspondente da propriedade *dct:location*. Neste caso, a URI <<https://sws.geonames.org/12217728>> conduzirá ao registro do dicionário geográfico *Geonames* referente à localidade ‘Amazônia Legal’, conforme ilustrado na Figura 36(a).

Da mesma forma, a partir desse registro do *Geonames*, ou a partir do valor correspondente da propriedade *dcat:keyword* (http://dbpedia.org/resource/Amaz%C3%B4nia_Legal), será possível chegar ao recurso da *DBpedia* sobre a ‘Amazônia Legal’, conforme ilustrado na Figura 36(b). E, caso o usuário queira ainda saber mais informações sobre o publicador do *dataset* e do conjunto de metadados, basta clicar no valor correspondente da propriedade *dct:publisher* (<http://viaf.org/viaf/168786964>) para ser levado ao registro sobre o ‘Instituto Nacional de Pesquisas Espaciais (INPE)’ naquela base de dados de autoridades, conforme ilustrado na Figura 36(c).

A seguir será apresentada a outra forma de disponibilização dos conjuntos de metadados semânticos gerados pelo *framework*, que é a partir de uma plataforma de catálogo de dados abertos.

5.3.2.5.2 Disponibilização por meio de plataforma de catálogo de dados abertos

A plataforma de catálogo de dados abertos escolhida para disponibilização do conjunto de metadados dessa prova de conceito foi a *Comprehensive Knowledge Archive Network* (CKAN)⁴⁰, desenvolvida pela *Open Knowledge Foundation* (OKF), e que possui os requisitos técnicos para indexação na *Web* pelo motor de busca do *Google*. O CKAN é a plataforma usada para a operacionalização do ‘Portal Brasileiro de Dados Abertos (dados.gov.br)’⁴¹. Dessa forma, dado que o conjunto de metadados utilizado como amostra (PRODES/INPE) foi obtido de uma IDE (plataforma TerraBrasilis) operada por um Instituto de Pesquisas do Poder Executivo Federal, sua publicação também deve ser feita nesse portal, conforme mostra a Figura 37.

Para disponibilizar o conjunto de metadados em RDF (LOD) obtido da execução das camadas anteriores do *framework*, o arquivo (*Incremento_Prodes_RDF_LOD.ttl*) contendo as triplas serializadas na sintaxe *Turtle* foi cadastrado, conforme mostrado na Figura 37, como um recurso (Incremento do Desmatamento do Bioma Amazônia à...) de um registro⁴² do catálogo do INPE, chamado ‘Metadados em RDF (LOD) da Plataforma Terrabrasilis’. Assim, ao consultar o referido registro no portal dados.gov.br, o usuário poderá fazer o *download* daquele arquivo. Da mesma forma, o registro poderá ser encontrado por meio de buscadores na *Web*, assim que for indexado pelos motores dos mecanismos de busca na *Web*.

⁴⁰ <https://ckan.org>

⁴¹ <https://dados.gov.br>

⁴² <https://dados.gov.br/dataset/prodesrdf>

Figura 37 – Conjunto de metadados em RDF disponibilizado no portal *dados.gov.br*.

(a) Metadados em RDF (LOD) da plataforma TerraBrasilis

Conjuntos de metadados do catálogo da plataforma TerraBrasilis/INPE convertidos em formato RDF e ligados à Web de dados (LOD). EXPERIMENTAL.

Estes dados estão disponíveis como o esperado?

0 5m 6h Não 0

Dados e recursos

Incremento de desmatamento do Bioma Amazônia à ...

Metadados descritivos do dataset "Incremento de desmatamento do Bioma..."

Informações Adicionais

Campo	Valor
Fonte	http://terrabrasilis.dpi.inpe.br/igone/work/srweb/catalog/search/home
Autor	Clayton Martins Pereira
Mantenedor	Clayton Martins Pereira
Versão	1.0
Estado	active
Última Atualização	20 de Setembro de 2022, 17:11 (BRT)
Criado	20 de Setembro de 2022, 16:58 (BRT)

(b) Incremento de desmatamento do Bioma Amazônia à ...

URL: http://mrc-m21d.sid.inpe.br/attachement.cgi?id=inpe-brimrc-m21d20220920.19.52/doc/Incremento_Prodes_LOD_TTL.rdf

Metadados descritivos do dataset "Incremento de desmatamento do Bioma Amazônia à partir de 2008" no formato RDF e ligados à Web de dados (LOD).

Não há visões criadas para este recurso ainda.

@ não está vendo as visões que esperava? Clique aqui para saber mais ...

Recursos

Incremento de

Informações Adicionais

Campo	Valor
Ultima atualização	20/Setembro/2022
Criado	20/Setembro/2022
Formato	RDF
Licença	Creative Commons Atribuição e Compartilhamento pela mesma Licença
created	5 minutos atrás
format	RDF
id	a5b7c263-0702-4046-a0b4-e74900f0b442
package id	54645c5e-e17b-4d36-b5a9-d4ba13d32b52
revision id	c79d3704-516d-478c-a0b4-05a987af275c
state	active

Fonte: Elaborado pelo autor.

Com isso, conclui-se a execução da prova de conceito e a respectiva validação do *framework* conforme proposto inicialmente. Na subseção a seguir será feita a análise dos resultados desta prova de conceito, à luz dos objetivos e questões de pesquisa deste trabalho.

5.3.3 Análise dos resultados

Neste experimento, que serviu de prova de conceito para validação do *framework* proposto nesse trabalho, foi possível acompanhar, passo a passo, a execução de cada um dos procedimentos integrantes de cada uma de suas camadas. Isso, de forma a comprovar sua eficácia ao receber um registro de metadados geoespaciais sintáticos, oriundo de uma plataforma de IDE em operação, e entregar como resultado final o registro de metadados convertido para o formato de triplas RDF conectadas à Web de dados (nuvem LOD) e devidamente publicadas na Web, seguindo os conjuntos de boas práticas indicados e os princípios do *Linked Data*, como proposto no objetivo principal.

Na primeira camada (Coleta e Conversão) foi possível verificar que a coleta de registros de metadados (procedimento de Coleta) pode se dar de forma manual, como demonstrado, ou ainda de forma automatizada, por meio da formulação de consultas CSW ou de *dump* da base de dados utilizando aplicativos *open source* existentes, tal como a própria plataforma VOS aplicada na camada de ‘Disponibilização’.

Além disso, a realização de *crosswalks* entre os modelos de dados ISO 19115 e RDF no procedimento de ‘Conversão de modelo de dados’, conforme visto em Nogueras-Iso (2005), se mostrou viável e sem ocasionar perda significativa de contexto ao mapear, um para um, as classes e propriedades do primeiro modelo para o segundo, utilizando para isso o vocabulário DCAT-2, como complemento ao vocabulário *Dublin Core*. Conforme mencionado anteriormente, e apontado por REID et al. (2012), apesar de que nem todos os elementos do registro de metadados de origem encontraram elementos correspondentes no vocabulário DCAT-2, os elementos mapeados no *crosswalk* cobrem a maior parte dos elementos comumente encontrados em registros de metadados geoespaciais.

Já no terceiro procedimento desta camada inicial (procedimento de Conversão de Formato de Serialização), o usuário pode decidir manter o formato de serialização padrão (RDF/XML) para os metadados convertidos em triplas, ou convertê-los para outro formato de serialização, por meio da aplicação de ferramentas existentes na *Web*, conforme demonstrado neste experimento, onde as triplas de metadados foram serializadas na sintaxe *Turtle*, conforme recomendado por Heath e Bizer (2011) para leitura e escrita humana de triplas.

Ao final desses três procedimentos, a primeira camada entregou, como resultado de sua execução, um arquivo de metadados geoespaciais semânticos no formato RDF (quatro estrelas na classificação de Berners-Lee, 2006), interoperáveis semanticamente, e legíveis por máquinas e agentes de software na *Web*, de forma que fica demonstrada e respondida a primeira questão de pesquisa levantada no início deste trabalho (*Como efetuar a conversão de padrões e formatos dos metadados geoespaciais catalogados em uma IDE, de forma a torná-los interoperáveis em nível semântico e legíveis também por máquinas e agentes de software na Web?*).

Na execução da segunda camada do *framework* foi possível verificar que a substituição dos valores ‘literais’ de algumas propriedades do conjunto de metadados em RDF, recebido da camada anterior, por URIs desreferenciáveis para outras fontes de dados na *Web* (procedimento de Ligação), conforme recomendado por Heath e Bizer (2011) e seguindo os princípios do *Linked Data* (Berners-Lee, 2006), proporcionou o enriquecimento semântico dos metadados, que passaram a ser sensíveis ao contexto, além de permitir que o usuário possa explorar novos recursos relacionados, bem como em outros idiomas, expandindo a busca inicialmente formulada, de forma a contribuir para a expansão da rede global de dados mencionada em Po et al. (2020).

Com isso, a segunda camada entregou, como resultado de sua execução, um novo arquivo contendo os mesmos metadados na forma de triplas, recebidos da camada anterior, porém com algumas das triplas ligadas, por meio de URIs, à *Web* de dados (nuvem LOD),

incluindo URIs para recursos da *DBpedia*, conforme recomendado por Ronzhin et al. (2019), de forma que fica demonstrada e respondida, diante do apresentado no Quadro 40, a segunda questão de pesquisa deste trabalho (*Quais elementos de metadados geoespaciais poderiam ser ligados a outras fontes de dados na Web, de forma a enriquecê-los semanticamente e possibilitar a descoberta de informações adicionais pelos consumidores desses metadados?*).

Por fim, a execução da terceira camada do *framework* promoveu a publicação na *Web* do conjunto de metadados geoespaciais enriquecidos semanticamente, através de duas plataformas, uma de *triplestore* (VOS), e outra de catálogo de dados abertos (CKAN). A plataforma VOS foi instalada e configurada especificamente para este experimento, enquanto no caso da plataforma CKAN foi utilizado o ‘Portal Brasileiro de Dados Abertos (dados.gov.br)’, atualmente em operação e disponível na *Web*.

A instalação, configuração e utilização da plataforma VOS se mostrou viável para aplicação no *framework*, uma vez que sua console SPARQL possui compatibilidade com a extensão GeoSPARQL, tal como indicado por Huang et al. (2019), permitindo a consulta de triplas contendo elementos geoespaciais, além de oferecer uma console de gerenciamento *Web* com interface gráfica, o que torna seu uso intuitivo. Quanto à plataforma CKAN, essa possui compatibilidade com recursos no formato RDF, bem como oferece um coletor automatizado (*dcat_rdf_harvester*) para recursos nesse formato, o que a torna uma opção viável de plataforma de catálogo de dados abertos a ser utilizada na implementação do *framework*.

A automatização do *crosswalk* efetuado no procedimento de ‘Conversão de Modelo de Dados’, na primeira camada, pode ser efetivada por meio da implementação de um ‘Processador de Transformação XML’, baseado na ‘Linguagem de Folhas de Estilo (XSL – *eXtensible Stylesheet Language*)’, também conhecido como ‘Processador XSLT (*eXtensible Stylesheet Language transformation*)’, conforme proposto em Fugazza et al., (2021). Tal implementação poderia ocorrer na forma de uma aplicação, que faria as associações entre os elementos de ambos os modelos de dados, por meio de uma folha de estilo XSL a ser aplicada ao arquivo XML (ISO 19115) coletado no procedimento inicial, gerando como saída um novo arquivo XML estruturado com as triplas RDF formadas.

A construção deste *framework* para publicação de metadados geoespaciais, na forma de um fluxo organizacional, com fases (camadas) compostas de várias tarefas (procedimentos), com suas entradas e saídas, seguindo o modelo proposto por Santarem Segundo (2018), se mostrou eficaz para atender os propósitos e objetivos desejados, permitindo ainda sua implementação em módulos que podem ser adaptados às necessidades dos usuários.

Além disso, na implementação dos procedimentos de ‘Conversão de Modelo de Dados’ e de ‘Ligação’ é recomendado contar com a colaboração de um profissional da informação, especialista em vocabulários, tesouros, diretórios e dicionários, a fim de assegurar o correto mapeamento e ligação dos elementos de metadados durante sua execução.

A aplicação das boas práticas do W3C (2017) para dados na *Web*, e do W3C e OGC (2017) para dados espaciais na *Web*, combinadas com as boas práticas e princípios do *Linked Data*, e da conexão dos elementos de metadados geoespaciais ao *Geonames* e *DBpedia*, conforme visto em Vilches-Blásquez e Saadrevia (2019), permitiram a transformação e publicação de elementos geográficos na forma de ‘*Linked GeoData*’, tornando-os úteis à *Web* de dados como elementos de ligação e hub para outros conjuntos de dados.

Com isso, diante dos resultados apresentados e analisados, conclui-se a presente Seção, que cumpriu, com sucesso, os objetivos específicos: IV - Construir um *framework* conceitual para a publicação de metadados geoespaciais na *Web*, descrevendo as etapas e respectivos procedimentos e atores envolvidos; e V - Validar o *framework* por meio da realização de uma prova de conceito na qual serão publicados os metadados catalogados na IDE de um provedor de dados geoespaciais.

Na próxima Seção são apresentadas as considerações finais deste trabalho.

6 CONSIDERAÇÕES FINAIS

O *framework* desenvolvido e apresentado neste trabalho é fruto de uma pesquisa que surge da oportunidade de contribuir para o desenvolvimento da próxima geração de Infraestruturas de Dados Espaciais, ao investigar a possibilidade e benefícios de integrá-las com as tecnologias e ferramentas da Web Semântica, de forma a permitir que os registros de metadados catalogados nessas plataformas, e os respectivos conjuntos de dados geoespaciais que descrevem, possam ser indexados, encontrados e recuperados na Web. Isso beneficia principalmente aqueles usuários não especialistas no domínio da geoinformação, bem como agentes de software, de forma a promover a interoperabilidade e estimular o reuso desse tipo de dado. Apesar disso, este é um tema ainda pouco explorado na comunidade de geoinformação, o que foi constatado através da revisão de literatura efetuada para posicionar este trabalho em relação a outros trabalhos relacionados.

Neste sentido, o desenvolvimento e evolução das pesquisas no campo da Web Semântica, em especial do paradigma do *Linked Data*, notados predominantemente na última década (2011-2020), produziram um referencial teórico e ferramental muito relevante para explorar as enormes possibilidades de aplicação nas mais diversas áreas (medicina, biblioteconomia, engenharias, geoinformática, entre outras), o que ficou evidenciado pela pesquisa bibliográfica e documental realizada neste trabalho, com destaque para as contribuições de Berners-Lee (2001; 2006) e Heath e Bizer (2011), além dos trabalhos e documentos de referência e recomendações produzidos pelos grupos de trabalho do W3C, muito consultados neste trabalho, e que contaram com a colaboração destes e de outros autores.

Quanto aos estudos para aplicação da Web Semântica e do *Linked Data* na área de geoinformação, esta pesquisa evidenciou as contribuições de Becker e Bizer (2009) e Yuan et al. (2013) para o chamado ‘*Linked GeoData*’, além dos trabalhos e documentos de referência e recomendações produzidos pelos grupos de trabalho da OGC que, em parceria com o W3C, elaboraram o documento de ‘Boas práticas para dados espaciais na Web’.

Diante disso, a Seção 2 deste trabalho trouxe a conceituação de dados e metadados geoespaciais, bem como a descrição da arquitetura, produtos e serviços de informação geoespacial disponibilizados pelas Infraestruturas de Dados Espaciais (IDE), com destaque para o serviço de catálogo de metadados (CSW), através do qual ocorre o acoplamento do *framework* aqui proposto, de forma a cumprir com o objetivo específico I apresentado inicialmente: “Apresentar os conceitos de dados e metadados geoespaciais, bem como a arquitetura e os

componentes de uma Infraestrutura de Dados Espaciais, com ênfase no Serviço de Catálogo de Metadados para a Web (CSW)”.

Na sequência, a Seção 3 deste trabalho trouxe os conceitos e a evolução da *Web Semântica*, bem como apresentou algumas de suas tecnologias e ferramentas, além de descrever as boas práticas do *Linked Data* e de dados espaciais na *Web*. Aqui convém destacar a abordagem ao termo ‘*Linked GeoData*’, que corresponde à aplicação do paradigma do *Linked Data* na área de Geoinformação. Com isso foi cumprido o objetivo específico II apresentado inicialmente: “*Conceituar e descrever as tecnologias e ferramentas da Web Semântica e as boas práticas de Linked Data e de dados espaciais na Web*”.

A partir de todo esse referencial bibliográfico e documental encontrado na literatura foi possível construir o *framework* apresentado na Seção 4 deste trabalho, com destaque para a aplicação, em sua camada de ‘Coleta e Conversão’, do método de *crosswalk* de padrões de metadados, que se mostrou uma solução ótima para a conversão dos metadados geoespaciais, originalmente do padrão ISO 19115, para o modelo de dados RDF, estruturado pela aplicação do vocabulário DCAT-2. Além disso, a observação das boas práticas para dados espaciais na *Web* e dos princípios do *Linked Data* na construção do *framework* podem contribuir para um melhor sequenciamento de tarefas e otimização das etapas que o compõe, de forma a proporcionar um nível satisfatório de qualidade aos metadados que serão publicados na *Web* por meio dele. Com isso foi respondida a primeira questão de pesquisa levantada neste trabalho: “*Como efetuar a conversão de modelo de dados e de formato de serialização dos metadados geoespaciais catalogados em uma IDE, de forma a torná-los interoperáveis em nível semântico e legíveis também por máquinas e agentes de software na Web?*”.

Na segunda camada do *framework* (camada de ‘Ligação’), a partir das boas práticas do *Linked Data*, alguns elementos de metadados selecionados foram ligados a outros recursos semelhantes ou complementares na *Web* de dados por meio de URIs desreferenciáveis para tais recursos. Isso, somado à estruturação e contextualização dos elementos de metadados a partir de vocabulários, podem consignar uma maior eficiência e eficácia nos processos de descoberta e recuperação da informação geoespacial para atender as necessidades informacionais dos usuários. Com isso, foi respondida a segunda questão de pesquisa formulada neste trabalho: “*Quais elementos de metadados geoespaciais poderiam ser ligados a outras fontes de dados na Web, de forma a enriquecê-los semanticamente e possibilitar a descoberta de informações adicionais pelos consumidores desses metadados?*”.

Já na terceira camada do *framework* (camada de ‘Disponibilização’), optou-se pela proposta de uma solução dupla para efetivar a publicação dos metadados geoespaciais

semânticos na *Web*, sendo: a primeira consistindo da adoção de uma plataforma de *triplestore*, compatível com a extensão GeoSPARQL, a fim de possibilitar o armazenamento e consulta (por usuários humanos e máquinas), através da linguagem de consulta semântica SPARQL, dos conjuntos de triplas de metadados produzidos nas camadas anteriores do *framework*; e a segunda consistindo da adoção de uma plataforma de catálogo de dados abertos, com o intuito de publicar os conjuntos de metadados geoespaciais semânticos produzidos, de forma que possam ser indexados, encontrados e recuperados por buscadores na *Web*. Com isso, foi respondida a terceira questão de pesquisa formulada: “*Como disponibilizar os conjuntos de metadados geoespaciais enriquecidos semanticamente, de forma a possibilitar sua indexação e facilitar sua consulta e recuperação na Web?*”.

Ao final da descrição das três camadas desse *framework*, fica cumprido o objetivo específico III deste trabalho: “*Construir um framework conceitual para a publicação de metadados geoespaciais semânticos na Web, descrevendo as etapas e respectivos procedimentos que as compõem.*”.

E, na Seção 5 deste trabalho, o *framework* construído na seção anterior foi validado por meio de uma prova de conceito, na qual foram executados, passo a passo, cada um dos procedimentos que o compõem, utilizando como amostra um registro de metadados extraído do catálogo de metadados da plataforma de IDE ‘TerraBrasilis/INPE’. A partir dessa prova de conceito foi possível constatar a viabilidade de implementação de cada procedimento proposto no *framework*, tais como o *crosswalk* para conversão de padrões de metadados, a estruturação das triplas RDF com o uso do vocabulário DCAT-2, o uso de ferramenta *Web* para conversão do formato de serialização (de XML para *Turtle*), a inserção de URIs para recursos da *DBpedia*, *Geonames*, entre outras fontes, ligando as triplas à *Web* de dados.

Finalizada a execução das três camadas desse *framework* foram obtidas triplas de metadados em RDF, as quais foram armazenadas com sucesso no banco de triplas da plataforma *Virtuoso Open Source* (VOS), além de serem disponibilizadas no Portal Brasileiro de Dados Abertos (dados.gov.br, baseado na plataforma CKAN). Dessa forma, estando validado o *framework* construído neste trabalho, resta cumprido o objetivo específico IV deste trabalho: “*Validar o framework por meio da realização de uma prova de conceito na qual serão publicados os metadados geoespaciais catalogados na IDE de um provedor de dados geoespaciais.*”.

Além disso, por meio da prova de conceito, foi possível constatar que este *framework* tem potencial para propor uma solução a um problema, ainda existente, de interoperabilidade de metadados entre a Infraestrutura Nacional de Dados Espaciais (INDE) e a Infraestrutura

Nacional de Dados Abertos (INDA, portal dados.gov.br), no qual os metadados dos *datasets* disponibilizados pela INDE não são replicados para a INDA, de forma que seja cumprida a exigência de catalogar na INDA todos os conjuntos de dados produzidos nas atividades do Governo Federal, incluindo os dados geoespaciais mantidos na INDE.

Ao serem cumpridos todos os objetivos específicos, resta cumprido também o objetivo geral deste trabalho: “*Propor um framework que estabeleça um fluxo organizacional para orientar a execução das etapas de um processo de publicação na Web, em formato semântico, dos metadados geoespaciais catalogados em Infraestruturas de Dados Espaciais (IDEs), seguindo as boas práticas de Linked Data.*”.

Diante disso, e tendo sido respondidas as três perguntas de pesquisa apresentadas, é possível tornar válida a hipótese de que “*se um fluxo organizacional puder ser estabelecido e validado para orientar a execução das etapas de um processo de publicação na Web dos metadados catalogados em IDEs seguindo as boas práticas de Linked Data, então estes metadados geoespaciais semânticos poderão ser indexados e descobertos diretamente na Web pelos usuários, quer sejam humanos ou máquinas.*”. E, uma vez validada a hipótese levantada, sustenta-se neste trabalho a tese de que ‘a integração das IDEs com a Web Semântica pode ser alcançada por meio da publicação de seus catálogos de metadados geoespaciais na Web seguindo as boas práticas de *Linked Data*’.

6.1 Trabalhos Futuros

A partir dos resultados obtidos nesta pesquisa, como trabalhos futuros são sugeridos a automatização dos procedimentos que compõem o *framework* proposto, por meio da construção de *scripts* e aplicações de software. Nisso inclui-se a implementação da coleta automatizada por meio de *dump* da base de dados do servidor de catálogo de metadados de IDEs, onde deve ser verificada a viabilidade de uso da plataforma D2RQ (<http://d2rq.org>) como padrão, ou ainda, a construção de *scripts* para customizar a criação de arquivos de mapeamento e de conexão à base de dados utilizando a linguagem R2RML. Também é sugerida a realização de estudos de implementação da camada de ‘Disponibilização’ com outras plataformas de *triplestore* e de catálogo de dados abertos.

Por fim, é sugerida a adaptação do *framework* para permitir a publicação na Web, em formato semântico, de outros padrões de metadados, tais como os bibliográficos, de forma a contribuir para a aplicação do *Linked Data* em bibliotecas. Para isso, torna-se necessário elaborar novos mapeamentos para *crosswalk* entre padrões de metadados.

REFERÊNCIAS

- AFONSO, M. H. F.; SOUZA, J. V. de; ENSSLIN, S. R.; ENSSLIN, L. Como construir conhecimento sobre o tema de pesquisa? Aplicação do processo PROKNOW-C na busca de literatura sobre avaliação do desenvolvimento sustentável. **Revista de Gestão Social e Ambiental**, São Paulo, v. 5, n. 2, p. 47-62, mai./ago. 2012. DOI: 10.5773/rgsa.v5i2.424.
- AKLI-ASTOUATI, K.; GUEBAILI, R.; MOKHTARI, A. Geographic metadata: data quality, uncertainty and imprecision. **International Journal of Reasoning-based Intelligent Systems**, Geneva, v. 3, n. 3–4, p. 164–172, 2011. DOI: 10.1504/IJRIS.2011.043541.
- ALLEMANG, D.; HENDLER, J. **Semantic Web for the working ontologist: effective modeling in RDFS and OWL**. Burlington: Morgan Kaufmann Publishers, 2008. 330 p.
- ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.
- ANGELIS, J. de. **Episódios Diários**. São Paulo: Leal, 1985. 57 p.
- ANKOLEKAR, A.; KRÖTZSCH, M.; TRAN, T.; VRANDECIC, D. The two cultures: mashing up Web 2.0 and the Semantic Web. **Web Semantics: Science, Services and Agents on the World Wide Web**, Amsterdam, v. 6, n. 1, p. 70-75, 2008. DOI: 10.1145/1242572.1242684.
- ANTONIOU, G.; VAN HARMELEN, F. **A Semantic Web primer**. 2 ed. Cambridge: MIT Press, 2008. 264 p.
- ARAKAKI, A. C. S.; ARAKAKI, F. A. Dados e metadados: conceitos e relações. **Ciência da Informação**, Brasília, v. 49, n. 3, 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/5504>. Acesso em: 7 ago. 2021.
- ARIA, M.; CUCCURULLO, C. Bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, Amsterdam, v. 11, n. 4, p. 959–975, 2017. DOI: 10.1016/j.joi.2017.08.007.
- ASSIS, L. F. F. G.; FERREIRA, K. R.; VINHAS, L.; MAURANO, L.; ALMEIDA, C.; CARVALHO, A.; RODRIGUES, J.; MACIEL, A.; CAMARGO, C. TerraBrasilis: A spatial data analytics infrastructure for large-scale thematic mapping. **ISPRS International Journal of Geo-Information**, Basel, v. 8, n. 11, p. 513, 2019. DOI: 10.3390/ijgi8110513.
- AUER, S.; LEHMANN, J.; NGONGA NGOMO, A.-C. Introduction to linked data and its lifecycle on the Web. In: POLLERES, A.; D'AMATO, C.; ARENAS, M.; HANDSCHUH, S.; KRONER, P.; OSSOWSKI, S.; PATEL-SCHNEIDER, P. (ed.). Reasoning Web. Semantic Technologies for the Web of Data. **Lecture Notes in Computer Science**, Berlin, v. 6848, p. 1–75, 2011.
- BAPTISTA, C. S.; MENEZES, L. C.; OLIVEIRA, M. G.; FALCÃO, A. G. R.; MARINHO, L. B. On building semantically enhanced location-based social networks. In: SIMPÓSIO

BRASILEIRO DE GEOINFORMÁTICA, 14., 2013, Campos do Jordão. **Anais [...]** São José dos Campos: INPE, 2013. p. 19-30.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 1979. 226p.

BECKER, C.; BIZER, C. Exploring the Geospatial Semantic Web with DBpedia Mobile. **Journal of Web Semantics**, Amsterdam, v. 7, n. 4, p. 278–286, 2009.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, New York, v. 284, n. 5, p. 34–43, 2001.

BERNERS-LEE, T. Linked Data - Design Issues. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 22 nov. 2020.

BIDNEY, M.; CLAIR, K. Harnessing the Geospatial Semantic Web: toward place-based information organization and access. **Cataloging & Classification Quarterly**, New York, v. 52, n. 1, p. 69–76, 2014.

BORKO, H. Information Science: What is it? **American Documentation**, Washington, v. 19, n. 1, p. 3–5, 1968.

BRASIL. Comissão Nacional de Cartografia. Comitê de Estruturação de Metadados Geoespaciais. **Perfil de metadados geoespaciais do Brasil (Perfil MGB)**: conteúdo de metadados geoespaciais em conformidade com a norma ISO 19115:2003. Versão homologada. Rio de Janeiro: Concar, 2009. 194 p. Acompanha 1 CD-ROM. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=283693>. Acesso em: 3 ago. 2020.

BRASIL. Comitê de Planejamento da Infraestrutura Nacional de Dados Espaciais. **Plano de ação para implantação da Infraestrutura Nacional de Dados Espaciais**. Rio de Janeiro: CONCAR, 2010. 203 p. Disponível em: <https://inde.gov.br/pdf/PlanoDeAcaoINDE.pdf>. Acesso em: 24 nov. 2020.

BRASIL. Decreto n. 6.666, de 27 de novembro de 2008. Institui, no âmbito do poder executivo federal, a Infraestrutura Nacional de Dados Espaciais (INDE), e dá outras providências. **Diário Oficial da União**: seção 1, Brasília, ano 145, n. 232, p. 57, 28 nov. 2008. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6666.htm. Acesso em: 3 ago. 2020.

BRASIL. Instituto Brasileiro de Geografia e Estatística. **Introdução à Infraestrutura Nacional de Dados Espaciais**: Unidade 4 – Geosserviços. Brasília: IBGE, 2022. 161 slides. Disponível em: <https://treinamento.inde.gov.br/treinamento/pdf/INDE-geosservicos-2022-04.pdf>. Acesso em: 10 set. 2022.

BRASIL. Instituto Nacional de Pesquisas Espaciais. **Monitoramento da Ocorrência de Focos de Fogo Ativo**. São José dos Campos: INPE, v. 2, n. 7. 2021. 1 p. Disponível em: <http://urlib.net/ibi/8JMKD3MGP3W34T/45CTN3B>. Acesso em: 20 set. 2021.

BREITMAN, K. K.; CASANOVA, M. A.; TRUSZKOWSKI, W. **Semantic Web**: concepts, technologies, and applications. London: Springer, 2007. 328 p.

BRINK, L. Van den; JANSSEN, P.; QUAK, W.; STOTER, J. Linking spatial data: automated conversion of geo-information models and GML data to RDF. **International Journal of Spatial Data Infrastructures Research**, Ispra, v. 9, n. 0, p. 59–85, 2014.

BUCKLAND, M. INFORMATION AS THING. **Journal of the American Society of Information Science**, New York, v. 5, n. 42, p. 351–360, 1991.

CAMBOIM, S. P.; SLUTER, C. R. Uso de ontologias para busca de dados geoespaciais: uma ferramenta semântica para a Infraestrutura Nacional de Dados Espaciais. **Revista Brasileira de Cartografia**, Uberlândia, v. 65, n. 6, p. 1127–1142, 2013. Disponível em: <http://www.seer.ufu.br/index.php/revistabrasileiracartografia/article/view/43885>. Acesso em: 26 out. 2020.

CARDOSO, J. **Semantic Web services: theory, tools and applications**. Hershey: IGI Global, 2007. 353 p.

CASANOVA, M. A.; BRAUNER, D. F.; CÂMARA, G. Integração e interoperabilidade entre fontes de dados geográficos. **Bancos de Dados Geográficos**. Curitiba: MundoGEO, 2005. 36 p.

CHAN, L. M.; ZENG, M. L. Metadata Interoperability and Standardization - A Study of Methodology Part I: Achieving Interoperability at the Schema Level. **D-Lib Magazine**, Virginia, v. 12, n. 6, 2006. Disponível em: <http://www.dlib.org/dlib/june06/chan/06chan.html>. Acesso em: 18 out. 2021.

CHEATHAM, M.; KRISNADHI, A.; AMINI, R.; HITZLER, P.; JANOWICZ, K.; SHEPHERD, A.; NAROCK, T.; JONES, M.; JI, P. The GeoLink knowledge graph. **Big Earth Data**, London, v. 2, n. 2, p. 131–143, 2018. DOI: 10.1080/20964471.2018.1469291.

CONEGLIAN, C. S.; DIEGER, R.; SEGUNDO, J. E. S.; CAPTREZ, M. O papel estratégico da Web Semântica no contexto do Big Data. In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 1., 2017, Florianópolis. **Anais [...]**. Florianópolis: UFSC, 2017. p. 1-6.

COWEN, D. J. Geospatial Metadata. In: RICHARDSON, D.; CASTREE, N.; GOODCHILD, M. F.; KOBAYASHI, A.; LIU, W.; MARSTON, R. A. (ed.). **International Encyclopedia of Geography: People, the Earth, Environment and Technology**. Oxford: Wiley, 2017. p. 1–6.

DAVIS JÚNIOR, C. A.; SOUZA, L. A. de; BORGES, K. A. V. Disseminação de dados geográficos na Internet. **Bancos de Dados Geográficos**. Curitiba: MundoGEO, 2005. 26 p.

DAVIS JÚNIOR, C. A.; ALVES, L. L. Infraestruturas de Dados Espaciais: Potencial para Uso Local. **Revista Informática Pública**, Belo Horizonte, v. 8, n. 1, p. 65–80, 2006.

DIBS, H. Elements, issues and challenges in implementation of NSDI. **Coordinates**, Delhi, v. 3, p. 42–52, 2013.

DO, B. L.; TRINH, T. D.; WETZ, P.; KIESLING, E.; ANJOMSHOAA, A.; TJOA, A. M. Multiscale exploration of spatial statistical datasets: a linked data mashup approach. In: INTERNATIONAL WORKSHOP ON SEMANTIC STATISTICS, 2., 2014, Riva del Garda.

Proceedings [...] Riva del Garda: CEUR, 2014. [s.p.]. Disponível em: <http://ceur-ws.org/Vol-1550/article-08.pdf>. Acesso em: 24 fev. 2021.

DUBLIN CORE METADATA INITIATIVE. **DCMI Metadata Terms**. 2020. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#http://purl.org/dc/terms/format>. Acesso em: 3 ago. 2020.

EGENHOFER, M. J. Toward the semantic geospatial *Web*. In: ACM INTERNATIONAL SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 10., 2002, McLean. **Proceedings** [...] New York: ACM Press, 2002. p. 1-4. DOI: 10.1145/585147.585148.

FAN, J.; YAN, J.; MA, Y.; WANG, L. Big Data integration in Remote Sensing across a distributed metadata-based spatial infrastructure. **Remote Sensing**, Basel, v. 10, n. 2, p. 7, 2017. DOI: 10.3390/rs10010007.

FEITOSA, A. **Organização da informação na Web**: das tags à Web semântica. Brasília: Thesaurus, 2006. 131p.

FERNÁNDEZ-ORDÓÑEZ, Y. M.; MEDINA-RAMÍREZ, R. C.; SORIA-RUIZ, J. Geographic metadata and ontology-based satellite image management. In: IEEE GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2014, Quebec City. **Proceedings** [...] New York: IEEE, 2014. p. 117-120. DOI: 10.1109/IGARSS.2014.6946370.

FGDC - FEDERAL GEOGRAPHIC DATA COMMITTEE. **A strategy for the national Spatial Data Infrastructure**. Reston: FGDC, 1997. Disponível em: <http://www.fgdc.gov/policyandplanning/A%20Strategy%20for%20the%20NSDI%201997.pdf>. Acesso em: 10 mar. 2021.

FGDC - FEDERAL GEOGRAPHIC DATA COMMITTEE. **National Spatial Data Infrastructure**: Geospatial Metadata. Reston: FGDC, 1998. Disponível em: <http://www.fgdc.gov/publications/documents/metadata/metafact.pdf>. Acesso em: 10 mar. 2021.

FORMENTON, D.; FERREIRA DE CASTRO, F.; DE SOUZA GRACIOSO, L.; FURNIVAL, A. C. M.; DE MELO SIMÕES, M. da G. Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. **Biblios**, Pittsburgh, n. 68, p. 82–95, 2017.

FUGAZZA, C.; PEPE, M.; OGGIONI, A.; TAGLIOLATO, P.; CARRARA, P. Raising semantics-awareness in geospatial metadata management. **ISPRS International Journal of Geo-Information**, Basel, v. 7, n. 9, p. 370, 2018. DOI: 10.3390/ijgi7090370.

FUGAZZA, C.; PEPE, M.; OGGIONI, A.; TAGLIOLATO, P.; CARRARA, P. Streamlining geospatial metadata in the Semantic *Web*. **IOP Conference Series: Earth and Environmental Science**, Bristol, v. 34, p. 012009, 2016. DOI: 10.1088/1755-1315/34/1/012009.

FUGAZZA, C.; PEPE, M.; OGGIONI, A.; TAGLIOLATO, P.; PAVESI, F.; CARRARA, P. Describing Geospatial Assets in the *Web* of Data: A Metadata Management Scenario. **ISPRS**

International Journal of Geo-Information, Basel, v. 5, n. 12, p. 229, 2016. DOI: 10.3390/ijgi5120229.

FUGAZZA, C.; TAGLIOLATO ACQUAVIVA D'ARAGONA, P.; OGGIONI, A.; CARRARA, P. Decentralized geospatial metadata management. **Earth Science Informatics**, Dublin, v. 14, n. 1, p. 1579-1596, 2021. DOI: 10.1007/s12145-020-00559-1.

FUGAZZA, C.; TAGLIOLATO, P.; FRIGERIO, L.; CARRARA, P. Web-Scale Normalization of Geospatial Metadata Based on Semantics-Aware Data Sources. **ISPRS International Journal of Geo-Information**, Basel, v. 6, n. 11, p. 354, 2017. DOI: 10.3390/ijgi6110354.

FURNER, J. "Data": The data. In: KELLY, M.; BIELBY, J. (ed.). **Information cultures in the digital age**. Wiesbaden: Springer VS, 2016. p.287-306. DOI:10.1007/978-3-658-14681-8_17.

GIANNOPOULOS, G.; MAROULIS, T.; SKOUTAS, D.; KARAGIANNAKIS, N.; ATHANASIOU, S. FAGI-tr: a tool for aligning geospatial RDF vocabularies. In: ESWC 2014 SATELLITE EVENTS, 2014, Cham. **Proceedings [...]** Basel: Springer International Publishing Switzerland, 2014. p. 355-361. DOI: 10.1007/978-3-319-11955-7_48.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5.ed. São Paulo: Atlas, 2010. 175p.

GILL, T. Metadata and the Web. In: BACA, M. (org.). **Introduction to Metadata**. 3 ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/metadata-and-the-Web>. Acesso em: 3 ago. 2020.

GILLILAND, A. J. Setting the Stage. In: BACA, M. (org.). **Introduction to Metadata**. 3 ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/setting-the-stage>. Acesso em: 3 ago. 2020.

GOODCHILD, M. F.; FU, P.; RICH, P. Sharing geographic information: An assessment of the geospatial one-stop. **Annals of the Association of American Geographers**. London, n. 97, v. 2, p. 250–266, 2007. DOI: 10.1111/j.1467-8306.2007.00534.x.

GRAUPNER, A.; NÜST, D. Serverless GEO labels for the semantic sensor Web. In: INTERNATIONAL CONFERENCE ON GEOGRAPHIC INFORMATION SCIENCE, 11., 2020, Dagstuhl. **Proceedings [...]** Dagstuhl: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020. p. 4:1-4:14. DOI: 10.4230/LIPIcs.GIScience.2021.I.4.

GSDI - GLOBAL SPATIAL DATA INFRASTRUCTURE ASSOCIATION. **Developing Spatial Data Infrastructures: the SDI Cookbook**. V.2, 2004. Disponível em: http://gsdi.org/images/publications/cookbooks/SDI_Cookbook_GSDI_2004_ver2.pdf. Acesso em 10 out. 2021.

HEATH, T.; BIZER, C. Linked Data: evolving the Web into a global data space. **Synthesis Lectures on the Semantic Web: Theory and Technology**, London, v. 1, n. 1, p. 1–136, 2011.

HEATH, T.; BIZER, C. Semantic Annotation and Retrieval: Web of Data. In: DOMINGUE, J. (ed.). **Handbook of semantic Web technologies**. Berlin: Springer, 2011b. p. 191-229.

HENRIQUES, H. B. **Anotação automática de dados geográficos baseada em bancos de dados abertos e interligados**. Campina Grande: UFCG, 2015. Disponível em: <http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/594>. Acesso em: 26 out. 2020.

HUANG, W.; RAZA, S. A.; MIRZOV, O.; HARRIE, L. Assessment and benchmarking of spatially enabled RDF stores for the next generation of Spatial Data Infrastructure. **ISPRS International Journal of Geo-Information**, Basel, v. 8, n. 7, p. 310, 2019.

HYLAND, B.; ATEMEZING, G.; VILLAZÓN-TERRAZAS, B. **Best practices for publishing Linked Data**. Cambridge: World Wide Web Consortium. 2014. Disponível em: <http://www.w3.org/TR/2014/NOTE-ld-bp-20140109>. Acesso em: 03 set. 2021.

ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19115-1:2014**: geographic information - metadata - part 1: fundamentals. Geneva: ISO, 2014. 167 p.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São Paulo: Novatec, 2015. 175p.

JAFFE, R. Library guides: metadata creation: 3. Dublin Core Metadata Schema. 2017. Disponível em: <https://guides.library.ucsc.edu/c.php?g=618773&p=4306386>. Acesso em: 25 out. 2021.

JIE YUAN; PENG YUE; JIANYA GONG; MINGDA ZHANG. A Linked Data Approach for Geospatial Data Provenance. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 51, n. 11, p. 5105–5112, 2013. DOI: 10.1109/TGRS.2013.2249523.

KALANTARI, M.; RAJABIFARD, A.; OLFAT, H.; WILLIAMSON, I. Geospatial Metadata 2.0 – An approach for volunteered geographic information. **Computers, Environment and Urban Systems**, New York, v. 48, p. 35–48, 2014. DOI: 10.1016/j.compenvurbsys.2014.06.005.

KIRYAKOV A.; DAMOVA M. Storing the Semantic Web: Repositories. In: DOMINGUE, J. (ed.). **Handbook of semantic Web technologies**. Berlin: Springer, 2011. p. 233-297.

LAUFER, C. **Guia de Web Semântica**. São Paulo: Projeto SPUK, 2015. 132p. Disponível em: <https://nic.br/media/docs/publicacoes/13/Guia_Web_Semantica.pdf>. Acesso em: 23 nov. 2019.

LI, Z.; CHIANG, Y.-Y.; TAVAKKOL, S.; SHBITA, B.; UHL, J. H.; LEYK, S.; KNOBLOCK, C. A. An automatic approach for generating rich, linked geo-metadata from historical map images. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY & DATA MINING, 26., 2020, New York. **Proceedings [...]** New York: ACM, 2020. p. 3290-3298. DOI: 10.1145/3394486.3403381.

LISBOA FILHO, J.; IOCHPE, C. **Introdução a Sistemas de Informações Geográficas com Ênfase em Banco de Dados**. Viçosa: UFV, 1996. 52p.

LOPEZ-PELLICER, F. J.; FLORCZYK, A. J.; NOGUERAS-ISO, J.; MURO-MEDRANO, P. R.; ZARAZAGA-SORIA, F. J. Exposing CSW Catalogues as Linked Data. **Lecture Notes in**

Geoinformation and Cartography, Berlin, v. 0, p. 183-200, 2010. DOI: 10.1007/978-3-642-12326-9_10

LÓSCIO, B. F.; BURLE, C.; OLIVEIRA, M. I. S.; CALEGARI, N. **Fundamentos para publicação de dados na Web**. São Paulo: Comitê Gestor da Internet no Brasil, 2018. 64p.

LOTI, L. B. S.; MEDEIROS, N. G.; SANTOS, A. P.; LISBOA-FILHO, J. Atualização da norma ISO 19115 e os impactos no Perfil de Metadados Geoespaciais do Brasil. *In*: SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA, 20., 2019, São José dos Campos. **Anais [...]** São José dos Campos: INPE, 2019. p. 218-223.

MARCONDES, C. H. “Linked data” – dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na *Web*. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 17, n. 34, p. 171–192, 2012. DOI: 10.5007/1518-2924.2012v17n34p171.

MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 21, n. 2, p. 61–83, 2016.

MCCRAE. The Linked Open Data Cloud. 2021. Disponível em: <https://lod-cloud.net/#>. Acesso em: 5 mar. 2022.

MCKEE, G. The Map as a Search Box: using Linked Data to create a geographic discovery system. **Information Technology and Libraries**, Chicago, v. 38, n. 1, p. 40–52, 2019. DOI: 10.6017/ital.v38i1.10592.

MENEZES, L. C. de; BAPTISTA, C. de S.; FALCÃO, A. G. R.; OLIVEIRA, M. G. de; MARINHO, L. B. DYSCS: A platform to build geographically and semantically enhanced social content sites. **Journal of Systems and Software**, New York, v. 94, p. 39–49, 2014. DOI: 10.1016/j.jss.2014.03.059.

MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, Washington, v. 2, n. 1, p. 20–32, 1951.

MOREIRA, F. M.; SANT’ANA, R. C. G.; SANTOS, P. L. V. A. C.; ZAFALON, Z. R. Metadados para descrição de *datasets* e recursos informacionais do “Portal Brasileiro de Dados Abertos”. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 22, n. 3, p. 158–185, 2017. DOI: 10.1590/1981-5344/2947.

MUELLER, S. P. M. **Métodos para a pesquisa em ciência da informação**. Série: Ciência da informação e da comunicação. Brasília: Thesaurus, 2007. 190 p.

NHACUONGUE, J. A.; ROZSA, V.; DUTRA, M. L. Linked Data e Ciência da Informação: diretrizes para a publicação de *datasets* institucionais abertos. **Biblios: Journal of Librarianship and Information Science**, [S.l.], n. 73, p. 20-34, fev. 2019. DOI: 10.5195/biblios.2018.429.

NILSSON, M.; POWELL, A.; JOHNSTON, P.; NAEVE, A. **Expressing Dublin Core metadata using the Resource Description Framework (RDF)**. Leesburg: Dublin Core

Metadata Initiative, 2007. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dc-rdf/2007-06-04/comments-received>. Acesso em: 10 jan. 2022.

OGC - OPEN GEOSPATIAL CONSORTIUM. **Geography Markup Language**. Arlington: OGC, 2007. Disponível em: <https://www.ogc.org/standards/gml>. Acesso em: 7 ago. 2020.

OGC - OPEN GEOSPATIAL CONSORTIUM. **OpenGIS Catalog Service Specification**. Arlington: OGC, 2007b. Disponível em: https://portal.ogc.org/files/?artifact_id=20555. Acesso em: 7 ago. 2020.

PABÓN, O. S. Propuesta para extender semánticamente el proceso de recuperación de información. **Revista EIA**, Antioquia, v. 11, n. 22, p. 51-65, 2014. DOI: 10.14508/reia.2014.11.22.49-63reia.2014.11.22.51-65.

PAGANI, R. N.; KOVALESKI, J. L.; RESENDE, L. M. Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citations, and year of publication. **Scientometrics**, Amsterdam, v. 105, n. 3, p. 2109–2135, 2015. DOI: 10.1007/s11192-015-1744-x.

PATROUMPAS, K.; GEORGOMANOLIS, N.; STRATIOTIS, T.; ALEXAKIS, M.; ATHANASIOU, S. Exposing INSPIRE on the Semantic Web. **Journal of Web Semantics**, Amsterdam, v. 35, n. 1, p. 53–62, 2015. DOI: 10.1016/j.Websem.2015.09.003.

PEREIRA, C. M.; FERNEDA, E.; SANTAREM SEGUNDO, J. E. Análise do processo de recuperação da informação em bases de dados publicadas como dados abertos ligados utilizando a abordagem RDB2LOD. **Em Questão**, Porto Alegre, v. 26, n. 3, p. 94–120, 2020.

PINTO, D. M.; BRANDÃO, V. V. S.; DRUCKER, D. P.; DOMPIERI, M. H. G.; RASCHE, F.; BETTIOL, G. M.; CUSTÓDIO, D. O.; VICTORIA, D. C.; GONÇALVES, L. M. P. B. Descrição dos dados da pesquisa geoespacial: a experiência da Embrapa. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**. Rio de Janeiro, v. 11, p. 1–16, 2017.

PO, L.; BIKAKIS, N.; DESIMONI, F.; PAPASTEFANATOS, G. Linked Data Visualization: Techniques, Tools, and Big Data. **Synthesis Lectures on the Semantic Web: Theory and Technology**, San Rafael, v. 10, n. 1, p. 1–157, 2020. DOI: 10.2200/S00967ED1V01Y201911WBE019.

POUCHARD, L. C.; BRANSTETTER, M. L.; COOK, R. B.; DEVARAKONDA, R.; GREEN, J.; PALANISAMY, G.; ALEXANDER, P.; NOY, N. F. A Linked Science investigation: enhancing climate change data discovery with semantic technologies. **Earth Science Informatics**, Berlin, v. 6, n. 3, p. 175–185, 2013. DOI: 10.1007/s12145-013-0118-2.

RAJABIFARD A.; BINNS A.; MASSER I.; WILLIAMSON, I. The role of sub-national government and the private sector in future spatial data infrastructures. **International Journal of Geographical Information Science**, London, v. 20, n. 7, p. 727-741, 2006. DOI: 10.1080/13658810500432224.

RAJABIFARD, A.; FEENEY, M.-E. F.; WILLIAMSON, I. P. Future directions for SDI development. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 4, n. 1, p. 11–22, 2002.

RAMALHO, R. A. S.; VIDOTTI, S. A. B. G.; FUJITA, M. S. L. *Web semântica: uma investigação sob o olhar da ciência da informação*. **DataGramaZero**, v. 8, n. 6, 2007. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/6192>. Acesso em: 22 out. 2022.

REED, C.; REICHARDT, M. From Spatial Data Infrastructure to the Geospatial *Web*: advancing the SDI vision in a changing technology environment. *In: GSDI World Conference Proceedings*, 10., 2008, St. Augustine. **Proceedings** [...]. New York: GSDI, 2008. p. 30.

REID, J.; WAITES, W.; BUTCHART, B. An infrastructure for publishing geospatial metadata as open linked metadata. *In: INTERNATIONAL CONFERENCE ON GEOGRAPHIC INFORMATION SCIENCE*, 15., 2012, Avignon. **Proceedings** [...] Dresden: Association of Geographic Information Laboratories in Europe, 2012. p. 99-104, 2012.

RENTERIA-AGUALIMPIA, W.; LÓPEZ-PELLICER, F. J.; LACASTA, J.; ZARAZAGA-SORIA, F. J.; MURO-MEDRANO, P. R. Identifying hidden geospatial resources in catalogues. *In: INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, MINING AND SEMANTICS*, 3., 2013, New York. **Proceedings** [...] New York: ACM, 2013. p. 1-7. DOI: 10.1145/2479787.2479812.

RILEY, J. **Understanding Metadata**: What is metadata, and what is it for? Baltimore: NISO, 2017. 49p. Disponível em: <http://www.niso.org/publications/understanding-metadata-2017>. Acesso em: 18 out. 2021.

RITSCHER, B.; MENDE, V.; GERICKE, L.; KORNMESSER, R.; PFEIFFER, S. *Web approach for ontology-based classification, integration, and interdisciplinary usage of geoscience metadata*. **Data Science Journal**, London, v. 11, p. S140–S153, 2012. DOI: 10.2481/dsj.IGY-014.

RONZHIN, S.; FOLMER, E.; LEMMENS, R. Technological aspects of (linked) open data. *In: VAN LOENEN, B.; VANCAUWENBERGHE, G.; CROMPVOETS, J. (ed.). **Open Data Exposed**: Information Technology and Law Series*. Hague: T.M.C. Asser Press, 2018. p. 173–193.

ROSA, R. **Introdução ao Geoprocessamento**. Uberlândia: UFU, 2013. 142p.

SANTARÉM SEGUNDO, J. E. *Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas*. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 8, n. 2, p.219-239, 2015. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/359/359>. Acesso em: 22 out. 2022.

SANTAREM SEGUNDO, J. E. *Web Semântica: fluxo para publicação de dados abertos e ligados*. **Informação em Pauta**, Fortaleza, v. 3, número especial, 2018, p. 117-140. DOI: <https://doi.org/10.32810/2525-3468.ip.v3iEspecial.2018.39721.117-140>.

SANTOS, B. P.; FEIJÓ, A. M.; VIEGAS, A. C.; LISE, F.; SCHWARTZ, E. Classificação das pesquisas. *In: LISE, F.; SOUZA, B. M.; SCHWARTZ, E.; GARCIA, F. R. M. (org.). **Etapas da construção científica**: da curiosidade acadêmica à publicação dos resultados*. Pelotas: Ed. UFPel, 2018. p.74-91.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. Traduzido por Ana Maria P. Cardoso. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, 1996.

SHVAIKO, P.; FARAZI, F.; MALTESE, V.; IVANYUKOVICH, A.; RIZZI, V.; FERRARI, D.; UCELLI, G. Trentino government linked open geo-data: a case study. *In*: INTERNATIONAL SEMANTIC WEB CONFERENCE, 11., 2012, Boston. **Proceedings [...]** Berlin: Springer, 2012. p. 196-211. DOI: 10.1007/978-3-642-35173-0_13.

SILVEIRA, D. T.; CÓRDOVA, F. P. A pesquisa científica. *In*: GERHARDT, T. E.; SILVEIRA, D. T. (org.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009. p.31-42.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 132–141, 2004.

US OFFICE OF MANAGEMENT AND BUDGET. **Circular No. A-16 Revised**: Coordination of geographic information and related spatial data activities. 2002. Disponível em: http://www.whitehouse.gov/omb/circulars_a016_rev#2b2. Acesso em: 12 set. 2021.

VALLE, E. D.; CERI, S. Querying the Semantic Web: SPARQL. *In*: DOMINGUE, J. (ed.). **Handbook of semantic Web technologies**. Berlin: Springer, 2011. P. 299-363.

VILCHES-BLÁZQUEZ, L. M.; SAAVEDRA, J. A framework for connecting two interoperability universes: OGC Web Feature Services and Linked Data. **Transactions in GIS**, Cambridge, v. 23, n. 1, p. 22–47, 2019.

VINHAS, L. **Representações para dados geográficos**. Disciplina CAP 349 - Bancos de dados geográficos. São José dos Campos: INPE, 2020. 58 slides.

WIEMANN, S.; BERNARD, L. Spatial data fusion in Spatial Data Infrastructures using Linked Data. **International Journal of Geographical Information Science**, London, v. 30, n. 4, p. 613–636, 2016. DOI: 10.1080/13658816.2015.1084420.

WOODLEY, M. S. Metadata matters: connecting people and information. *In*: BACA, M. (org.). **Introduction to Metadata**. 3ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/metadata-matters>. Acesso em: 3 ago. 2020.

W3C - WORLD WIDE WEB CONSORTIUM. Data on the Web Best Practices. W3C Recommendation. 2017. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 3 ago. 2020.

W3C - WORLD WIDE WEB CONSORTIUM; OGC - OPEN GEOSPATIAL CONSORTIUM. Spatial Data on the Web Best Practices. W3C Working Group Note. 2017. Disponível em: <https://www.w3.org/TR/sdw-bp/#bp-metadata>. Acesso em: 10 maio 2021.

YUAN, J.; YUE, P.; GONG, J.; ZHANG, M. A Linked Data approach for geospatial data provenance. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 51, n. 11, p. 5105-5112, 2013. DOI: 10.1109/TGRS.2013.2249523.

YUE, P.; GONG, J.; DI, L.; HE, L. Automatic geospatial metadata generation for Earth Science virtual data products. **GeoInformatica**, Boston, v. 16, n. 1, p. 1–29, 2012.

YUE, P.; GUO, X.; ZHANG, M.; JIANG, L.; ZHAI, X. Linked Data and SDI: the case on Web geoprocessing workflows. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 114, p. 245–257, 2016.

ZHANG, M.; YUAN, J.; GONG, J.; YUE, P. An interlinking approach for linked geospatial data. *In*: INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES, 2013, Antalya. **Proceedings** [...]. Hannover: Copernicus GmbH, 2013. p.283-287. DOI: 10.5194/isprsarchives-XL-7-W2-283-2013.

ANEXO A

Íntegra do registro de metadados ‘Incrementos de Desmatamento para o Bioma Amazônia a partir do ano de 2008’, obtido da plataforma TerraBrasilis/INPE, utilizado como amostra para validação do framework (íntegra do registro apresentado na Figura 33).

```
<?xml version="1.0" encoding="UTF-8"?>
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd"
  xmlns:gml="http://www.opengis.net/gml"
  xmlns:gco="http://www.isotc211.org/2005/gco"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.isotc211.org/2005/gmd
    http://www.isotc211.org/2005/gmd/gmd.xsd
    http://www.isotc211.org/2005/gmx
    http://www.isotc211.org/2005/gmx/gmx.xsd
    http://www.isotc211.org/2005/srv
    http://schemas.opengis.net/iso/19139/20060504/srv/srv.xsd">
  <gmd:fileIdentifier>
    <gco:CharacterString>b75b83db-8026-43f9-9537-ee1dfa308158
    </gco:CharacterString>
  </gmd:fileIdentifier>
  <gmd:language>
    <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/"
      codeListValue="por"/>
  </gmd:language>
  <gmd:characterSet>
    <gmd:MD_CharacterSetCode codeListValue="utf8"
      codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_1913
        9_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_CharacterSetCode"/>
  </gmd:characterSet>
  <gmd:contact>
    <gmd:CI_ResponsibleParty>
      <gmd:individualName>
        <gco:CharacterString>André Fernando Araújo de Carvalho
        </gco:CharacterString>
      </gmd:individualName>
      <gmd:organisationName>
        <gco:CharacterString>Divisão de Processamento de Imagens - DPI
        </gco:CharacterString>
      </gmd:organisationName>
      <gmd:positionName gco:nilReason="missing">
        <gco:CharacterString/>
      </gmd:positionName>
      <gmd:contactInfo>
        <gmd:CI_Contact>
          <gmd:phone>
            <gmd:CI_Telephone>
              <gmd:voice gco:nilReason="missing">
                <gco:CharacterString/>
              </gmd:voice>
              <gmd:facsimile gco:nilReason="missing">
                <gco:CharacterString/>
              </gmd:facsimile>
            </gmd:CI_Telephone>
          </gmd:phone>
          <gmd:address>
            <gmd:CI_Address>
              <gmd:deliveryPoint gco:nilReason="missing">
                <gco:CharacterString/>
              </gmd:deliveryPoint>
              <gmd:city gco:nilReason="missing">
                <gco:CharacterString/>
              </gmd:city>
            </gmd:CI_Address>
          </gmd:address>
        </gmd:CI_Contact>
      </gmd:contactInfo>
    </gmd:CI_ResponsibleParty>
  </gmd:contact>
</gmd:MD_Metadata>
```

```

        </gmd:city>
        <gmd:administrativeArea gco:nilReason="missing">
          <gco:CharacterString/>
        </gmd:administrativeArea>
        <gmd:postalCode gco:nilReason="missing">
          <gco:CharacterString/>
        </gmd:postalCode>
        <gmd:country gco:nilReason="missing">
          <gco:CharacterString/>
        </gmd:country>
        <gmd:electronicMailAddress>
          <gco:CharacterString>andre.carvalho@inpe.br
          </gco:CharacterString>
        </gmd:electronicMailAddress>
      </gmd:CI_Address>
    </gmd:address>
  </gmd:CI_Contact>
</gmd:contactInfo>
<gmd:role>
  <gmd:CI_RoleCode codeListValue="pointOfContact"
    codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#CI_RoleCode"/>
</gmd:role>
</gmd:CI_ResponsibleParty>
</gmd:contact>
<gmd:dateStamp>
  <gco:DateTime>2022-06-15T00:14:49</gco:DateTime>
</gmd:dateStamp>
<gmd:metadataStandardName>
  <gco:CharacterString>ISO 19115:2003/19139</gco:CharacterString>
</gmd:metadataStandardName>
<gmd:metadataStandardVersion>
  <gco:CharacterString>1.0</gco:CharacterString>
</gmd:metadataStandardVersion>
<gmd:referenceSystemInfo>
  <gmd:MD_ReferenceSystem>
    <gmd:referenceSystemIdentifier>
      <gmd:RS_Identifier>
        <gmd:code>
          <gco:CharacterString>EPSG:4674</gco:CharacterString>
        </gmd:code>
      </gmd:RS_Identifier>
    </gmd:referenceSystemIdentifier>
  </gmd:MD_ReferenceSystem>
</gmd:referenceSystemInfo>
<gmd:identificationInfo>
  <gmd:MD_DataIdentification>
    <gmd:citation>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>Incremento de desmatamento do Bioma
            Amazônia a partir de 2008</gco:CharacterString>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:DateTime>2022-06-13T12:00:00</gco:DateTime>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode codeListValue="publication"
                codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#CI_DateTypeCode"/>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
      </gmd:CI_Citation>
    </gmd:citation>
  </gmd:MD_DataIdentification>
</gmd:identificationInfo>

```

<gmd:abstract>

<gco:CharacterString>Área desmatada a partir de 2008 discretizadas por ano e recortadas pelos limites do bioma Amazônia(*).

O mapeamento utiliza imagens do satélite Landsat ou similares, para registrar e quantificar as áreas desmatadas maiores que 6,25 hectares. O PRODES considera como desmatamento a supressão da vegetação nativa, independentemente da futura utilização destas áreas.

Os limites dos biomas brasileiros foram alterados conforme publicação do IBGE de 30/10/2019. Este conjunto de dados foi ajustado para o novo recorte.
<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/25798-ibge-lanca-mapa-inedito-de-biomas-e-sistema-costeiro-marinho>

(*) Consulte os metadados de "Limites do Bioma Amazônia" para informações sobre os limites do bioma que estamos usando nesta operação:
<http://terrabrasilis.dpi.inpe.br/geonetwork/srv/eng/catalog.search#/metadata/481e439e-b0fa-4cb0-890b-11941264db69>

Descrição dos atributos do dado composto por: nome do atributo, tipo do dado e descrição

 uuid - uuid - identificador único universal de cada feição
 uid - número inteiro - identificador para rastreabilidade da feição na origem/produção do dado
 state - texto - estado/unidade da federação
 path_row - texto - código da cena formado por linha/coluna da grade de passagem do satélite Landsat
 main_class - texto - nome da classe principal atribuída à feição
 class_name - texto - nome da classe específica atribuída à feição
 def_cloud - número - Ano que indica desde quando esta área está coberta por nuvens. Fazendo a diferença do ano corrente onde se detectou o desmatamento e def_cloud, temos o número de anos coberto por nuvens.
 julian_day - número - dia juliano
 image_date - texto - data da cena usada para obter a feição
 year - número - ano do desmatamento, usado para facilitar as consultas ao banco de dados
 area_km - número - área calculada para a feição em km²
 scene_id - número - identificador da cena no banco de dados, usado para consultas
 publish_year - data - usado para permitir a publicação do dado no GeoServer com dimensão temporal
 source - texto - origem do polígono (amazonia, cerrado, caatinga, mata_atlantica, pampa, pantanal), caso o polígono tenha sido incorporado de outro projeto, por exemplo, devido à alteração do mapa de biomas ou ainda às zonas da Amazônia legal no cerrado e pantanal
 satellite - texto - nome do satélite. Em geral será o landsat mas em alguns casos, outros satélites podem ser utilizados
 sensor - texto - nome do sensor que obteve a imagem
 geom - feição composta por um ou mais polígonos - geometria obtida por interpretação visual de imagem de satélite

A lista de atributos é padronizada com o dado de desmatamento, classe principal do PRODES. Para as demais classes o preenchimento ocorre apenas quando aplicável.</gco:CharacterString>

</gmd:abstract>

<gmd:purpose>

<gco:CharacterString>O dado apresenta os incrementos de desmatamento para o Bioma Amazônia a partir de 2008.</gco:CharacterString>

</gmd:purpose>

<gmd:status>

<gmd:MD_ProgressCode codeListValue="onGoing" codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#MD_ProgressCode"/>

</gmd:status>

<gmd:pointOfContact>

<gmd:CI_ResponsibleParty>

```

    <gmd:individualName>
      <gco:CharacterString>André Fernando Araújo de Carvalho
    </gco:CharacterString>
    </gmd:individualName>
    <gmd:organisationName>
      <gco:CharacterString>Divisão de Processamento de Imagens -
        DPI/OBT/INPE</gco:CharacterString>
    </gmd:organisationName>
    <gmd:positionName gco:nilReason="missing">
      <gco:CharacterString/>
    </gmd:positionName>
    <gmd:contactInfo>
      <gmd:CI_Contact>
        <gmd:phone>
          <gmd:CI_Telephone>
            <gmd:voice gco:nilReason="missing">
              <gco:CharacterString/>
            </gmd:voice>
            <gmd:facsimile gco:nilReason="missing">
              <gco:CharacterString/>
            </gmd:facsimile>
          </gmd:CI_Telephone>
        </gmd:phone>
        <gmd:address>
          <gmd:CI_Address>
            <gmd:deliveryPoint gco:nilReason="missing">
              <gco:CharacterString/>
            </gmd:deliveryPoint>
            <gmd:city gco:nilReason="missing">
              <gco:CharacterString/>
            </gmd:city>
            <gmd:administrativeArea gco:nilReason="missing">
              <gco:CharacterString/>
            </gmd:administrativeArea>
            <gmd:postalCode gco:nilReason="missing">
              <gco:CharacterString/>
            </gmd:postalCode>
            <gmd:country gco:nilReason="missing">
              <gco:CharacterString/>
            </gmd:country>
            <gmd:electronicMailAddress>
              <gco:CharacterString>andre.carvalho@inpe.br
            </gco:CharacterString>
            </gmd:electronicMailAddress>
          </gmd:CI_Address>
        </gmd:address>
      </gmd:CI_Contact>
    </gmd:contactInfo>
    <gmd:role>
      <gmd:CI_RoleCode codeListValue="publisher"
        codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#CI_RoleCode"/>
    </gmd:role>
  </gmd:CI_ResponsibleParty>
</gmd:pointOfContact>
<gmd:pointOfContact>
  <gmd:CI_ResponsibleParty>
    <gmd:individualName>
      <gco:CharacterString>PRODES</gco:CharacterString>
    </gmd:individualName>
    <gmd:organisationName>
      <gco:CharacterString>INPE/MCTIC</gco:CharacterString>
    </gmd:organisationName>
    <gmd:contactInfo>
      <gmd:CI_Contact>
        <gmd:address>
          <gmd:CI_Address>

```

```

        <gmd:electronicMailAddress>
          <gco:CharacterString>prodes@dpi.inpe.br
        </gco:CharacterString>
        </gmd:electronicMailAddress>
      </gmd:CI_Address>
    </gmd:address>
  </gmd:CI_Contact>
</gmd:contactInfo>
<gmd:role>
  <gmd:CI_RoleCode codeList="http://standards.iso.org/ittf/
    PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist
    /ML_gmxCodelists.xml#CI_RoleCode" codeListValue="author"/>
  </gmd:role>
</gmd:CI_ResponsibleParty>
</gmd:pointOfContact>
<gmd:resourceMaintenance>
  <gmd:MD_MaintenanceInformation>
    <gmd:maintenanceAndUpdateFrequency>
      <gmd:MD_MaintenanceFrequencyCode codeListValue="annually"
        codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_MaintenanceFrequencyCode"/>
    </gmd:maintenanceAndUpdateFrequency>
  </gmd:MD_MaintenanceInformation>
</gmd:resourceMaintenance>
<gmd:gr4114444aphicOverview>
  <gmd:MD_BrowseGraphic>
    <gmd:fileName>
      <gco:CharacterString>http://terrabrasilis.dpi.inpe.br:/
        geonetwork/srv/api/records/b75b83db-8026-43f9-9537-
        eeldfa308158/attachments/Captura%20de%20tela%20de%202021-05-
        21%2011-40-36.png</gco:CharacterString>
    </gmd:fileName>
    <gmd:fileDescription>
      <gco:CharacterString>Print of yearly deforestation layer
    </gco:CharacterString>
    </gmd:fileDescription>
  </gmd:MD_BrowseGraphic>
</gmd:graphicOverview>
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>prodes</gco:CharacterString>
    </gmd:keyword>
    <gmd:keyword>
      <gco:CharacterString>desmatamento</gco:CharacterString>
    </gmd:keyword>
    <gmd:keyword>
      <gco:CharacterString>incremento</gco:CharacterString>
    </gmd:keyword>
    <gmd:keyword>
      <gco:CharacterString>Bioma Amazônia</gco:CharacterString>
    </gmd:keyword>
    <gmd:keyword>
      <gco:CharacterString>features</gco:CharacterString>
    </gmd:keyword>
  </gmd:MD_Keywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>Brazil</gco:CharacterString>
    </gmd:keyword>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
</gmd:descriptiveKeywords>

```

```

</gmd:keyword>
<gmd:type>
  <gmd:MD_KeywordTypeCode codeListValue="place"
    codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_KeywordTypeCode"/>
</gmd:type>
</gmd:MD_Keywords>
</gmd:descriptiveKeywords>
<gmd:resourceConstraints>
  <gmd:MD_LegalConstraints>
    <gmd:accessConstraints>
      <gmd:MD_RestrictionCode codeListValue="license"
        codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_RestrictionCode"/>
    </gmd:accessConstraints>
    <gmd:useConstraints>
      <gmd:MD_RestrictionCode codeListValue="license"
        codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_RestrictionCode"/>
    </gmd:useConstraints>
    <gmd:otherConstraints>
      <gco:CharacterString>Creative Commons - Atribuição-CompartilhaIgual 4.0 Internacional</gco:CharacterString>
    </gmd:otherConstraints>
  </gmd:MD_LegalConstraints>
</gmd:resourceConstraints>
<gmd:spatialRepresentationType>
  <gmd:MD_SpatialRepresentationTypeCode codeListValue="vector"
    codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_SpatialRepresentationTypeCode"/>
</gmd:spatialRepresentationType>
<gmd:spatialResolution>
  <gmd:MD_Resolution>
    <gmd:equivalentScale>
      <gmd:MD_RepresentativeFraction>
        <gmd:denominator>
          <gco:Integer>250000</gco:Integer>
        </gmd:denominator>
      </gmd:MD_RepresentativeFraction>
    </gmd:equivalentScale>
  </gmd:MD_Resolution>
</gmd:spatialResolution>
<gmd:language>
  <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/"
    codeListValue="por"/>
</gmd:language>
<gmd:characterSet>
  <gmd:MD_CharacterSetCode codeListValue="utf8"
    codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_CharacterSetCode"/>
</gmd:characterSet>
<gmd:topicCategory>
  <gmd:MD_TopicCategoryCode>imageryBaseMapsEarthCover
</gmd:MD_TopicCategoryCode>
</gmd:topicCategory>
<gmd:topicCategory>
  <gmd:MD_TopicCategoryCode>environment</gmd:MD_TopicCategoryCode>
</gmd:topicCategory>
<gmd:topicCategory>
  <gmd:MD_TopicCategoryCode>geoscientificInformation
</gmd:MD_TopicCategoryCode>
</gmd:topicCategory>
<gmd:extent>

```

```

    <gmd:EX_Extent>
      <gmd:temporalElement>
        <gmd:EX_TemporalExtent>
          <gmd:extent>
            <gml:TimePeriod gml:id="dl8857e395a1051934">
              <gml:beginPosition>2008-04-01</gml:beginPosition>
              <gml:endPosition>2021-07-31</gml:endPosition>
            </gml:TimePeriod>
          </gmd:extent>
        </gmd:EX_TemporalExtent>
      </gmd:temporalElement>
    </gmd:EX_Extent>
  </gmd:extent>
  <gmd:extent>
    <gmd:EX_Extent>
      <gmd:geographicElement>
        <gmd:EX_GeographicBoundingBox>
          <gmd:westBoundLongitude>
            <gco:Decimal>-74</gco:Decimal>
          </gmd:westBoundLongitude>
          <gmd:eastBoundLongitude>
            <gco:Decimal>-44</gco:Decimal>
          </gmd:eastBoundLongitude>
          <gmd:southBoundLatitude>
            <gco:Decimal>-17.5</gco:Decimal>
          </gmd:southBoundLatitude>
          <gmd:northBoundLatitude>
            <gco:Decimal>6</gco:Decimal>
          </gmd:northBoundLatitude>
        </gmd:EX_GeographicBoundingBox>
      </gmd:geographicElement>
    </gmd:EX_Extent>
  </gmd:extent>
  <gmd:supplementalInformation>
    <gco:CharacterString>Página Web com informações detalhadas sobre o
    projeto, o processo de produção dos dados e opção para download do
    arquivo de mídia digital no formato compactado ZIP contendo os
    arquivos do formato Shapefile. Acesse http://www.inpe.br/ e procure
    por PRODES.</gco:CharacterString>
  </gmd:supplementalInformation>
</gmd:MD_DataIdentification>
</gmd:identificationInfo>
<gmd:distributionInfo xmlns:che="http://www.geocat.ch/2008/che">
  <gmd:MD_Distribution>
    <gmd:distributionFormat>
      <gmd:MD_Format>
        <gmd:name>
          <gco:CharacterString>vetorial</gco:CharacterString>
        </gmd:name>
        <gmd:version gco:nilReason="missing">
          <gco:CharacterString/>
        </gmd:version>
      </gmd:MD_Format>
    </gmd:distributionFormat>
  </gmd:distributionInfo>
  <gmd:distributor>
    <gmd:MD_Distributor>
      <gmd:distributorContact>
        <gmd:CI_ResponsibleParty>
          <gmd:individualName>
            <gco:CharacterString>André Fernando Araújo de Carvalho
            </gco:CharacterString>
          </gmd:individualName>
          <gmd:organisationName>
            <gco:CharacterString>Divisão de Processamento de Imagens
            - DPI</gco:CharacterString>
          </gmd:organisationName>
          <gmd:contactInfo>
            <gmd:CI_Contact>

```

```

        <gmd:address>
            <gmd:CI_Address>
                <gmd:electronicMailAddress>
                    <gco:CharacterString>andre.carvalho@inpe.br
                </gco:CharacterString>
                </gmd:electronicMailAddress>
            </gmd:CI_Address>
        </gmd:address>
    </gmd:CI_Contact>
</gmd:contactInfo>
<gmd:role>
    <gmd:CI_RoleCode codeList="http://standards.iso.org/ittf/
        PubliclyAvailableStandards/ISO_19139_Schemas/resources/c
        odelist/ML_gmxCodelists.xml#CI_RoleCode"
        codeListValue="pointOfContact"/>
</gmd:role>
</gmd:CI_ResponsibleParty>
</gmd:distributorContact>
</gmd:MD_Distributor>
</gmd:distributor>
<gmd:transferOptions>
    <gmd:MD_DigitalTransferOptions>
        <gmd:onLine>
            <gmd:CI_OnlineResource>
                <gmd:linkage>
                    <gmd:URL>http://terrabrasilis.dpi.inpe.br/app/map/
                        deforestation</gmd:URL>
                </gmd:linkage>
                <gmd:protocol>
                    <gco:CharacterString>WWW:LINK-1.0-http-link
                </gco:CharacterString>
                </gmd:protocol>
                <gmd:name>
                    <gco:CharacterString>Portal TerraBrasilis
                </gco:CharacterString>
                </gmd:name>
                <gmd:description>
                    <gco:CharacterString>Mapa para acesso online aos dados de
                        desmatamento na Amazônia e demais biomas
                </gco:CharacterString>
                </gmd:description>
            </gmd:CI_OnlineResource>
        </gmd:onLine>
    </gmd:onLine>
    <gmd:CI_OnlineResource>
        <gmd:linkage>
            <gmd:URL>http://terrabrasilis.dpi.inpe.br/geoserver/
                prodes-amazon-nb/wms</gmd:URL>
        </gmd:linkage>
        <gmd:protocol>
            <gco:CharacterString>OGC:WMS-1.3.0-http-get-capabilities
            </gco:CharacterString>
        </gmd:protocol>
        <gmd:name>
            <gco:CharacterString>yearly_deforestation_biome
            </gco:CharacterString>
        </gmd:name>
        <gmd:description>
            <gco:CharacterString>Yearly deforestation biome since
                2008</gco:CharacterString>
        </gmd:description>
        <gmd:function>
            <gmd:CI_OnLineFunctionCode codeList="http://standards.iso
                .org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/re
                sources/codelist/ML_gmxCodelists.xml#CI_OnLineFunctionCod
                e" codeListValue="order"/>
        </gmd:function>
    </gmd:CI_OnlineResource>

```

```

    </gmd:onLine>
    <gmd:onLine>
      <gmd:CI_OnlineResource>
        <gmd:linkage>
          <gmd:URL>http://terrabrasilis.dpi.inpe.br/downloads/
            #cat_5</gmd:URL>
        </gmd:linkage>
        <gmd:protocol>
          <gco:CharacterString>WWW:DOWNLOAD-1.0-http-download
          </gco:CharacterString>
        </gmd:protocol>
        <gmd:name>
          <gmx:MimeType xmlns:gmx="http://www.isotc211.org/
            2005/gmx" type="application/octet-stream">Desmatamento -
            Bioma Amazônia</gmx:MimeType>
        </gmd:name>
        <gmd:description>
          <gco:CharacterString>Lista de arquivos Shapefile
            compactados em formato ZIP para os temas relacionados ao
            desmatamento. Acesse para ver os detalhes.
          </gco:CharacterString>
        </gmd:description>
        <gmd:function>
          <gmd:CI_OnlineFunctionCode codeList="http://standards.iso
            .org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/r
            esources/codelist/ML_gmxCodelists.xml#CI_OnlineFunctionC
            ode" codeListValue="download"/>
        </gmd:function>
      </gmd:CI_OnlineResource>
    </gmd:onLine>
  </gmd:MD_DigitalTransferOptions>
</gmd:transferOptions>
</gmd:MD_Distribution>
</gmd:distributionInfo>
<gmd:dataQualityInfo>
  <gmd:DQ_DataQuality>
    <gmd:scope>
      <gmd:DQ_Scope>
        <gmd:level>
          <gmd:MD_ScopeCode codeListValue="dataset"
            codeList="http://standards.iso.org/ittf/PubliclyAvailableStand
            ards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#
            MD_ScopeCode"/>
        </gmd:level>
      </gmd:DQ_Scope>
    </gmd:scope>
    <gmd:lineage>
      <gmd:LI_Lineage>
        <gmd:statement>
          <gco:CharacterString>O dado foi criado em 1988 e vem sendo
            mantido desde então. A série publicada com frequência anual
            inicia em 2008 e vem sendo mantida desde então.
          </gco:CharacterString>
        </gmd:statement>
      </gmd:LI_Lineage>
    </gmd:lineage>
  </gmd:DQ_DataQuality>
</gmd:dataQualityInfo>
</gmd:MD_Metadata>

```