



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

MARCELO TOMITA

**TECNOLOGIAS PARA APLICAÇÃO DA WEB SEMÂNTICA  
NAS UNIDADES DE INFORMAÇÃO**

---

Londrina  
2015

MARCELO TOMITA

**TECNOLOGIAS PARA APLICAÇÃO DA WEB SEMÂNTICA  
NAS UNIDADES DE INFORMAÇÃO**

Dissertação apresentada ao Programa de Pós Graduação em Ciência da Informação do Departamento de Ciência da Informação da Universidade Estadual de Londrina para obtenção do título de Mestre no Curso de Mestrado em Ciência da Informação.

Orientadora: Profa. Dra. Maria Elisabete Catarino.

Londrina  
2015

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca  
Central da Universidade Estadual de Londrina**

**Dados Internacionais de Catalogação-na-Publicação (CIP)**

T657t Tomita, Marcelo.

Tecnologias para aplicação da web semântica nas unidades de informação / Marcelo Tomita. – Londrina, 2015.  
148 f. : il.

Orientador: Maria Elisabete Catarino.

Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual de Londrina, Centro de Educação, Comunicação e Artes, Programa de Pós-Graduação em Ciência da Informação, 2015.

Inclui bibliografia.

1. Framework (Programa de computador) – Teses. 2. Web semântica – Teses. 3. Unidades de informação – Teses. 4. Ciência da informação – Teses. I. Catarino, Maria Elisabete. II. Universidade Estadual de Londrina. Centro de Educação, Comunicação e Artes. Programa de Pós-Graduação em Ciência da Informação. III. Título.

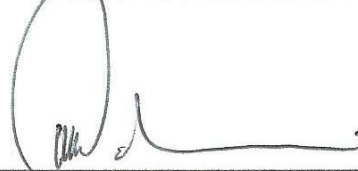
CDU 02:519.68.04

MARCELO TOMITA

**TECNOLOGIAS PARA APLICAÇÃO DA WEB SEMÂNTICA NAS  
UNIDADES DE INFORMAÇÃO**

Dissertação apresentada ao Programa de Pós Graduação em Ciência da Informação do Departamento de Ciência da Informação da Universidade Estadual de Londrina para obtenção do título de Mestre no Curso de Mestrado em Ciência da Informação.

**BANCA EXAMINADORA**



---

Orientadora: Profa. Dra. Maria Elisabete  
Catarino  
Universidade Estadual de Londrina



---

Profa. Dra. Silvana Drummond Monteiro  
Universidade Estadual de Londrina



---

Prof. Dr. Rodolfo Miranda de Barros  
Universidade Estadual de Londrina

Londrina, 23 de Outubro de 2015.

À Deus, que sempre abençoou-me nos  
momentos de dificuldades

À Tiemi, minha filha e fiel amiga

Aos meus pais, pelo apoio, ensinamentos e  
valores construídos

## **AGRADECIMENTOS**

Agradeço à minha orientadora Maria Elisabete Catarino, professora e profissional que estava presente em todos os momentos (finais de semana, feriado e de madrugada). Sempre com palavras de incentivo e ideias novas que encaixavam exatamente na escrita da dissertação. Ficarei perene a imagem de uma profissional que terei muita satisfação em espelhar-me.

À professora Silvana Drumond, em que tive a oportunidade de apropriar conhecimentos importantes para o desenvolvimento da pesquisa e pelas contribuições na qualificação.

Ao professor Rodolfo Miranda que apresentou uma visão tecnológica no projeto com a sugestão da criação do Framework que enriqueceu significativamente o trabalho.

Ao departamento da Ciência da Informação, na figura de seus professores, funcionários e alunos.

Aos meus pais que oportunizaram este momento ao permitir que eu me desligasse de todos os problemas e focasse somente neste trabalho.

À minha filha Tiemi, que manteve-se firme, teve paciência, enfrentou junto as dificuldades e entendeu todos os finais de semana e feriados que precisei ficar concentrado na escrita.

Ao SENAC, na figura de Sidnei de Oliveira, diretor de RH e Denny Enzo, atual gestor de Londrina, possibilitou-me por meio de ajustes de horários, ajuda de custo em viagens, cessão de espaço da instituição para reuniões com colegas e orientações e nas conversas de incentivo.

À Lucimara, companheira sensata que soube no momento oportuno, intervir para participar ativamente dos meus planos.

Já ancorado na Antártica, ouvi ruídos que pareciam de fritura. Pensei: Será que até aqui existem chineses fritando pastéis?

Eram cristais de água doce congelada que faziam aquele som quando entravam em contato com a água salgada. O efeito visual era belíssimo. Pensei em fotografar, mas falei pra mim mesmo - Calma, você terá muito tempo para isso... Nos 637 dias que seguiram o fenômeno não se repetiu. As oportunidades são únicas.

AmyrKlink

TOMITA, Marcelo. **Tecnologias para aplicação da Web Semântica nas Unidades de Informação**. 2015. 148f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual de Londrina, Londrina.

## RESUMO

O projeto *Web Semântica* requer o emprego de várias tecnologias para a organização e estruturação dos conteúdos na *Web*, que são imprescindíveis para o armazenamento, localização e recuperação da informação nos ambientes das unidades de informação. Este estudo descreve as tecnologias, recomendações e padrões necessários para a aplicação da *Web Semântica* nas unidades de informação. Realizou-se uma abordagem qualitativa como proposta metodológica do trabalho, que permitiu uma percepção minuciosa das características do objeto de pesquisa. As técnicas adotadas foram os estudos bibliográficos e documentais, aliadas à metodologia de estudo de caso, que proporcionou uma percepção mais precisa das tecnologias que abrangem o projeto da *Web Semântica*, sendo a biblioteca virtual Europeia, o objeto de estudo de caso selecionado. Para a realização da investigação do estudo de caso organizou-se em cinco categorias que tornaram-se itens elementares da pesquisa durante todo o processo de análise de resultados. Elaborou-se um *framework* com o propósito de representar o processo de estabelecimento da *Web Semântica* nas unidades de Informação. Os resultados da análise possibilitaram compreender que as unidades de informação devem formatar os métodos de catalogação, normalizações, indexação, classificação e resumos em padrões recomendados pela W3C. Ficou evidente que existem diversas soluções para a implementação da *Web Semântica* nas unidades de informação. As tecnologias, recomendações e padrões podem ser distintos para cada unidade, no entanto, entende-se que o RDF, para representar significado aos dados, a linguagem XML para transporte da estrutura dos dados, a linguagem de programação para o desenvolvimento de ontologias e vocabulários, um banco de dados orientado a documentos e uma linguagem de busca como a SPARQL, são elementos essenciais para consolidação da *Web Semântica* nas unidades de informação.

**Palavras-chave:** *Europeana*. *Framework*. Organização da Informação. Unidades de Informação. *Web Semântica*.

TOMITA, Marcelo. **Technologies for application of Semantic Web in information units.** 2015. 148f. Dissertation (Master's Degree in Information Science) – State University of Londrina, Londrina.

## ABSTRACT

The Semantic Web project requires the use of various technologies for the organization and structuring of content on the Web, which are essential for storage, location and retrieval information in the environments of the information units. This study describes the technologies, recommendations and standards necessary for the application of Semantic Web in information units. It was conducted a qualitative approach as a methodological proposal of work, allowing a thorough insight into the research object characteristics. The adopted techniques were the bibliographic and documentary studies, with the case study methodology, which provided better accurate perception of the technologies that cover the Semantic Web project, being the virtual library Europeana the object of selected case study. To carry out the case study research it was organized into five categories that made up the basic items of research results throughout the analysis process. A framework was developed for the purpose of representing the process of the Semantic Web in information units. Analysis results made it possible to understand that the information units must format the cataloging methods, normalization, indexing, classification and summaries standards recommended by the W3C. It became clear that there are various solutions for the implementation of the Semantic Web in information units, The technologies, recommendations and patterns can be different for each unit, however, it is understood that the RDF to represent meaning to the data, the XML language to transport the data structure, the programming language to develop ontologies and vocabularies, a database oriented to documents and a query language like SPARQL are essentials elements to consolidate the Semantic Web in information units.

**Keywords:** Europeana. Framework. Information organization. Information units. Semantic Web.

## LISTA DE FIGURAS

<b>Figura 1</b>	–	Técnicas de OI nas Unidades de Informação.....	43
<b>Figura 2</b>	–	Pseudocódigo da Tripla RDF.....	51
<b>Figura 3</b>	–	Gráfico RDF.....	52
<b>Figura 4</b>	–	Tecnologias, recomendações, linguagens e padrões da WS.....	58
<b>Figura 5</b>	–	Estatueta de Pedra (Chipre).....	64
<b>Figura 6</b>	–	As três classes núcleo do EDM.....	66
<b>Figura 7</b>	–	Arquitetura Técnica Europeia.....	72
<b>Figura 8</b>	–	Estrutura Básica dos Recursos de Rede do EDM.....	73
<b>Figura 9</b>	–	Agregadores no modelo Europeia.....	76
<b>Figura 10</b>	–	Organizações envolvidas no Harvesting.....	77
<b>Figura 11</b>	–	Processo para submissão de dados à Europeia.....	79
<b>Figura 12</b>	–	Referências.....	84
<b>Figura 13</b>	–	Documentos Incorporados.....	85
<b>Figura 14</b>	–	Código em SPARQL para retornar cinco itens.....	87
<b>Figura 15</b>	–	Resultado da consulta de cinco itens.....	88
<b>Figura 16</b>	–	Código em SPARQL para listar três registros em que o idioma seja “Inglês” e Assunto “Law”.....	89
<b>Figura 17</b>	–	Resultado da consulta para exibir registros com idioma em português.....	90
<b>Figura 18</b>	–	Enriquecimento - Castelo de Bragança.....	97
<b>Figura 19</b>	–	Geographical Coverage.....	98
<b>Figura 20</b>	–	What.....	99
<b>Figura 21</b>	–	When.....	100
<b>Figura 22</b>	–	Web Semântica em Unidades de Informação.....	101
<b>Figura 23</b>	–	Web Semântica em Unidades de Informação.....	103
<b>Figura 24</b>	–	Seção Unidade Informacional.....	104
<b>Figura 25</b>	–	Seção Fornecedores de dados.....	106
<b>Figura 26</b>	–	Seção Agregadores.....	107
<b>Figura 27</b>	–	Seção Biblioteca Virtual.....	108
<b>Figura 28</b>	–	Realizar Consulta.....	109

## LISTA DE QUADROS

<b>Quadro 1</b>	–	Categorias de Estudo .....	21
<b>Quadro 2</b>	–	Metodologia Utilizada .....	35
<b>Quadro 3</b>	–	Instrumentos de pesquisa.....	41
<b>Quadro 4</b>	–	Normas Internacionais.....	41
<b>Quadro 5</b>	–	Elementos do DC .....	48
<b>Quadro 6</b>	–	Recurso, Propriedade e Valor .....	52
<b>Quadro 7</b>	–	História da Fundação Europeia .....	61
<b>Quadro 8</b>	–	Exemplo de um Registro ESE Europeia.....	63
<b>Quadro 9</b>	–	Propriedades das Classes Núcleo.....	66
<b>Quadro 10</b>	–	Propriedade das Classes Contextuais.....	68

## LISTA DE ABREVIATURAS SIGLAS

AACR2	Anglo-American Cataloguing Rules
API	Application Programming Interface
BDTD	Biblioteca Digital de Teses e Dissertações
CC	Ciência da Computação
CI	Ciência da Informação
EDM	Europeana Data Model
EFG	Europeana Film Gateway
ePMG	Padrão de Metadados do Governo Eletrônico
ESE	Europeana Semantic Elements
DC	Dublin Core
HTML	Hyper Text Markup Language
INESC-ID	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa
IRI	International Resource Identifier
ISAAR (CPF)	Norma Internacional de Registro de Autoridades Arquivísticas para Pessoas Coletivas, Pessoas Singulares e Famílias
ISAD(G)	Norma Geral Internacional de Descrição Arquivística
ISBD	International Standard Bibliographic Description
MPEG	Moving Picture Experts Group
MTD-BR	Padrão Brasileiro de Metadados para Teses e Dissertações
NOBRADE	Norma Brasileira de Descrição Arquivística
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
OI	Organização da Informação
OPAC	On-line Public Access Catalogue
OSI	Open Source Initiative
OWL	Web Ontology Language
RDA	Resource Description Access
RDF	Resource Description Framework
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Server Query Language

TEL	The European Library
URI	Universal Resource Identifier
URL	Uniform Resource LocatorW3C World Wide Web Consortium
WS	Web Semântica
XML	eXtensible Markup Language
XLST	Extensible Stylesheet Language Transformations

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>14</b>
1.1	DEFINIÇÃO DO PROBLEMA.....	15
1.2	OBJETIVOS .....	16
1.2.1	Objetivo Geral.....	16
1.2.2	Objetivos específicos.....	17
1.3	METODOLOGIA.....	17
1.4	JUSTIFICATIVA.....	22
1.5	ESTRUTURA DO TRABALHO .....	23
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>25</b>
2.1	A INFORMAÇÃO .....	25
2.2	CIÊNCIA DA INFORMAÇÃO.....	26
2.3	ORGANIZAÇÃO DA INFORMAÇÃO .....	28
2.4	GESTÃO DE DOCUMENTOS.....	29
2.5	ÁREAS QUE SE INTERAGEM .....	30
2.6	INTERNET, A <i>WEB</i> E A <i>WEB SEMÂNTICA</i> .....	32
<b>3</b>	<b>APRESENTAÇÃO E ANÁLISE DOS RESULTADOS</b> .....	<b>35</b>
3.1	TÉCNICAS DA OI NAS UNIDADES DE INFORMAÇÃO .....	36
3.2	ANALISAR AS TECNOLOGIAS, RECOMENDAÇÕES, LINGUAGENS E PADRÕES QUE COMPÕEM A <i>WEB SEMÂNTICA</i> .....	43
3.2.1	Estrutura para uma Web Semântica.....	44
3.2.2	Metadados .....	45
3.2.3	<i>Dublin Core</i> .....	47
3.2.4	RDF ( <i>Resource Description Framework</i> ).....	49
3.2.5	Ontologias .....	53
3.2.6	Linguagens da <i>Web Semântica</i> .....	54

3.3	APLICABILIDADE DA WEB SEMÂNTICA EM AMBIENTES DAS UNIDADES DE INFORMAÇÃO, POR MEIO DE UM ESTUDO DE CASO E UM <i>FRAMEWORK</i> .....	60
3.3.1	As Categorias de Estudo .....	62
3.3.1.1	Metadados na Europeana .....	62
3.3.1.2	<i>Harvesting</i> , a coleta de dados do portal Europeana .....	75
3.3.1.3	Linguagem de desenvolvimento e banco de dados .....	82
3.3.1.4	Linguagem de busca .....	85
3.3.1.5	Ontologias e vocabulários para enriquecimento de dados .....	92
3.3.2	Europeana em Síntese .....	101
3.3.3	Um Framework para representar a Web Semântica nas Unidades de Informação.....	103
<b>4</b>	<b>CONSIDERAÇÕES</b> .....	<b>110</b>
4.1	RESULTADOS OBTIDOS.....	110
4.2	TRABALHOS FUTUROS .....	113
4.3	CONSIDERAÇÕES FINAIS .....	114
	<b>REFERÊNCIAS</b> .....	<b>115</b>
	<b>APÊNDICES</b> .....	<b>127</b>
	APÊNDICE A.....	128
	APÊNDICE B.....	145

## 1 INTRODUÇÃO

É nítido como a *Web* desenvolve-se verticalmente. Os conteúdos armazenados e disponíveis são incalculáveis, visto que nunca na humanidade produziu-se como hoje e as estatísticas atuais não são capazes de dimensionar com exatidão o tamanho real da *Web*.

A *Web* está tornando-se imprescindível para armazenar todo o conhecimento exigido nas relações profissionais e pessoais. É como uma extensão dos nossos pensamentos, pois é praticamente impossível armazenar, organizar e assimilar todas as informações necessárias.

Quando um *website* é acessado, é possível identificar por meio de uma rápida visualização, dezenas de elementos nessa página, distinguir seus componentes, compreender suas informações, permitindo várias interpretações sobre esse conteúdo. Mas um computador ou *Gadget*<sup>1</sup>, mesmo com seus algoritmos mais recentes e inovadores, ainda é incapaz de realizar de forma consistente, essa interpretação.

Existem diversos tipos de estoques informacionais, dentre eles os Arquivos e Bibliotecas, os quais são objetos de estudo nesta pesquisa, em que se observa a problemática do volume de informações armazenadas e organizadas para serem recuperadas, principalmente na *Web*.

O consórcio *World Wide Web Consortium (W3C)*<sup>2</sup> tem sido a principal organização que fomenta a padronização da *Web* ao desenvolver especificações técnicas e orientações para que as recomendações possam estar ao alcance das pessoas que produzem e disponibilizam conteúdos na *Web*. Dentre as recomendações para a organização estrutural da *Web*, destaca-se o projeto da *Web Semântica*, que de acordo com os pesquisadores Berners-Lee, Hendler e Lassila (2001, p.1) “[...] é uma extensão da *Web* atual, na qual é dado à informação um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação.”, e segundo Levy (2012) vai permitir que máquinas analisem os dados de maneira mais orgânica.

---

<sup>1</sup> *Gadget*: um dispositivo pequeno, muitas vezes mecânico ou eletrônico com um uso prático, muitas vezes visto como uma novidade (DICIONÁRIO Merriam-Webster, 2014, tradução nossa).

<sup>2</sup> Comunidade internacional que desenvolve padrões com o objetivo de garantir o crescimento da web. (WORLD WIDE WEB CONSORTIUM, 2014b).

Para aperfeiçoar essa *Web*, existem diversos recursos tecnológicos recomendados que podem ser empregado: a linguagem de marcação para estruturação e descrição de dados, *eXtensible Markup Language* (XML); a linguagem para construção de ontologias, *Web Ontology Language* (OWL); o modelo *Resource Description Framework* (RDF), que permite a estruturação dos dados e a interoperabilidade entre sistemas; a linguagem de buscas *Protocol and RDF Query Language* (SPARQL), criada especificamente para a localização de dados na *Web Semântica*, entre outros (WORLD WIDE WEB CONSORTIUM, 2011b).

Por outro lado, na Ciência da Informação temos a Organização da Informação com suas tradicionais tecnologias que organizam e estruturam as informações, por intermédio das práticas de Representação Descritiva, tais como Catalogação para identificação e descrição dos documentos, metadados que sintetizam os dados para registro eletrônico; a Representação Temática que retrata os conteúdos e auxilia na recuperação da informação, empregando a indexação para extrair os termos mais representativos de documentos e referenciá-los para melhor recuperação, os tesouros com seus vocabulários controlados dentro de um domínio específico, as taxonomias que permitem criar estruturas classificatórias para organizar e reunir documentos de forma lógica e as ontologias que representam modelos de dados, abstraindo as especificações dos objetos.

A *Web Semântica* é um termo incomum para o usuário final, visto que não é uma interface que interage diretamente com pessoas, embora sua aplicação seja essencial para uma organização, busca, armazenamento e recuperação da informação na *Web*. O projeto da *Web Semântica* envolve diversas tecnologias que permeiam as várias áreas do conhecimento. Desse modo, o tema dessa pesquisa é apresentar e descrever o projeto do W3C e exemplificar por meio de um Estudo de Caso a aplicação dessas tecnologias nas Bibliotecas e Arquivos.

## 1.1 DEFINIÇÃO DO PROBLEMA

Dominar os conceitos, a estrutura e as tecnologias envolvidas no projeto da *Web Semântica* é uma tarefa complexa e pretensiosa, visto que envolvem vários elementos que tratam de conteúdos específicos e transitam entre diversos domínios

científicos, entre eles a Ciência da Computação e a Ciência da Informação, citando somente os campos mais evidentes e atuantes.

O projeto *Web Semântica* do W3C permite que os dados estejam estruturados e disponíveis na *Web* para que possam ser lidos e interpretados por máquinas. No contexto da ciência da informação, a área da organização se preocupa com a representação da informação (descritiva e temática) das unidades de informação<sup>3</sup>.

As unidades de informação que adquirem, armazenam e organizam diversos materiais de mídias distintas, sendo virtuais ou físicos, têm se ajustado às mudanças de comportamentos dos usuários frente às novas tecnologias. Estas mudanças não envolvem somente a manipulação dos objetos que deixaram de ser concretos, mas nas técnicas de tratamento das informações. No entanto, com a ratificação da *Web Semântica*, os processos tradicionais de indexação e classificação ainda são efetivos e eficazes para a realidade contemporânea.

Sendo o propósito deste estudo as unidades de informação dos tipos arquivos e bibliotecas, questiona-se: De que maneira pode-se aplicar as tecnologias, recomendações, linguagens e padrões da WS nas unidades de informação?

## 1.2 OBJETIVOS

Considerando-se que a questão levantada para esta pesquisa foia seguinte: “de que maneira pode-se aplicar as tecnologias, recomendações, linguagens e padrões da WS nas unidades de informação?”, foram definidos objetivos que permitiram responder a esta pergunta. A seguir apresentam-se os objetivos geral e específicos.

### 1.2.1 Objetivo Geral

---

<sup>3</sup> Unidade de Informação: entidade encarregada de adquirir, processar, armazenar e disseminar informações com o objetivo de satisfazer as necessidades do usuário. Em muitos casos é sinônimo de Biblioteca, agência de informação, [...] (CUNHA; CAVALCANTI, 2008, p.370).

O Objetivo Geral deste trabalho consiste em descrever as tecnologias necessárias para a aplicação da *Web Semântica* nas unidades de informação, especificamente neste estudo, nos arquivos e bibliotecas.

Utilizou-se um estudo de caso do site da biblioteca da Fundação *Europeana*, o que possibilitou caracterizar a realidade do ambiente de uma biblioteca virtual. Para este propósito, o trabalho foi organizado em três objetivos específicos:

### 1.2.2 Objetivos específicos

- Identificar as técnicas de Organização da Informação nas unidades de informação.
- Analisar tecnologias, recomendações, linguagens e padrões que compõem a *Web Semântica*.
- Demonstrar a aplicabilidade da *Web Semântica* em ambientes das unidades de informação, por meio de um estudo de caso e um framework.

### 1.3 METODOLOGIA

Para a realização deste estudo optou-se pelo emprego da abordagem qualitativa, uma vez que não houve a necessidade de medir ou enumerar eventos na análise dos dados. A abordagem qualitativa possibilita obter os dados descritivos mediante contato direto e interativo do pesquisador com a situação do objeto em estudo (NEVES, 1996). A adoção dessa abordagem permitiu uma visão minuciosa e uma melhor compreensão do ambiente da *Web Semântica* desde a concepção até a sua aplicação. Na afirmação de Maanen (1979a, p.520) a abordagem qualitativa:

Compreende um conjunto de diferentes técnicas interpretativas que visam a descrever e a decodificar os componentes de um sistema complexo de significados. Tem por objetivo traduzir e expressar os sentidos dos fenômenos do mundo social; trata-se de reduzir a distância entre o indicador e indicado, entre teoria e dados, entre contexto e ação.

Com o propósito de compreender os detalhes que possibilitam o funcionamento da *Web Semântica* e identificar as principais contribuições científicas desse tema, foram adotadas a pesquisa bibliográfica, a pesquisa documental e o método de estudo de caso.

A pesquisa bibliográfica de acordo com Gil (2002, p. 44)

[...] é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos. Embora em quase todos os estudos seja exigido algum tipo de trabalho dessa natureza, há pesquisas desenvolvidas exclusivamente a partir de fontes bibliográficas. As pesquisas sobre ideologias, bem como aquelas que se propõem à análise das diversas posições acerca de um problema, também costumam ser desenvolvidas quase exclusivamente mediante fontes bibliográficas.

A pesquisa bibliográfica é uma fonte confiável e precisa, por intermédio da qual localiza-se os livros, artigos e periódicos necessários para a construção do referencial teórico.

A pesquisa documental, segundo Gil (2002, p. 45),

[...] assemelha-se muito à pesquisa bibliográfica. A diferença essencial entre ambas está na natureza das fontes. Enquanto a pesquisa bibliográfica se utiliza fundamentalmente das contribuições dos diversos autores sobre determinado assunto, a pesquisa documental vale-se de materiais que não recebem ainda um tratamento analítico, ou que ainda podem ser reelaborados de acordo com os objetos da pesquisa.

A pesquisa documental é uma fonte com materiais que podem ainda não estarem prontos, ou que não foram publicados em meios científicos, mas que possuem um rico acervo, necessários para a constituição do trabalho. Pretende-se usufruir da pesquisa documental para consultar manuais, testes de desempenho, planos de trabalhos, relatórios e reportagens atuais.

O estudo de caso possibilita compreender os fenômenos sociais complexos com mais profundidade sob todos seus aspectos, valendo-se de variadas técnicas de pesquisa para identificar determinadas situações e relatar suas complexidades. Dessa forma, preserva-se as características globais e significativas dos acontecimentos da vida real, tais como os processos organizacionais e administrativos, mudanças ocorridas em regiões urbanas, relações internacionais e maturação dos setores econômicos (MARCONI; LAKATOS, 2008; YIN, 2005).

Para Ludke Menga e Marli (1986, 18-20), no Estudo de Caso, algumas características são fundamentais, como:

- a) visar a descoberta;
- b) enfatizar a interpretação do contexto;
- c) retratar a realidade de forma ampla;
- d) valer-se de fontes diversas de informações;
- e) permitir substituições;
- f) representar diferentes pontos de vista em dada situação;
- g) usar linguagem simples.

O estudo de caso possibilita a observação direta dos acontecimentos que estão sendo estudados e “[...] conta com o diferenciador de lidar com uma ampla variedade de evidências – documentos, artefatos, entrevistas e observações.” (YIN, 2005, p. 26).

Na técnica de estudo de caso, de acordo com Yin (2005, p. 40-41), “[...] os projetos de pesquisa de estudo de caso ainda não foram esquematizados [...]” e “[...] cada tipo de pesquisa empírica possui um projeto de pesquisa implícito.”. Entende-se que não há uma sistematização que deve ser seguida no desenvolvimento da pesquisa, as etapas podem ser redefinidas e ajustadas durante o progresso do estudo. No entanto, para que houvesse uma parametrização na realização do plano de pesquisa foram adotadas as etapas a seguir, conforme Nascimento et al.(2013, p.54-56) sugere:

1. **Preparação Teórico Metodológica:** consistirá em uma revisão da literatura disponível, buscando o embasamento teórico e conhecimento dos elementos pesquisados por meio de experimentos já concluídos.
2. **Seleção do Caso:** considerará as questões teóricas e práticas, procurando neutralizar ao máximo os fatores externos que possam interferir no processo a ser estudado. A seleção do exemplo não é feita por qualquer método estatístico de amostragem.
3. **Coleta de Dados:** atividade central do estudo de caso, a técnica permitirá levantar a maior quantidade de informações sobre o caso, com o objetivo de compreender os detalhes do objeto de estudo.
4. **Análise dos Dados:** possibilitará organizar e avaliar os dados coletados para que sejam detectadas as informações necessárias para a síntese e compreensão do estudo.

A adoção dessas etapas permitiu abordar os aspectos mencionados nos objetivos específicos para o desenvolvimento da pesquisa.

Para a preparação teórico metodológica do estudo de caso, realizou-se uma revisão de literatura para a familiarização da estrutura da *Web Semântica* e as tecnologias envolvidas. O estudo constituiu de pesquisas bibliográficas em artigos, dissertações, teses e livros e nas pesquisas documentais, em manuais, documentos

eletrônicos, reportagens de especialistas conceituados em jornais e revistas e *sites* da *Web*.

Na seleção de caso desta pesquisa, foi selecionado como estudo de caso o *site* da biblioteca virtual Europeia, uma produção da fundação Europeia com apoio dos países da Comunidade Europeia que organiza conteúdos de domínio público. O site apresenta as características essenciais da *Web Semântica* e atende aos requisitos básicos que possibilitem ilustrar sua aplicação nos conteúdos dos arquivos e bibliotecas.

A coleta de dados consistiu em levantar e documentar as aplicações da *Web Semântica* no *site* da Fundação Europeia, empregando fontes de evidências, conforme Yin (2010, p.124) sugere. Ainda, segundo o autor supracitado, é relevante utilizar “[...] duas ou mais fontes, convergindo sobre os mesmos fatos ou descobertas.”, visto que possibilita contemplar os tópicos exigidos. Para este estudo de caso, alinhado com as ideias de Yin (2010), foram utilizadas 4 (quatro) fontes de evidências:

1. **Documentação:** materiais produzidos pelas instituições parceiras do site Europeia;
2. **Registro em arquivo:** registros de serviço e organizacionais, relatórios, plano de trabalho;
3. **Observação Direta:** observar a tecnologia no seu ambiente é inestimável para compreender os limites e ou problemas dessa nova tecnologia;
4. **Artefatos Digitais:** Produtos digitais gerados a partir do contexto.

Para a fase de análise de dados foi elaborada uma avaliação das tecnologias empregadas, para detectar quais delas são imprescindíveis à aplicação nas atividades das Bibliotecas e Arquivos. Diante desta proposta, utilizou-se o método de Análise de Conteúdo, definida por Bardin (2010, p.44) como:

Um conjunto de técnicas de análise das comunicações visando obter, por procedimentos, sistemáticos e objectivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos as condições de produção/recepção (variáveis inferidas) destas mensagens.

A metodologia de Análise de Conteúdo é organizada em três fases principais: **pré-análise, exploração do material e tratamento dos resultados**

**obtidos e interpretação.** A pré-análise caracteriza-se pela organização, tem por objetivo tornar operacional e sistematizar as ideias iniciais. A exploração do material consiste em operações de codificação do material que corresponde à transformação dos dados em bruto, segundo regras específicas, a decomposição ou a enumeração dos dados seguindo regras já formuladas. Na fase de tratamentos dos resultados e interpretação, os resultados brutos são tratados para tornarem-se significativos e válidos, e possibilitam a produção de quadros de resultados, diagramas, figuras e modelos que condensam e destacam as informações recebidas da análise (BARDIN, 2010).

Bardin(2010, p.147) classifica os elementos em agrupamentos conforme suas características comuns o qual permite a investigação e as relações entre eles. A autora sustenta que ao codificar o material, será produzido um sistema de categorias que permitirá “[...] fornecer, por condensação, uma representação simplificada dos dados.”. Nesta análise, foram definidas as categorias iniciais para estudo, conforme os objetivos da pesquisa:

#### **Quadro 1 – Categorias de Estudo**

<b>Categoria</b>	<b>Descrição</b>	<b>Fontes de Pesquisa</b>
Metadados	Tipo e função dos metadados e os requisitos básicos.	Relatórios, planos de trabalho do <i>websites</i> da Fundação Europeia Manual de Normas em: <a href="http://www.europeana-libraries.eu/documents/868553/b1608294-839e-42d2-8eb2-e20ecf4938ba">http://www.europeana-libraries.eu/documents/868553/b1608294-839e-42d2-8eb2-e20ecf4938ba</a>  <a href="http://pro.europeana.eu/edm-documentation">http://pro.europeana.eu/edm-documentation</a>  <a href="http://pro.europeana.eu/technical-requirements">http://pro.europeana.eu/technical-requirements</a>  <a href="http://pro.europeana.eu/moving-to-edm">http://pro.europeana.eu/moving-to-edm</a>
Linguagem de Marcação e ou programação e Banco de dados	Quais as linguagens e banco de dados empregados para o desenvolvimento do site e as motivações técnicas para sua adoção.	Relatórios, planos de trabalho do <i>website</i> da Fundação Europeia.  <a href="https://code.google.com/p/europeana4j/">https://code.google.com/p/europeana4j/</a>  <a href="http://labs.europeana.eu/api/getting-started/">http://labs.europeana.eu/api/getting-started/</a>  <a href="http://labs.europeana.eu/api/record-jsonld/">http://labs.europeana.eu/api/record-jsonld/</a>

Linguagem de Busca	Qual linguagem de recuperação e quais estratégias de busca no banco de dados são utilizadas.	<a href="http://labs.europeana.eu/api/">http://labs.europeana.eu/api/</a> <a href="http://labs.europeana.eu/api/introduction/">http://labs.europeana.eu/api/introduction/</a> <a href="http://labs.europeana.eu/api/linked-open-data/introduction/">http://labs.europeana.eu/api/linked-open-data/introduction/</a> <a href="http://labs.europeana.eu/api/query/">http://labs.europeana.eu/api/query/</a>
Ontologias e Vocabulários	Como os conceitos estão representados dentro da área de domínio e sua interoperabilidade, aliado aos termos padronizados.	<a href="http://pro.europeana.eu/documents/866205/331925/EuV1_WP3_Amsterdam_October_2010_03_Meghini_EDM5.2.pdf">http://pro.europeana.eu/documents/866205/331925/EuV1_WP3_Amsterdam_October_2010_03_Meghini_EDM5.2.pdf</a> <a href="http://conference.ifla.org/past-wlic/2010/149-doerr-en.pdf">http://conference.ifla.org/past-wlic/2010/149-doerr-en.pdf</a>
Método de coleta de dados ( <i>harvest</i> )	Técnicas utilizadas para reunir os metadados, formatos e serviços necessários para os repositórios de dados.	Manual de Normas em: <a href="http://www.europeana-libraries.eu/documents/868553/b1608294-839e-42d2-8eb2-e20ecf4938ba">http://www.europeana-libraries.eu/documents/868553/b1608294-839e-42d2-8eb2-e20ecf4938ba</a> <a href="http://pro.europeana.eu/providers-faq">http://pro.europeana.eu/providers-faq</a>

**Fonte:** o próprio Autor.

As fontes de pesquisa relacionadas no Quadro 1 foram as iniciais e seguramente não representam todos os locais em que foram realizadas as consultas para o desenvolvimento do estudo, no entanto, entende-se como os principais insumos que permitiram enriquecer a pesquisa.

Com esta organização dos dados, foi possível realizar inferências sobre os conteúdos e visualizar a aplicabilidade dessas tecnologias nas Bibliotecas e Arquivos.

#### 1.4 JUSTIFICATIVA

As evidências atuais indicam que a internet modifica as relações do cotidiano, tornando muitas vezes desnecessário estar presente fisicamente para realizar compras, vender produtos ou contratar serviços. Possibilita presenciar vários acontecimentos em qualquer parte do planeta ao mesmo tempo, vinte e quatro horas

por dia. Ainda, com um clique localiza-se uma quantidade inesgotável de materiais para pesquisa.

A facilidade, extensibilidade e agilidade em buscar conteúdos tornou a rede mundial de computadores uma ferramenta obrigatória de consultas, que aos poucos tem aposentado as volumosas, mas valiosas enciclopédias em papel. No entanto, a cada momento as pessoas estão mais seletivas e exigentes nos resultados das pesquisas na *Web*, não é mais aceitável desperdiçar o tempo em minerar os resultados para localizar os materiais significativos. A exigência é que a resposta seja precisa e possibilite total aproveitamento do material.

Pesquisar, explorar, analisar e testar os recursos da *Web Semântica* foi a proposta deste trabalho, que pretendeu compreender o ciclo virtuoso dessa arquitetura, para difundir a estrutura funcional, modelos e padrões da *Web Semântica* nos meios científicos, o que permitirá sua adoção pelos desenvolvedores e mantenedores da *Web* atual.

É perceptível que os estudos atuais sobre o assunto referem-se especificamente a área da Computação ou Ciência da Informação. O desafio dessa pesquisa é integrar as duas ciências, permitindo que as duas visões sejam incorporadas ao trabalho.

Propiciar que arquivistas e bibliotecários possam identificar, através do estudo de caso, métodos que possibilitem organizar, tratar, armazenar, preservar, recuperar e disseminar a informação de forma a aplicar o projeto da *Web Semântica*, se constitui em outra contribuição deste estudo.

Ainda, a realização desta pesquisa contribuirá para a área da Ciência da Informação no sentido de proporcionar um referencial teórico aos profissionais da área no tema *Web Semântica*.

## 1.5 ESTRUTURA DO TRABALHO

Este trabalho está organizado em quatro capítulos, iniciando pelo capítulo 1, que trata da problemática que motivou o estudo, ou seja, o objetivo geral proposto e os objetivos específicos. Apresenta as argumentações da justificativa e contribuições do trabalho, descrevendo os procedimentos metodológicos empregados para o desenvolvimento do estudo.

No capítulo 2 é apresentado o referencial teórico que trata da história da Informação e Ciência da Informação, a representação descritiva e temática na Organização da Informação, a gestão de documentos realizados pelos arquivistas e a história da *Web*.

O capítulo 3 mostra a análise de resultados, baseados nos três objetivos específicos, e organiza as tecnologias da *Web Semântica* do estudo de caso Europeia em cinco categorias. Neste capítulo produziu-se um *Framework* com o propósito de representar a *Web Semântica* nas unidades de informação.

No capítulo 4 são apresentadas as considerações finais do trabalho, assim como a análise dos resultados obtidos. Também há referências a futuros trabalhos e a conclusão obtida com a pesquisa.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão contextualizados conceitos de informação, Ciência da Informação, Organização da Informação, Gestão de Documentos e a história da *Web* para que o capítulo de apresentação e análise de resultados tenha um material linear e não adentre em história.

### 2.1 A INFORMAÇÃO

O termo informação “tem raízes latinas (*informatio*)”, inicialmente era definido como “moldado por; posteriormente, veio a significar relatos recebidos por”, conceitua Capurro (1985, p.155-158). O “moldado por”, expressão utilizado no século XVI, definia a ação de “formação ou moldagem da mente ou o do caráter, treinamento, instrução, ensinamento, comunicação [...]” (WURMAN, 1991, p.42) e “relatos recebidos por” era em razão da forma como a informação era transmitida, por um canal elétrico ou mecânico. A informação está presente em todas as áreas do saber, cada qual com conceitos dentro de seu próprio contexto. Bogdan (1994, p.53) define que:

A noção de informação tem sido usada para caracterizar uma medida de organização física (ou sua diminuição, na entropia), um padrão de comunicação entre fonte e receptor, uma forma de controle e feedback, a probabilidade de uma mensagem ser transmitida por um canal de comunicação, o conteúdo de um estado cognitivo, o significado de uma forma linguística ou a redução de uma incerteza. Estes conceitos de informação são definidos em várias teorias como a física, a termodinâmica, a teoria da comunicação, a cibernética, a teoria estatística da informação, a psicologia, a lógica indutiva e assim por diante. Parece não haver uma ideia única de informação para a qual estes vários conceitos convirjam e, portanto, nenhuma teoria proprietária da informação.

Elemento essencial para o desenvolvimento de uma nação, a informação possibilitaria que pessoas pudessem viver de maneira mais saudável, políticos teriam condições de identificar as carências de uma sociedade, gestores administrariam empresas de forma mais precisa, médicos, professores, pesquisadores e estudantes seriam capazes de determinar novos métodos ou técnicas para aplicação nas soluções de problemas, como aborda Capurro e Hjørland (2007, p.149) “É lugar comum considerar a informação como condição

básica para o desenvolvimento econômico juntamente com o capital, o trabalho e a matéria prima”.

Le Coadic (2004, p.4) expressa que:

A informação comporta um elemento de sentido. É um significado transmitido a um ser consciente por meio de uma mensagem inscrita em suporte espacial-temporal: impresso, sinal elétrico, onda sonora, etc. Inscrição feita graças a um sistema de signos (linguagem), signo este que é um elemento da linguagem que associa um significante a um significado.

O autor supracitado também define que “Documento é o termo genérico que designa os objetos portadores de informação”. Sendo assim, um documento é todo material que “represente ou expressa um objeto, uma ideia ou uma informação por meio de signos gráficos e icônicos (palavras, imagens, diagramas, mapas, figuras, símbolos), sonoros e visuais (gravados em suporte de papel ou eletrônico)” (LE COADIC, 2004, p.5).

A informação é a fonte de todas as ciências, a matéria prima para a expansão de todas as atividades, que após serem registradas propiciarão um repositório de conteúdos com informações técnicas e científicas. É o fomentador responsável pelas grandes descobertas, a essência da ciência. Sem a informação, a pesquisa seria inútil e não haveria conhecimento, deve ser continuamente produzido e renovado e circular livremente por todos os meios acadêmicos, científicos e profissionais (LE COADIC, 2004).

## 2.2 CIÊNCIA DA INFORMAÇÃO

Uma ciência é um sistema de conhecimentos que promove adquirir mais conhecimentos baseados em métodos científicos. Uma ciência se estabelece quando existe dificuldade em se esclarecer problemas e na carência de informação de um assunto.

Na década de 40, a comunidade científica sentiu falta da abordagem da informação de uma maneira mais relevante, pois a sociedade da informação carecia de uma ciência que estudasse as características da informação e os processos de sua construção, comunicação e uso (LE COADIC, 2004).

Os conteúdos informacionais, naquele momento da história, cresciam de maneira colossal e descontrolada, tornando clara a necessidade de organizar e gerenciar a informação. Foi quando um artigo de nome *As we may think*, de Vannevar Bush, um renomado cientista do Instituto de Tecnologia de Massachusetts, identifica um problema crítico, que era a necessidade de disponibilizar um acervo de conhecimento em estado crescente para um número maior de pessoas. O autor propõe a utilização dos recursos da tecnologia da informação como solução, utilizando uma máquina denominada MEMEX, que tinha como princípio, armazenar informações e para posterior recuperação (SARACEVIC, 1996).

Assim, em 1948, aproximados 340 cientistas e documentalistas de várias partes do mundo participaram por 10 dias da *Royal Society Information Conference*, quando discutiram as ideias de Bush e apresentaram propostas para solucionar as questões da gestão da informação (BARRETO, 2002, p.3).

A conferência foi o despertar para a concepção da Ciência da Informação, no qual cientistas e pesquisadores de vários países, principalmente dos EUA e Europa, compartilharam seus conhecimentos e esforços para a elaboração do escopo científico da Ciência da Informação (CI).

A CI possui diversas definições que partem de vários autores. Para Borko:

Ciência da Informação é a disciplina que investiga as propriedades e o comportamento informacional, as forças que governam os fluxos de informação, e os significados do processamento da informação, visando à acessibilidade e a usabilidade. A Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação, e utilização da informação. Isto inclui a pesquisa sobre a representação da informação em ambos os sistemas, tanto naturais quanto artificiais, o uso de códigos para a transmissão eficiente da mensagem, bem como o estudo do processamento e de técnicas aplicadas aos computadores e seus sistemas de programação (BORKO, 1968, p. 1, tradução nossa).

De acordo com Borko(1968, p.1), a CI preocupa-se com o comportamento informacional desde sua “[...] origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação, e utilização da informação”. Assim, é uma ciência que trata todo o ciclo da informação, em muitos momentos excedendo de seus limites, transitando no domínio de outras ciências, e em razão disso, a CI é também conhecida como interdisciplinar.

Le Coadic (2004, p.20) escreveu que:

A ciência da informação é uma das novas interdisciplinas, um desses novos campos de conhecimento onde colaboram entre si, principalmente: psicologia, linguística, sociologia, informática, matemática, lógica, estatística, eletrônica, economia, direito, filosofia, política e telecomunicações. [...] A interdisciplinaridade traduz-se por uma colaboração entre diversas disciplinas, que leva a interações, isto é, uma certa reciprocidade nas trocas, de modo que haja, em suma, enriquecimento mútuo.

O autor ressalta que a interdisciplinaridade somente se efetiva quando existe uma interação entre as disciplinas, a partir do diálogo e consolidação dos conceitos, teorias, métodos e campos de investigação nas disciplinas. Para Saracevic (1996, p.48) “A interdisciplinaridade foi introduzida pela própria variedade da formação de todas as pessoas que se ocuparam com os problemas descritos”, e ainda, segundo o autor supracitado, muitos profissionais de várias ciências enriqueceram o arcabouço da CI, como “[...] engenheiros, bibliotecários, químicos, linguistas, filósofos, psicólogos, matemáticos, cientistas da computação [...]”.

### 2.3 ORGANIZAÇÃO DA INFORMAÇÃO

Uma unidade de informação deve estar preparada para atender às necessidades dos usuários, mantendo seus itens informacionais padronizados de maneira que facilite seu acesso.

A padronização envolve processos que organizam todos os setores da unidade de informação que definem desde a aquisição, tratamento e recuperação do item.

O propósito da OI é tornar acessível o conhecimento contido na informação, por meio de um ambiente informacional que possibilite seu fácil acesso, empregando procedimentos que compreendem a descrição física e de conteúdo dos itens informacionais para a sua posterior localização e recuperação em sistemas de informação (BRÄSCHER; CAFÉ, 2008).

Para as autoras Catarino e Souza, a representação descritiva e a representação temática são definidas como:

A descritiva representa as características específicas do documento que permite sua individualização e também define e padroniza os pontos de acesso responsáveis pela busca e recuperação da informação e pela reunião de documentos semelhantes. A temática detém-se na atribuição de assuntos aos documentos a partir da classificação bibliográfica, da indexação e da elaboração de resumos, facilitando a recuperação de materiais relevantes que dizem respeito a temas semelhantes (CATARINO; SOUZA, 2012, p.84).

A representação descritiva refere-se à descrição física dos documentos e dos elementos que o identificam, essa atividade é também denominada de catalogação, ou catalogação descritiva, para os atuantes em bibliotecas, e como descrição bibliográfica para os profissionais dos serviços de informação científica (ORTEGA, 2011). As normalizações objetivam aplicar diretivas padronizadas na descrição do item. Para a arquivologia emprega-se a descrição arquivística, que possui o intuito de destacar os elementos que representem o item informacional.

A representação temática, ou representação do conteúdo de um documento envolve os processos de classificação e indexação e que permite interagir com outras áreas do conhecimento, em especial a Linguística, a Terminologia e a Ciência da Computação (CATARINO, 2013). A representação temática também envolve a catalogação por assunto e os resumos.

## 2.4 GESTÃO DE DOCUMENTOS

Organizar documentos é uma prática que demanda diversas técnicas apropriadas e padronizadas por organizações nacionais e internacionais. Por meio da gestão de documentos é possível avaliar documentos que, por possuir valor futuro, devem ser conservados.

O crescimento da produção de documentos nas unidades informacionais, ocasionou uma perda de controle na organização dos documentos existentes, o que levou profissionais da área a encontrarem alternativas para gerir os materiais guardados nos arquivos (PAES, 2002).

O Dicionário de Terminologia Arquivística (BRASIL, 2005, p.100), estabelece que a gestão de documentos pode ser definida como:

Conjunto de procedimentos e operações técnicas referentes à produção, tramitação, uso, avaliação e arquivamento de documentos em fase corrente e intermediária, visando sua eliminação ou recolhimento. Também chamado de administração de documentos.

A gestão de documentos tem o propósito de administrar itens informacionais o que possibilita o controle desde sua produção até o arquivamento, possibilitando que a recuperação desse item nas instituições seja realizada de forma precisa. A avaliação, classificação e descrição são os procedimentos empregados que garantem a longevidade e preservação dos itens informacionais.

Para o estudo desta pesquisa, cujo o tema está focado na *Web Semântica*, três campos são essenciais e mantêm estreita relação com a CI, esses campos são a Biblioteconomia, Arquivologia e a Ciência da Computação, que são descritos a seguir.

## 2.5 ÁREAS QUE SE INTERAGEM

A Biblioteconomia é uma área que tem forte relação com a CI, visto que ambastem papel fundamental como facilitador aos acessos dos registros gráficos.

[...] remontando a três mil anos, devotada à organização, à preservação e ao uso dos registros gráficos humanos. Essas atividades são realizadas pelas bibliotecas não apenas como uma organização particular ou um tipo de sistema de informação, mas principalmente, como uma instituição social, cultural e educacional. (SARACEVIC, 1996, p.49).

Biblioteconomia, segundo Le Loadic (2004, p.12), é a “[...] união de duas palavras, a biblioteca e economia (esta no sentido de organização, administração e gestão) [...]”. E o termo foi utilizado pela primeira vez em 1839 na obra publicada por Léopold Auguste Constantin Hesse com nome de *Bibliothéconomie: instructions sur l'arrangement, la conservation e l'administration des bibliothèques* (ORTEGA, 2002).

No Brasil, os primeiros bibliotecários foram os jesuítas, que eram os responsáveis pelas primeiras bibliotecas, além de exercerem as atividades específicas de organizar e manter os livros seguros.

Outras missões religiosas foram relevantes para incorporação de bibliotecas, como os franciscanos, as carmelitas e os beneditinos (FONSECA, 1979).

Essa disciplina da CI tem como um dos propósitos preparar técnicas mais eficazes para organizar os registros humanos físicos ou virtuais de maneira que sua recuperação seja rápida e precisa, possibilitando sua disseminação.

A Arquivologia<sup>4</sup> aproxima-se da CI e surge em meados do século XX, motivada pelas mudanças nos processos de tratamento e disseminação da informação (SOUTO; CALDERON; SILVA; 2013). Os autores Silva e Guimarães (2007, p. 9), a partir da definição proposta por Chauí (2002), definem que:

---

<sup>4</sup> Também denominada Arquivística (BRASIL, 2005).

A Arquivologia é um conhecimento científico concebido para solucionar problemas práticos, em consonância com a teoria, em busca de esclarecer fatos através das relações dos mesmos e que, para contextualizar a análise desses fatos, é necessária uma teoria.

Neste sentido, Fuzer Ruiz (1999) relaciona a Arquivologia ao seu objeto de estudo, ou seja, os arquivos. Dessa maneira, ela estaria relacionada aos processos que envolvem a confecção de um documento, que seria a materialização ou registro da informação, estudando técnicas e procedimentos adequados para uma gestão documental (avaliação, classificação, descrição), administração e tratamento dos arquivos.

Se comparada a outras áreas, a Ciência da Computação (CC) é muito nova, os registros dos primeiros equipamentos de características digitais indicam a Segunda Guerra Mundial, como marco inicial. Por outro lado o crescimento foi vertical, oferecendo grandes quantidades de novas tecnologias, conceitos, ideias e fornecendo soluções para as dificuldades de empresas e pessoas (FONSECA FILHO, 2007).

A Sociedade Brasileira de Computação (SBC, 2003, p.1) define que “A Ciência da Computação estuda a fundamentação teórica das construções computacionais, bem como suas aplicações em dispositivos tecnológicos e sistemas de computação”, sendo os cientistas da computação os profissionais

[...]responsáveis pelo desenvolvimentocientífico (teorias, métodos, linguagens, modelos, entre outras) etecnológico da Computação. Eles constroem ferramentas que são normalmente utilizadas por outros profissionais da área deComputação, responsáveis pela construção de software/hardwarepara usuários finais. Eles são também responsáveis pelainfraestrutura de software dos computadores (sistemasoperacionais, compiladores, banco de dados, navegadores entreoutras) e software para sistemas embarcados, sistemas móveis,sistemas de computação nas nuvens entre outros. [...] Sabem fazer uso dainterdisciplinaridade, na medida em que conseguem combinarciências, dando a elas um tratamento computacional. (SBC, 2003, p.10).

É possível inferir, que a CC tem como propósito estudar as tecnologias que envolvem os equipamentos computacionais e o *software* representado pelos algoritmos que descrevem e transferem informações, no intuito de automatizar as tarefas comuns dos usuários. Os estudos da CC possuem grande visibilidade à sociedade, pois suas pesquisas impactam de forma direta a vida das pessoas,

possibilitando o surgimento de novas tecnologias que facilitam as tarefas nas áreas de comunicação e serviços.

No eixo da CC, é oportuno evidenciar Sistemas de Informação, uma das subáreas da Computação, que lida com sistemas complexos que exigem amplos conhecimentos técnicos e organizacionais para serem concebidos, desenvolvidos e geridos. Estuda os modelos e as áreas de negócio, a arquitetura da informação para a padronização na descrição de significados de palavras, símbolos e expressões através da análise de cenários (SBC, 2003, p.18).

Na leitura do último parágrafo, é possível identificar algumas similaridades de Sistemas de Informação com a CI, visto que ambas estudam o mesmo objeto, a informação, porém com enfoques distintos. Sistemas de Informação trata a informação tendo os equipamentos e programas como eixo principal de pesquisa, enquanto que CI trata da natureza da informação e sua comunicação para uso pelos humanos.

## 2.6 INTERNET, A *WEB* E A *WEB SEMÂNTICA*

Os termos Internet e *Web* são vocábulos comuns utilizados para identificar um espaço no qual é possível armazenar, localizar e recuperar conteúdos de qualquer espécie. Porém, são termos distintos que identificam tecnologias peculiares.

A internet é uma complexa rede de computadores e dispositivos interligados pelo planeta utilizando protocolos que possibilitam a execução de diversos serviços, permitindo a distribuição e recuperação de informações para todos os usuários conectados. Localizar essa informação com precisão e agilidade é o propósito de todos os usuários que recorrem a essa rede.

No entanto, a internet inicial foi concebida sem qualquer organização estrutural, foi elaborada como uma rede para simples compartilhamento de informação e que cresceu de maneira descontrolada, como citam Souza e Alvarenga (2004, p.133).

Embora tenha sido projetada para possibilitar o fácil acesso, intercâmbio e a recuperação de informações, a Internet foi implementada de forma descentralizada e quase anárquica; cresceu de maneira exponencial e caótica e se apresenta hoje como um imenso repositório de documentos que deixa muito a desejar quando precisamos recuperar aquilo de que temos necessidade.

O termo *World Wide Web* (WWW) conforme o dicionário Webster a WWW “[...] é uma parte da Internet, acessada por uma interface gráfica e contém documentos

que são conectados por links, também chamados de Web”.(DICIONÁRIO Merriam-Webster, 2014, tradução nossa).

Robredo e Brascher(2010, p.14) em um artigo denominado “Ciência da informação e *Web*Semântica: Linhas convergentes ou linhas paralelas?”, refere-se a história da *Web* em “Web 1.0, Web 2.0 e Web 3.0”. A primeira *Web*, conforme Robredo(2010, p.14), “[...] caracterizava-se por suas páginas estáticas, comportando textos, imagens e links que somente podiam ser alterados pelo Webmaster”. Foi um ambiente em que não havia muita interatividade entre participantes e seu caráter era conectar as informações. Para a *Web* 2.0 Robredo (2010, p.16) destaca que os serviços de distribuição e comunicação foram aperfeiçoados, a interoperabilidade entre os sistemas ficaram mais transparentes e iniciou uma “[...] proliferação de redes comunitárias e sociais, hospedagem de serviços e aplicações, compartilhamento de vídeos, wikis e blogs”. Breitman (2010) menciona essa *Web* como “Web Sintática”, na qual descreve que os computadores realizam compartilhamento da informação, visto que a interpretação dos dados é um processo desempenhado pelas pessoas.

Tim Berners Lee publicou um artigo publicado na Scientific American, no qual manifestou seu entendimento sobre a *Web* Semântica, também denominada como *Web* 3.0:

[...] a *Web* Semântica não é uma *Web* separada, mas uma extensão da atual, na qual é dada à informação um significado bem definido, o que permite às pessoas e aos computadores trabalharem em cooperação.[...] Num futuro próximo, esses desenvolvimentos deverão conduzir a significativas novas funcionalidades à medida que as máquinas irão se tornando bem mais capazes de processar e —entender os dados que, por enquanto, se limitam somente a apresentar. (BERNERS LEE; HENDLER; LASSILA, 2001, p.1, tradução nossa).

Diante do conceito de Berner Lee, entende-se que a *Web* Semântica é uma proposta para tornar a *Web* um ambiente mais organizado, com os dados interconectados, permitindo que os computadores possam realizar ligações entre conteúdos e criar suas próprias inferências.

Na visão da Inteligência artificial, “A *Web* Semântica é um conjunto de iniciativas para promover uma *Web* futura, cujas páginas serão organizadas, estruturadas e codificadas para que os computadores sejam capazes de fazer inferências e raciocinar seu conteúdo” (CODINA; ROVIRA, 2006, p.2).

O projeto da *Web Semântica* emprega muitas tecnologias para a organização e estruturação do conteúdo da *Web*, necessárias para facilitar a localização e recuperação da informação. Alvarenga e Souza (2004, p.134) entende que:

Para atingir tal propósito, 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 e que esta possa ser ‘consumida’ por outros usuários humanos ou não, de maneira automática e não ambígua. Com a existência da infra-estrutura tecnológica comum da Internet, o primeiro passo para este objetivo está sendo a criação de padrões para descrição de dados e de uma linguagem que permita a construção e codificação de significados compartilhados.

Esta padronização possivelmente seja a maior dificuldade para o estabelecimento da *Web Semântica*, visto que não há como exigir que desenvolvedores aceitem ou adotem as recomendações da W3C no momento de produzir e publicar páginas na *Web*.

A *Web Semântica* é a *Web* dos dados lincados<sup>5</sup> para o W3C, que visa criar um ambiente em que aplicações possam consultar dados e realizar inferências com o uso de vocabulários, além de possibilitar que os computadores realizem tarefas úteis aos usuários e que proporcionem a interoperabilidade dos dados entre os sistemas ao empregar tecnologias como RDF, SPARQL, OWL e SKOS (W3C, 2011).

Breitman (2010, p. 6) ressalta que para a “construção” da *Web Semântica*, alguns “temas estão presentes em todas as discussões”, destacando-se os “Metadados, Ontologias, Linguagens da *Web Semântica*, Construção de Modelos Semânticos, *Web Services*, Agentes e Ferramentas.”

Todo o referencial escrito representa apenas um embasamento primário com acontecimentos históricos da OI, arquivologia, biblioteconomia e *Web Semântica* e uma pequena introdução dos conceitos decorrentes da área. As técnicas, recomendações, normas, tecnologias empregadas nos tópicos mencionados serão descritas na seção de Apresentação e Discussão dos Resultados.

---

<sup>5</sup> “Dados lincados é o nome que se dá às coleções de dados relacionados na *Web*.” (CATARINO; SOUZA, 2012, p.79).

### 3 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

A *Web* atual emprega diversas tecnologias no intuito de facilitar a tarefa do usuário na busca de um documento ou item informacional. Essas tecnologias invisíveis aos usuários finais possibilitam que a localização seja ágil e precisa. A *Web Semântica* é uma proposta que visa tornar a *Web* um ambiente mais organizado e estruturado, que possibilitará aos computadores compreender os dados armazenados e realizar conexões entre esses dados.

As unidades de informações têm grande interesse nessa organização e estruturação de dados, visto que esse é o trabalho realizado nas bibliotecas, arquivos, museus e agora tem-se a oportunidade de expandir para a *Web*.

Como pesquisador da área de Organização da Informação da CI, este estudo tem o propósito de apresentar e descrever as tecnologias necessárias para aplicação da *Web Semântica* nas unidades de informação. Para isso, foram apresentados os seguintes objetivos: 1) identificar as técnicas de Organização da Informação nas unidades de informação; 2) analisar as tecnologias, recomendações, linguagens e padrões que compõem a *Web Semântica*; 3) Demonstrar a aplicabilidade da *Web Semântica* em ambientes das unidades de informações, por meio de um estudo de caso e um framework.

O Quadro 2 apresenta os objetivos específicos.

**Quadro 2** – Objetivos Específicos

OBJETIVOS ESPECÍFICOS		
Objetivos	Descrição	Aplicação
Objetivo 1	Identificar as técnicas de Organização da Informação nas unidades de informação.	Pesquisa bibliográfica
Objetivo 2	Analisar tecnologias, recomendações, linguagens e padrões que compõem a <i>Web Semântica</i> .	Pesquisa bibliográfica e documental
Objetivo 3	Demonstrar a aplicabilidade da <i>Web Semântica</i> em ambientes das unidades de informação, por meio de um estudo de caso e um framework.	Pesquisa bibliográfica, documental e estudo de caso

**Fonte:** o próprio Autor.

No objetivo específico 1 foi aplicada somente a pesquisa bibliográfica, visto que realizou-se a busca e estudo em materiais consolidados e publicados em revistas, livros, periódicos, etc. Para o objetivo 2 aplicaram-se as pesquisas bibliográfica e documental. As duas pesquisas complementaram-se, sendo a documental essencial, já que muitos materiais foram localizados em relatórios de grupos de trabalhos, manuais, planos de negócios e recomendações de organizações. No objetivo 3, as pesquisas bibliográfica, documental e o método estudo de caso foram empregados. As duas primeiras, conforme já descritas nos itens acima e no estudo de caso, tinham como objetivo identificar, compreender e representar a *Web Semântica* nas unidades de informação.

A seguir é realizado uma apresentação dos resultados deste estudo para cada objetivo específico.

### 3.1 TÉCNICAS DA OI NAS UNIDADES DE INFORMAÇÃO

Na OI as operações que tratam da identificação para individualização do item, sua descrição física e dos elementos que o caracterizam unicamente, são definidas como representação descritiva. Já as operações que possibilitam descrever o conteúdo, o assunto do item, define-se como representação temática (CATARINO; SOUZA, 2012).

A arquivologia e a Biblioteconomia são áreas da OI, no entanto, possuem definições e conceitos peculiares e assim serão apresentados separadamente.

A representação descritiva é caracterizada pelo serviço de catalogação descritiva, que é um processo para descrever formalmente um item informacional por meio do registro do título, autor(es), tradutor(es), número da edição, local e data da publicação, número de páginas, volumes, dimensões, entre outros proporcionando ao usuário final a condição de encontrar, identificar, selecionar e obter o item informacional descrito (SOUZA, 2009).

No entender de Catarino (2013, p.31), um item informacional é considerado catalogado quando “[...] independente de seu suporte, puder ser recuperado, automática ou manualmente, não apenas por seus autores (principais ou secundários), títulos e edições, mas também por termos que representem seu

conteúdo”. Entende-se dessa forma que a catalogação é um importante instrumento mediador entre o usuário e o acervo.

A catalogação descritiva extrai a partir do registro de um conjunto de dados, informações que descrevem fisicamente o documento. Essa extração de dados deve seguir um critério definido por normas locais, nacionais ou internacionais, de acordo com as particularidades de cada unidade de informação.

As normatizações visam a estabelecer diretivas padronizadas, consistentes e apropriadas para a descrição dos itens, no intuito de possibilitar sua efetiva localização, como também promover o intercâmbio de informações por meio dos metadados em âmbito nacional e internacional.

Para as normas internacionais, destacam-se algumas que possuem aplicações nas unidades de informação, como a segunda edição da Anglo-American Cataloguing Rules (AACR2<sup>6</sup>), o padrão RDA que está substituindo o AACR2, o formato de registro bibliográfico Machine Readable Cataloging (MARC<sup>7</sup>), a International Standard Bibliographic Description (ISBD<sup>8</sup>) e aos registros bibliográficos por meio da catalogação coletiva dos On-line Public Access Catalogue (OPAC<sup>9</sup>), disponibilizado por diversas bibliotecas (CATARINO, 2013).

Fontes externas de referências são empregadas em situações que a fonte principal de informação não possui os dados necessários para a catalogação descritiva do item, que podem ser distintos para cada tipo e área da descrição.

A representação temática é caracterizada por expressar o conteúdo de um documento e compreende os processos de indexação, catalogação por assunto, classificação e resumos, que resultarão nos índices, aos catálogos de assunto, números de classificação e aos resumos que tornam possível a recuperação dos documentos de acordo com os interesses dos utilizadores (RUBI, 2008).

A indexação é definida por Robredo (1982, p. 236) como “[...] operação de representar o conteúdo dos documentos, qualquer que seja o método utilizado” e

---

<sup>6</sup> AACR2 – Regras de Catalogação Anglo Americana, projetada para uso na construção de catálogos e outras listas em bibliotecas gerias de todos os tamanhos. A segunda edição é baseada em uma conciliação dos textos britânicos e norte-americanos da edição de 1967 (AACR2, 2008)

<sup>7</sup> MARC – Catálogo Legível por Máquinas, padrão para a representação e comunicação de informações bibliográficas e relacionados em forma legível por máquina (MARC21, 2007).

<sup>8</sup> ISBD – Padrão Internacional de Descrição Bibliográfica, destina-se a tornar o principal padrão para promover o controle bibliográfico universal, por meio do reconhecimento e exibição dos elementos de dados, independente da linguagem de descrição (IFLA, 2015).

<sup>9</sup> OPAC – Catálogo de Acesso Público *On-Line*, catálogo em tempo real com registros bibliográficos em vários formatos mantidos por uma biblioteca ou grupos de bibliotecas (ALIC, 2015).

propicia descrever os tópicos principais de um documento por meio de termos retirados do texto original ou de termos selecionados de uma linguagem de informação ou indexação.

Para a realização desse processo, a indexação emprega o uso de instrumentos normativos e linguísticos, sendo exemplo a linguagem natural e documentárias. A linguagem natural utiliza termos da forma como se apresentam nos documentos, é facilmente legível pelo homem, já que utiliza os mesmos sinais. A linguagem documentária é um conjunto determinado de regras, símbolos e termos que produz uma linguagem artificial, também definida como metalinguagem, que pode transformar uma linguagem em outra (FEITOSA, 2006).

A localização de um item informacional em uma base de dados de uma unidade de informação depende da qualidade como ocorreu sua descrição. A indexação produz as palavras chaves que são determinantes para a localização precisa de um item, conseqüentemente, uma indexação bem realizada é decisiva para que o software de busca, comercial ou interno da instituição atinja os resultados requisitados.

A catalogação por assunto tem como propósito prover ao usuário formas de identificar os conteúdos particulares dos documentos e o estabelecimento de relações que possuam características comuns a outros itens documentais, possibilitando que sejam recuperados de modo coeso (LANCASTER, 2004).

A origem da catalogação por assunto está ligada aos catálogos de bibliotecas, e é organizada por meio do cabeçalho de assunto que são como enunciados de assuntos, formados como resultado da composição ordenada de palavras. Os cabeçalhos de assunto são formados por termos que expressam o conteúdo temático do documento. O objetivo da catalogação é representar o assunto de um item informacional por meio de conceitos significativos, evidenciando todos os pontos que destaquem o item informacional (SILVA; FUJITA, 2004).

Classificar é um ato que todos realizam ao longo da vida. Por meio da classificação torna-se possível organizar grupos, distinguir seus elementos, criar subgrupos com o propósito de facilitar nossas relações. Burke (2003) destaca que o conhecimento pode ser classificado por grupos diferentes e de maneiras diferentes. Svenonius (1985) destaca que existem as classificações que são provenientes de filósofos, enciclopedistas, epistemólogos e outros voltados para a organização de

documentos, localização física e recuperação, denominado de classificação bibliográfica.

A classificação vem evoluindo para as novas tecnologias e desenvolvendo novos instrumentos para organizar o arranjo dos documentos e o arquivamento com a finalidade de facilitar a recuperação desses conteúdos, seja no ambiente tradicional quanto no eletrônico. Lago (2009, p.15) descreve classificação como:

O agrupamento de documentos semelhantes, distribuídos em classes e representados por símbolos (números, letras, sinais gráficos) dentro de um determinado sistema de classificação, seja CDD, CDU. Assim, os documentos de um assunto deverão estar reunidos num mesmo local.

Os sistemas ou esquemas de classificação, conforme Souza (2000), estabeleceu-se com o CDD (Classificação Decimal de Dewey), esquema enumerativo que abrange todas as áreas do conhecimento, seguido do semi-enumerativo ou semi-facetado CDU (Classificação Decimal Universal) e a dos esquemas analítico-sintético ou facetados CC (Classificação de Dois Pontos). Numa etapa seguinte, despertado pelas teorias da classificação facetada, iniciou-se a classificação por áreas do conhecimento.

A classificação produzirá um número ou código que diferenciará cada um dos itens informacionais, denominado de identificador único na área de desenvolvimento de sistemas, que será empregado para localizá-lo na base de dados das unidades informacionais. O identificador único é um campo que descreve um item informacional e não possui esse valor repetido na base de dados, o que possibilita integridade no momento de localizar o item informacional.

Um resumo, de acordo com Lancaster (2004), é uma representação sintética e precisa do conteúdo de um documento, que facilita o processo de escolha de documentos por parte do usuário e auxilia na compreensão de documentos de língua estrangeira, visto que exhibe um texto sucinto sobre o assunto tratado. Tem importante papel nos sistemas de recuperação de informação ao facilitar a identificação dos conteúdos e proporcionar acesso aos itens armazenados (LANCASTER, 2004).

Os mecanismos de busca atuais possuem a capacidade de localizar palavras e frases nos websites, determinadas pelo usuário no momento da busca,

com o auxílio dos metadados. Logo, é possível perceber a relevância em preparar resumos estruturados que ficarão armazenados em metadados, o qual possibilitará compartilhar e promover materiais nas unidades informacionais.

Na arquivologia, a gestão de documentos prepara os itens informacionais para que todas as etapas do ciclo de vida, desde sua produção até o arquivamento e manutenção sejam realizadas respeitando as diretrizes da norma vigente, assegurando sua longevidade. Os procedimentos adotados como prática para organização dos itens informacionais são a avaliação, classificação e a descrição arquivística.

A avaliação é um processo de análise dos itens informacionais de arquivos, no qual define-se o prazo de armazenamento e a destinação de acordo com os valores atribuídos (SOUTO, 2014). Para Silvino Filho (1996), a avaliação considera três diretrizes básicas: a) levantamento da produção documental, concebido e desenvolvido por funções, atividades e rotinas com o objetivo de evidenciar o contexto em que os itens informacionais se inserem e a inter-relação entre os produzidos e os recebidos pelos órgãos de um instituição; b) tabela de temporalidade, que se elabora a partir do levantamento da produção documental para a orientação dos profissionais de arquivo; c) plano de destinação dos itens informacionais onde deve mostrar a inter-relação entre os itens informacionais.

A classificação trata e organiza os itens informacionais, e tem como objetivo “[...] dar visibilidade as funções e as atividades do organismo produtor do arquivo, deixando clara as ligações entre os documentos” (GONÇALVES, 1998). Considera-se importante o arquivista possuir conhecimentos da estrutura organizacional da instituição e dos itens informacionais por meio do levantamento da produção documental, que permitirá conhecer os itens informacionais produzidos e acumulados pela instituição (SOUTO, 2014). É pertinente destacar que o objetivo da classificação na arquivologia é dar visibilidade às funções e às atividades do produtor. Na biblioteconomia, o objetivo é o assunto.

A descrição arquivística tem como propósito identificar e compreender o contexto e o conteúdo dos itens informacionais. Para Cruz Mundet (1996), é uma coleção de procedimentos juntamente com os elementos formais e de conteúdo que possibilitam a identificação de itens informacionais e a elaboração de instrumentos de pesquisa. Hagen (1998) sugere que a descrição seja um procedimento pensado

em todas as fases da gestão documental, desde a produção do item informacional, para que possibilite a recuperação e o processo de arranjo.

Os instrumentos de pesquisa, já salientado por Cruz Mundet (1996) e reforçado por Belloto (2007), expressam que estes instrumentos devem seguir uma hierarquia, para que completem o trabalho de um arquivo, proporcionando uma ligação entre a informação e o pesquisador, possibilitando o acesso aos itens informacionais, a consulta e a divulgação do acervo.

No Quadro 3 são relacionados os principais instrumentos de pesquisa.

### Quadro 3 – Instrumentos de pesquisa

INSTRUMENTOS	NÍVEL	BASE DA DESCRIÇÃO
Guia	Instituição	Conjunto informacionais amplos
Inventário	Fundos, grupos e coleções	Séries
Catálogo	Séries	Unidades informacionais
Repertório e Índice	Unidades informacionais selecionadas pertencentes a uma ou mais origens	Assunto: recorte temático
Tabela de equivalência	Unidades informacionais	Equivalência de anotações

Fonte: adaptado do autor (SOUTO, 2014)

O Quadro 3 apresenta os cinco instrumentos de pesquisa da descrição arquivística, estabelecidos em seus específicos ambientes organizacionais.

A descrição arquivística necessita de prudência nos momentos de elaboração e execução, haja vista a complexidade de seu emprego, portanto requer atenção às suas normas nacionais e internacionais.

As normas internacionais que atendem a descrição arquivística são resultado de um esforço em comum pela convergência dos termos, conceitos e práticas no campo em auxiliar a comunicação e o acesso às informações (SILVA, 1999). No Quadro 4 estão relacionadas as principais normas internacionais e suas datas de publicação:

### Quadro 4 – Normas Internacionais

NORMA	EDIÇÃO	DATAS DE DESENVOLVIMENTO	DATA DE PUBLICAÇÃO
Declaração de Princípios		(1988)1989 – 1992	1992
ISAD (G)	1 <sup>a</sup>	1990 – 1993	1994
ISAAR (CPF)	1 <sup>a</sup>	1993 – 1995	1996
ISAD (G)	2 <sup>a</sup>	1996 – 2000	1999
ISAAR (CPF)	2 <sup>a</sup>	2000 – 2004	2004
ISDF	1 <sup>a</sup>	2005 – 2007	2007
ISDIAH	1 <sup>a</sup>	2005 – 2008	2008

Fonte: (GUEGUEN et al, 2013, p.102)

A Norma Brasileira de Descrição Arquivística (NOBRADE), publicada em 2006 é a norma nacional aplicada aos itens informacionais. Ela reúne as principais informações das normas internacionais ISAD(G) e ISAAR(CF), adaptando a realidade brasileira, com o propósito de facilitar o acesso e intercâmbio de informações (BRASIL, 2006).

A descrição arquivística elabora as representações de um item informacional mediante a aplicação dos modelos como o ISAD(G) e a NOBRADE. As normas nacionais e internacionais são diretrizes para padronização da descrição dos itens informacionais que consistem em certificar e garantir descrições consistentes, que possam acompanhar o avanço tecnológico atual.

A Figura 1exibe as técnicas identificadas da Organização da Informação nas Unidades de Informação. O item informacional representa livros, mídias, artigos, revistas, documentos, etc, armazenados ou guardados em bibliotecas e arquivos.

**Figura 1** – Técnicas de OI nas Unidades de Informação



**Fonte:** o próprio Autor.

As bibliotecas e arquivos também são conhecidos por unidades de informação que são encarregados de adquirir, processar, armazenar e disseminar informações [...] (CUNHA; CAVALCANTI, 2008). No caso, são as unidades que estarão compartilhando os conteúdos para a *Web Semântica*.

### 3.2 ANALISAR AS TECNOLOGIAS, RECOMENDAÇÕES, LINGUAGENS E PADRÕES QUE COMPÕEM A *WEB SEMÂNTICA*.

A *WebSemântica*, conforme seus idealizadores, é uma extensão da *Web* atual, uma proposta para embutir significado aos dados, possibilitando que máquinas não sejam apenas apresentadores de conteúdo, que tenham competência

para ler, interpretar e compreender o significado dos dados (BERNERS LEE; LASSILA; HENDLER, 2001).

Entretanto, possibilitar que as máquinas realizem esse procedimento é uma tarefa que exige o emprego de tecnologias, recomendações, linguagens e padrões que estão em etapa de consolidação, visto que a quantidade de materiais que tratam dos temas especificamente ainda é escassa, em razão da dificuldade de se localizar materiais bibliográficos para serem utilizados nesta pesquisa.

Para a *Web* de dados linkados transformar-se em realidade é preciso que os dados da *Web* e seus relacionamentos estejam em um formato padrão, acessível e gerenciável por ferramentas da *Web* Semântica.

### 3.2.1 Estrutura para uma *Web* Semântica

Segundo Berners Lee, Hendler e Lassila (2001, tradução nossa), o grande desafio para *Web* Semântica é fornecer uma linguagem que expresse dados e regras para que haja um raciocínio, inferências sobre os dados pelas máquinas e que as regras existentes nos sistemas dos provedores de acesso possam exportar para a *Web*. Essa lógica seria uma mistura de decisões matemáticas e engenharias que deve ser adicionado à *Web* para que os objetos tenham suas complexas propriedades descritas. Para que a *Web* Semântica se estabeleça, existem três componentes importantes, a linguagem XML, o RDF e as Ontologias.

Berners Lee, Hendler e Lassila (2001, tradução nossa) definem em seu artigo que a linguagem XML e o modelo RDF são essenciais para estabelecer semântica na *Web*. A XML codifica as páginas da *Web* utilizando *tags* ou etiquetas, criadas pelos próprios desenvolvedores, para que os programas ou *scripts* que realizam a leitura das páginas façam uso dessas *tags*. Entretanto, nem sempre é possível identificar o significado delas, visto que a XML não representa seu significado.

O significado é expresso pelo RDF, que transforma as *tags* em conjunto de triplas, sendo cada tripla composta por um sujeito, verbo e objeto de uma frase elementar. O RDF amarra as relações e cria afirmações de que coisas (*sites* da *Web*, pessoas) têm propriedades (é publicado em, casado com) com determinados valores (outro site da *Web*, outra pessoa). Para que os sistemas da *Web* possam

individualizar cada um desses elementos, o sujeito, o verbo e o objeto são identificados cada um por um indicador universal chamado de *Universal Resource Identifier*<sup>10</sup> (URI), semelhante aos *Uniform Resource Locator* (URL) (BERNERS LEE; HENDLER; LASSILA, 2001, tradução nossa).

Ainda segundo os autores supracitados, essas relações entre os elementos com *URIs* possibilitam que se tornem definições únicas em toda *Web* e principalmente disponível para todos que necessitam dessa informação. No entanto, existe a possibilidade de dois bancos de dados utilizarem identificadores distintos para um mesmo termo, evidentemente os sistemas devem ter recursos para diferenciar tais significados comuns.

A solução é alcançada por meio do terceiro componente denominado Ontologia, que é o documento ou arquivo que define formalmente as relações entre os termos. As ontologias podem melhorar muito a precisão dos mecanismos de busca, visto que estes programas procurariam apenas as páginas que se referem a um conceito preciso, sem sair em busca de palavras chaves ambíguas (BERNERS LEE; HENDLER; LASSILA, 2001, tradução nossa).

No texto supracitado, os autores descrevem as recomendações essenciais para a implantação da *Web* dos dados lincados e uma descrição sucinta do seu funcionamento. A seguir, serão apresentadas estas e outras recomendações relevantes para a realização deste estudo.

### 3.2.2 Metadados

Para recuperação de um item informacional na *Web* por meio de um navegador, é vital os conteúdos estarem descritos de forma clara para que os *sites* de busca possam localizá-los. Os metadados possibilitam a realização dessa tarefa, visto que o seu propósito é descrever os itens informacionais, conforme Takahashi (2000, p.72) define:

Dados a respeito de outros dados, ou seja, qualquer dado usado para auxiliar na identificação, descrição e localização de

---

<sup>10</sup> Identificador Universal de Recursos, é conjunto de caracteres empregado para identificar ou denominar um recurso na internet. Semelhante à URL, porém é mais genérica, retornando menos informações. A URI garante um identificador único em RDF para uma página na Web ou documento (POLLOCK, 2010, p. 154).

informações. Trata-se em outras palavras, de dados estruturados que descrevem as características de um recurso de informação.

Pollock (2010, p.118) afirma que os metadados são:

[...] formas de enriquecer os dados para que os sistemas de software possam interagir com a informação. Os metadados sobre os modelos, vocabulários, e até mesmo as linguagens de programação são simplesmente maneiras de fornecer 'dados sobre dados' para que um intérprete, processador ou algoritmo saiba o que fazer.

A World Wide Web Consortium (2001, p.1) tem uma definição ainda mais específica para metadados e voltada para a *Web Semântica* que define como “[...] informações para a *Web* que podem ser compreendidos por máquinas [...]”. Embora Pollock e a W3C possuam um conceito muito fechado, uma vez que estão focados na *Web* e para o desenvolvimento de *software*, estes se ajustam no objeto de estudo deste projeto.

Muitos aspectos dos metadados podem ser poucos claros para as pessoas não envolvidas na *Web Semântica*, Breitman (2010, p.17) descreve algumas características relativas ao conceito e utilização dos metadados:

1. **Metadados não precisam ser necessariamente digitais** – profissionais de museologia, arqueologia e biblioteconomia tem utilizado metadados há muito tempo para gerenciar suas coleções.
2. **Metadados vão além de fornecer dados sobre um objeto** – embora estejamos mais familiarizados com a utilização de metadados na classificação e na catalogação de objetos, metadados são largamente utilizados para processamento e na preservação de objetos.
3. **Metadados podem ser obtidos a partir de uma variedade de fontes** – podem ser fornecidos por seres humanos, extraídos de grandes bases de dados ou obtidos automaticamente. Um grande projeto na internet hoje são os portais de informação auto generativos, capazes de atualizar suas informações de forma automática.
4. **Metadados evoluem durante a vida útil do sistema de informação ou objeto a que se referem** – metadados são criados, modificados e até mesmo descartados durante a vida útil dos recursos a que se referem.

Para descrever os recursos informacionais os metadados utilizam formatos padronizados que parametrizam a captura de atributos dos recursos informacionais. Os metadados *Moving Picture Experts Group* (MPEG), Padrão de Metadados do Governo Eletrônico (ePMG), Padrão Brasileiro de Metadados para Teses e Dissertações (MTD-BR) e MARC21 são metadados muito empregados, no

entantona *Web Semântica*, segundo Breitman (2010) o formato Dublin Core é um dos padrões mais representativos.

### 3.2.3 Dublin Core

O Dublin Core (DC) é um padrão de Metadados muito empregado em repositórios. Seu aproveitamento deve-se a alta compatibilidade com o protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH)<sup>11</sup> que é o responsável pelo compartilhamento de dados entre os provedores de dados e de serviços (CATARINO, 2009).

A Dublin Core Metadata Initiative (DCMI) é a Instituição responsável pela organização, disseminação, a adoção de padrões e desenvolvimento de vocabulários especializados, como o DC que tem o objetivo de criar mecanismos que auxiliem na recuperação de informações na *Web*, utilizando os padrões de metadados. O padrão apresenta quinze elementos de metadados que podem ser todos utilizados, ou em parte, sem respeitar sua ordem, conforme **Quadro 5**.

A DCMI foi criada em 1994, a partir da conferência da *World Wide Web* realizada em Chicago com representantes de diversas entidades, como Yuri Rubinsky da *SoftQuad*; Stuart Weibel, Eric Miller e Terry Noreault do *OnLine Computer Library Centre (OCLC)*; e Joseph Hardin da *National Center for Supercomputing Applications (NCSA)*, que resultou na realização de um *Workshop* em Dublin, Ohio, em 1995. Nesse evento, foram discutidos quais elementos poderiam ser considerados como um conjunto mínimo de elementos para a classificar a *Web* e facilitar a busca e recuperação das informações da *Web* (DCMI, 2004).

---

<sup>11</sup> Protocolo desenvolvido pela Open Archives Initiative que permite colher metadados em qualquer formato pelos provedores de dados (OPEN ARCHIVES, 2014).

**Quadro 5 – Elementos do DC**

Nome	Descrição
<b>Assunto (subject)</b>	O tópico abordado pelo trabalho
<b>Título (title)</b>	Nome do objeto
<b>Criador (creator)</b>	Pessoa(s) responsável(eis) pelo conteúdo intelectual do objeto
<b>Descrição (description)</b>	Descrição do conteúdo do objeto
<b>Editor (publisher)</b>	Agente ou agência responsável por disponibilizar o objeto
<b>Outro Agente (contributor)</b>	Pessoa(s) que fez (fizeram) contribuições significativas para o objeto
<b>Data (date)</b>	Data de publicação
<b>Tipo de Objeto (type)</b>	Gênero do Objeto, se ficção, novela, poema ou dicionário
<b>Formato (format)</b>	Manifestação física do objeto. Exemplo são arquivos executáveis, do tipo texto ou PDF
<b>Identificador (identifier)</b>	Cadeia ou número utilizado para identificar unicamente aquele objeto
<b>Relacionamento (relation)</b>	Relacionamento com outros objetos
<b>Fonte (source)</b>	Outros objetos, eletrônicos ou físicos, dos quais este foi derivado (caso seja aplicável)
<b>Linguagem (language)</b>	Linguagem do conteúdo intelectual
<b>Cobertura (coverage)</b>	Localizações espaciais e durações temporais características do objeto
<b>Direitos (rights)</b>	Informação sobre os direitos acerca do objeto

**Fonte:** Bretman (2010, p.18)

O padrão de metadados DC é um simples, mas eficaz, conjunto de elementos que permitem descrever grande parte dos recursos da rede. O DC pode ser dividido em dois níveis: o Simples e o Qualificado. O Simples, compreende os 15 termos básicos, o Qualificado inclui três elementos adicionais, o Audiência (classe de entidade para a qual o recurso é dirigido ou útil), Proveniência (Declaração de qualquer mudança na propriedade e custódia dos recursos, desde a sua criação) e Detentor de Direitos (Pessoa ou organização detentora ou gestora de direitos sobre o recurso) que refinam a semântica dos elementos para uma representação mais completa. A semântica do DC foi estabelecida por um grupo internacional e interdisciplinar de profissionais de Biblioteconomia, Ciência da Computação, codificadores de texto, museólogos e outras áreas afins (HILLMANN, 2005).

O nível Simples pode conter subpropriedades que permitem uma descrição mais específica. A propriedade Data pode ser subdividida em outras descrições que possuam propriedades relacionadas com o termo Data. Dessa maneira, permite-se atingir uma descrição mais apurada do objeto em estudo, subdividindo em data de criação, modificação, aceitação, submissão, validade, entre outros (CATARINO, 2009).

Com essas características, de acordo com Hillman (2005, p.5-6), o DC pretende atingir os seguintes propósitos:

- Simplicidade de criação e manutenção: o DC é mantido tão simples quanto possível para permitir que não especialistas possam descrever facilmente os recursos eletrônicos.
- Semântica comum e universal: o DC pode auxiliar um pesquisador não especialista a achar seu modo de manter um conjunto de elementos, a semântica entendida universalmente.
- Alcance Internacional: o DC é originalmente desenvolvido em Inglês, mas tem versões em diversos idiomas.
- Extensibilidade: o DC tem considerado a importância de fornecer um mecanismo para estender o conjunto de elementos DC para necessidades adicionais que surjam, considerando que outras comunidades podem desenvolver conjunto de elementos adicionais específicos para suas necessidades específicas.

Embora o DC tenha sido originalmente desenvolvido com o enfoque na descrição de documentos como objetos (os recursos de textos tradicionais são bem mais legíveis) eles também podem ser aplicados a outros recursos, dependendo de como se assemelham aos metadados do padrão DC. (HILLMANN, 2005).

### 3.2.4 RDF (*Resource Description Framework*)

O RDF, Modelo para Descrição de Recursos, é uma estrutura e ou modelo para representar informações recomendada pelo W3C que “[...] fornece uma maneira padronizada de utilizar o XML para representar metadados no formato de sentenças sobre propriedades e relacionamento entre itens na web.”. Dessa maneira, com a linguagem de marcação XML, o RDF consegue descrever as

propriedades do objeto. Para os itens na *Web*, entende-se como “recursos”, que “[...] podem ser virtualmente qualquer objeto (texto, figura, vídeo e outros).” (BREITMAN, 2010, p.20).

Para Pollock (2010, p.77), “O RDF foi concebido desde o início para o acesso e utilização ao longo da *World Wide Web*, e é projetado para proporcionar uma base fundamental para linguagens de dados mais avançadas.”

Cunha (2002, p.16) define que o propósito do RDF seria:

[...] definir um mecanismo para descrever recursos não vinculados a um domínio específico de aplicação. RDF facilita o intercâmbio de informações, que podem ser interpretadas por máquinas, entre aplicativos via *Web*, permite adicionar semântica formal para a *Web* e também, o compartilhamento de conhecimento.

O modelo RDF, conforme Cunha, tem a finalidade de adicionar semântica para a *Web* para facilitar o intercâmbio de informações e possibilitar que possam ser interpretadas por máquinas. O intercâmbio de informações pode-se entender como interoperabilidade, ou, padrões que possibilitem trocas entre diferentes plataformas e sistemas heterogêneos de forma transparente. Siqueira (2003, p.103) define interoperabilidade como:

[...] a capacidade de compartilhamento de informações entre softwares, independente da estrutura de armazenamento dos dados usada em seu banco de dados, ou seja, a estrutura de armazenamento dos dados não impede a troca de informações entre instituições.

Diante disso, entende-se que o RDF, em consonância com a linguagem XML, tem uma importância relevante para a *Web Semântica*, visto que uma das características da Internet é a diversidade de plataformas, além de incorporar a semântica aos conteúdos compartilhados.

Para representar as propriedades de recursos informacionais e seus valores, o RDF “[...] possui um modelo de framework baseado na ideia de uma tripla”, composto de “Recurso, Propriedade e Valor” (POLLOCK, 2010, p.152). Breitman (2010, p.22) descreve os elementos do RDF com esses conceitos: Recurso, ou sujeito: pode ser qualquer coisa (livro, revista, pessoa, documento), desde que possua uma URI. Propriedade: informações que representam as características do recurso, são os atributos que permitem distinguir um recurso de outro. Valor: são os dados que representam a informação a ser descrita, um literal, valor numérico ou cadeia de caracteres.

Nas Figura 2 e 3, a W3C (WORLD WIDE WEB CONSORTIUM, 2014a) exemplifica de maneira sintética como o RDF é expresso em código e em sua forma gráfica. A Figura 2 reproduz um exemplo com código adaptado para facilitar a leitura e entendimento, denominado de pseudocódigo. Nesse exemplo, Bob é o recurso de quatro triplas e Mona Lisa é o recurso de uma tripla e valor de outras duas. Essa possibilidade de o mesmo elemento estar como recurso em uma tripla e objeto em outra, torna possível a conexão entre triplas, caracterizando o poderio do RDF.

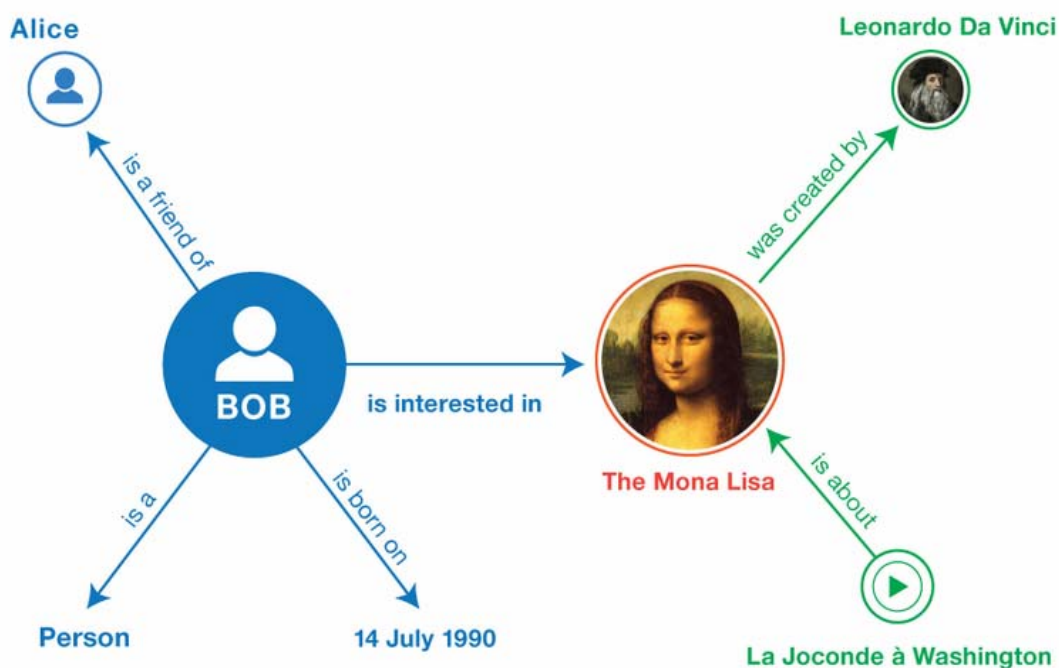
**Figura 2** – Pseudocódigo da Tripla RDF

```
<Bob> <is a> <person>.
<Bob> <is a friend of> <Alice>.
<Bob> <is born on> <the 4th of July 1990>.
<Bob> <is interested in> <the Mona Lisa>.
<the Mona Lisa> <was created by> <Leonardo da Vinci>.
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>
```

**Fonte:** W3C (WORLD WIDE WEB CONSORTIUM, 2014a)

A Figura 3 exibe as triplas em um gráfico conectado e composto de nós e arcos. Os recursos e valores compõem os nós dos gráficos e as propriedades formam os arcos. Com as ligações entre os elementos realizados, pode-se utilizar uma linguagem de busca como o SPARQL, para consultar informações referentes aos nós do gráfico.

**Figura 3 – Gráfico RDF**



Fonte: W3C (WORLD WIDE WEB CONSORTIUM, 2014a)

No Quadro 6 estão representadas as informações descritas nas Figura 2 e Figura 3, em um formato que possibilite visualizar o que é um recurso, uma propriedade e um valor.

**Quadro 6 – Recurso, Propriedade e Valor**

Recurso	Propriedade	Valor
Bob	Is a	Person
Bob	Is a friend of	Alice
Bob	Is born on	14 jul 1990
Bob	Is interested in	The Mona Lisa
The Mona Lisa	Was Created by	Leonardo da Vinci
The Vídeo La Joconde à Washington	Is about	The Mona Lisa

Fonte: próprio autor

Na versão RDF Primer 1.1, a W3C (WORLD WIDE WEB CONSORTIUM, 2014a) utiliza três tipos de declaração para o RDF data, que são *IRIs*, *literais* e *nós em branco*.

O IRI é um Identificador de Recursos Internacional, sendo o *Uniform Resource Locators (URL)* o utilizado para acessar sites da *Web* e o *URI*. Este sem

limitações de utilização de caracteres, o que permite ser empregado em diferentes linguagens.

Conforme as orientações da W3C o IRI pode estar em três posições de uma tripla.

- utilizado para identificar recursos como um documento, pessoas, objetos físicos e conceitos abstratos, por exemplo o IRI de Leonardo da Vinci na *Dbpedia*:  
[http://dbpedia.org/resource/Leonardo da Vinci](http://dbpedia.org/resource/Leonardo_da_Vinci)
- utilizado para identificar um vídeo sobre Mona Lisa de nome “La Joconde à Washington” no *site* da Europeana:  
<http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619>
- utilizado como identificadores globais, para que outras pessoas possam utilizar essa *IRI* para conhecer outras pessoas:  
<http://xmlns.com/foaf/0.1/knows>

Além de representar graficamente uma informação por meio de grafos, o modelo RDF é escrito em linguagem XML, o que possibilita sua interoperabilidade e multiplataforma.

### 3.2.5 Ontologias

De acordo com sua origem etimológica, a ontologia significa estudo do “ente” ou estudo do “ser”. Desse modo, ontologia é o estudo da existência de todos os tipos de entidades, abstratas ou concretas que fazem parte do mundo. A palavra tem origem grega e é um termo de domínio da Filosofia (LIMA-MARQUES, 2006). Muito embora tenham surgido na Filosofia há muitos anos, as ontologias como um modelo de representação do conhecimento são estudadas há pouquíssimo tempo. Amplamente empregadas nas Ciências da Computação, as ontologias são também um objeto de estudo na Ciência da Informação (SALES; CAFÉ, 2008).

A ontologia é uma estrutura de dados que representam conceitos e relações entre entidade em um domínio de forma compartilhada e harmônica no intuito de possibilitar a interoperabilidade entre sistemas de informação. Esta

estrutura deve ser formal, legível e compreendida pelos computadores para que haja o reuso deste conteúdo entre os sistemas (BRÄSCHER; CARLAN, 2010).

Uma ontologia possui elementos que compõem sua estrutura e podem ser organizados conforme a declaração de Gómez-Pérez (1999) em Classe, Relações, Funções, Axiomas e Instâncias.

- classe: representação de um objeto ou uma atividade, podem ser concretos ou abstratos, pode ser uma tarefa, uma função, ação, estratégia;
- relações: são os tipos de interações entre classes de um domínio. Elas podem ser formalmente definidas como qualquer subconjunto de um produto;
- funções: são casos especiais de relações no qual o enésimo elemento da relação é o único para os n-1 elementos precedentes;
- axiomas: são usados para modelar sentenças que sempre são verdadeiras;
- instâncias: são empregados para representar elementos.

A estrutura descrita são os elementos para organizar a preparação e o inter-relacionamento das entidades que compõem o domínio de uma ontologia. Uma das propriedades de uma ontologia é de se reutilizar as informações, visto que estes conteúdos estão classificados, ordenados e disponíveis para consultas. Uma ontologia de um domínio qualquer pode ser incorporada a outros domínios de áreas equivalentes ou distintas, mas que possuem alguma relação em comum, permitindo interconexões, produzindo expansões e tornando-o maior.

Diante disso, fica evidente a diversidade de áreas em que a ontologia atua, Breitman (2010, p.31) cita as várias “comunidades” em que as ontologias são úteis: “Inteligência artificial, Representação do Conhecimento, Processamento de Linguagem Natural, *Web Semântica*, Engenharia de Software, entre outras”.

### 3.2.6 Linguagens da *Web Semântica*

Uma linguagem de Programação é um “[...] conjunto de símbolos e regras para combiná-los de modo a expressar operações lógicas, matemáticas e de recepção e emissão de dados para dispositivos.” (OLIVEIRA, 1999, p.175). Pode-se compreender os dispositivos como computadores, *tablets*, *smartphones*, entre outros equipamentos de comunicação, que necessitam de um conjunto padronizado de instruções para manter suas operações.

Uma linguagem semântica é uma linguagem de codificação que emprega métodos padronizados que possibilitam a “publicação de ontologias” de forma que os computadores possam compreender e processar a informação. Essas linguagens também são reconhecidas como Linguagens de Ontologias do tipo *mark-up* ou Linguagem de Marcação e são baseadas no HTML, XML e RDF (BREITMAN, 2010, p.7; 48).

A linguagem de marcação foi criada no final da década de 60 com o propósito de criar um sistema portátil, independente de *hardware* e *software*, para compartilhar e processar documentos. Com a internet, um ambiente muito heterogêneo, as linguagens de marcação tornaram-se um padrão amplamente utilizado, com as seguintes características básicas: “Possui marcação descritiva; suas marcações podem ser tipadas e independência de plataforma [...]” (FARIA, 2005, p.8).

As marcações descritivas permitem utilizar códigos de nomes para identificar diferentes partes de um documento, as instruções identificarão facilmente o que são marcações e dados. A tipagem das marcações permite através dos *Document Type Definition*(DTD), que os softwares não transportem as informações dos tipos de dados de um documento, tornando-o mais específico. A independência de plataforma assegura que os dados sejam mantidos, não interessando a plataforma que o *software* ou *hardware* esteja em execução (FARIA, 2005).

A XML – *eXtensible Markup Language* é uma linguagem de marcação que “permite a troca de informações de forma estruturada através da Internet. [...] permite aos programadores transportar dados de um servidor para outro na rede, de forma transparente e organizada.” (APDSI, 2005, p.48). Faria (2005, p.11) também complementa que o XML utiliza o padrão UNICODE que permite armazenar qualquer tipo de caractere ou símbolo, representando informações em qualquer idioma ou cultura.

A linguagem XML pode ser utilizada junto com a HTML, no entanto, as duas têm propósitos distintas. A HTML opera sobre a formatação para a exibição de uma palavra ou trecho de um texto, a XML cria as estruturas para representar seu significado (FURGERI, 2006).

Ambas são linguagens de marcação, ou seja, operam com identificadores denominados tags que ficam entre os sinais de “<>” e que especificam sua

característica. A HTML utiliza tags próprios, como <b> que assinala o início de uma formatação em negrito, enquanto que no XML poderia representar um valor, um objeto ou uma estrutura.

Um documento gerado em XML tem embutido informações de sua estrutura e significado, o que possibilita que esse material possa ser compartilhado com outras empresas, provendo a automatização do processo de troca de dados. Esse material estruturado armazenado estaria pronto para ser recuperado, sendo possível afirmar que traria benefícios para quem produz ou quem o utiliza, tornando o custo quase nulo. Ao elaborar documentos em formatos XML, em detrimento dos formatos “doc” ou “pdf”, também seria possível um material mais acessível (FURGERI, 2006).

O W3C tem empenhado muitos esforços para que a *Web* de dados seja uma realidade para os exigentes consumidores da *Web*, possibilitando que os equipamentos possam realizar consultas mais precisas em bases de dados estruturados e organizados. A *Web* de dados utiliza o RDF como base para publicação e *links* entre os dados, e para que as consultas nas bases de dados sejam realizadas com integridade é necessária uma linguagem de consulta própria, desenvolvida especificamente para ser utilizada na *Web* Semântica, a SPARQL (WORLD WIDE WEB CONSORTIUM, 2011b).

A linguagem de consulta SPARQL e seus protocolos torna possível enviar consultas e recuperar os resultados ao utilizar os protocolos comuns da Internet. As consultas são baseadas nos padrões das triplas do RDF que representam as propriedades dos recursos informacionais e seus valores. A máquina de busca SPARQL retorna os resultados para todas as triplas correspondentes, o que possibilita aos consumidores da *Web* de dados extraírem os resultados complexos que podem ser devolvidos em formato de tabelas, para serem incorporados em outras páginas da *Web* (WORLD WIDE WEB CONSORTIUM, 2013a).

A OWL ou linguagem de Ontologia da *Web* é uma linguagem para instanciar classes e definir ontologias na *Web*, foi projetada para que máquinas pudessem processar os conteúdos da *Web* com a colaboração das tecnologias XML e RDF. A OWL foi idealizada pelo Departamento de Defesa dos EUA e pela comunidade de Defesa Europeia que demandavam por uma linguagem de dados

auto descritiva e dinâmica para que os agentes de *softwares* pudessem agir de forma autônoma nesses dados (POLLOCK, 2010).

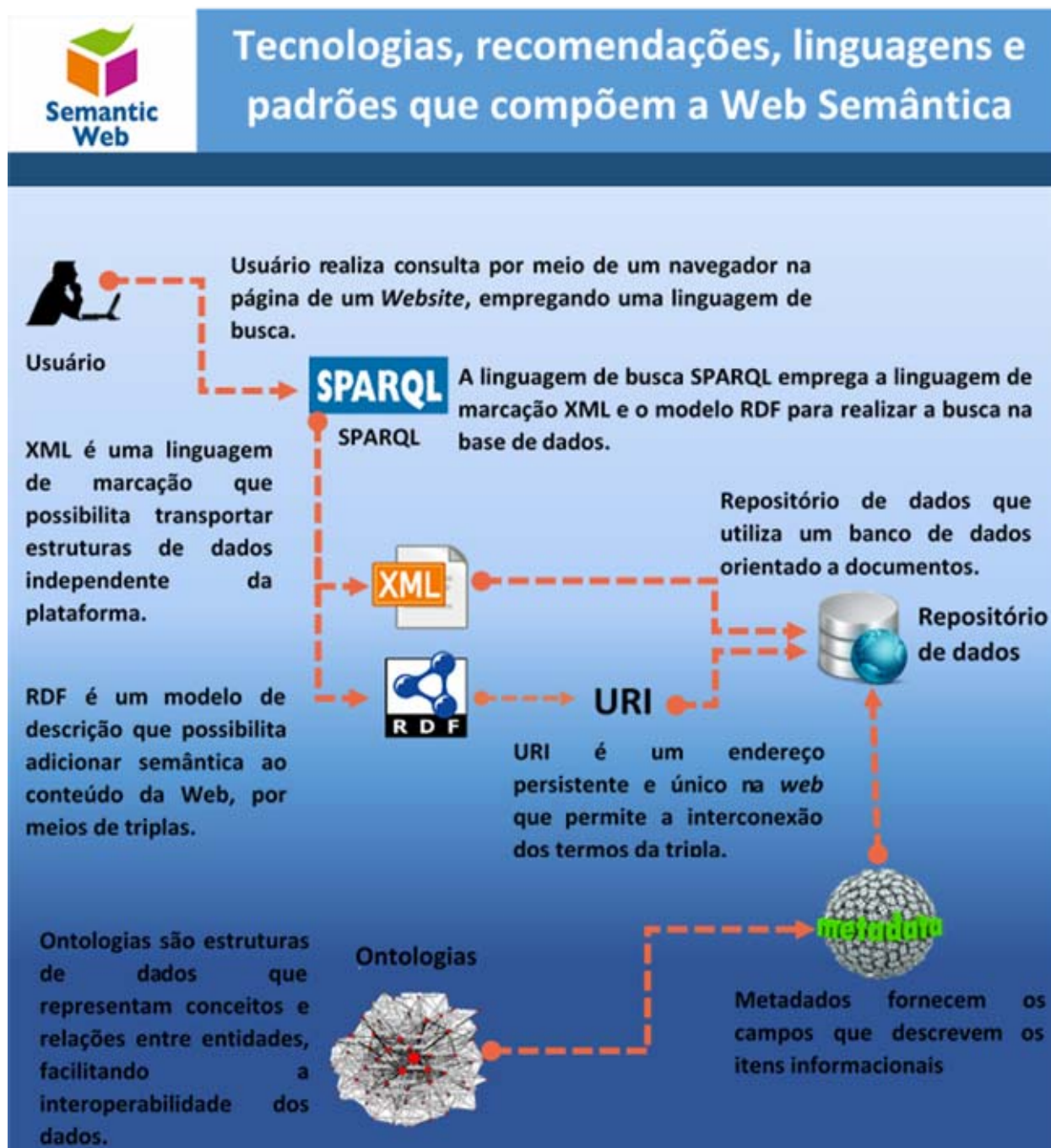
A linguagem de ontologia é relevante para a *Web Semântica* por tornar os conteúdos da *Web* mais acessíveis para os processos automatizados, adicionando informações nos recursos da *Web*. À medida que a rede mundial de computadores cresce, a OWL deve acompanhar e estender suas conexões por todas as fontes distribuídas, por intermédio das ontologias relacionadas (SMITH; WELTY; MCGUINNESS, 2004).

Embora a *Web Semântica* não seja considerada uma revolução ou uma nova *Web*, possibilitou que novas linguagens de representação e distribuição de dados, consultas e criação de ontologias fossem concebidas ou aprimoradas no decorrer de seu projeto. Essa movimentação sempre traz diversas modificações no campo da tecnologia, o que possibilita a mudança para novos princípios e conceitos contribuindo para o aperfeiçoamento da *Web*.

Na Figura 4 criou-se um infográfico para representar, de maneira simplificada, algumas etapas que envolvem o projeto da *Web Semântica*.

O usuário da Figura 4 representa um indivíduo que necessita realizar uma busca de um item informacional (artigo, slide, vídeo, tese, etc) em um repositório de dados que emprega os princípios da *Web Semântica*. Por meio de um ambiente próprio em uma página da *Web*, o usuário escreve os comandos em linguagem SPARQL para a recuperação do conteúdo no repositório de dados. A linguagem de marcação XML é a estrutura que acomoda os códigos SPARQL e o modelo RDF e o provável retorno do resultado da pesquisa também poderá ser no formato XML.

Figura 4 – Tecnologias, recomendações, linguagens e padrões da WS



Fonte: o próprio autor

O modelo de descrição RDF emprega o conceito de triplas e tem a finalidade de adicionar semântica à *Web* e facilitar o intercâmbio de informações e possam ser interpretadas por máquinas (CUNHA, 2002). No entanto, para que ocorra interconexões, o RDF deve expressar o sujeito ou recurso por meio de URI's, o que garante o intercâmbio e unicidade da informação.

O repositório de dados, que mantém os itens informacionais, é preferencialmente aberto, o que torna seu conteúdo de acesso público. Emprega-se o banco de dados orientado a documentos, que possui uma estrutura de dados composta por campos que podem incluir outros documentos, matrizes e matrizes de documentos. A biblioteca virtual Europeia utiliza o banco de dados MongoDB.

Os repositórios de dados devem fornecer informações a respeito dos itens informacionais no momento da consulta por parte do usuário, essas informações são recebidas dos diversos provedores de conteúdo que integram a rede do repositório. Os provedores de conteúdo disponibilizam ou enviam as descrições dos itens informacionais em um padrão de metadados determinado pelo provedor de conteúdo. A padronização por um modelo de metadados, possibilita a interoperabilidade entre os diversos provedores de conteúdo, visto que os nomes dos elementos utilizados na descrição do item informacional são comuns a todos.

As ontologias possibilitam o enriquecimento do item informacional localizado pelo usuário. Estas ontologias, ou os vocabulários, considerado um modelo mais simples de ontologia, conectam os termos com termos de vocabulários distintos, permitindo que ocorra compartilhamentos, comparações ou combinações de informações de outros domínios. Dessa forma, na página de resultado da busca, realizado pelo usuário, exibe-se informações recuperadas da base de dados do repositório origem, complementados pelos vocabulários estendidos de outros repositórios.

O Projeto da *Web Semântica* envolve diversos elementos que devem ter sincronia em suas ações para que possam ser aproveitadas todas as funcionalidades e benefícios. No entanto, existem dificuldades que devem ser superadas para sua aplicação. Como exemplo, é possível citar as ontologias, estruturas determinantes para a padronização dos dados na *Web*, que são complexas e exige muito tempo para sua preparação. A linguagem SPARQL, possibilita localizar termos na *Web* com o emprego de triplas conectadas à URI, porém, exige um conhecimento em uma linguagem de programação para a realização de buscas e as URI's nem sempre são persistentes.

Em suma, entende-se que a *Web Semântica* ou a *Web* dos dados lincados enfrenta inúmeras dificuldades para sua consolidação, em especial no que tange a organização dos dados na *Web*, visto que a *Web* é um espaço aberto e

democrático, no qual pessoas ou organizações não são obrigadas a respeitar e seguir um padrão, normas ou recomendações para a produção de *Websites*.

### 3.3 APLICABILIDADE DA *WEB* SEMÂNTICA EM AMBIENTES DAS UNIDADES DE INFORMAÇÃO, POR MEIO DE UM ESTUDO DE CASO E UM *FRAMEWORK*

Com os dados linkados, os itens informacionais das unidades de informação passam a estar compartilhados na *Web*, estendendo seus domínios, permitindo uma infinita possibilidade de acessos, interagindo com outras unidades de informação e enriquecendo seus conteúdos.

A biblioteca virtual Europeia contribuiu com o projeto da *Web Semântica* ao tornar-se um provedor de conteúdos para toda Europa, formando uma cadeia de fornecedores de dados, promovendo o intercâmbio de ideias e informações fundamentado nos princípios dos dados linkados.

A Biblioteca Digital Europeia foi liberada para acesso público em 20 de novembro de 2008, para tornar o patrimônio cultural mais acessível através dos meios digitais, promovendo o intercâmbio de ideias e informações, para assim possibilitar compreender melhor a diversidade cultural dos países envolvidos. Tão logo da sua inauguração, o *site* da Europeia teve de ser suspenso logo após seu funcionamento, devido ao excessivo tráfego registrado, retornando novamente em 22 de dezembro, quando quadruplicaram a capacidade de tráfego da rede dos servidores.

No Quadro 7 descreve-se a história da Fundação Europeia, a partir do ano de 2005 até 2014.

### Quadro 7 – História da Fundação Europeia

Ano	Descrição
2005	Uma carta para seis líderes de estado (França, Alemanha, Itália, Espanha e Hungria) sugere a criação da Biblioteca Digital Europeia.
2007	A rede de Biblioteca Digital Europeia – EDLnet – inicia a criação de um protótipo, financiado pela i2010
2008	Um protótipo da Europeia é lançado em 20 de Novembro por Viviane Reding, Comissária Europeia para a Sociedade da Informação e Mídia, e do Presidente da Comissão José Manuel Barroso.
2009	Coleção da Europeia atinge cinco milhões de itens.
2010	Um relatório do Parlamento Europeu aprovada em fevereiro, pede mais conteúdo e financiamento para a Europeia. É aprovada por unanimidade e em julho, a coleção da Europeia atinge 10 milhões de itens. O protótipo se torna um serviço operacional financiado no âmbito da Competitividade da CE e a Inovação (CIP ICT-PSP).
2011	Um novo relatório, em Maio de 2011, o “Comité des Sages” apoia a Europeia como um ponto de referência para a cultura europeia on-line. Em Outubro, a CE publica recomendações pedindo os Estados membros da UE a intensificar os seus esforços para contribuir com a Europeia.
2012	Europeana libera todos os metadados, tornando-se livremente disponível para reutilização.
2013	Europeana continua a promover sua posição como catalisador para a inovação e ao empreendedorismo digital no apoio da Agenda Digital para a Europa.
2014	Preparar a Europeia para a transição do projeto de 2015-2020, para atender as demandas das indústrias de criações nos materiais de herança cultural que é o combustível para novas aplicações, como também experimentar novas ideias e maneiras de tornar a Europeia sustentável a longo prazo.

**Fonte:** Europeia(2014b, tradução nossa)

Com um vasto material, com a expectativa de atingir 30 milhões de objetos até 2015, a biblioteca digital tem o compromisso de manter preservadas a cultura e a história do “velho continente” que moldou o mundo e participa de maneira acentuada nos acontecimentos atuais.

O projeto Europeia trabalha para tornar esses objetos digitais das 19 principais bibliotecas europeias de livre acesso nos *sites* da Biblioteca Europeia. As Fotografias da Guerra Civil Espanhola, manuscritos otomanos, filmes históricos médicos e milhares de teses, revistas e artigos acadêmicos estão entre os recursos digitais que a Europeia Bibliotecas está trabalhando.

Quando este projeto terminar, ele terá criado:

- um recurso valioso para os estudantes, para pesquisas de materiais;
- uma rede nacional sólida, universitárias e pesquisas para bibliotecas;
- um modelo de agregação eficiente que pode ser utilizado por bibliotecas de pesquisa em toda a Europa (EUROPEANA, 2014<sup>a</sup>, tradução nossa).

A Fundação Europeia é a operadora responsável pelo serviço da Europeia, incluindo o portal de busca *Europeana.eu* e os serviços relacionados com dados, o qual é governado por um comitê executivo e um conselho. A rede Europeia é um fórum aberto, é especializada e composta por mantenedores e agregadores de conteúdo, juntamente com os provedores de conhecimentos técnicos, jurídicos e estratégicos.

A Europeia foi o estudo de caso selecionado para o caso proposto. Para a realização da análise dos dados, classificou-se os assuntos mais importantes em categorias de estudo, conforme recomenda Bardin(2010).

### 3.3.1 As Categorias de Estudo

A organização do estudo em agrupamentos ou categorias, possibilitou a classificação dos tópicos mais relevantes da *Web Semântica* na biblioteca virtual Europeia tornando a pesquisa mais objetiva, permitindo investigar e analisar com propriedade as peculiaridades de cada categoria.

Foram elencadas cinco categorias consideradas essenciais para o estabelecimento da *Web Semântica* nas unidades de informação: Metadados; Harvesting; Linguagem de Marcação e Banco de Dados; Linguagem de Busca e Ontologias.

Os tópicos a seguir descrevem cada uma das categorias estudadas.

#### 3.3.1.1 Metadados na Europeia

No propósito de possibilitar a interoperabilidade entre os dados de diferentes provedores de conteúdo que possuem e fornecem conjuntos de

metadados distintos, a biblioteca *Europeana* optou por estabelecer uma estrutura de metadados com conteúdo comum para toda comunidade fornecedora de conteúdo.

Dessa maneira, inicialmente a *Europeana* implantou a estrutura de metadados *Europeana Semantic Elements* (ESE), que consiste em um perfil de aplicação<sup>12</sup> que utiliza um subconjunto de elementos do esquema Dublin Core com elementos específicos da *Europeana*. No entanto, a estrutura ESE apresentou vários problemas que dificultaram a transformação dos registros de metadados em dados linkados, visto que os elementos representados pelo esquema são em sua maioria textuais, que impede a sua ligação a outros objetos ou entidades contextuais (ISAAC; CLAYPHAN; HASLHOFER, 2012).

#### Quadro 8 – Exemplo de um Registro *ESE Europeana*

ESE Field	Value
dc:identifier	02-Lemba-Lakkous-Lady.tif
dc:title	Stone figurine
dc:subject	Cyprus--Antiquities
dc:description xml:lang="en"	Stone vessel from Choirokoitia with embossed decoration...
dc:publisher	Cyprus Archaeological Museum
dc:type	Image
dc:format	image/tiff
dcterms:temporal	mid 3rd millenium B.C.
dc:rights	Cyprus Archaeological Museum
dc:title	Λίθινο ειδώλιο
dc:description	Λίθινο ειδώλιο από τη θέση Λέμπα ...
Europeana:provider	Cyprus Library
Europeana:type	IMAGE
Europeana:rights	<a href="http://creativecommons.org/licenses/by/3.0/nl/">http://creativecommons.org/licenses/by/3.0/nl/</a>
Europeana:dataProvider	Cyprus Archaeological Museum
Europeana:isShownBy	<a href="http://www.mcw.gov.cy/mcw/da/da.nsf/All/8308A7AB9CA6CB5EC22574CC003FF5E1/\$file/02-Lemba-Lakkous-Lady.jpg?OpenElement">http://www.mcw.gov.cy/mcw/da/da.nsf/All/8308A7AB9CA6CB5EC22574CC003FF5E1/\$file/02-Lemba-Lakkous-Lady.jpg?OpenElement</a>
Europeana:uri	<a href="http://www.europeana.eu/resolve/record/00000/E2AAA3C6DF09F9FAA6F951FC4C4A9CC80B5D4154">http://www.europeana.eu/resolve/record/00000/E2AAA3C6DF09F9FAA6F951FC4C4A9CC80B5D4154</a>
Europeana:country	cyprus
Europeana:language	gr

Fonte: (HASLHOFER; ISAAC, 2011, p.98)

O Quadro 8 apresenta um exemplo de registro de metadados elaborado com o modelo *ESE*, o qual possui elementos *Dublin Core* identificados pela sigla

<sup>12</sup> Perfil de Aplicação: esquema que representa os metadados de um conjunto de metadados e tipos de valores permitidos. Define os termos para a anotação de dados com metadados. (RATANAJAIPAN; NANTAJEEWARAWAT; WUWONGSE, 2007)

“dc” antecipando o nome do campo, e metadados específicos da Europeana como ***dataProvider***, ***URI*** e ***country*** como parte de um esforço em solucionar as dificuldades da interoperabilidade e conexões com outros objetos. O registro do metadados representa uma peça de cerâmica mantida no museu arqueológico de Chipre (*Cyprus*), conforme Figura 5.

**Figura 5** – Estatueta de Pedra (Chipre)



Fonte: (EUROPEANA, 2015)

Entende-se que o modelo ESE provocou a perda de informações recebidas, em razão da diversidade de modelos de metadados adotados pelos provedores de conteúdo. Muitos desses modelos não possuíam compatibilidade suficiente com os elementos padrões de dados da *Europeana*, o que ocasionou empobrecimento dos dados agregados pela Europeana.

Com o propósito de minimizar as perdas no momento da coleta de dados, a Europeana elaborou um novo modelo de estrutura de metadados, o *Europeana Data Model* (EDM), que é um modelo para estruturação dos dados que a Biblioteca Europeana tem recebido, organizado e publicado. O modelo de dados é uma melhoria do ESE, que não foi descartado. Os dados continuam a serem agregados em ESE, convertidos posteriormente para EDM para reverter o esquema redutor do modelo anterior. Este novo modelo suporta a representação de objetos mais complexos ou de estrutura hierárquica (partes de um livro, arquivo, etc), tornando-o mais extensível (ISAAC et al., 2011).

O modelo de estrutura de dados concebido pela Europeana não é construído baseado em padrões de uma determinada comunidade, o EDM adota

uma estrutura baseada em uma *Web Semântica* aberta, que possibilita o enriquecimento de dados a partir de uma variedade de fontes de terceiros. Um item informacional do provedor A pode ser contextualmente enriquecido por metadados do provedor B, o qual pode ser enriquecido pela adição de dados mantidos pelo provedor C e por um tesouro oferecido pela editora D. O modelo EDM apoia esta riqueza de ligações, e exibe a proveniência de todos os dados que ligam ao item informacional (ISAAC et al, 2011).

Ao recolher os dados dos provedores de conteúdo, a Europeia os reúne em classes<sup>13</sup>, que são classificadas em diversos elementos denominados propriedades<sup>14</sup>, que são ordenadas e tratadas. Existem sete classes principais no EDM, sendo estruturadas em três classes núcleo e quatro classes contextuais.

As classes núcleo são constituídas por (THE EDM MAPPING GUIDELINES, 2014):

- edm:ProvidedCHO – Provedor de objetos do patrimônio cultural;
- edm:WebResource – Representação Digital;
- edm:Aggregation – Agregação que agrupa as classes.

As classes contextuais são constituídas por:

- edm:Agent – Quem;
- edm:Place – Onde;
- edm:TimeSpan – Quando;
- skos:Concept – O quê

Nas **Classes Núcleo**, os valores dos metadados das classes do provedor de itens do patrimônio cultural (edm:ProvidedCHO) são separados de sua Representação Digital (edm:WebResource) para que os valores possam ser associados de maneira adequada. Ao permitir essa organização, o EDM define três classes para representar o item informacional. Por exemplo, a Mona Lisa é representada pela edm:ProvidedCHO e sua imagem digital pela classe edm:WebResource. Isto permite que as propriedades mais relevantes de metadados possam ser aplicadas a cada classe.

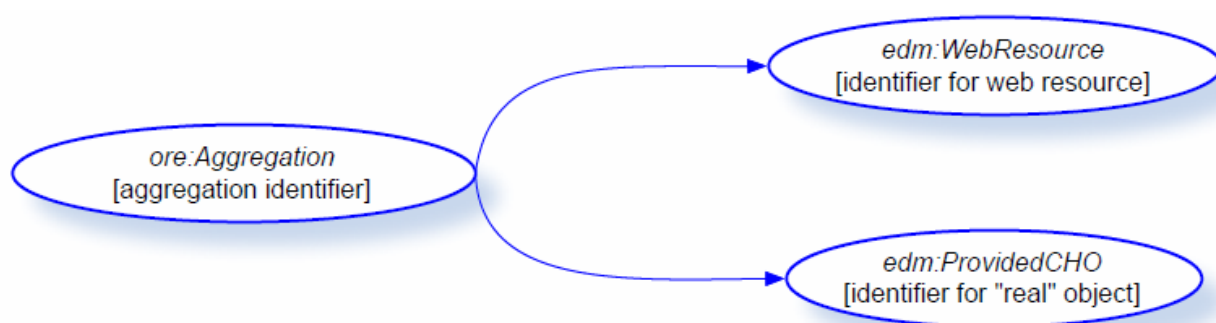
---

<sup>13</sup> Nome atribuído a um grupo de elementos que possuem propriedades em comum (THE EDM MAPPING GUIDELINES, 2014, pág.46).

<sup>14</sup> Um elemento que expressa o relacionamento entre dois recursos. Pode ser o atributo ou característica de um recurso (THE EDM MAPPING GUIDELINES, 2014, pág.46).

Ainda no exemplo, Mona Lisa teria na classe `edm:ProvidedCHO` a propriedade `dc:creator` com o valor “Leonardo da Vinci” e na classe `edm:WebResource` na propriedade `dc:format` o valor “jpg”. Para associar as duas classes utiliza-se a classe `edm:Aggregation`, o qual possui propriedades como o `edm:dataProvider` que permite conectar e associar as classes (THE EDM MAPPING GUIDELINES, 2014). Observe na figura 6 como as três classes estão interligadas.

**Figura 6** – As três classes núcleo do EDM



**Fonte:** (THE EDM MAPPING GUIDELINES, 2014, p.6).

O quadro 9 apresenta as três classes núcleo do EDM e as suas propriedades internas.

**Quadro 9** – Propriedades das Classes Núcleo

Propriedades do <code>edm:ProvidedCHO</code>	
<code>dc:contributor</code>	<code>dcterms:isReplacedBy</code>
<code>dc:coverage</code>	<code>dcterms:isRequiredBy</code>
<code>dc:creator</code>	<code>dcterms:issued</code>
<code>dc:date</code>	<code>dcterms:isVersionOf</code>
<code>dc:description</code>	<code>dcterms:medium</code>
<code>dc:format</code>	<code>dcterms:provenance</code>
<code>dc:identifier</code>	<code>dcterms:references</code>
<code>dc:language</code>	<code>dcterms:replaces</code>
<code>dc:publisher</code>	<code>dcterms:requires</code>
<code>dc:relation</code>	<code>dcterms:spatial</code>

dc:rights	dcterms:tableOfContents
dc:source	dcterms:temporal
dc:subject	edm:currentLocation
dc:title	edm:hasMet
dc:type	edm:hasType
dcterms:alternative	edm:incorporates
dcterms:conformsTo	edm:isDerivativeOf
dcterms:created	edm:isNextInSequence
dcterms:extent	edm:isRelatedTo
dcterms:hasFormat	edm:isRepresentationOf
dcterms:hasPart	edm:isSimilarTo
dcterms:hasVersion	edm:isSuccessorOf
dcterms:isFormatOf	edm:realizes
dcterms:isPartOf	edm:type
dcterms:isReferencedBy	owl:sameAs
<b>Propriedades do ore:Aggregation</b>	
edm:aggregatedCHO	edm:object
edm:dataProvider	edm:provider
edm:hasView	dc:rights
edm:isShownAt	edm:right's
edm:isShownBy	edm:ugc
<b>Propriedades do edm:WebResource</b>	
dc:creator	dcterms:hasPart
dc:description	dcterms:isFormatOf
dc:format	dcterms:isPartOf

dc:rights	dcterms:issued
dc:source	edm:isNextInSequence
dcterms:conformsTo	edm:rights
dcterms:created	owl:sameAs
dcterms:extent	

Fonte: (THE EDM MAPPING GUIDELINES, 2014).

Estas três classes núcleo ou “Core” proporcionam que o modelo EDM realize ligações dos elementos do item real com sua descrição digital através das agregações.

As **Classes Contextuais** são úteis, visto que diversos fornecedores de conteúdo possuem os dados enriquecidos devido à utilização de arquivos de autoridade, vocabulários controlados e tesouros. A inclusão de recursos contextuais permite a exploração destes dados enriquecidos e mantém os dados separados do item de descrição. Por exemplo, um provedor poderia criar uma instância de uma classe edm:Agent e instanciá-la simplesmente inserindo o texto “William Shakespeare” como dc:creator, poderia fornecer o link (URI) de Shakespeare para um arquivo de autoridade e habilitar o uso dos dados relacionados enriquecidos.

A Europeana pode utilizar estas URI para localizar mais informações dos recursos externos, se estes estiverem disponíveis como dados linkados. Este sim, portanto, é o principal método para realizar o enriquecimento dos dados, pois adicionam detalhes que não estão presentes no provedor de conteúdo primário (THE EDM MAPPING GUIDELINES, 2014).

O **Quadro 10**, descreve as quatro classes contextuais do EDM e as suas propriedades internas.

**Quadro 10** – Propriedade das Classes Contextuais

Propriedades do edm:Agent	
skos:prefLabel	foaf:name
skos:altLabel	rdaGr2:biographicalInformation
skos:note	rdaGr2:dateOfBirth
dc:date	rdaGr2:dateOfDeath

dc:identifier	rdaGr2:dateOfEstablishment
dcterms:hasPart	rdaGr2:dateOfTermination
dcterms:isPartOf	rdaGr2:gender
edm:begin	rdaGr2:placeOfBirth
edm:end	rdaGr2:placeOfDeath
edm:hasMet	rdaGr2:professionOrOccupation
edm:isRelatedTo	owl:sameAs
<b>Propriedades do edm:Place</b>	
wgs84_pos:lat	skos:note
wgs84_pos:long	dcterms:hasPart
wgs84_pos:alt	dcterms:isPartOf
skos:prefLabel	edm:isNextInSequence
skos:altLabel	owl:sameAs
<b>Propriedades do edm:TimeSpan</b>	
skos:prefLabel	edm:begin
skos:altLabel	edm:end
skos:note	edm:isNextInSequence
dcterms:hasPart	owl:sameAs
dcterms:isPartOf	
<b>Propriedades do skos:Concept</b>	
skos:prefLabel	skos:relatedMatch
skos:altLabel	skos:exactMatch
skos:broader	skos:closeMatch
skos:narrower	skos:note
skos:related	skos:notation

skos:broadMatch	skos:inScheme
skos:narrowMatch	

**Fonte:** (THE EDM MAPPING GUIDELINES, 2014).

As classes contextuais não descrevem o item em si, mas possibilitam a conexões com outras classes e provedores de conteúdo ao utilizar os recursos de dados linkados, promovendo a interoperabilidade e o enriquecimento dos dados.

O *Simpleknowledge Organization System* (SKOS) desenvolve especificações e normas para apoiar o uso de sistemas de organização de conhecimento, como os tesouros, esquemas de classificação sistemas de cabeçalhos de assuntos e taxonomias dentro da estrutura da *Web Semântica*. O SKOS fornece um padrão para representar sistemas de organização do conhecimento codificando em RDF, o que possibilita sua leitura por máquinas, vinculação com outros dados, intercâmbio entre outros aplicativos e publicação na *Web* (W3C, 2009).

O SKOS é aplicado nas quatro classes contextuais, o que possibilita o intercâmbio entre outros provedores por meio das propriedades em comum, favorecendo a interoperabilidade e o enriquecimento dos dados.

O *Resource Description Access* (RDA) é um padrão de catalogação que fornece um conjunto de orientações e instruções sobre a descrição de recursos e de acesso que abrange todos os tipos de conteúdo e mídia. Está sendo desenvolvido para aplicação em bibliotecas, no entanto existe um movimento para que seja empregado em outros domínios (arquivos, museus, editoras, etc.) o que possibilitaria um alinhamento dos padrões RDA com as estruturas utilizadas nesses domínios (JOINT STEERING COMMITTEE, 2009).

O

**APÊNDICE A** possui um quadro com as classes núcleo e contextuais envolvidos na coleta de dados e seus respectivos atributos, com a descrição de suas funções.

Existem requisitos considerados fundamentais para o EDM representar os diferentes itens informacional recebidos das variadas comunidades que compõem a biblioteca Europeana:

- distinguir entre um item fornecido de uma representação digital;
- distinguir entre um item e descrição do registro de metadados;
- permitir a ingestão de múltiplos registros para um mesmo item, mesmo contendo declarações contraditórias entre eles;
- prover suporte para recursos contextuais, incluindo conceitos de vocabulários controlados (ISAAC; CLAYPHAN; HASLHOFER, 2012).

Para o EDM fornecer suporte aos vários modelos de itens distribuídos foi importante prover o modelo com mecanismos que pudessem distinguir as diversas perspectivas e representações desse item, especialmente as estruturas hierárquicas.

A aplicação do modelo da *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE), possibilitou a integração dos modelos de itens na questão das entidades e das relações a que os metadados se referem. O OAI-ORE desenvolve e promove padrões para descrições e troca de agregações nos recursos da *Web* e da interoperabilidade que auxiliam na eficiência da disseminação de conteúdo. Esses padrões fornecem a base para aplicações e serviços que podem visualizar, preservar, transferir, resumir e melhorar o acesso às agregações que os usuários utilizam em sua interação com a *Web*. O núcleo da estrutura de dados do EDM é baseado no modelo OAI-ORE (LAGOZE; SOMPEL, 2008a).

As agregações estão presentes na rotina das pessoas, ao colecionar fotos em álbuns, lerem jornais que são uma coleção de artigos ou gravar um disco com diversas músicas. Essa prática também se estende à *Web*, visto que é hábito armazenar vários endereços de *websites* na lista dos favoritos do navegador e mantê-los organizados por assuntos.

No entanto, a existência dessa agregação é muito instantânea, não há um padrão para identificar uma agregação. É corriqueiro empregar a URI de uma página de um agregado para identificar toda a agregação, porém esse URI identifica apenas

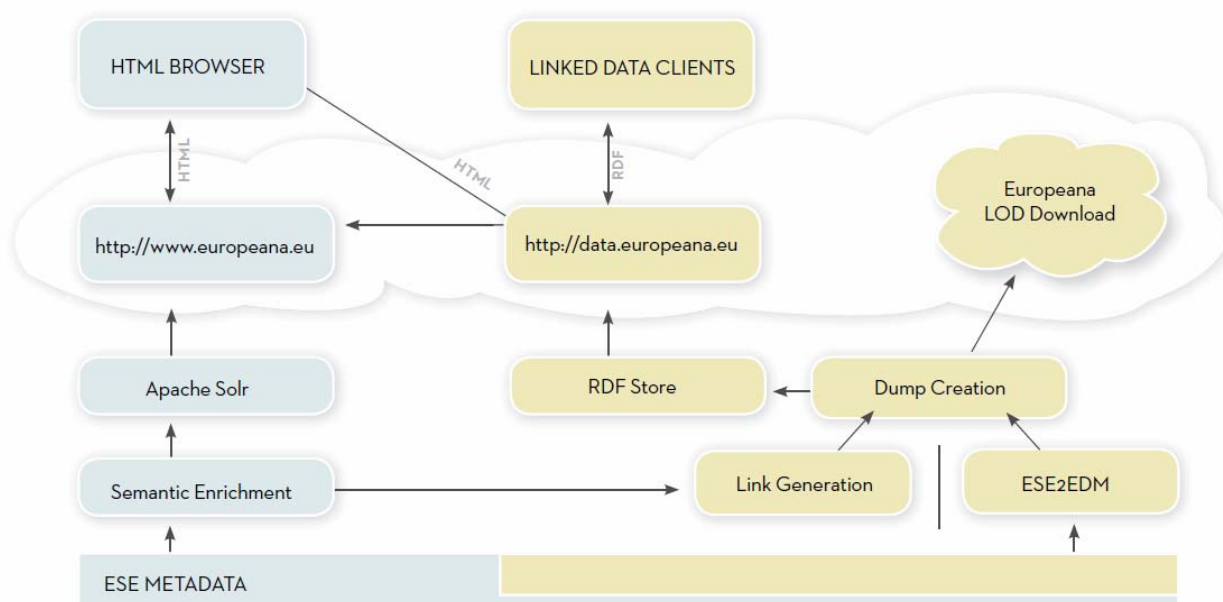
uma página do *Website*, e não o conjunto de páginas que compõem todo o documento ou de imagens. Essa é uma dificuldade, uma vez que não existe um padrão para descrever as páginas que constituem a agregação ou seus limites, e é esse o objetivo que o OAI-ORE visa a proporcionar (LAGOZE; SOMPEL, 2008b).

Agregações ORE são utilizadas para representar uma contribuição de um provedor de conteúdo para Europeana, o qual consiste no elemento fornecido juntamente com sua visão digital. Cada um deles pode vir com seu próprio metadados, usando um conjunto de propriedades DC e outro específico da Europeana (HASLHOFER; ISAAC, 2011).

O EDM propicia a representação de informações contextuais com as diferentes perspectivas de um dado item cultural, na forma de entidades, o que possibilita que vários elementos EDM não utilizem agrupamentos nos registros, como ocorria com o ESE, sendo possível identificar a que entidade se aplica determinada propriedade.

Preparar os metadados recebidos dos fornecedores e parceiros é um processo que envolve muitas etapas. Afigura 7 é um projeto piloto que possibilitará compreender como os conteúdos são trabalhados para que a Fundação Europeia possa disponibilizar seu material.

**Figura 7 – Arquitetura Técnica Europeia**



**Fonte:** (ISAAC; CLAYPHAN; HASLHOFER, 2012, p.36)

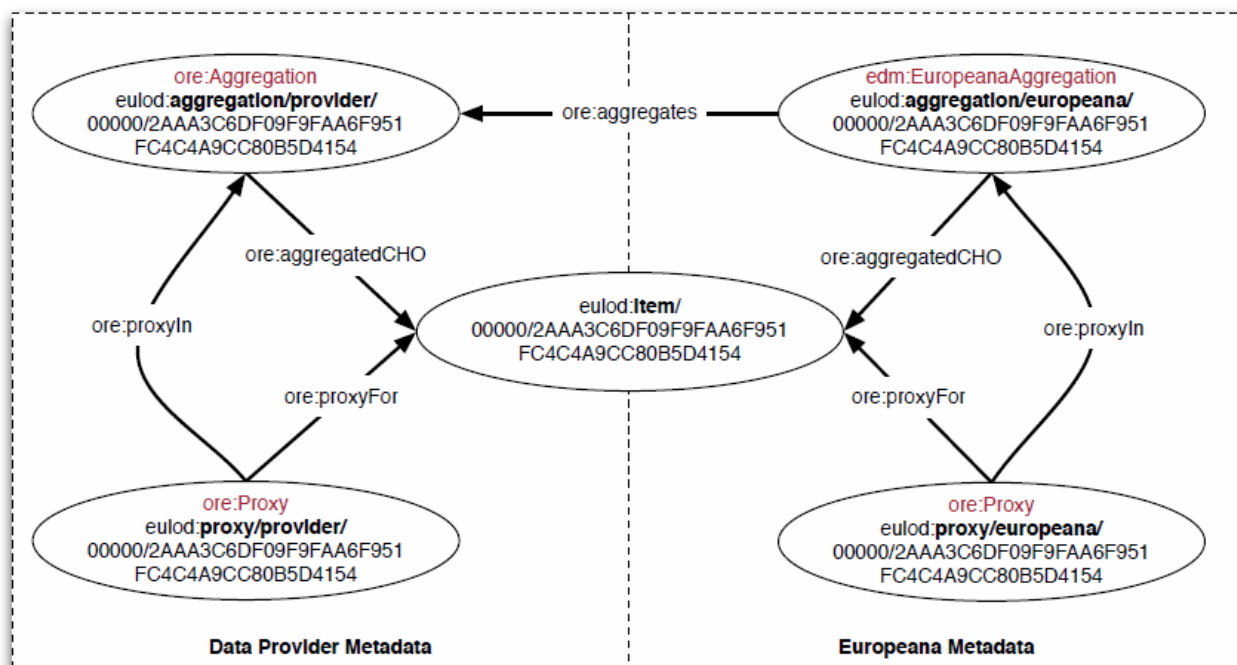
A primeira etapa consiste em extrair um subconjunto de metadados de origem ESE com base nos provedores que participam do projeto. Na segunda etapa,

este conjunto de metadados é convertido para o modelo de dados Europeana, o EDM, e atribuído URI's para os objetos resultantes (agregações, proxies, mapas de recursos). O mapeamento entre ESE e EDM, é implementado em XML e executado através do XSLT. Como resultado, obtém-se uma representação RDF/XML de metadados de cada provedor de dados. Por fim, na terceira etapa, os conteúdos são enriquecidos com anotações semânticas criadas no escritório da Europeana, através da ferramenta *Annocultor*<sup>15</sup> (HASLHOFER; ISAAC, 2011).

Este enriquecimento é obtido por meio da combinação de valores de metadados correspondentes. Por exemplo: *Dublin Core dc:spatial*, com os valores vindo de vocabulários controlados, os quais possuem quatro locais de referência: *Geonames* para lugares, *GEMET* para tópicos gerais, a ontologia *Semium Time* para os períodos de tempo e *DBpedia* para as pessoas (ISAAC; CLAYPHAN; HASLHOFER, 2012).

A estrutura do registro EDM é resultado da conversão dos registros ESE, conforme figura 8.

**Figura 8** – Estrutura Básica dos Recursos de Rede do EDM



**Fonte:** (HASLHOFER; ISAAC, 2011, p.99)

O recurso no centro representa o elemento do mundo real. Isto é, uma figura de pedra da *Cyprus Archaeological Museum*, que é descrito pelos metadados

<sup>15</sup> Software que converte banco de dados e arquivos XML para o RDF, e os dados semanticamente enriquecidos com links para vocabulários, para ser publicado como dados lincados na Web Semântica (LINKED HERITAGE, 2011).

vindos de um provedor de conteúdo e também pelos metadados específicos da Europeia. As descrições dos provedores de conteúdo, anexados em *ORE:Proxy*, estão agregados em um simples provedor *ORE:Aggregation*. Europeia associa essa agregação, bem como o específico da Europeia *ORE:Proxy resource*, para uma única estrutura *edm:EuropeanaAggregation*, a cada um destes itens, sempre é atribuído um URI (ISAAC; CLAYPHAN; HASLHOFER, 2012).

O mecanismo *ORE proxy* observado na Figura 8 permite diversas descrições para um mesmo item fornecido. É especialmente útil para rastrear metadados de diferentes provedores, ou seja, quando diferentes provedores de conteúdo contribuem com descrições distintas de um mesmo recurso, ou a Europeia realiza o enriquecimento semântico, criando assim metadados que diferem dos dados dos provedores originais.

Existe um *proxy* específico para cada descrição de um item, quando o *proxy* estiver relacionado a um recurso, utiliza-se a propriedade *ore:proxyFor* e quando o *proxy* estiver relacionado a uma agregação, utiliza-se *ore:proxyIn* (HASLHOFER; ISAAC, 2011).

A Figura 8 possibilita identificar as origens dos conteúdos que formam as classes dos metadados principais de cada item e suas transformações e enriquecimentos até atingir o objeto final para ser disponibilizada para consulta dos usuários.

Em geral, haverá um conjunto de metadados para os dados dos provedores descritivos de um item e outro para os criados pela Europeia (PATRICIO, 2012).

- Proxy do fornecedor de conteúdo: classe que separa o “item” das declarações descritivas ESE realizadas pelo fornecedor de dados. Esta separação possibilita distinguir diversas descrições de um item.
- Proxy da Europeia: classe que separa o “item” das declarações descritivas ESE realizadas pela Europeia. Este proxy possui declarações que permitem conexões com lugares, pessoas, conceitos, banco de dados externos.

A mudança do modelo ESE para EDM foi uma evolução necessária para que a biblioteca Europeia disponibilizasse seus conteúdos utilizando os recursos e tecnologias da *Web Semântica*. O ESE, mesmo obsoleto, não foi descartado, visto

que seus metadados ainda são agregados e posteriormente convertidos em EDM, no intuito de promover a compatibilidade com os vários provedores de conteúdo. A organização dos dados capturados dos provedores em classes, possibilita que os metadados sejam classificados de acordo com sua funcionalidade, já que as classes núcleo e contextual desmembram os metadados para classes que descrevem o objeto e classes que realizam a interoperabilidade.

A aplicação do modelo OAI-ORE proporcionou a padronização dos modelos de objetos e nas relações entre os metadados, que favoreceu a interoperabilidade, o núcleo da estrutura de dados do EDM é baseado no modelo OAI-ORE.

### 3.3.1.2 Harvesting, a coleta de dados do portal Europeana

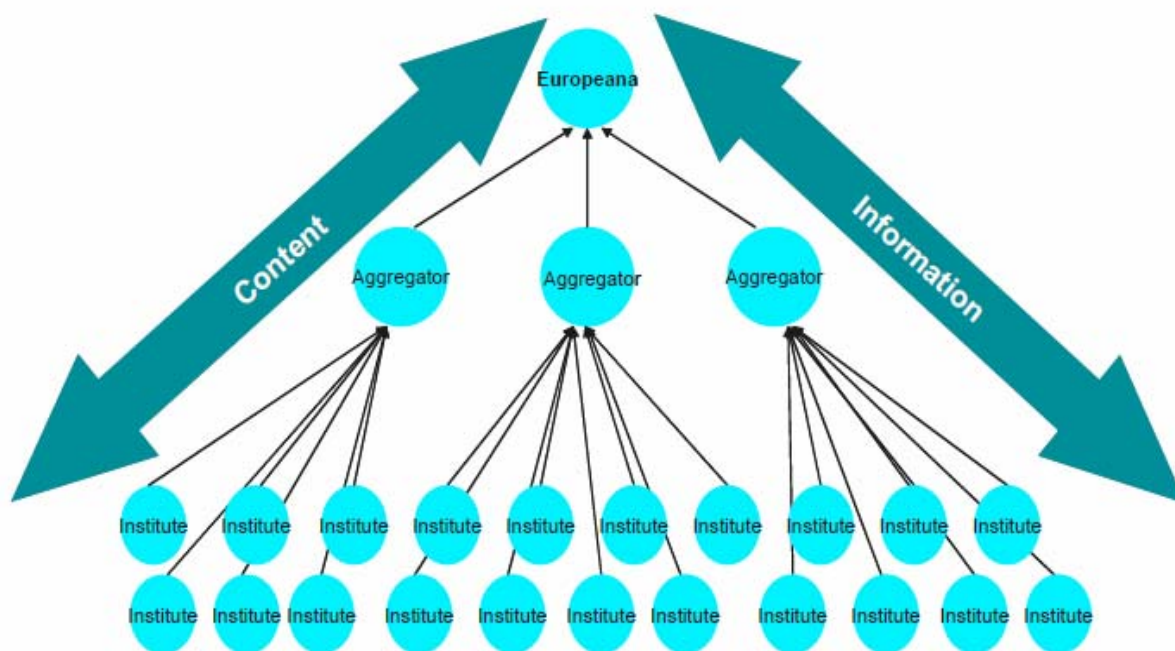
A estratégia de coleta de dados empregada pela Europeana consiste em receber os dados das entidades agregadoras e provedoras de conteúdo organizados e padronizados, de acordo com os critérios determinados com o modelo adotado e aceitos pelos parceiros. O termo *Harvesting* de acordo com Pedrosa et al. (2010, p.7) “[...] é um processo de recolher um conjunto de dados a partir de um provedor”.

A Europeana somente recolhe e indexa os metadados coletados, enquanto o item informacional permanece na instituição de origem. No entanto, a quantidade, o tipo de conteúdo, formato de saída e a documentação podem variar significativamente entre os vários provedores de conteúdo. Esta diversidade é uma das grandes dificuldades da Europeana, visto que atender individualmente cada provedor não seria o procedimento apropriado. A solução adotada para minimizar esta questão foi a introdução de um elemento intermediário entre a Europeana e os provedores de conteúdo, denominado de Agregador (EUROPEANA AGGREGATORS, 2011).

Um agregador no contexto da Europeana é uma organização que coleta os metadados de um grupo de provedores de conteúdo e transmite-os para a Europeana. Agregadores reúnem material de organizações individuais, padronizam o formato dos arquivos e metadados e direciona-os de acordos com os padrões e

procedimentos Europeana (EUROPEANA AGGREGATORS, 2011). A **Figura 9** ilustra a posição ocupada pelos agregadores.

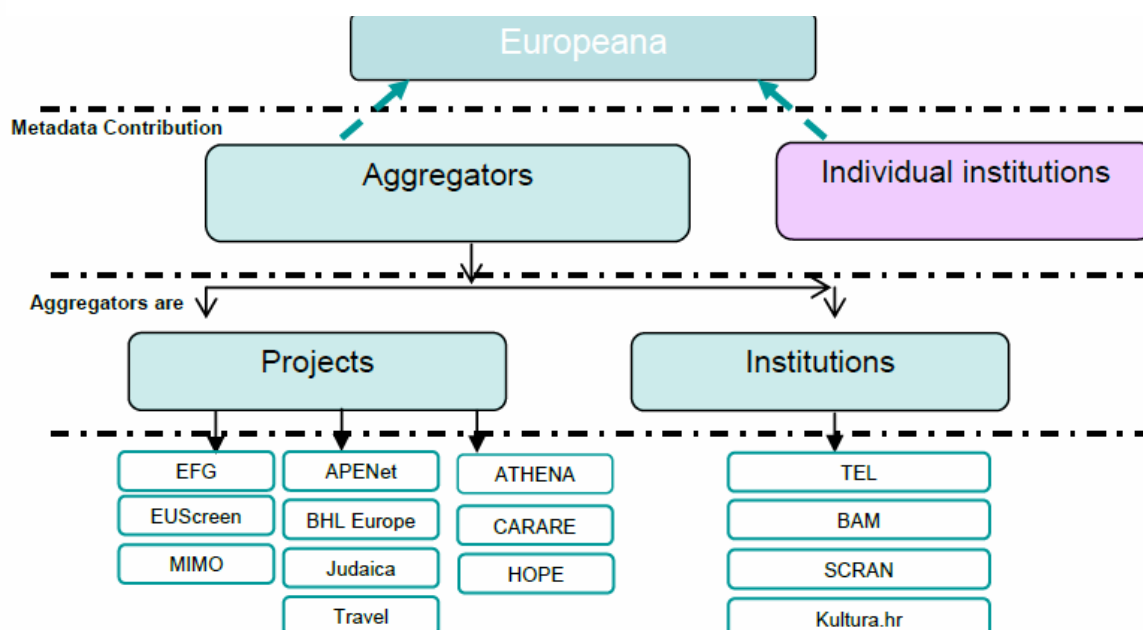
**Figura 9** – Agregadores no modelo Europeana



**Fonte:** EUROPEANA AGGREGATORS (2010, p.3)

Os agregadores são, portanto, organizações que reúnem vários provedores de conteúdo, os quais possuem dados a serem exportados para o Portal Europeana. Os agregadores e provedores de conteúdo podem ser *websites* que possuem acesso público e compartilham seus materiais. Procedimentos para coleta de dados Europeana realiza o *harvesting* com o auxílio dos agregadores, estes importam os dados dos provedores de conteúdo que os disponibilizou e preparou conforme os requisitos técnicos para submissão de dados. Qualquer instituição cultural Europeia que possuir coleções digitalizadas pode prover conteúdos para Europeana. A **Figura 10** ilustra os atores e suas relações.

**Figura 10** – Organizações envolvidas no Harvesting



**Fonte:** EUROPEANA AGGREGATORS (2010, p.7)

Os *Projects* apresentados na **Figura 10** referem-se aos projetos liderados pela Fundação Europeia que envolvem parceiros na produção de novas tecnologias ou conteúdos para a expansão de coleções do patrimônio histórico cultural. Dentre os vários projetos pode-se citar o *Europeana Film Gateway* (EFG), que envolveu vinte e dois parceiros de dezesseis países europeus os quais desenvolveram um portal provendo o acesso a seiscentos mil itens informacionais, entre filmes, fotos, pôsteres, desenhos e documentos (EFG, 2011).

As *Institutions* são instituições que possuem portais na *Web* utilizados como repositórios de coleções em bibliotecas, arquivos e museus. Como exemplo, pode-se mencionar a Instituição *The European Library* (TEL) que é uma biblioteca sem fins lucrativos com a missão de tornar-se um distribuidor de dados abertos, ao compartilhar recursos digitais e registros bibliográficos fortalecendo bibliotecas de todo o continente europeu (TEL, 2013).

A biblioteca virtual Europeia reúne em torno de trinta milhões de itens informacionais em todos os projetos da fundação. A estratégia adotada para receber estes itens possibilitou que todos estes conteúdos estejam organizados e padronizados dentro dos critérios técnicos estabelecidos pela fundação.

Para compreender o processo de submissão dos itens, a Figura 11 ilustra as etapas e a atuação de cada ator a partir do acordo dos termos adotados para as normas e padrões técnicos de envio, para ingestão dos dados pela Europeana.

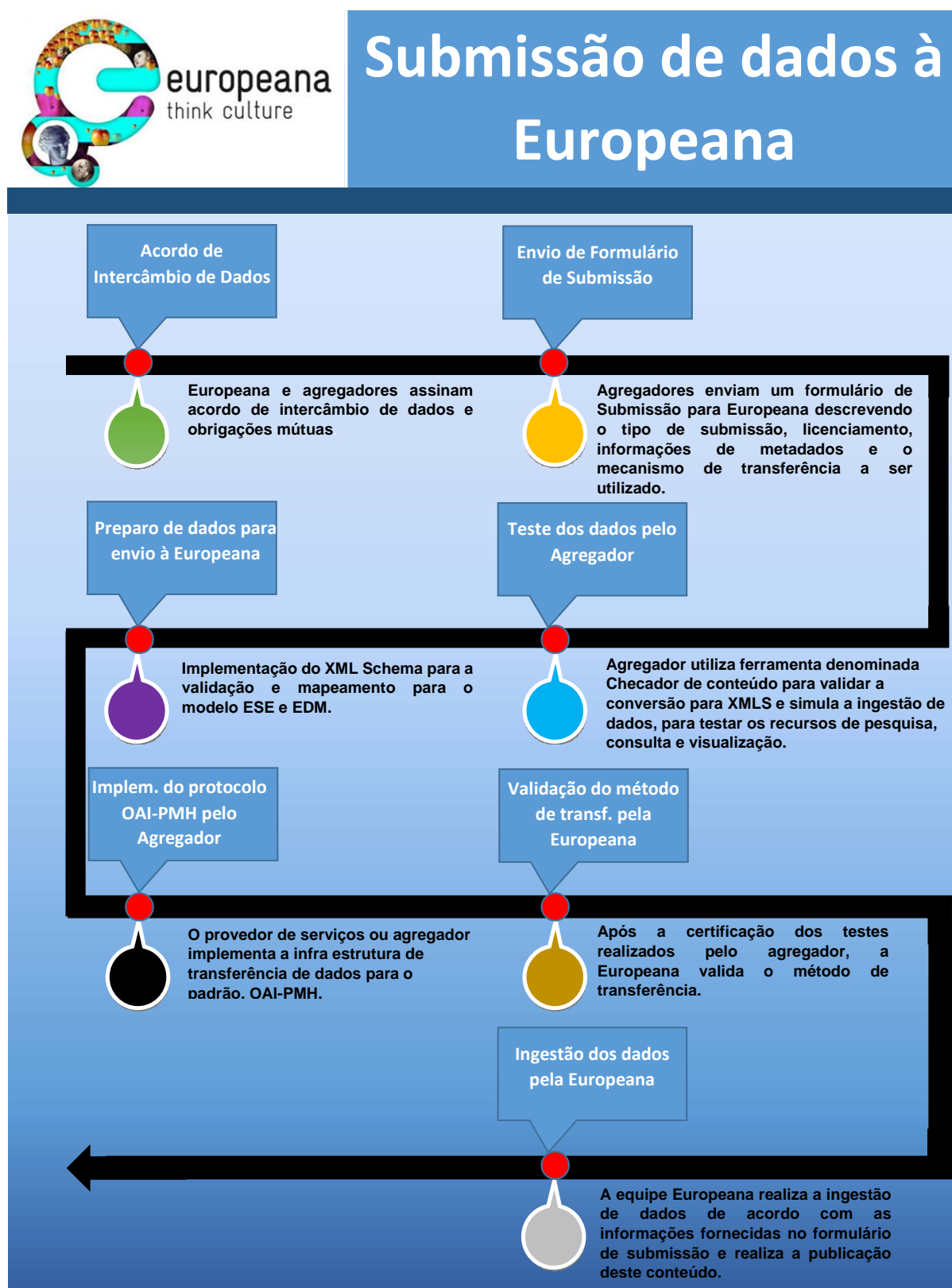
Na Figura 11 é possível visualizar como os agregadores possuem um papel de extrema importância no processo de submissão dos itens, pois são responsáveis por realizar validações e disponibilizar os dados para a Europeana, no entanto existem outras operações realizadas pelos agregadores:

- Avaliar os metadados recebidos dos provedores de conteúdo para que estejam de acordo com o modelo EDM e que foi desenvolvido para esse propósito.
- Gerar um link persistente (URI's) do item para possibilitar à Europeana acessá-lo e disponibilizar no portal.
- Gerar um identificador único e persistente por registro, de acordo com os metadados, utilizando um número do catálogo ou o `dc:identifier` (EUROPEANA AGGREGATORS, 2011).

O portal Europeana exibe imagens e metadados dos itens e seus relacionamentos, e fornece um link de acesso ao *website* onde estão armazenados o conteúdo original do item informacional e podem ser encontrados mais informações a respeito deste item específico.

A Europeana realiza o *harvesting*, armazena e indexa os metadados em um índice central, procedimento elementar para o enriquecimento dos dados e a exibição de resultados de forma integrada (EUROPEANA AGGREGATORS, 2011). O método de *harvesting* adotado é o *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), definido pela OAI como um protocolo que define um mecanismo de registro de coleta de metadados em provedores de conteúdo ou agregadores. O OAI-PMH é considerado uma opção técnica simples para os provedores de conteúdo em tornar seus metadados disponíveis para os agregadores, com base nos padrões aberto *Hypertext Transport Protocol* (HTTP) e *XML* (OPEN ARCHIVES, 2003).

Figura 11 – Processo para submissão de dados à Europeana



Fonte: EUROPEANA AGGREGATORS (2010, p.10) (Adaptado pelo autor)

O OAI-PMH foi construído para suportar a coleta de grandes quantidades de dados, com o uso de “fichas de reinício”, que são essencialmente os valores adotados para paginar o número de registros de metadados recuperados, visto que o coletor recebe os dados em blocos, não em fluxo controlado, o que pode ocasionar uma falha de transmissão. Nessa situação, a coleta continua ativa, reiniciando a partir da falha (TZANAKIS, 2015).

Com uma estrutura simples comparada a outros protocolos, o OAI-PMH permite aos participantes do projeto OAI compartilhar seus metadados, para outros provedores de conteúdo, agregadores ou aplicações externas que se interessam na coleta de dados. O protocolo possui duas propriedades importantes que devem ser consideradas (OLIVEIRA; CARVALHO, 2009):

- Interoperabilidade: característica que possibilita aos metadados serem visíveis e permita conexões a outras estruturas de dados, adota o Dublin Core como padrão, o que possibilita que outros repositórios OAI possam trocar metadados.
- Extensibilidade: característica que possibilita a criação ou utilização de outros metadados diferentes do Dublin Core, para satisfazer necessidades especiais e manter a interoperabilidade entre os repositórios.

Para auxiliar os provedores de conteúdo e agregadores a gerenciar e coletar os metadados, a *EuropaConnect*, em parceria com o Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC-ID), produziu uma solução que resultou no software *REPOX*. *REPOX* é uma infraestrutura para armazenar, preservar e gerenciar conjuntos de metadados de maneira transparente e independente do esquema e formato. Para o gerenciamento dos provedores de conteúdo OAI-PMH o software *REPOX* oferece as seguintes funcionalidades:

- o registro dos agregadores, dos provedores de conteúdo, suas descrições de coleta e as configurações para a coleta dos metadados;
- coleta manual ou automática das coleções por OAI-PMH;
- monitoramento da qualidade de serviço dos servidores OAI-PMH, incluindo relatórios estatísticos (PEDROSA et al., 2010).

No âmbito de gerenciamento dos registros de metadados coletados, oferece as seguintes funcionalidades:

- suporte a múltiplos formatos de metadados;
- um serviço de repositório de metadados para tornar os registros de metadados coletados disponíveis para a Europeia;
- solução em escalabilidade, capaz de manter um grande número de agregadores, provedores de dados OAI-PMH com um número ilimitado de registros (PEDROSA et al., 2010).

Os agregadores e provedores de conteúdo organizam os metadados com o apoio do software *REPOX*. Esta ferramenta de interface gráfica possibilita aos parceiros da Europeia realizar a submissão dos metadados, por meio do protocolo OAI-PMH ou por meio de uma pasta no serviço de arquivo no formato ISO2709<sup>16</sup>, MarcXML<sup>17</sup>, MarcXchange<sup>18</sup>, ESE, EDM ou qualquer formato XML (PEDROSA et al., 2010).

O *harvesting*, ou coleta de dados realizado pela Europeia, conta com um parceiro intermediário que media o preparo e a submissão dos metadados por parte dos provedores de conteúdo, conforme as normas e validações do modelo empregado. O agregador, organização que media o processo de exportação de metadados para a Europeia, pode tratar-se de um repositório, com *website*, ou portal com acesso público ou não.

O protocolo OAI-PMH empregado pela Europeia, como o método para coleta dos metadados, permite a interoperabilidade entre seus parceiros e tem se consolidado entre os países em desenvolvimento como um padrão que proporciona a visibilidade e a integração de informações com custos acessíveis.

O software *REPOX* foi desenvolvido para armazenar, preservar e gerenciar conjuntos de metadados coletados dos provedores de conteúdo. Trata-se

---

<sup>16</sup> Norma que especifica os requisitos de formato para o intercâmbio de registros bibliográficos, que descreve diversos tipos de documentos quanto à descrição bibliográfica. É uma estrutura generalizada projetada especialmente para a comunicação entre sistemas de processamento de dados (ISO 2709, 2011).

<sup>17</sup> Estrutura desenvolvida pela biblioteca do Congresso e o padrão MARC para trabalhar com o MARC no ambiente XML. Essa estrutura tem como propósito ser extensível e flexível para permitir aos usuários trabalhar com os dados MARK de diversas maneiras, com o emprego de vários componentes como os Schemas, Folhas de Estilos e ferramentas de software (MARCXML, 2014).

<sup>18</sup> É um padrão internacional que fornece uma ferramenta para intercâmbio dos registros MARC em XML como um suplemento para o intercâmbio dos registros MARC com a ISO 2709. Descreve uma estrutura geral, um modelo primário desenhado para a comunicação entre sistemas de processamento de dados (ISO/DIS 25577, 2006).

de uma ferramenta que possibilita realizar o intercâmbio de metadados de maneira transparente, entre todas as entidades envolvidas no *harvesting* da Europeia, independente dos seus esquemas ou formatos.

### 3.3.1.3 Linguagem de desenvolvimento e banco de dados

O portal Europeia abriga uma interface de busca na *Web* para a localização dos itens informacionais dos provedores de conteúdo parceiros da Fundação. A construção do portal foi complexa e para sua concepção foram empregados somente software *Open Source* ou código aberto.

A *Open Source Initiative* (OSI) é uma entidade sem fins lucrativos que promove e protege softwares de código aberto. Define-se software de código aberto, o software que pode ser usado livremente, modificado, compartilhado por qualquer pessoa, pode ser desenvolvido por qualquer pessoa e distribuído sob licença da OSI e deve apresentar as seguintes características (OPEN SOURCE INITIATIVE, 2006):

- Livre distribuição: a licença não deve restringir qualquer parte do software, como seu desempenho e não deve exigir nenhum valor ou taxa de venda;
- Código fonte: o programa deve incluir o código fonte e sua distribuição, sem qualquer forma deliberada de dificultar seu entendimento ou emprego;
- Trabalho derivado: A licença deve permitir modificações e trabalhos derivados, que permita a distribuição sob os mesmos termos do software original;
- Integridade do código fonte do autor: a licença deve permitir explicitamente a distribuição do software construído a partir do código fonte modificado, podendo ter nomes ou número de versão diferentes do original;
- Não à discriminação contra pessoas ou grupos: a licença não pode discriminar nenhuma pessoa ou grupo de pessoas;
- Não à discriminação contra campos de atuação: não deve restringir o uso do programa para qualquer campo específico de atuação, como por exemplo para desenvolvimento genético ou em negócios comerciais;
- Distribuição de licença: os direitos devem aplicar a todos para quem o programa é redistribuído, sem necessidade de licença adicional;
- Licença não pode ser específica para um produto: todo o programa ou parte dele que for redistribuído deve ter os mesmos direitos do original;
- Licença não deve restringir outro software: a licença não deve colocar restrições em outro software que é distribuído juntamente com o software licenciado;

- Licença de ser tecnologicamente neutro: nenhuma disposição da licença pode ser baseada em qualquer tecnologia ou estilo de interface.(OPEN SOURCE INITIATIVE, 2006; tradução nossa):

A preferência pelo código aberto no desenvolvimento de todos os projetos da fundação Europeia é coerente, visto que o propósito do portal é transformá-lo em uma referência multicultural e facilitar o acesso e a utilização do diversificado patrimônio cultural e científico da Europa. O software aberto possibilita a transparência de todo o processo de desenvolvimento, uma vez que o código fonte poderá ficar disponível para os utilizadores que tenham interesse em compreender em detalhes o mecanismo de funcionamento do portal.

Um *site* como o portal da biblioteca virtual Europeia pode utilizar diversas linguagens de programação ou de marcação para seu desenvolvimento, visto que existem softwares específicos para determinadas etapas da produção. A linguagem de programação de código aberto adotado pelo portal para o desenvolvimento foi o Java versão 1.7.0\_55 e a 1.7.0\_65 e o Banco de dados empregado foi o PostgreSQL v.9.3 e o MongoDB v.2.4.9 (GITHUB, 2014).

Java é uma linguagem de programação multiplataforma que possibilita desenvolver softwares corporativos, aplicativos móveis e conteúdo baseado na *Web* que podem ser executados em *datacenters*, supercomputadores científicos, computadores pessoais, console de games, *smartvs*, *tablets* ou *smartphones* (JAVA, 2015).

O PostgreSQL é um sistema gerenciador de banco de dados de código aberto. É compatível com as plataformas de sistemas operacionais mais conhecidas como *GNU/Linux*, *Unix(AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris e Tru64)* e *MS Windows*. Possui interface nativa de programação para as linguagens *C/C++*, *Java*, *.Net*, *Perl*, *Phyton*, *Ruby*, *TCL*, *ODBC*, entre outros (POSTGRESQL, 2015). O banco de dados armazena os nomes dos usuários, senhas e chaves de API usados na aplicação “*My Europeia*” (GITHUB, 2015).

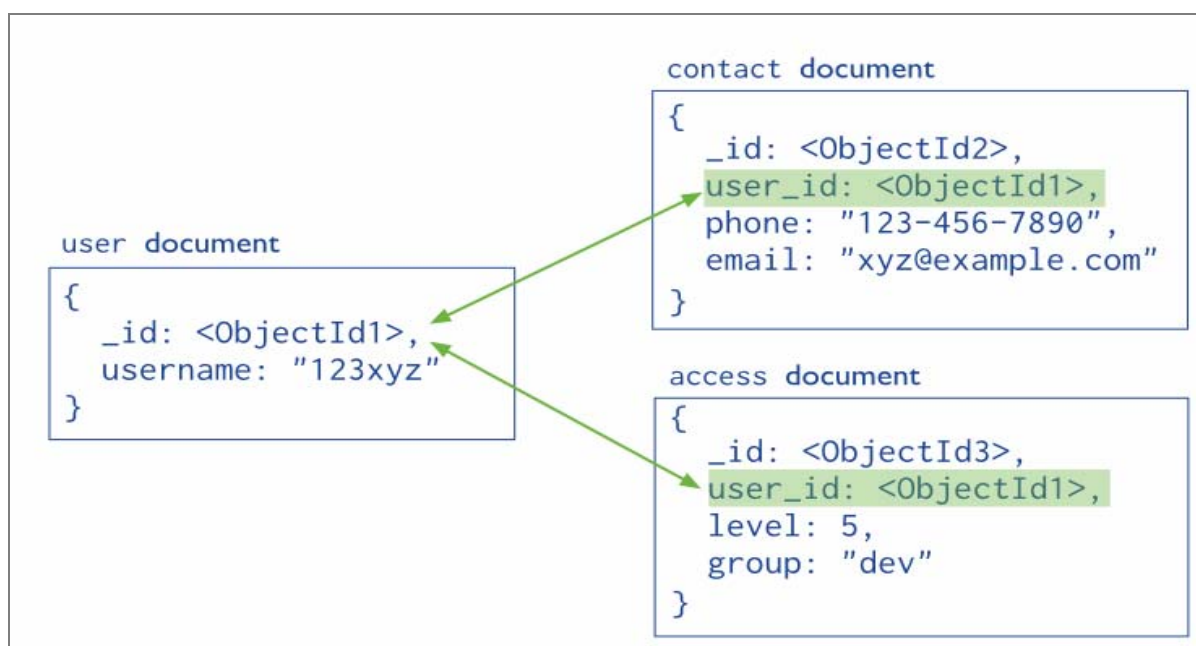
MongoDB é um banco de dados orientado a documentos de código aberto, desenvolvido na linguagem C++. É um documento que se constitui numa estrutura de dados composta por campos e pares de valores (MONGODB, 2015a). Os valores dos campos podem incluir outros documentos, matrizes e matrizes de documentos. Ao contrário de um banco de dados SQL, o qual deve-se determinar e declarar o esquema de uma tabela antes de inserir os dados, as coleções MongoDB

não forçam a estrutura do documento. Esta flexibilidade facilita o mapeamento de documentos a uma entidade ou um objeto (MONGODB, 2015b).

A estrutura do documento e a maneira como o aplicativo representa as relações entre os dados são os pontos chave para a concepção do modelo de dados para as aplicações MongoDB. Existem duas ferramentas que permitem aos aplicativos representarem esses relacionamentos: as Referenciadas e os Documentos Incorporados (MONGODB, 2015b):

- **Referenciadas:** armazena os relacionamentos entre dados com a inclusão de links ou referências de um documento para outro. Aplicações podem entender essas referências acessando os dados relacionados.(MONGODB, 2015b, tradução nossa).

**Figura 12 - Referências**



Fonte: (MONGODB, 2015b)

A representação do tipo referência, ilustrada na Figura 12, possibilita compreender que dessa maneira, os dados de um documento se relacionam através de links ou referências com outros documentos.

- **Documentos Incorporados:** capturam relacionamentos entre dados por meio dos dados relacionados armazenados em um documento de estrutura simples. Os documentos MongoDB torna isto possível ao incorporar estruturas de documentos em um campo ou matriz. Esses modelos de dados permitem que os aplicativos possam recuperar e manipular dados relacionados em única operação de banco de dados.(MONGODB, 2015b, tradução nossa).

**Figura 13** – Documentos Incorporados

**Fonte:** (MONGODB, 2015b) adaptado pelo autor.

A representação Documentos Incorporados, ilustrada na Figura 13, possibilita compreender que aplicativos armazenam partes de um documento em um mesmo registro de banco de dados, o que resulta em menos consultas e atualizações para concluir as operações mais comuns.

O banco de dados MongoDB é utilizado para armazenar os metadados dos registros (modelo ESE/EDM) como também os vocabulários e os eventos de API, que requerem três bases de dados: *europæana*, *vocabulary* e *apilog*(GITHUB, 2015).

O portal Europeana corrobora os princípios da licença gratuita e ideologia ao empregar software aberto, visto que o propósito de todo o projeto Europeana é promover um registro cultural e científico do patrimônio cultural Europeu acessível a todos. A acessibilidade estende na transparência em todo o processo de desenvolvimento, uma vez que todo o código fonte utilizado está disponível para compartilhamento e apropriação de conhecimento. Da mesma maneira, as classes e propriedades dos bancos de dados e as etapas para sua configuração.

### 3.3.1.4 Linguagem de busca

O banco de dados do portal Europeana possui em torno de 30 milhões de registros com conteúdo organizados em documentos textos, imagens, vídeos e sons. Os dados são representados pela linguagem de descrições RDF e

estruturados segundo o modelo EDM, armazenados em documentos no formato XML. Considerando que os dados armazenados tendem a crescer cada vez mais, faz-se necessário conceber mecanismos de busca capazes de recuperar informações de maneira ágil, relevante para o usuário e em um ambiente amigável.

O portal Europeia possibilita a pesquisa de conteúdo por meio de dois mecanismos de busca empregados de maneira e situações distintas: o SOLR que é uma solução código aberto consolidada e preparada para ser incorporada a estrutura do site e o SPARQL, uma linguagem de busca específica para pesquisa em ambientes semânticos (EUROPEANA LABS, 2015b; EUROPEANA LABS, 2015a).

O SOLR é um software de indexação e busca, desenvolvido na linguagem Java pela *Apache Software Foundation* que oferece diversas ferramentas para auxiliar o usuário a realizar consultas em diversas línguas, possibilitando desde consultas com palavras-chave, frases e caracteres curinga (SOLR, 2014).

O cenário para consulta pode ser apresentado da seguinte maneira: o usuário define o idioma no qual o portal Europeia deva ser exibido, após a expressão ou frase ser digitada no campo de busca, o software de busca analisa a expressão (verifica se é uma palavra ou frase e o idioma), realiza algumas transformações (localiza sinônimos por meio de um arquivo interno, ou descrições por meio de bibliotecas externas), e cria uma representação interna própria (ajusta para um padrão sintático por meio do agrupamento dos termos) .

SPARQL é um conjunto de especificações para consultas e manipulações de conteúdo em forma de grafos RDF na *Web* ou em um banco de dados local e possui características específicas de linguagem e protocolos que diferem de outros software similares (SPARQL, 2013):

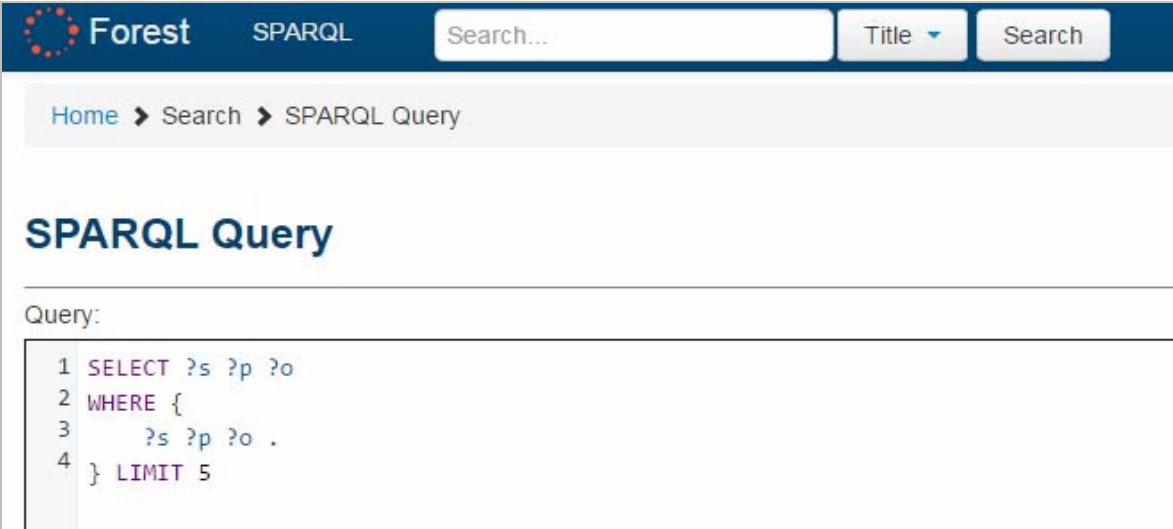
- uma linguagem de consulta próprio para o RDF;
- uma especificação que determina a semântica de consultas SPARQL, empregando RDF Schema ou OWL;
- um protocolo que define os meios para transmissão de consultas SPARQL para o serviço que o consultou;
- uma especificação que define um método de busca e um vocabulário para descrever os serviços SPARQL.

O portal Europeana não utiliza o SPARQL efetivamente em suas buscas, no entanto, disponibiliza o link <<http://europeana.ontotext.com/sparql>> para que os usuários tenham acesso aos metadados do banco de dados Europeana e realizem consultas empregando as especificações da linguagem.

A pesquisa com SPARQL por meio do link, exige conhecimentos dos comandos da linguagem, visto que a consulta é realizada em um ambiente específico no qual é preciso digitar os comandos de acordo com a sintaxe da linguagem. A seguir, apresenta-se exemplos de como o SPARQL possibilita recuperar conteúdo.

Para o SPARQL localizar e recuperar conteúdo em um ambiente semântico é preciso utilizar declarações RDF que são compostos dos três objetos básicos: recurso, propriedade e valor. Na Figura 14, foi criada uma consulta para retornar cinco itens aleatórios que contenham estes três objetos.

**Figura 14** – Código em SPARQL para retornar cinco itens



The screenshot shows the Forest SPARQL interface. At the top, there is a navigation bar with the Forest logo, the text 'Forest SPARQL', a search input field, a 'Title' dropdown menu, and a 'Search' button. Below the navigation bar, there is a breadcrumb trail: 'Home > Search > SPARQL Query'. The main heading is 'SPARQL Query'. Underneath, there is a 'Query:' label and a text area containing the following SPARQL query:

```
1 SELECT ?s ?p ?o
2 WHERE {
3     ?s ?p ?o .
4 } LIMIT 5
```

**Fonte:** (FOREST, 2014)

A cláusula `SELECT` define quais dados serão exibidos após a consulta, que utiliza as variáveis `?s`, `?p` e `?o`, identificadas pelo símbolo de “?”. A cláusula `WHERE` define os dados a serem recuperados no grafo RDF e “`LIMIT`” define a quantidade de itens que deverão ser exibidos no resultado, conforme a Figura 15.

**Figura 15** – Resultado da consulta de cinco itens

s	p	o
rdf:type	rdf:type	rdf:Property
rdfs:subPropertyOf	rdf:type	rdf:Property
rdfs:subClassOf	rdf:type	rdf:Property
rdfs:domain	rdf:type	rdf:Property
rdfs:range	rdf:type	rdf:Property

**Fonte:** (FOREST, 2014)

O resultado apresentado na Figura 15 exibe o conteúdo existente nos grafos RDF para as variáveis selecionadas.

Para compreender o processo de consulta e a correta sintaxe da linguagem SPARQL é relevante identificar alguns componentes básicos da estrutura da linguagem, que é constituído por (SPARQL, 2013):

- PREFIX: declaração de prefixos, declara o esquema usado na consulta e abreviar URI's;
- FROM: identificação das fontes que devem ser consultadas;
- SELECT: define os resultados que deverão ser exibidos;
- WHERE: determina o que será consultado na base RDF;
- FILTER: filtra os registros de acordo com as definições.

**Figura 16** – Código em SPARQL para listar três registros em que o idioma seja “Inglês” e Assunto “Law”

```

1 # Listar cinco registros em que o idioma seja Inglês e assunto igual a "law"
2 PREFIX edm: <http://www.europeana.eu/schemas/edm/>
3 PREFIX dc: <http://purl.org/dc/elements/1.1/>
4 SELECT ?Titulo ?Descricao ?Data ?Assunto ?Idioma ?Url
5 WHERE {
6     ?Europeana dc:title ?Titulo .
7     ?Agregador edm:isShownBy ?Url .
8     ?Europeana dc:description ?Descricao .
9     ?Europeana dc:date ?Data .
10    ?Europeana dc:subject ?Assunto .
11    ?Europeana dc:language ?Idioma .
12    FILTER(?Assunto = 'law' && ?Idioma = 'en')}
13 LIMIT 3
14

```

**PREFIX para abreviar URI's**

**SELECT define o que exibir**

**WHERE define o que será consultado**

**FILTER define o conteúdo da busca**

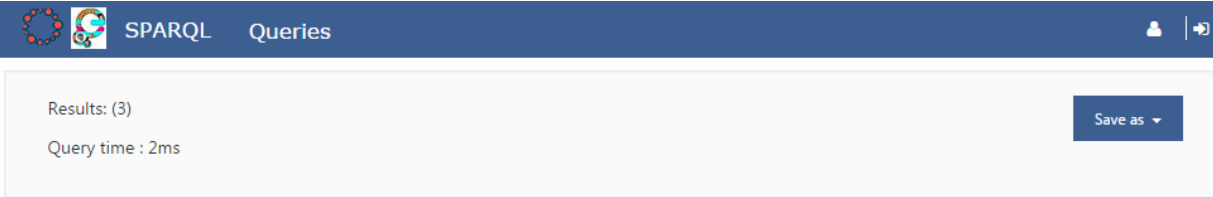
**Fonte:** próprio autor.

Os códigos exibidos na Figura 16 foram digitados em um ambiente próprio da Europeana para consultas em SPARQL para produzir um resultado que exiba somente três registros em que o assunto seja “law” e o idioma registrado no metadados `language` representado pela variável `Idioma`, seja “en”, *English*. Importante destacar que a linguagem é sensível a caracteres maiúsculos e minúsculos.

Para a realização da consulta visualizada na Figura 16, foram empregadas as classes de metadados EDM e DC, conforme as linhas 2 e 3. Na cláusula `SELECT`, da linha 4, foram determinados os metadados que serão visualizados na coluna, na forma de variáveis. Com a cláusula `WHERE`, emprega-se as triplas para a realização da consulta no formato recurso, propriedade e valor, visualizada nas linhas 6 a 11. A cláusula `FILTER` foi incluída para que somente os registros que possuam o idioma Português fossem selecionados. Por fim, para evitar um resultado com uma quantidade desnecessária, limitou-se o total a ser exibida em 3 registros, com a cláusula `LIMIT`.

O resultado da consulta pode ser visualizado na Figura 17, com o conteúdo das variáveis temporárias Título, Descrição, Data, Assunto, Idioma e URL origem sendo exibidos.

**Figura 17** – Resultado da consulta para exibir registros com idioma em português



The screenshot shows a SPARQL query interface with a dark blue header containing the SPARQL logo and the text 'SPARQL Queries'. Below the header, it indicates 'Results: (3)' and 'Query time : 2ms'. A 'Save as' button is visible on the right. The main content is a table with six columns: Título, Descrição, Data, Assunto, Idioma, and Url. The table contains three rows of results, all with the same title and description, but different URLs.

Título	Descrição	Data	Assunto	Idioma	Url
360 Degrees of Segregation: A Historical Perspective Of Segregation-era School Equalization Programs In the Southern United States	en	Amsterdam Law Forum; issue.vol 2, issue.no 3 (2010): Future of Legal Education	law	en	<a href="http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/2919.jpg">http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/2919.jpg</a>
360 Degrees of Segregation: A Historical Perspective Of Segregation-era School Equalization Programs In the Southern United States	en	Amsterdam Law Forum; issue.vol 2, issue.no 3 (2010): Future of Legal Education	law	en	<a href="http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/2809.jpg">http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/2809.jpg</a>
360 Degrees of Segregation: A Historical Perspective Of Segregation-era School Equalization Programs In the Southern United States	en	Amsterdam Law Forum; issue.vol 2, issue.no 3 (2010): Future of Legal Education	law	en	<a href="http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/2739.jpg">http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/2739.jpg</a>

**Fonte:** (FOREST, 2014)

Nas linhas a seguir, o código em XML gerado pelo *website* da Europeana, na URL <http://europeana.ontotext.com/> é exibido em sua forma original.

```
<?xml version='1.0' encoding='UTF-8'?>
<sparql xmlns='http://www.w3.org/2005/sparql-results#'>
  <head>
    <variable name='Título' />
    <variable name='Descrição' />
    <variable name='Data' />
    <variable name='Assunto' />
    <variable name='Idioma' />
    <variable name='Url' />
  </head>
  <results>
    <result>
      <binding name='Título'>
        <literal>360 Degrees of Segregation: A
Historical Perspective Of Segregation-era School Equalization
Programs In the Southern United States</literal>
      </binding>
      <binding name='Descrição'>
```

```

        <literal>en</literal>
    </binding>
    <binding name='Data'>
        <literal>Amsterdam Law Forum; issue.vol 2,
issue.no 3 (2010): Future of Legal Education</literal>
    </binding>
    <binding name='Assunto'>
        <literal>law</literal>
    </binding>
    <binding name='Idioma'>
        <literal>en</literal>
    </binding>
    <binding name='Url'>
        <uri>http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/291
9.jpg</uri>
    </binding>
</result>
<result>
    <binding name='Titulo'>
        <literal>360 Degrees of Segregation: A
Historical Perspective Of Segregation-era School Equalization
Programs In the Southern United States</literal>
    </binding>
    <binding name='Descricao'>
        <literal>en</literal>
    </binding>
    <binding name='Data'>
        <literal>Amsterdam Law Forum; issue.vol 2,
issue.no 3 (2010): Future of Legal Education</literal>
    </binding>
    <binding name='Assunto'>
        <literal>law</literal>
    </binding>
    <binding name='Idioma'>
        <literal>en</literal>
    </binding>
    <binding name='Url'>
        <uri>http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/280
9.jpg</uri>
    </binding>
</result>
<result>
    <binding name='Titulo'>
        <literal>360 Degrees of Segregation: A
Historical Perspective Of Segregation-era School Equalization
Programs In the Southern United States</literal>
    </binding>
    <binding name='Descricao'>

```

```

        <literal>en</literal>
    </binding>
    <binding name='Data'>
        <literal>Amsterdam Law Forum; issue.vol 2,
issue.no 3 (2010): Future of Legal Education</literal>
    </binding>
    <binding name='Assunto'>
        <literal>law</literal>
    </binding>
    <binding name='Idioma'>
        <literal>en</literal>
    </binding>
    <binding name='Url'>
        <uri>http://www.musei.uniroma1.it/dbinfo/RMSMUS12/JPEG/273
9.jpg</uri>
    </binding>
</result>
</results>
</sparql>

```

No código é possível visualizar o propósito da linguagem XML em possibilitar o intercâmbio de dados na *Web*, visto que o código transporta consigo informações de sua estrutura e significado (nome do campo, tipificação), estando pronto para ser recuperado por qualquer sistema independente de plataforma.

Sem dúvida, a consulta por meio da linguagem SPARQL é complexa para os usuários finais quando utilizada em forma de códigos, visto que exige conhecimentos básicos em lógica de programação. No entanto, a linguagem SPARQL possibilita que a busca dos itens informacionais possa ser realizada por qualquer um dos metadados do modelo EDM, como data de criação, direitos autorais, tipo de imagem, local, entre outros, tornando a busca minuciosa. Por outro lado, se o resultado é uma busca simples, empregando uma frase para retornar um item informacional, a busca por meio do SOLR é suficiente.

### 3.3.1.5 Ontologias e vocabulários para enriquecimento de dados

As ontologias são estruturas de dados que representam conceitos e relações entre entidades, facilitando a interoperabilidade dos dados entre sistemas de informação, conforme já definida por Bräscher; Carlan(2010). No entanto, o custo

de produção de uma ontologia é alto, motivando o emprego de ontologias prontas e reconhecidas pelo mercado, o que agiliza o processo de desenvolvimento das ferramentas utilizadas no ambiente semântico.

O W3C descreve os vocabulários como os responsáveis por definir os conceitos e relacionamentos entre as áreas do conhecimento, sendo empregado para classificar os termos que podem ser utilizados em uma aplicação específica, detalhar as relações e suas restrições. E esses podem ser muito complexos, com milhares de termos, ou simples, com apenas um ou dois conceitos. O W3C entende que não existe uma clara diferenciação entre os vocabulários e ontologias e orienta que sejam utilizadas ontologias para os modelos de dados mais complexos<sup>19</sup> e formal dos termos, enquanto que vocabulários sejam empregados quando não houver necessidade de formalismo (WORLD WIDE WEB CONSORTIUM, 2013b).

Embora a ontologia possua um papel relevante para o desenvolvimento de sistemas na *Web Semântica*, os responsáveis pelo desenvolvimento da biblioteca Europeia não mencionam a utilização de ontologias e somente o emprego de dois vocabulários próprios como modelo de padronização e enriquecimento de metadados, a *Europeana Photography*<sup>20</sup> e *Europeana Fashion*<sup>21</sup>, entende-se que o tempo gasto e o custo foram fatores decisivos para a preferência de vocabulários externos.

Considera-se que o enriquecimento proporciona melhora no acesso ao material, define as relações entre os itens informacionais e permite uma recuperação de dados em diversos idiomas. Para o enriquecimento de metadados, a Europeia utiliza o software *Annocultor*, que foca a ligação dos itens informacionais da Europeia com outros recursos. As principais etapas do processo de enriquecimento estão descritos a seguir (EUROPEANA NETWORK, 2014):

- **o enriquecimento interno:** os campos das classes EDM dos quais as conexões são derivadas e principalmente relacionando o valor destes campos com os rótulos dos recursos contextuais;

---

<sup>19</sup> Para Quin e Paling (2001) modelo de dados mais complexos apresentam um nível mais alto de concepção e descrição dos vocabulários, acarretando um material semântico com mais detalhes para as relações do tipo classe/subclasse e suas relações, além da reusabilidade das ontologias.

<sup>20</sup> <http://www.europeana-photography.eu/>

<sup>21</sup> <http://www.europeanafashion.eu/>

- **o enriquecimento externo:** o conjunto de itens da Europeana que estarão ligados com vocabulários externos;
- **regras:** especifica como uma relação é obtida entre o campo de metadados fonte e os rótulos do recurso contextual.

No **enriquecimento interno**, as classes contextuais do modelo de dados estruturado pela Europeana possuem os dados enriquecidos mediante a utilização de vocabulário controlado<sup>22</sup> e dicionários, com a inclusão desses recursos contextuais através de links que estarão ligados a outros dados externos, mantendo-os separados do item de descrição. Pode-se detalhar as classes contextuais sob aspecto de enriquecimento (ISAAC et al., 2014):

- Classe `edm:Place` refere-se a um **local** com localização espacial identificadas pelo fornecedor e nomeado de acordo com algum vocabulário ou convenção local;
- Classe `skos:Concept` refere-se a um **conceito** definido como uma unidade de pensamento ou significado que vem de uma base de conhecimento organizado, o qual URI's ou identificadores locais foram criados para representar cada conceito;
- Classe `edm:Agent` refere-se a um **agente** que compreende as pessoas, tanto individual como em grupos, que tem potencial para executar ações intencionais no qual podem ser responsabilizados;
- Classe `edm:TimeSpan` refere-se a um **período de tempo** definidos por intervalos de início e fim e sua duração.

Essas quatro classes contextuais não possuem descrições ou características dos itens, o propósito é realizar a ligação com outras classes e fornecedores de conteúdo para a interoperabilidade dos dados.

O **enriquecimento externo** é realizado por meio de um vocabulário externo, que para a Europeana é uma base de metadados armazenada em ambiente *Open Source*, disponível em dados ligados, devidamente documentado,

---

<sup>22</sup> Um vocabulário controlado é uma lista de termos autorizados por uma entidade específica, para controlar sinônimos, diferenciar palavras homógrafos e reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si. (LANCASTER, 1991. p.14)

interoperável e multi-idioma possibilitando aos itens da biblioteca estarem linkados. Sustentado por esses requisitos, a Europeana estabeleceu empregar os serviços de algumas instituições para enriquecimento de sua base de metadados. Segue abaixo um exemplo de vocabulário adotado, entre diversos, para cada um dos itens referenciados (ISAAC et al., 2014):

- para **local**, adota-se a base de dados do vocabulário GeoNames<sup>23</sup>, que oferece uma ampla cobertura de lugares com conteúdo detalhado;
- para **conceito**, adota-se a base de dados dos vocabulários *General Multilingual Environmental Thesaurus* (GEMET<sup>24</sup>) que possui um rico acervo em diversos idiomas e DBpedia<sup>25</sup> com o aproveitamento de algumas informações da Primeira Guerra Mundial;
- para **agente**, adota-se a base de dados do DBpedia para uma identificação única das pessoas;
- para **período de tempo**, adota-se a base de dados do *Semium Time Vocabulary*<sup>26</sup> para expressões de períodos de tempo.

Os metadados obtidos das instituições parceiras são convertidos para a base de dados do modelo EDM, armazenados como arquivos RDF em separado, e finalmente incluído como parte do recurso. Isto significa que se os vocabulários externos sofrem modificações significativas, os arquivos RDF devem ser atualizados e uma nova versão deve ser montada e implantada (ISAAC et al, 2014).

As **regras** aplicadas no modelo de estrutura semântica serão conforme a classe contextual na qual o item será enriquecido. As regras são aplicadas apenas para algumas propriedades do item. Para cada um dos campos de origem, será aplicada uma verificação de letras maiúsculas e minúsculas em comparação com os

---

<sup>23</sup> GeoNames: Banco de dados que disponibiliza metadados geográficos e contém um conjunto de vocabulários para adicionar semântica para distribuição de dados (GEONAMES, 2015).

<sup>24</sup> GEMET: Banco de dados que possui um dicionário de sinônimos com termos gerais compilados de vários vocabulários multi-idiomas (EIONET, 2012).

<sup>25</sup> DBpedia: banco de dados colaborativo de informações extraídas da *Wikipédia*, que utiliza os princípios da Web Semântica, como a consulta via SPARQL (DBPEDIA, 2015).

<sup>26</sup> Semium Time Vocabulary: Vocabulário que estabelece uma referência única para o período de tempo com datas de início e fim e conectar nomes com períodos históricos (EUROPEANA PRO, 2015).

rótulos dos vocabulários externos. Para o caso dos Agentes em particular, os campos de origem são processados prioritariamente para remover datas de nascimento e morte entre parênteses ou funções entre colchetes (ISAAC et al, 2014).

Com a ferramenta *Annocultor*, desenvolvida pela Europeana, é possível retirar as referências URI's dos vocabulários e buscar todos os dados multi-idíomas e semânticos diretamente da central de serviço de dados dos vocabulários externos. As referências de URI's das instituições são retiradas no momento da realização da ingestão dos itens na Europeana (CHARLES; DEVARENNE, 2015).

O enriquecimento de um item pode ser visualizado no portal de busca da biblioteca Europeana, no exemplo em que busca-se dados sobre o castelo de Bragança, em Portugal, no endereço do *website*: <http://www.europeana.eu/portal/record/2023706/6B701270DC2AC7EF8CC9766888CF8F036071B956.html> em que utiliza-se os vocabulários do *Geonames*, *GEMET* e *Semium Time*. Veja resultado na Figura 18.

**Figura 18** – Enriquecimento - Castelo de Bragança

**Castelo de Bragança**

**Título:**  
Bragança Castle

**Descrição:**  
Castelo de Bragança;  
Bragança Castle

**Criador:**  
[Lusa](#)

**Cobertura geográfica:**  
[Bragança](#)  
Latitude: 41.81667; Longitude: -6.75

**Data:**  
1987-02-01

**Data de criação:**  
1987-02-01

**Tipo:**  
[fotografia \(imagem\)](#)

**Formato:**  
image/jpeg

**Identificador:**  
LUS12519750

**Direitos:**  
EuroPhoto LUSA provider

**Editor:**  
EuroPhoto LUSA provider

**Fornecedor de conteúdos:**  
[EuroPhoto LUSA provider](#)

**Agregador:**  
[EuroPhoto](#)

**Providing country:**  
Portugal

[Etiquetas geradas automaticamente ▾](#)

[Geographical coverage ▶](#)

[What ▶](#)

[When ▶](#)

**Fonte:** (EUROPEANA, 2015) adaptado pelo autor.

Ao final da Figura 18 é possível identificar o texto `Etiquetas geradas automaticamente`, circundada por uma elipse para efeito de destaque. Ao clicar sobre o item, serão exibidos links para `Geographical coverage`, `What` e `When`, delineados por um retângulo.

O link `Geographical coverage`, fornece metadados geográficos da base de dados do vocabulário *Geonames*, para o *link* do termo `What`, fornece

metadados de conceitos e sinônimos da base de dados do vocabulário *GEMET* e no link *When*, fornece metadados da base de dados do vocabulário *Semium Time*.

Ao clicar sobre o *link* *Geographical coverage*, visualiza-se as informações na Figura 19, enriquecidas pelo vocabulário *Geonames*, identificado pelo URI base <http://sws.geonames.org>, com *links* para o mapa do país Portugal e do ponto turístico, local de origem, formas de grafia do termo Portugal em diversos idiomas e coordenadas do castelo.

**Figura 19 – Geographical Coverage**

**Geographical coverage** ▾

**Lugar:**  
<http://sws.geonames.org/2264397/>

**Lugar:**  
[Portugal] (no); [Portegal] (fy); [Portugal] (nn); [An Phortaingéil] (ga); [Portugal] (oc); [Portugali] (fi); [Portugal] (fr); [Португали] (os); [Πορτογαλ] (he); [Portugal] (gl); [Yn Phortugal] (gv); [Portugalia] (pl); [ປໍລຕຸກັລ] (lo); [Bồ Đào Nha] (vi); [പോർട്ടുഗാൾ] (dz); [Portugåle] (lv); [Portugalija] (lt); [Portugal] (de); [Portugal] (uz); [Lusitania, PT, Portuguese Republic, Republic of Portugal] (def); [Португалија] (mk); [പോർട്ടുഗാൾ] (ml); [Portugal] (eu); [Португалия, Португалия] (uk); [पोर्तुगाल] (mr); [پورتوگالیہ] (ug); [Portugall] (mt); [Feringgi, Portugal] (ms); [پرتگال] (ur); [پرتغال] (fa); [Põititi] (ty); [Πορτογαλία] (el); [Portugal] (nb); [पोर्तुगल] (ne); [Portugaliao, Portugalujo] (eo); [Portugal, Portuguese Republic] (en); [Portugal] (et); [Portugal] (es); [Portugal] (nl); [Potukalii] (to); [Portugal] (ca); [Portugal] (tl); [Portekiz] (tr); [Португалия] (tg); [Portugal] (bs); [Portugal] (br); [പ്രതലാലപ്രമല, സാമ്രാജ്യപ്രമല, പ്രമല] (th); [পোর্টুগাল] (bn); [പോർട്ടുഗാൾ] (bo); [போபர்ச்சக்கல், போபർத்துகல்] (ta); [Portugal] (sv); [Португалия] (bg); [පොර්තුගාලය] (ka); [Ureno] (sw); [Партугалия] (be); [Portugalska] (sl); [Portyngal] (kw); [Portugalsko] (sk); [Portugal] (da); [ਪੁਰਤੁਗਲ] (ks); [Bortuqaal] (so); [Portekiz] (ku); [Португал, Португалија] (sr); [Portugali, Portugalia] (sq); [포르투갈, 포트갈] (ko); [Portiugal] (cy); [Portugallu] (sc); [Portugália] (se); [Portugal] (sh); [പോർട്ടുഗാൾ] (km); [Portugalsko] (cs); [Portugal] (li); [Portugallu] (co); [ਪੁਰਤੁਗਲ] (sa); [Lusitania] (la); [Portugal] (lb); [Португалия] (ru); [Portugal] (hr); [葡萄牙] (zh); [Portugalia] (ro); [Portugal] (rm); [Pòtigal] (ht); [Portugália] (hu); [ਪੁਰਤੁਗਲ] (hi); [Portugal, Portugis] (id); [Portugal] (ia); [Purtugal] (qu); [Ἰντερνאַಷιῶναλ] (hy); [Portugal] (is); [Portogallo] (it); [البرتغال, برتغال] (ar); [Portugal] (io); [පොርටුගාල] (am); [Portugal] (an); [Portugal, República Portuguesa] (pt); [ポルトガル, ポルトガル共和国] (ja); [پورتگال] (ps); [פּאָרטוגאַל, פּאָרטוגאַל] (yi); [Portugal] (af)

**Coordenadas:**  
null; null

**Lugar:**  
<http://sws.geonames.org/2742027/>

**Lugar:**  
[Bragança] (def); [Bragance] (fr); [Браганца] (ru)

**Coordenadas:**  
41.81667; -6.75

Part of: [Portugal](#)

[What](#) ▸

[When](#) ▸

**Fonte:** (EUROPEANA, 2015) adaptado pelo Autor.

Ao clicar sobre o *link* *What*, visualiza-se as informações na Figura 20, enriquecidas pelo vocabulário *GEMET*, identificado pelo URI base

<http://eionet.europa.eu/gemet/concept>, com *links* para definições mais amplas de termos utilizados nos metadados do item.

**Figura 20 – What**

[Geographical coverage ▸](#)

**What ▾**

**Conceito (termo):**  
<http://www.eionet.europa.eu/gemet/concept/2284>


**Conceito (etiqueta):**  
[dokumentacija] (sl); [dokumentácia] (sk); [dokumentatze; dokumentazio] (eu); [dokumentation] (da); [documentație] (ro); [documentazione] (it); [dokümantasyon] (tr); [dokumentazzjoni] (mt); [dokumentasjon] (no); [dokumentálás] (hu); [dokumentācija] (lv); [توثيق] (ar); [dokumentacija] (lt); [dokumentace] (cs); [Dokumentation] (de); [τεκμηρίωση] (el); [documentation, 文档化] (def); [dokumentointi, asiakirja-aineisto] (fi); [documentação] (pt); [dokumentacija] (pl); [Документиране] (bg); [documentation] (fr); [dokumentation] (sv); [documentation] (en); [документация] (ru); [dokumentatsioon] (et); [documentación] (es); [documentatie] (nl)

**Conceito mais lato (etiqueta):**  
<http://www.eionet.europa.eu/gemet/concept/4303>

**Conceito (termo):**  
<http://www.eionet.europa.eu/gemet/concept/6205>

**Conceito (etiqueta):**  
[fotografija] (sl); [fotografia] (sk); [fotografi] (da); [fotografia; argazki] (eu); [fotograf] (ro); [fotografia (immagine)] (it); [fotograf] (tr); [ritratt] (mt); [fotografi (bilde)] (no); [fénykép] (hu); [fotogrāfs] (lv); [صورة فوتوغرافية] (ar); [nuotrauka] (lt); [fotografie] (cs); [Photo] (de); [φωτογραφία] (el); [photograph, 照片] (def); [valokuva] (fi); [fotografia] (pl); [fotografia (imagem)] (pt); [Фотография, фотографско изображение, снимка] (bg); [fotografi] (sv); [photo] (fr); [photograph] (en); [фотография] (ru); [foto] (et); [fotografía] (es); [foto] (nl)

**Conceito mais lato (etiqueta):**

Broader concept: <http://www.eionet.europa.eu/gemet/concept/4303> 

Broader concept: [documentação](#)

**When ▸**

**Fonte:** (EUROPEANA, 2015) adaptado pelo Autor.

Ao clicar sobre o *link* *when*, visualizam-se as informações na Figura 21 enriquecidas pelo vocabulário *Semium Time*, identificado pelo URI base <http://semium.org/time>, com *links* de acontecimentos ocorridos em um período de tempo determinado. Um problema ocorre no *link* do *website Semium Time*, visto que retorna um erro no momento em que é acessado.

**Figura 21 – When**

Geographical coverage ▶
What ▶
When ▼
<b>Period Term:</b> <a href="http://semium.org/time/19xx_4_quarter">http://semium.org/time/19xx_4_quarter</a>
<b>Period Label:</b> [4e quart 20e siècle] (fr); [4 quarter of the 20th century] (en); [4-я четверть 20-го века] (ru)
<b>Period Begin:</b> Thu Jan 01 01:00:00 CET 1976
<b>Period End:</b> Sun Dec 31 01:00:00 CET 2000
<b>Period Term:</b> <a href="http://semium.org/time/1987">http://semium.org/time/1987</a>
<b>Period Label:</b> [1987] (def)
<b>Period Begin:</b> Thu Jan 01 01:00:00 CET 1987
<b>Period End:</b> Thu Dec 31 01:00:00 CET 1987

**Fonte:** (EUROPEANA, 2015) adaptado pelo Autor

O enriquecimento de metadados realizado pela Europeana possibilita estender as descrições do referido item, explorando as relações semânticas e traduções por meio de uma rede de vocabulários. Os metadados são complementados por meio de outros vocabulários abertos e multi-idiomas, além daqueles citados nos exemplos, a Europeana incentiva o emprego de diversos vocabulários como *AAT* da *Getty vocabularies*, *GND Ontology*, *ICONGLASS*, *VIAF*, *MIMO*, *Partage Plus*, etc ou qualquer outro que siga as recomendações EDM para metadados sobre os recursos contextuais.

No APÊNDICE B está disponível uma relação de vocabulários externos utilizados pela Europeana. Essa lista provém de um estudo de pesquisadores da Europeana denominado “*Europeana Tech Task Force on a Multilingual and Semantic Enrichment Strategy: final report*” (EUROPEANA NETWORK, 2014) que analisaram itens da base de dados e realizaram uma estatística identificando as falhas ocorridas no enriquecimento de dados e soluções adequadas.

### 3.3.2 Europeana em Síntese

Neste estudo de caso, utilizando a biblioteca Europeana como objeto de estudo, concebeu-se um infográfico. A Figura 22 sintetiza os momentos essenciais do funcionamento da Web Semântica nas unidades de informação.

**Figura 22** – Web Semântica em Unidades de Informação



**Fonte:** o próprio Autor.

Observa-se na Figura 22, a Unidade Informacional representada pelas bibliotecas e arquivos, responsáveis por descrever o item informacional de acordo com normas, padrões e técnicas pertinentes de cada área, conforme já apresentado na Figura 1. Dessa maneira, neste momento as bibliotecas aplicam as normas e

técnicas da representação descritiva e temática, e os arquivos empregam os procedimentos de gestão de documentos.

O item informacional, devidamente descrito e que são representados pelos livros, filmes, revistas, fotos, quadros, etc, é o objeto que será transmitido pelos provedores, agregadores ou descrito na própria biblioteca, aqui representada pela Europeia.

Os provedores são repositórios de dados que podem ou não possuir seu próprio *website* e são responsáveis por cadastrar os dados do item informacional para serem enviados aos agregadores, conforme o padrão de metadados definido pelo repositório final.

Os agregadores são repositórios de dados que agregam vários provedores e validam os metadados simulando sua ingestão por meio de uma ferramenta de interface gráfica, denominada REPOX, convertendo-os para XMLS e para o padrão de metadados empregado pela Europeia, o EDM. O software REPOX pode armazenar os metadados em servidor OAI-PMH remoto nos padrões da Europeia ou em um arquivo com as configurações do formato em uma pasta no sistema, para ser exportada.

O *website* do portal da biblioteca permite a realização da consulta por meio da SPARQL, uma linguagem de busca, na qual o usuário escreve os comandos da linguagem, empregando os conceitos do RDF em uma base XML. O *harvesting* é realizado por intermédio do protocolo OAI-PMH que possibilita a interoperabilidade dos dados.

Os vocabulários externos são obtidos por meio da ferramenta *Annocultor* em forma de metadados e convertidos para a base de dados do modelo em uso, armazenados como arquivos RDF. Com os vocabulários externos, a apresentação da busca torna-se mais rico, complementado com extensões de descrições do item de outros parceiros da rede semântica.

O modelo gráfico apresentado sintetiza um ambiente que enquadra-se nos princípios da *Web Semântica*, deve ser observado em conjunto com os capítulos anteriores o qual conceitua-se de maneira particular as normas, padrões e técnicas empregadas nas unidades informacionais e como o item informacional é descrito. Detalha o trabalho realizado pelos provedores e agregadores e as tecnologias

empregadas na gestão dos dados e o enriquecimento dos dados por meio dos vocabulários externos.

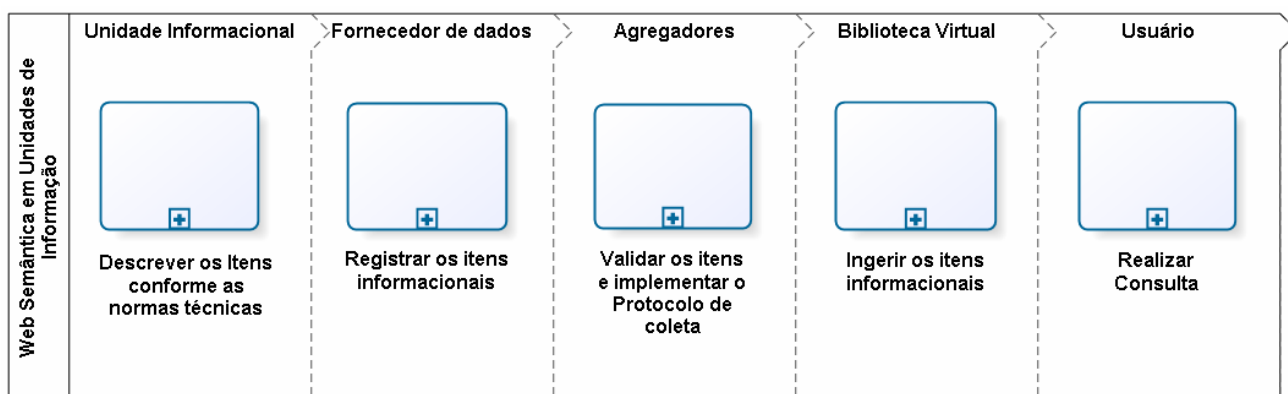
### 3.3.3 Um *Framework* para Representar a *Web Semântica* nas Unidades de Informação.

Um *Framework*, para Souza (1998), “[...] é um projeto genérico em um domínio que pode ser adaptado às aplicações específicas, servindo como um molde para a construção de aplicações.” Ao moldar um projeto ou aplicação, o *framework* produz um padrão que proporcionará as aplicações desenvolvidas acompanharem a mesma estrutura.

O *framework* possibilita o reaproveitamento de código e projeto de uma aplicação. A reutilização do projeto proporciona aos desenvolvedores um aproveitamento em larga escala, visto que utiliza-se a estrutura do domínio. A reutilização do projeto é considerada mais relevante que o código, visto que o projeto é mais complexo de criar e recriar (JOHNSON; FOOTE, 1998).

O *framework* foi elaborado com o propósito de representar os serviços considerados vitais nos procedimentos de instituir o projeto da *Web Semântica* nas Unidades de Informação e baseou-se no framework para Gerenciar os Riscos de Projeto por meio de Serviços, proposto por Barros e Gaffo (2012) e empregou o software de modelagem Bizagi<sup>(R)</sup>. Esses serviços que denominaremos como seção, estão sintetizados no framework da Figura 23 que determinam os procedimentos a serem empregados em cada um dos ambientes que compõem o domínio da *Web Semântica*.

**Figura 23** – Web Semântica em Unidades de Informação

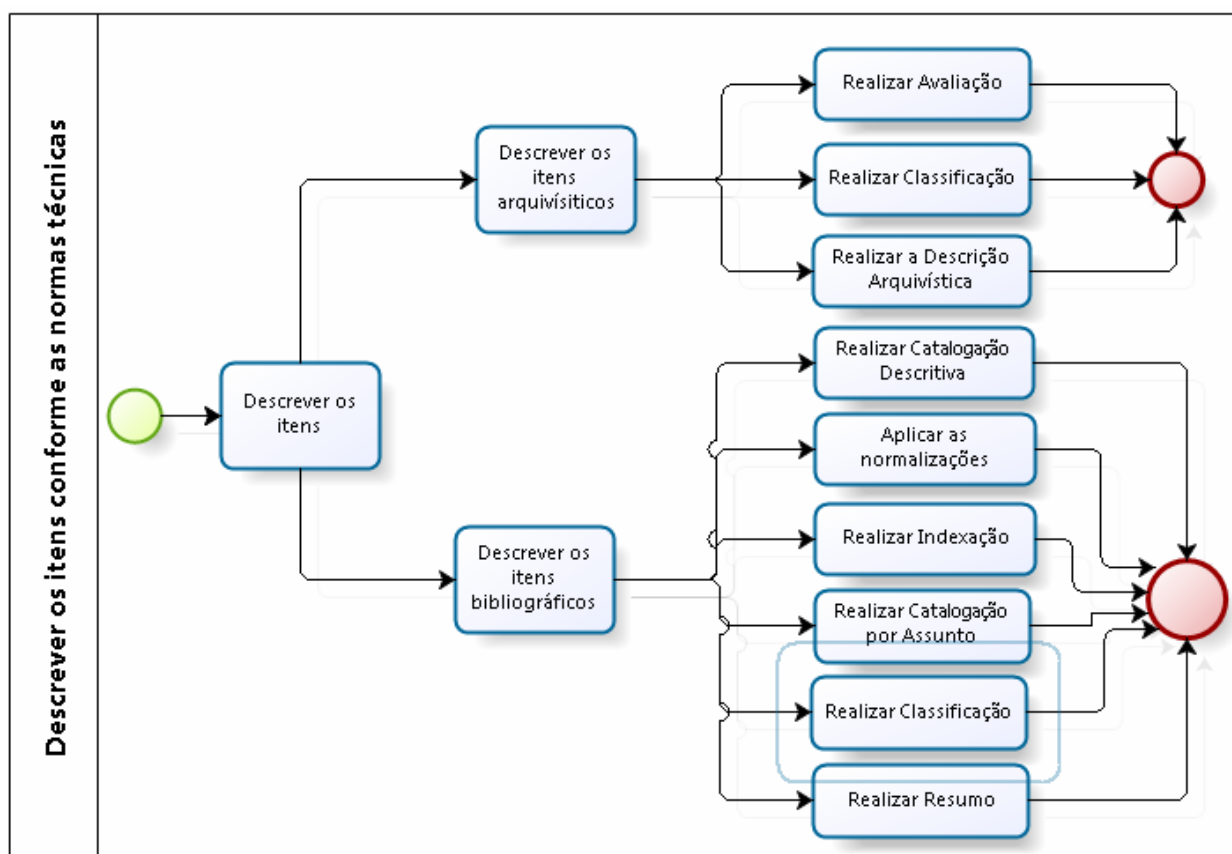


**Fonte:** o próprio autor.

Observa-se na Figura 23, a seção da Unidade Informacional que possui o processo Descrever os itens conforme as normas técnicas, é responsável por descrever o item informacional conforme normas, padrões e técnicas pertinentes de cada área. Nesta seção, conforme pode ser visualizado pelo Workflow na Figura 24, são realizadas, por intermédio dos Arquivos, a avaliação, a classificação e a descrição arquivística. Nas Bibliotecas, a catalogação descritiva e a normalização por meio da Representação Descritiva e a indexação, catalogação por assunto, classificação e o resumo por meio da Representação Temática.

Contudo para que esta descrição esteja de acordo com as propostas da Web Semântica, as normas, técnicas e padrões da OI devem estar formatadas pelas recomendações da W3C. Na Figura 24, é possível observar as práticas da OI e a seguir os detalhes sob a perspectiva da Web Semântica.

**Figura 24** – Seção Unidade Informacional



Fonte: o próprio autor.

Para a descrição adequada conforme os padrões e recomendações da Web Semântica, os metadados são empregados para os arquivos realizarem a descrição arquivística e as bibliotecas a catalogação descritiva e temática, para o

registro dos dados em banco de dados locais. Os metadados são representados pela estrutura do modelo RDF e escritos em XML, essa prática possibilita que os sistemas de informação possam interagir com a informação.

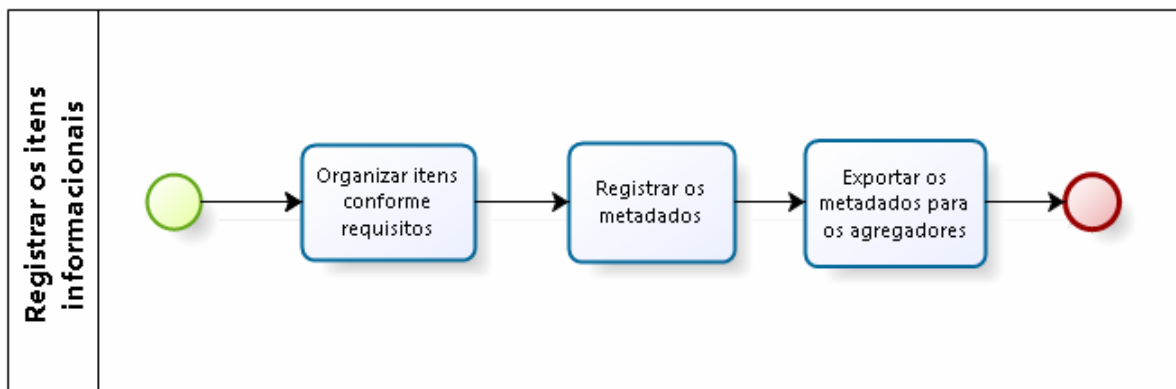
Conforme Catarino, Cervantes e Souza (2013), existem várias aplicações já realizadas neste sentido para que os arquivos e bibliotecas possam descrever seus itens de forma a estarem de acordo com as recomendações do W3C. As autoras destacam que para os dados bibliográficos transformarem-se em dados ligados é preciso converter os dados atuais descritos em MARC ou DC para a estrutura RDF, modificar o modelo conceitual FRBR e o código de catalogação RDA para bases RDF. Assim como escrever os vocabulários existentes em linguagens de desenvolvimento recomendado pela W3C, em particular o RDF Schema, SKOS e OWL.

Portanto, visualiza-se que, em resumo, para que a descrição dos itens dos arquivos e bibliotecas adequem-se ao projeto da *Web Semântica*, é preciso empregar o modelo RDF nos metadados padrão das unidades de informação, visto que o modelo é a chave para conversão dos dados das unidades de informação para a *Web Semântica*.

No Workflow da Figura 24, os fluxos dos processos *Descrever os itens arquivísticos e bibliográficos* direcionam-se individualmente para os processos realizados por cada um destes, visto que procura-se representar que os processos seguintes podem ocorrer unilateralmente e sem uma ordem de aplicação.

Na seção *Fornecedor de dados*, o *Workflow* da Figura 25, representa o processo *Registrar os itens informacionais*. Os Fornecedores de dados são repositórios de dados responsáveis por cadastrar o registro dos itens informacionais para serem enviados aos agregadores, conforme o padrão de metadados especificado pelo repositório final.

**Figura 25 – Seção Fornecedores de dados**



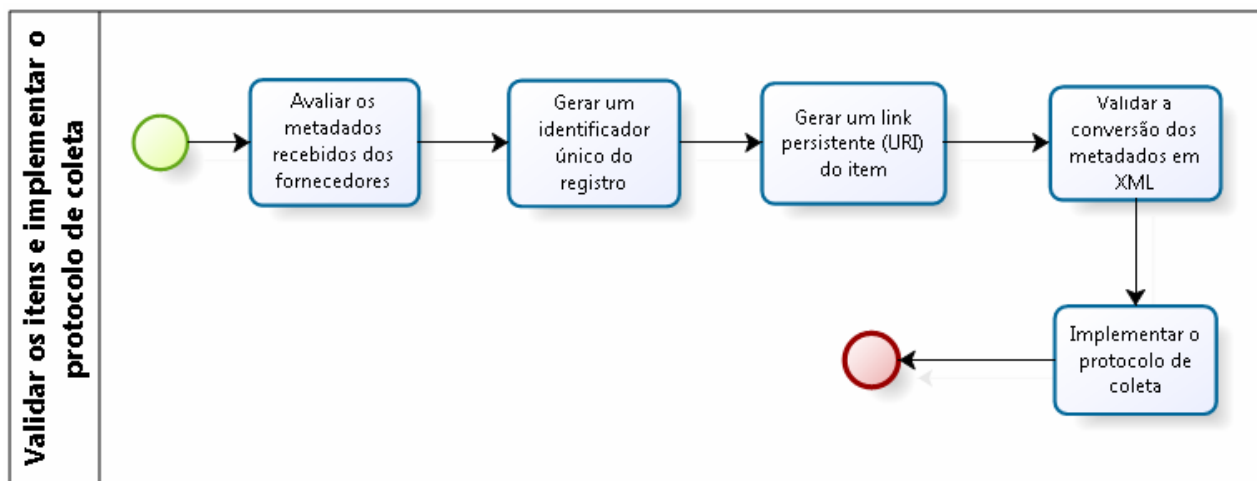
**Fonte:** o próprio autor.

A ação *Organizar Itens conforme requisitos* consiste em preparar os metadados do fornecedor de dados para o padrão de metadados da biblioteca virtual. Para a ação *Registrar os metadados*, é o momento em que define-se configurações de armazenamento, formato de entrada de metadados e URL do servidor OAI.

Os catálogos online (OPAC) compostos de metadados padrões como MARC e Dublin Core, ainda muito utilizado pelas unidades de informação, necessitam estar adequado aos princípios da *Web* de dados, para que possam ser lidos por máquinas. Os *software* de gestão das unidades de informação devem estar preparados para atender aos fundamentos da *Web* de dados, empregando o modelo RDF nos metadados padrões, escritos na linguagem XML.

A seção *agregadores*, do processo *Validar os itens* e *implementar o protocolo de coleta*, na *Figura 26*, compreende os repositórios de dados que reúnem vários fornecedores de dados e validam os dados simulando sua ingestão, convertendo-os para XMLS. Os agregadores também podem criar o conjunto de dados, empregando a mesma ferramenta dos fornecedores de dados.

Figura 26 – Seção Agregadores



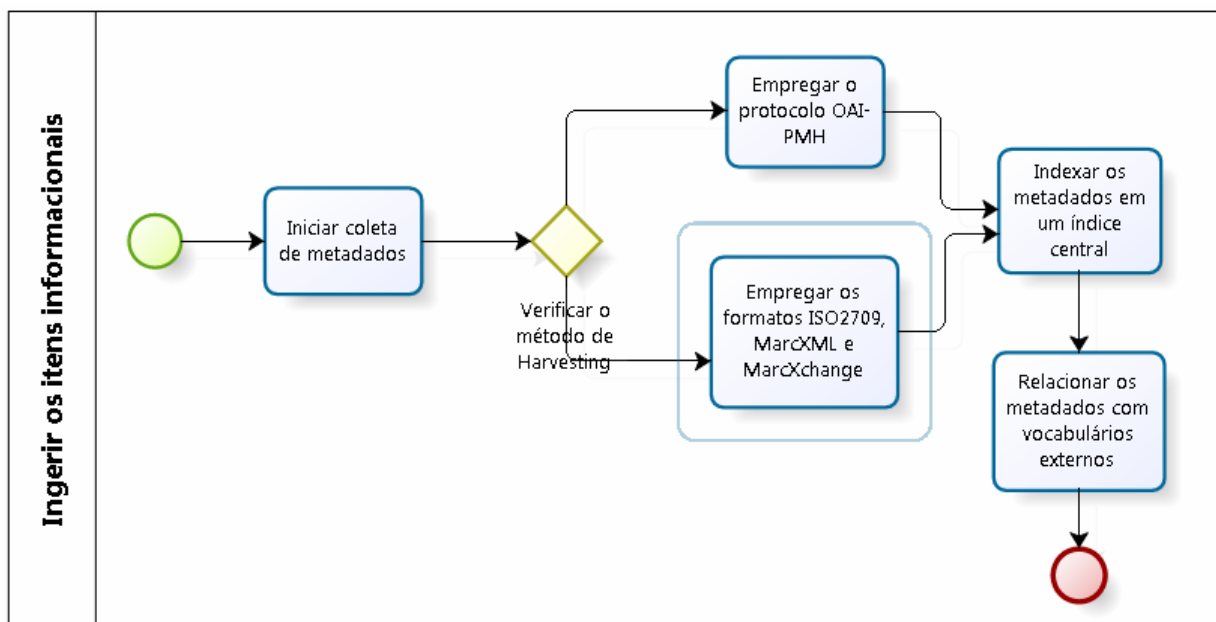
Fonte: o próprio autor.

Cada registro de metadados deve possuir um identificador único e persistente no metadados para gerar um link com um URI permanente para o registro no repositório de dados. Esse identificador único deve ser persistente, visto que caso haja alguma atualização no registro de metadados, o *link* deve continuar estável. Aconselha-se os agregadores realizarem uma simulação de ingestão para avaliar os metadados que serão enviados à Biblioteca Virtual, possibilitando identificar falhas e solicitar ao fornecedor origem corrigi-lo.

Agregador implementa o protocolo de coleta ou *harvesting* que será empregado para submeter os dados. O OAI-PMH é o protocolo empregado por diversos repositórios de dados, por ser aberto, gratuito e amplamente difundido.

A Figura 27 apresenta o *workflow* do processo Ingerir os itens informacionais, realizado pela biblioteca virtual, que consiste em identificar o formato em que os metadados serão coletados.

**Figura 27 – Seção Biblioteca Virtual**



**Fonte:** o próprio autor.

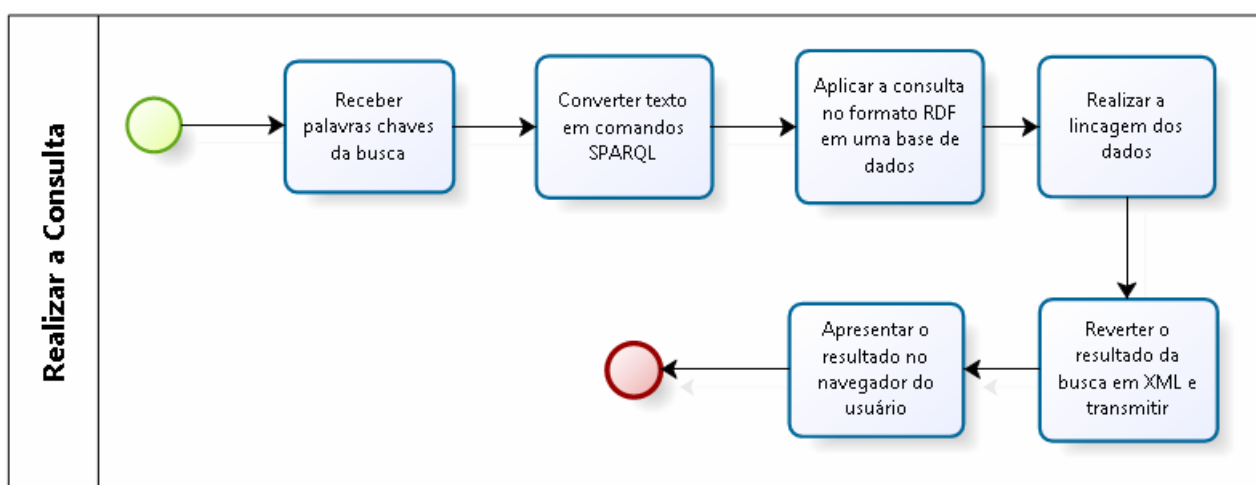
Na ação *Verificar o método de Harvesting*, a biblioteca virtual identificará o método utilizado pelo agregador para realizar a ingestão dos metadados. A ingestão pode ser realizada por meio do protocolo OAI-PMH, baseado em um servidor e por meio de uma pasta criada especificamente para realizar esse processo, que suporte vários formatos de metadados, entre eles o MTD-BR, empregado por universidades brasileiras para realizar o harvesting na Biblioteca Digital de Teses e Dissertações (BDTD).

Os metadados ingeridos devem ser organizados em uma base de dados principal e serem indexados para possibilitar que seu acesso realize-se de maneira ágil e precisa. O relacionamento dos metadados por fontes externas proporcionará que os conteúdos localizados pelos usuários possuam informações complementares tornando a busca mais completa.

Propõem-se que utilize um software específico que propicie acesso a vocabulários externos que acrescente dados a respeito de locais, datas, pessoas, objetos, entre outros. Uma das grandes dificuldades em interligar com vocabulários externos são as relações entre si, visto que não pode haver duplicidade dos metadados que realizam a ligação. A *Library of Congress* (LC), emprega uma lista de cabeçalho em URI, o que possibilita que cada cabeçalho seja uma URI e um identificador único. Dessa forma, essas URI's podem ser utilizadas nos namespaces das descrições RDF (CATARINO; CERVANTES; SOUZA, 2013).

No momento que o usuário irá Realizar a Consulta, Figura 28, o navegador deve enviar à biblioteca virtual ou ao repositório principal, as palavras chave que determinam a busca. Essas palavras chave preferencialmente foram digitadas em um ambiente gráfico desenvolvido para esse fim. O software deste ambiente gráfico deve realizar a conversão dessas palavras em código do comando SPARQL, poupando o usuário final da necessidade de escrever em código SPARQL.

**Figura 28 – Realizar Consulta**



**Fonte:** o próprio autor

A linguagem SPARQL, empregando os princípios do modelo RDF, possibilitará realizar a consulta na base de dados do repositório da biblioteca virtual, implementando a ligação dos dados com outros vocabulários de dados internos ou externos. O resultado da busca será convertido para o código da linguagem XML, que enviará tanto a estrutura como os dados do resultado para o navegador do usuário, indiferente da plataforma.

## 4 CONSIDERAÇÕES

Para a realização deste estudo, uma das características da Ciência da Informação foi essencial, visto que foi necessário realizar estudos em três áreas distintas: a Arquivologia, Biblioteconomia e Ciência da Computação.

Dessa forma a interdisciplinaridade da Ciência da Informação tornou-se evidente, já que, em muitos momentos, conceitos e técnicas completavam-se, aparentando serem todos de um domínio único.

Inicialmente, o estudo foi dedicado a definir as técnicas e processos aplicados pelos arquivos e bibliotecas para descrever o item informacional. A apresentação foi realizada com o apoio de um infográfico, que possibilitou uma visualização sintetizada de todos os procedimentos.

As tecnologias da *Web Semântica* foram descritas logo em seguida, com um destaque maior para os metadados e o *Dublin Core*, as triplas do RDF que possibilitam a interconexão de dados, as ontologias que por serem complexas e trabalhosas são substituídas por vocabulários e as linguagens de desenvolvimento empregadas na produção das ontologias, aplicação do RDF, busca de conteúdo, entre outros.

Para o estudo do site da biblioteca virtual Europeia, categorizou-se os tópicos consideradas relevantes, que foram definidos como: os metadados, *harvesting*, linguagem de desenvolvimento e banco de dados, linguagem de busca e vocabulários.

Concebeu-se um *Framework* com o propósito de identificar e visualizar os processos aplicados na implantação da *Web Semântica* nas Unidades de Informação, especificando por meio de *Workflow* cada um dos processos principais do *Framework*.

### 4.1 RESULTADOS OBTIDOS

A pesquisa teve como objetivo principal descrever as tecnologias necessárias para a aplicação da *Web Semântica* nas unidades de informação. Para tanto, foram definidos três objetivos específicos como meta.

O primeiro foi identificar as tecnologias da OI nas unidades de informação: elaborou-se um infográfico com as técnicas e normalizações empregadas na arquivologia e biblioteconomia para descrever um item informacional. Conceituou-se cada uma das técnicas e normas de acordo com a bibliografia estudada e foi possível identificar que a atuação dos arquivos e bibliotecas ocorre no início do processo da *Web Semântica*, ou seja no momento em que é realizada a descrição do item informacional, seja na forma física como no conteúdo.

Segundo: analisar as tecnologias, recomendações, linguagens e padrões que compõem a *Web Semântica*. Realizou-se um estudo por meio da pesquisa bibliográfica, que possibilitou apontar os metadados, RDF, Ontologias e linguagens de desenvolvimento como elementos essenciais à implantação da *Web Semântica*, fortalecido pelo artigo escrito por Berners Lee, Hendler e Lassila (2001) no qual descrevem a estrutura para a existência da *Web Semântica* para que as máquinas compreendam o significado dos dados.

Ainda segundo os autores supracitados, os computadores devem acessar uma coleção estruturada de dados e um conjunto de regras de inferências que possibilitem o raciocínio automatizado. A linguagem XML transporta a estrutura de dados pela *Web*, no entanto, sem expressar significado. O RDF expressa esse significado dos dados por meio das triplas que interliga diversos *websites* de ontologias ou vocabulários. Dessa forma o RDF amarra as relações e cria *links* individualizados por meio dos URI's.

Por fim, demonstrar a aplicabilidade da *Web Semântica* em ambientes das unidades de informação, por meio de um estudo de caso e um *framework*. Para o estudo deste objetivo, agrupou-se os temas em categorias, conforme Bardin(2010) recomenda, o que possibilitou uma organização do estudo e direcionamento dos esforços em eixos específicos e primordiais para o entendimento da *Web Semântica* nas Unidades de Informação.

O modelo de metadados EDM, concebido especificamente para a biblioteca Europeia, possibilitou que os dados coletados nos fornecedores de conteúdo tenham uma perda quase insignificativa, já que o EDM possui uma quantidade de metadados que permite descrever um item informacional com muitos detalhes. A consulta por meio do SPARQL exige o conhecimento dos metadados

EDM, visto que a linguagem realiza a busca conforme os metadados empregados no código da pesquisa. No exemplo da Figura 16 – Código em SPARQL para listar três registros em que o idioma seja “Inglês” e Assunto “Law”, necessitou-se consultar as definições dos metadados para a realização da consulta.

O *harvesting* na biblioteca Europeia é um processo que envolve diversos provedores e agregadores todos interconectados e submetendo dados ao repositório principal. As etapas podem ser visualizadas na Figura 11 – **Processo para submissão de dados à Europeia**. A biblioteca somente armazena os metadados dos itens informacionais, mantendo os arquivos de foto, vídeo, documentos armazenados no repositório de origem, criando somente um link para acesso.

Projetos institucionais como a Europeia tendem a empregar *software* abertos e gratuitos com o propósito de compartilhar conhecimentos por meio do acesso ao código fonte que também promove a visibilidade de todo o processo de desenvolvimento e como alternativa menos onerosa. Para um *Website* ser desenvolvido tanto no aspecto visual como estrutural, é preciso empregar linguagens de marcação e programação. A XML é uma linguagem de marcação que possibilita o emprego das linguagens de busca e o transporte da estrutura de dados entre provedores, sem distinção de plataforma.

O banco de dados da biblioteca segue os princípios de código aberto. O MongoDB é um banco orientado a documentos, apropriado para armazenar registros que ocupem grande quantidade de espaços. Os campos aceitam outros campos que podem ser matrizes. Os metadados EDM dos registros dos itens informacionais, como os vocabulários, são armazenados nessa base de dados.

A linguagem de busca SPARQL apresenta algumas dificuldades em seu emprego, visto que exige do usuário conhecimentos em comandos XML e dos metadados EDM para sua aplicação. No entanto, o SPARQL possibilita que as buscas realizem-se por meio de qualquer metadado EDM. Pode-se localizar o item informacional pela data de registro, fornecedor de dados, tipo de documento ou imagem, tamanho, qualidade da imagem, entre outros.

Os vocabulários possibilitam o enriquecimento das buscas. Em um item informacional pode adicionar-se ao seu resultado dados complementares localizados de outros fornecedores de dados e de instituições que fornecem esse serviço, como

o *Geonames*, *GEMET*, *DBpedia*, *Semium Time*, entre outros. Esse enriquecimento estende as descrições do item informacional, explorando as relações semânticas através de uma rede de vocabulários.

Um *framework* foi proposto para demonstrar os procedimentos importantes para a implantação do projeto *Web Semântica* nas unidades de informação. O *framework* organiza-se em cinco seções em que descreve por meio de um *Workflow*, ações realizadas nas Unidades de informação, Fornecedores de dados, Agregadores e usuários.

Por meio dos *Workflow* visualiza-se os momentos em que ocorrem a descrição do item informacional pelos arquivos e bibliotecas, o registro dos metadados, geração do URI nos registros, conversão para XML, implementação do protocolo OAI-PMH, o *harvesting*, o enriquecimento dos dados pelos vocabulários, a busca com SPARQL e aplicação do RDF.

#### 4.2 TRABALHOS FUTUROS

A *Web Semântica* ou a *Web* dos dados linkados é um projeto complexo, que exige a articulação de diversas áreas da comunidade científica, visto que a Ciência da Informação depende dos desenvolvedores de código da Ciência da Computação para implementar os *Websites* das bibliotecas, da Ciência da computação, da Ciência da Informação, dos bibliotecários e arquivistas para descrever os itens informacionais e concepção dos vocabulários e ontologias.

Existem muitas vertentes que podem ser exploradas na *Web Semântica*, visto que a tendência dos grandes portais é implementar uma solução de busca que possibilite ao usuário localizar o item de forma global, que o resultado seja rico em informações e interligado com diversos outros portais ou base de dados.

Com esse propósito, entende-se que desenvolver uma aplicação que empregue os conhecimentos da Ciência da Informação e Computação e possibilite realizar buscas em bases de dados orientados a documentos, aplicando as soluções de SPARQL, RDF, XML e a linguagem OWL para criar ontologias é uma possibilidade real e exequível, visto que muitos dessas tecnologias já são empregadas comercialmente.

### 4.3 CONSIDERAÇÕES FINAIS

Entende-se que o trabalho desenvolvido foi frutífero, visto que se estudou a *Web Semântica* sob vários aspectos, da arquivologia e biblioteconomia por parte da Ciência da Informação e sob o ângulo da Ciência da Computação.

Esse estudo permeou as duas áreas e proporcionou compreender como as ontologias e vocabulários dependem muito da tecnologia para serem desenvolvidos, visto que hoje existem ferramentas que facilitam a criação dessas estruturas e por outro lado para que os dados estejam estruturados na *Web*, os metadados exercem uma função determinante, sendo sua concepção e organização tarefa para os profissionais da Ciência da Informação.

As unidades de informação mediante a representação descritiva, temática e arquivística deverão formatar os tradicionais métodos de catalogação, normalizações, indexação, classificação e resumos em padrões recomendados pela W3C para estabelecimento da *Web Semântica*. Formatar no intuito de preparar os metadados dos itens informacionais para que sejam lidos por máquinas, tornem-se interoperáveis e carreguem sentido.

O emprego do modelo de descrição RDF é o fundamento para transformação dos dados bibliográficos das unidades de informação em dados ligados. Contudo, faz-se necessário também utilizar as linguagens XML como base para o RDF, as linguagens de desenvolvimento para criação das ontologias ou vocabulários, um banco de dados orientado a documentos e uma linguagem de busca fundamentado no RDF.

Por meio do estudo de caso da biblioteca virtual Europeia, pode-se demonstrar, com exemplos, a estrutura, os processos e as relações do funcionamento da *Web Semântica* nas unidades de informação. Os infográficos possibilitaram visualizar de forma gráfica os momentos determinantes, os papéis de cada ator, as aplicações empregadas para gestão e enriquecimento dos dados no ambiente da Europeia.

Por fim, espera-se que este estudo propicie aos arquivistas e bibliotecários compreender as fases necessárias para a implementação da *Web Semântica* nas unidades de informação para que os dados possam ser organizados, tratados, armazenados, preservados, recuperados e disseminados.

## REFERÊNCIAS

AACR2. **About ACCR2**. 2008. Disponível em: <<http://www.aacr2.org/about.html>>. Acesso em: 06 Jun. de 2015.

ALIC. **Archives Library Information Center Online Public Access Catalog**. 2015. Disponível em: <<http://www.archives.gov/research/alic/tools/online-public-access-catalog.html>>. Acesso em: 06 Jun. de 2015.

BARDIN, L. **Análise de Conteúdo**. 5.ed. Lisboa: Edições 70, 2010. 223p.

BARRETO, A. de A. A questão da informação. **São Paulo em Perspectiva**, São Paulo, v.8, n.4, p.3-8, out./dez. 1994. Disponível em: <<http://aldoibct.bighost.com.br/quest/quest2.pdf>>. Acesso em: 30 mar. 2014.

BARRETO, Aldo Albuquerque. A condição da informação. **São Paulo em Perspectiva**. São Paulo, v. 16, n. 3, p. 67-74, 2002.

BARROS, Rodolfo Miranda de. GAFFO, Fernando Henrique. **Gaia Riscos: Framework para Gerenciar os Riscos do Projeto por meio de Serviços**. 2012. Disponível em: <[http://gaia3.uel.br/projetos/gaia\\_riscos/#home](http://gaia3.uel.br/projetos/gaia_riscos/#home)>. Acesso em: 11 de ago. de 2015.

BELLOTO, H. L. **Arquivos permanentes: tratamento documental**. Rio de Janeiro: FGV, 2007.

BERNERS LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new Possibilities. **Scientific American**. Maio de 2001. Disponível em: <[http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American\\_%20Feature%20Article\\_%20The%20Semantic%20Web\\_%20May%202001.pdf](http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf)>. Acesso em: 21 mar 2014.

BERNERS-LEE, T; LASSILA, O; HENDLER, J. The Semantic Web. **Scientific American**, 2001. P. 34-43.

BOGDAN, R.J. **Grounds for cognition: How to goal-guided behavior shapes the mind**. Hillsdale, NJ: Lawrence Earlbaum, 1994.

BORKO, H. **Information science: what is it? American Documentation**, Jan. 1968. Disponível em: <<http://pt.scribd.com/doc/533107/Borko-H-v-19-n-1-p-35-1968>>. Acesso em: 08 mar. 2014.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: ENCONTRO NACIONAL DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo. ANCIB, 2008. Disponível em: <[http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%C3%89\(2008\)-1835.pdf](http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%C3%89(2008)-1835.pdf)>. Acesso em: 24 abr. 2014.

BRÄSCHER, Marisa; CARLAN, Eliana. Sistemas de organização do conhecimento: Antigas e novas linguagens. In: ROBREDO, Jaime; BRÄSCHER, Marisa. (Orgs.). **Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento**. Cap. 8. Brasília: DF: IBICT, 2010. Disponível em: <<http://www.ibict.br/publicacoes/eroic.pdf>>. Acesso em: 24 abr. 2014.

BRASIL. Arquivo Nacional. **Dicionário Brasileiro de Terminologia Arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. Disponível em: <<http://www.portalan.arquivonacional.gov.br/Media/Dicion%20Term%20Arquiv.pdf>>. Acesso em: 23 set. 2014.

\_\_\_\_\_. **NOBRADE: Norma Brasileira de Descrição Arquivística**. Rio de Janeiro: Arquivo Nacional, 2006. Disponível em: <<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/nobrade.pdf>>. Acesso em 09 jul. 2015.

BREITMAN, Karin Koogan. **Web Semântica: a internet do futuro**. Rio de Janeiro: LTC, 2010.

BURKE, Peter. **Uma história social do conhecimento: de Gutenberg a Diderot**. Rio de Janeiro: Zahar, 2003.

CAFÉ, L.; SALES, R. Organização da informação: conceitos básicos e breve fundamentação teórica. In: ROBREDO, J.; BRÄSCHER, M. (Org.). **Passeios no bosque da informação: estudos sobre representação e organização da informação e do conhecimento**. Brasília DF: IBICT, 2010, p. 115-129. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/36/1/eroic.pdf>>. Acesso em: 30 ago. 2014.

CAPURRO, Rafael. **Epistemology and Information Science**. Royal Institute of Technology Library, Stockholm, August 1985, Report TRITA-LIB-6023. Disponível em <<http://www.capurro.de/trita.htm>>. Acesso em 03 abr. 2014.

CAPURRO, Rafael; HJORLAND, Birger. O Conceito de Informação. **Perspectivas em Ciências da Informação**, v.12, n.1, p.148-207, jan./abr. 2007.

CATARINO, M. E. **Contribuições da Ciência da Informação para a Web Semântica**. 2013. 85p. Relatório de Projeto – Universidade Estadual de Londrina, Londrina.

\_\_\_\_\_. **Contribuições da Ciência da Informação para a Web Semântica**. 2013. 85p. Relatório de Projeto – Universidade Estadual de Londrina, Londrina.

\_\_\_\_\_. **Integração das folksonomias nos metadados: identificação de novos elementos como contributo para a descrição de recursos em repositórios**. 2009. 257f. Tese (Doutorado em Tecnologias e Sistemas de Informação) –Universidade do Minho. Disponível

em:<[http://repositorium.sdum.uminho.pt/bitstream/1822/9564/1/Tese\\_CatarinoMelisabete.pdf](http://repositorium.sdum.uminho.pt/bitstream/1822/9564/1/Tese_CatarinoMelisabete.pdf)>. Acesso em 07 set. 2014.

\_\_\_\_\_; SOUZA, T. B. A representação descritiva no contexto da *web* semântica. **TransInformação**, Campinas, v. 24, n. 2, p. 77-90, maio/ago., 2012. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewissue.php?id=33>>. Acesso em: 30 ago. 2014.

CHARLES, Valentine; DEVARENNE, Cécile. **Europeana enriches its data with the AAT**. 2015. Disponível em: <<http://pro.europeana.eu/europeana-aat/>>. Acesso em 15 maio 2015.

CHAUÍ, M. **Convite à Filosofia**. 12 ed. São Paulo: Ática, 2002.

CNPQ. **Áreas do conhecimento**: classificação. Brasília: CNPQ, 1984.

CODINA, Lluís; ROVINA, Cristòfol. La Web Semántica. In: TRAMULLAS, Jesús (Org.). **Tendencias em Documentacion Digital**. Guijón: Trea, 2006. Cap. 1, p. 9-54.

CRUZ MUNDET, J.R. **Manual de Archivística**. 2.ed. Madrid: Pirámide. 1996.

CUNHA, L. M. S. **Web Semântica**: um estudo preliminar. *Documentos/Embrapa Informática Agropecuária*, Campinas, v. 18, out. 2002. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/bitstream/doc/8670/1/doc18.pdf> >. Acesso em 30/01/2014.

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Brique de Lemos. 2008. 451p.

DBPEDIA. **O que é?** 2015. Disponível em: <<http://pt.dbpedia.org/pt>>. Acesso em 14 mai. 2015.

DCMI. **Sobre DCMI**. 2014. Disponível em : <<http://dublincore.org/about/history/>>. Acesso em: 07/09/2014. Disponível em: <<http://www.asocarchi.cl/DOCS/79.pdf>>. Acesso em: 20 jul. 2014.

EDM DEFINITION. **Definition of the Europeana Data Model v5.2.6**. 2014. Disponível em <[http://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation//EDM%20Definition%20v5.2.6\\_01032015.pdf](http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM%20Definition%20v5.2.6_01032015.pdf)>. Acesso em: 2 mar. 2015.

EFG. **The European Film Gateway**. 2011. Disponível em: <<http://www.efgproject.eu>>. Acesso em: 22 mar. 2015.

EIONET. **About GEMET – General Multilingual Environmental Thesaurus**. 2012. Disponível em: <<http://www.eionet.europa.eu/gemet/about?langcode=en>>. Acesso em 14 mai. 2015.

EUROPEANA AGGREGATORS. **Europeana Aggregators' Handbook**. 2010.

Disponível em:

<[http://version1.europeana.eu/c/document\\_library/get\\_file?uuid=94bcddb-3625-4e6d-8135-c7375d6bbc62&groupId=10602](http://version1.europeana.eu/c/document_library/get_file?uuid=94bcddb-3625-4e6d-8135-c7375d6bbc62&groupId=10602)>. Acesso em: 20 mar. de 2015.

EUROPEANA AGGREGATORS. **The Europeana Aggregator Handbook 2011, October**. 2011. Disponível em:

<[http://newpro.europeana.eu/files/Europeana\\_Professional/Publications/Aggregators%20Handbook.pdf](http://newpro.europeana.eu/files/Europeana_Professional/Publications/Aggregators%20Handbook.pdf)>. Acesso em: 20 mar. 2015.

EUROPEANA NETWORK. **EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy**: Final Report.2014. Disponível em:

<<http://pro.europeana.eu/europeanatech-task-forces/multilingual-and-semantic-enrichment-strategy>>. Acesso em 13 mai. 2015.

EUROPEANA PRO. **Provider FAQs**. 2015. Disponível em:

<<http://pro.europeana.eu/share-your-data/provider-faqs>>. Acesso em 14 mai. 2015.

EUROPEANA. **Europeana Libraries**. 2014a. Disponível em:<<http://www.europeana-libraries.eu/web/guest/home>>. Acesso em 25 maio 14.

EUROPEANA. **History of Europeana**. 2014b. Disponível em:

<<http://pro.europeana.eu/web/guest/history>>. Acesso em 25 maio 2014.

EUROPEANA. **Stone Figurine**. 2015. Disponível em:

<<http://www.europeana.eu/portal/record/04601/E2AAA3C6DF09F9FAA6F951FC4C4A9CC80B5D4154.html?start=6&query=cyprus+antiquities&startPage=1&qt=false&rows=24>>. Acesso em 02 Mar. 2014.

EUROPENA LABS. **Europeana Linked Open Data**. 2015a. Disponível em:

<<http://labs.europeana.eu/api/linked-open-data/introduction>>. Acesso em 28 abr. de 2015.

EUROPENA LABS. **Query Syntax**. 2015b. Disponível em:

<<http://labs.europeana.eu/api/query>>. Acesso em 28 abr. de 2015.

FARIA, Rogério Amorim de. **Treinamento Avançado em XML**. São Paulo: Digerati. 2005. 126p.

FEITOSA, Ailton. **Organização da informação na web**: das tags à web semântica. Brasília: Briquet de Lemos, 2006.

FERNANDES, José Palma. (Org). **Glossário da Sociedade da Informação**.

Portugal: APDSI, 2015. 114p.

FONSECA FILHO, Clézio. **História da Computação [recurso eletrônico ]**: O Caminho do Pensamento e da Tecnologia. Porto Alegre: EdiPUCRS, 2007. 205p.

FONSECA, Edson Nery da. **A Biblioteconomia brasileira no contexto mundial**. Rio de Janeiro: Tempo Brasileiro; Brasília: INL, 1979. 112p.

FOREST. **SPARQL Query**. 2014. Disponível em: <<http://europeana.ontotext.com/sparql>>. Acesso em 29 abr. de 2015.

FURGERI, Sérgio. O papel das linguagens de marcação para a Ciência da Informação. **TransInformação**. Campinas. Vol. 18, n.3. Set/Dez 2006. Disponível em <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/670>>. Acesso em 27 abr. 2014.

FUSTER RUIZ, F. Archivística, archivo, documento de archivo necesidad de clarificar los conceptos. **Anales de Documentación**, n. 2, p.103-120, 1999. Disponível em: <<http://revistas.um.es/analesdoc/article/view/2631/2611>>. Acesso em: 20 jul. 2012.

GADGET. In: **DICIONÁRIO Merriam-Webster**. Disponível em <<http://www.merriam-webster.com/dictionary/gadget>>. Acesso em 17 jun. 2014.

GEONAMES. **About**. 2015. Disponível em: <<http://www.geonames.org/about.html>>. Acesso em 14 maio 2015.

GETTY. **About the ATT**. 2014. Disponível em: <<http://www.getty.edu/research/tools/vocabularies/aat/about.html>>. Acesso em 15 maio 2015.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4.ed. São Paulo: Atlas, 2002. 176p.

GITHUB. **Environment**. 2014. Disponível em: <<https://github.com/europeana/portal/wiki/Environment>>. Acesso em 22 abr. 2015.

GITHUB. **Setup**. 2015. Disponível em: <<https://github.com/europeana/portal/wiki/Setup#PostGRES>>. Acesso em 23 abr.2015.

GÓMEZ-PÉREZ, A. **Ontological Engineering: A State of the Art**. Expert Update. Madri. P.33-44. 1999. Disponível em: <[http://oa.upm.es/6493/1/Ontological\\_Engineering\\_A\\_st.pdf](http://oa.upm.es/6493/1/Ontological_Engineering_A_st.pdf)>. Acesso em 05 maio 2014.

GUEGUEN, G. et al. Para um modelo conceitual internacional de descrição arquivística. **Acervo**. Rio de Janeiro, v.26, n2, 2013. Disponível em: <<http://revistaacervo.an.gov.br/seer/index.php/info/article/view/658>>. Acesso em 09 jul. 2015.

GUZMAN-LUNA; Jaime; PARDO, Durley Torres; GARCIA, Alba Nubia Lopez. Desarrollo de una ontología en el contexto de la web semántica a partir de un tesouro documental tradicional. **Revista Interamericana de Bibliotecología**.

Medellín, Colombia, v. 29, n. 2, jul./dez. 2006, p. 79-94. Disponível em: <<http://www.scielo.org.co/pdf/rib/v29n2/v29n2a05.pdf>>. Acesso em 24 abr. 2014.

HASLHOFER, Bernhard. ISAAC, Antoine; **The Europeana Linked Open Data Pilot**. 2011. Disponível em: <<http://dcpapers.dublincore.org/pubs/article/viewFile/3625/1851>>. Acesso em: 18 fev. 2015.

HILLMANN, D. **DCMI, Recommended Resource**. 2005. Disponível em: <<http://dublincore.org/documents/usageguide/>>. Acesso em 09 set. 2014.

IFLA. **International Standard Bibliographic Description**. 2015. Disponível em: <<http://www.ifla.org/publications/international-standard-bibliographic-description>>. Acesso em: 06 jun. 2015.

ISAAC, Antoine. CHARLES, Valentine; MAMAKIS, Yorgos; STILLER, Juliane. **Europeana Semantic Enrichment Framework**. 2014. Disponível em: <<https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y/edit?pli=1#>>. Acesso em 14 maio 2015.

\_\_\_\_\_; et al. **Europeana Data Model Primer**. 2011. Disponível em: <<http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5>>. Acesso em: 02 fev 2015.

\_\_\_\_\_; CLAYPHAN, Robina; HASLHOFER, Bernhard. Europeana: moving to Linked Open Data. **SQ Information Standards Quarterly**. Vol. 42, n° 2/3 (Spring/Summer 2012). Disponível em: <[http://www.niso.org/apps/group\\_public/download.php/9407/IP\\_Isaac-et-al\\_Europeana\\_isqv24no2-3.pdf](http://www.niso.org/apps/group_public/download.php/9407/IP_Isaac-et-al_Europeana_isqv24no2-3.pdf)>. Acesso em: 18 jan. 2015.

ISO 2709. **Information and documentation – Format for Information Exchange**. Disponível em: <[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=41319](http://www.iso.org/iso/catalogue_detail.htm?csnumber=41319)>. Acesso em: 27 mar. 2015.

ISO/DIS 25577. **Information and Documentation – MarcXchange**. 2006. Disponível em: <[http://www.loc.gov/standards/iso25577/ISO\\_DIS\\_25577\\_\\_E\\_.pdf](http://www.loc.gov/standards/iso25577/ISO_DIS_25577__E_.pdf)>. Acesso em: 27mar. 2015.

JAVA. **Obtenha Informações sobre a tecnologia Java**. 2015. Disponível em: <[https://www.java.com/pt\\_BR/about/](https://www.java.com/pt_BR/about/)>. Acesso em 23 abr. 2015.

JOINT STEERING COMMITTEE. **RDA: Resource Description and Access**. 2009. Disponível em: <<http://www.rda-jsc.org/rdaprospectus.html>>. Acesso em: 14 abr. 2015.

JOHNSON, R. E.; FOOTE, B. Designing reusable classes. *Journal of Object Oriented Programming*. p. 22-35. 1998.

KIRÁLY, Peter. **Query Translation in Europeana**. 2015. Code4Lib. Chicago. 21 Jan de 2015. Ed.28. . Disponível em: <<http://journal.code4lib.org/articles/10285>>. Acesso em 29 abr. de 2015.

LAGO, Elzula Santos do. **Desmitificando a classificação documentária: CDD e CDU**. Teresina: EDUFPI, 2009. 72p.

LAGOZE, Carl; SOMPEL, Herbert Van de. **OPEN ARCHIVES INITIATIVE OBJECT REUSE AND EXCHANGE**. 2008a. Disponível em: <<http://www.openarchives.org/ore/documents/ore-production-press-release.pdf> >. Acesso em 14 fev. 2015.

\_\_\_\_\_; SOMPEL, Herbert Van de. **ORE User Guide - Primer**. 2008b. Disponível em: <<http://www.openarchives.org/ore/1.0/primer>>. Acesso em 14 fev. 2015.

LANCASTER, F.W. **Indexação e resumo: teoria e prática**. Brasília. Briquet de Lemos. 1993. 347p.

LE COADIC, Y.-F. **A Ciência da Informação**. Brasília: Briquet de Lemos, 2004. 124p.

LEVY, Steven. **Google a Biografia**. São Paulo: Universo dos Livros. 2012. 464p.

LIMA-MARQUES, Mamede. **Ontologia: Da Filosofia à Representação do Conhecimento**. Thesaurus, 2006. 69p.

LINKED HERITAGE. **AnnoCultor**. 2011. Disponível em:<<http://www.athenaeurope.org/athenawiki/index.php/AnnoCultor> >. Acesso em: 14 abr. 2015.

LUDKE MENGA, André; MARLI, E. D. **Pesquisa em Educação: Abordagem Qualitativa**. São Paulo: EPU, 1986.

MAANEN, John, Van. Reclaiming qualitative methods for organizational research: a preface. In **Administrative Science Quarterly**, vol. 24, no. 4, December 1979a, pp. 520-526.

MARC21. **The MARC 21 Formats: Background and Principles**. 2008. Disponível em: <<http://www.loc.gov/marc/96principl.html>>. Acesso em: 06 Jun. 2015.

MARCONI, Marina de A.; LAKATOS, Eva Maria. **Metodologia Científica**. São Paulo: Atlas, 2008. 321p.

MARXML. **MARC 21 XML Schema**. Disponível em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 27 mar. 2015.

MONGODB. **Data Model Introduction**. 2015b. Disponível em: <[http://docs.mongodb.org/manual/core/data-modeling-introduction/?\\_ga=1.205795029.716352465.1428861889](http://docs.mongodb.org/manual/core/data-modeling-introduction/?_ga=1.205795029.716352465.1428861889)>. Acesso em 23 abr. 2015.

MONGODB. **Introduction to MongoDB**. 2015a. Disponível em: <<http://docs.mongodb.org/manual/core/introduction/>>. Acesso em 23 Abr. de 2015.

NASCIMENTO, Aline Vieira do; CRUZ, Emilia Barroso; TERRA, Leticia Labati; SILVA, Marina Ramalho; COSTA, Alexandre de Souza. O uso do método Estudo de Caso na Ciência da Informação no Brasil. InCID: **Revista de Ciência da Informação e Documentação**, v. 4, p. 49-69, 2013.

NEVES, J.L. Pesquisa Qualitativa – Características, Usos e Possibilidades. **Caderno de pesquisas em Administração**, São Paulo, V1, N° 3, 2º Sem. 1996. Disponível em <<http://www.ead.fea.usp.br/cad-pesq/arquivos/c03-art06.pdf>>. Acesso em 08 fev. 2014.

OLIVEIRA, Renan Rodrigues de. CARVALHO, Cedric Luiz de. **Implementação de Interoperabilidade entre Repositórios Digitais por Meio do Protocolo OAI-PMH**. 2009. 55p. Relatório Técnico – Universidade Federal de Goiás. Disponível em: <[http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_003-09.pdf](http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-09.pdf)>. Acesso em: 26 mar. 2015.

OLIVEIRA, Renato da Silva. **Minidicionário compacto de informática**. São Paulo: Editora Ridel. 1999. 354p.

OPEN ARCHIVES. **OAI for Beginners: Overview**. Disponível em: <<http://www.oaforum.org/tutorial/english/page1.htm>>. Acesso em: 25 mar. 2015.

OPEN ARCHIVES. **Sobre PMH**. Disponível em <<http://www.openarchives.org/pmh/>>. Acesso em: 07 set. 2014.

OPEN SOURCE INITIATIVE. **The Open source Initiative**. 2006. Disponível em: <<http://opensource.org/osd>>. Acesso em 21 abr. de 2015.

ORTEGA, C. D. **Do princípio monográfico à unidade documentária**: Exploração do Fundamentos da Catalogação. Liinc em Revista, v7, n.1, p.43-60, 2011. Disponível em < <http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/402/263> >. Acesso em 30 ago. 2014.

PAES, M. L. **Arquivo**: teoria e prática: Introdução aos estudos dos arquivos. 3. ed. Revista e Ampliada. Rio de Janeiro: FGV, 2002.

PATRÍCIO, Helena S. A Europeia e a agregação de metadados na Web: análise dos esquemas ESE/EDM e da aplicação de standards da web semântica a dados de bibliotecas. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 11, 2012, Lisboa. **Anais...** Lisboa. Biblioteca Nacional de Portugal, 2012.

PEDROSA, Gilberto; GEORG, Petz; CONCORDIA, Cesare; ALOIA, Nicola. **EUROPEANA CONECT**: Europeana OAI-PMH Infrastructure – Documentation and final propotype. Austria. 2010. Disponível em:

<[http://www.europeanaconnect.eu/documents/01a\\_Europeana\\_OAI\\_PMH\\_APPENDIX\\_User%20Manual.pdf](http://www.europeanaconnect.eu/documents/01a_Europeana_OAI_PMH_APPENDIX_User%20Manual.pdf)>. Acesso em: 20 mar. 2015

PINTO, M. C. M. F. Análise e representação de assuntos em sistemas de recuperação da informação: linguagens de indexação. **Revista da Escola de Biblioteconomia da UFMG**. Belo Horizonte, v. 14, n. 2, p. 169-186, set. 1985. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/reb/>>. Acesso em: 30 ago. 2014.

POLLOCK, Jeffrey T. **Web Semântica para Leigos**. Rio de Janeiro: Alta Books, 2010. 402p.

POSTGRESQL. **About**. 2015. Disponível em: <<http://www.postgresql.org/about/>>. Acesso em 23 abr. 2015.

QIN, J.; PALING, S. **Converting a controlled vocabulary into an ontology: the case of GEM**. Information Research, v. 6, no. 2, 2001. Disponível em: <<http://www.informationr.net/ir/6-2/paper94.html>>. Acesso em: 25 maio 2015.

RATANAJAIPAN, P.; NANTAJEEWARAWAT, E.; WUWONGSE, V. OWL/XDD: **A formal language for application profiles**. EIECE Transactions, 2007, p. 1611-1620.

ROBREDO, J. Ciência da informação e Web Semântica: Linhas convergentes ou linhas paralelas?. In: ROBREDO, J.; BRÄSCHER, M. (Org.). **Passeios no bosque da informação: estudos sobre representação e organização da informação e do conhecimento**. Brasília DF: IBICT, 2010, p. 13-47.

\_\_\_\_\_. **Aindexação automática de textos: o presente já entrou no futuro**. In: Machado, U.D. (editor). Estudos Avançados em Biblioteconomia e Ciência da Informação. Volume I. Brasília: ABDF, 1982.

\_\_\_\_\_; BRÄSCHER, Marisa. (Org.). Estudos sobre a representação e organização da informação e do conhecimento. In: **Passeios pelo Bosque da Informação**:. Brasília: IBICT, 2010. p. 2-40. Disponível em: <[http://repositorio.unb.br/bitstream/10482/7952/6/CAPITULO\\_CiencialInformacaoWebSemantica.pdf](http://repositorio.unb.br/bitstream/10482/7952/6/CAPITULO_CiencialInformacaoWebSemantica.pdf)>. Acesso em: 20 abr. 2014.

RUBI, M. P. **Política de indexação para construção de catálogos coletivos em bibliotecas universitárias**. 2008. 169 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista, Marília.

SALES, Rodrigo de; CAFÉ, Lígia. Semelhanças e diferenças entre tesauros e ontologias. **DataGramZero**, Rio de Janeiro, v. 9, n. 4, ago. 2008a. Disponível em: <[http://www.dgz.org.br/ago08/F\\_I\\_art.htm](http://www.dgz.org.br/ago08/F_I_art.htm)>. Acesso em: 24 abr. 2014.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.1, n.1, p. 41-62, jan./jun. 1996.

SBC. **Sociedade Brasileira de Computação**. *Diretrizes Curriculares – MEC Consulta Pública – versão 2003*. Disponível em: <[http://www.sbc.org.br/index.php?option=com\\_jdownloads&Itemid=195&task=finish&cid=186&catid=36](http://www.sbc.org.br/index.php?option=com_jdownloads&Itemid=195&task=finish&cid=186&catid=36)>. Acesso em: 01 dez. 2014.

SETZER, V.W. Dado, informação, conhecimento e competência. **DataGramaZero**. Rio de Janeiro: nº zero, Dez/1999. Disponível em <[http://www.dgz.org.br/dez99/Art\\_01.htm](http://www.dgz.org.br/dez99/Art_01.htm) >. Acesso em 05 abr. 2014.

SILVA, J. A. **A norma internacional de descrição arquivística: estágio de discussão no Brasil**. 1999. Disponível em: <<http://www.archivonacional.go.cr/memorias/1999/01.pdf> >. Acesso em 09 jul. 2015.

SILVA, J. S. da; GUIMARÃES, J. A. C. As relações teóricas da Arquivologia e a Ciência da Informação. In: CONGRESSO DE ARQUIVOLOGIA DO MERCOSUL, 7., 2007, Viña del Mar. **Anais....** Congresso de Arquivologia do Mercosul, 2007.

SILVA, M. dos R. da; FUJITA, M. S. L. A prática de indexação: análise evolutiva de tendências teóricas e metodológicas. **Transinformação**, Campinas, v. 16, n. 2, p. 133-161, 2004.

SILVINO FILHO, J. **Avaliação de documentos de arquivo**. Brasília: CORBI, 1996.

SMITH, Michael K.; WELTY, Chris; MCGUINNESS, Deborah L. **OWL Web Ontology Language Guide**. 2004. Disponível em < <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>>. Acesso em 06 out. 2014.

SOLR. **Solr Features**. 2014. Disponível em: <<http://lucene.apache.org/solr/features.html>>. Acesso em 28 abr. 2015.

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

SOUZA, Cleidson Ronald Botelho. **Um Framework para editores de diagramas cooperativos baseados em anotações**. 1998. 105f. Dissertação (Mestrado em Ciência da Computação) - Universidade Estadual de Campinas, Campinas. Disponível em: <<http://www.bibliotecadigital.unicamp.br/document/?code=000135377&fd=y>>. Acesso em: 04 de ago. de 2015.

SOUZA, Marcos Aurélio Lopes de; RODRIGUES, Georgete Medleg. O Arquivista como cientista da informação: formação e perspectivas profissionais. Identificação dos conteúdos curriculares e das definições dos papéis e funções da profissão de

arquivista na literatura em língua inglesa e francesa. **Programa de Iniciação Científica da UNB**. Brasília, 2002.

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, jan./abr. 2004.

SOUZA, R.F. A Classificação como Interface da Internet. **DatagramaZero**. Rio de Janeiro: v.2, n.2. abr/00. Disponível em: <[http://www.dgz.org.br/abr00/Art\\_01.htm](http://www.dgz.org.br/abr00/Art_01.htm)>. Acesso em: 05 set. 2014.

SOUZA, T.B. **O ensino de representação descritiva nos cursos da área de ciência da informação no Brasil e em Portugal: estudo comparativo**. 2009. Tese (Doutorado em Ciências Documentais) - Faculdade de Letras, Universidade do Porto, Porto.

SPARQL. **SPARQL 1.1 Query Language**. 2013. Disponível em: <<http://www.w3.org/TR/sparql11-query/>>. Acesso em 29 abr. de 2015.

SVENONIUS, Elaine. **Classification theory**. Los Angeles, 1985, mimeo.

TAKAHASHI, T. (Org.). Sociedade da Informação no Brasil: **O livro verde**. Brasília: Ministério da Ciência e Tecnologia, 2000. Disponível em: <[www.governoeletronico.gov.br/biblioteca/arquivos/livro-verde/download](http://www.governoeletronico.gov.br/biblioteca/arquivos/livro-verde/download)>. Acesso em 29 abr. de 2015.

TEL. **The European Library**. 2013. Disponível em: <<http://www.theeuropeanlibrary.org/tel4/aboutus>>. Acesso em: 22 mar. 2015.

THE EDM MAPPING GUIDELINES. **Europeana Data Model – Mapping Guidelines v2.2**. 2014. Disponível em: <[http://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Mapping\\_Guidelines\\_v2.2.pdf](http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.2.pdf)>. Acesso em 6 mar. 2015.

TZANAKIS, Simon. **Introducing REPOX: a tool to manage metadata spaces**. 2015. Disponível em: <<http://pro.europeana.eu/blogpost/introducing-repoxa-tool-to-manage-metadata-spaces>>. Acesso em: 31 mar. 2015.

WORLD WIDE WEB CONSORTIUM. **SKOS SIMPLE KNOWLEDGE ORGANIZATION SYSTEM PRIMER**. 2009. Disponível em: <<http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>>. Acesso em: 14 abr. 2015.

\_\_\_\_\_. **Metadata and Resource Description**. 2001. Disponível em: <<http://www.w3.org/Metadata/>>. Acesso em 31 jan. 2014.

\_\_\_\_\_. **Query**. 2013a. Disponível em: <<http://www.w3.org//standards/semanticweb/query>>. Acesso em 05 out. 2014.

\_\_\_\_\_. **RDF 1.1 Primer**. 2014a. Disponível em <http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624>. Acesso em 22 set. 2014.

\_\_\_\_\_. **Sobre W3C**. 2011a. Disponível em < <http://www.w3c.br/Sobre/>>. Acesso em 31 jan. 2014.

\_\_\_\_\_. **Vocabularies**. 2013b. Disponível em: <<http://www.w3.org/standards/semanticweb/ontology>>. Acesso em 04 out. 2014.

\_\_\_\_\_. **W3C Brasil**. 2014b. Disponível em <<http://www.w3c.br/Home/WebHome>>. Acesso em 31 jan. 2014.

\_\_\_\_\_. **Web Semântica**. 2011b. Disponível em <<http://www.w3c.br/Padroes/WebSemantica>>. Acesso em 31 jan. 2014.

WURMAN, Richard Saul. **Como transformar informação em compreensão**. São Paulo: Cultura: Editores Associados, 1991. 380 p.

YIN, Robert K. **Estudo de Caso - Planejamento e Métodos**. 3. ed. Porto Alegre. Bookman. 2005. 212p.

\_\_\_\_\_. **Estudo de Caso - Planejamento e Métodos**. 4. ed. Porto Alegre. Bookman. 2010. 248p.

## APÊNDICES

## APÊNDICE A

### As classes núcleo

Propriedades do Provided Cultural Heritage Object (edm:ProvidedCHO)

Os valores nessa propriedade relacionam-se com o objeto de patrimônio cultural original, não com a sua representação digital. Por exemplo os atributos de Monalisa e não sua imagem digitalizada.

Property	Note	Value type
<b>dc:contributor</b>	<p>Use for contributors to the CHO. If possible supply the identifier of the contributor from an authority source.</p> <p>Providers with richer role terms can elect to map a subset to dc:contributor and others to dc:creator. Repeat for multiple contributors.</p> <p>&lt;dc:contributor&gt;Maria Callas&lt;/dc:contributor&gt;</p> <p>or create a reference to an instance of the Agent class</p> <p>&lt;dc:contributor rdf:resource="http://www.example.com/MariaCallas"/&gt;</p>	literal or ref
<b>dc:coverage</b>	<p>The spatial or temporal topic of the CHO. Use the more precise dcterms:spatial or dcterms:temporal properties if the data will support it. One of dc:coverage or dc:subject or dc:type or dcterms:spatial <b>must</b> be provided.</p> <p>&lt;dc:coverage&gt;1995-1996&lt;/dc:coverage&gt; or &lt;dc:coverage&gt;Berlin&lt;/dc:coverage&gt;</p> <p>or create a reference to an instance of a contextual class, for example, a Place class</p> <p>&lt;dc:coverage rdf:resource="http://sws.geonames.org/2950159"/&gt;</p>	literal or ref
<b>dc:creator</b>	<p>For the creator of the CHO. If possible supply the identifier of the creator from an authority source. Repeat for multiple creators.</p> <p>&lt;dc:creator&gt;Shakespeare, William&lt;/dc:creator&gt;</p> <p>or create a reference to an instance of the Agent class</p> <p>&lt;dc:creator rdf:resource="http://viaf.org/viaf/96994048"/&gt;</p>	literal or ref
<b>dc:date</b>	<p>Use for a significant date in the life of the CHO. Consider using the sub-properties of dcterms:created or dcterms:issued.</p> <p>&lt;dc:date&gt;Early 20th century&lt;/dc:date&gt; or &lt;dc:date&gt;1919&lt;/dc:date&gt;</p> <p>or create a reference to an instance of the TimeSpan class</p> <p>&lt;dc:date rdf:resource="http://semium.org/time/19xx_1_third"/&gt;</p>	literal or ref

<b>dc:description</b>	<p>A description of the CHO. Either dc:description or dc:title <b>must</b> be provided.</p> <p>&lt;dc:description&gt;Illustrated guide to airport markings and lighting signals, with particular reference to SMGCS (Surface Movement Guidance and Control System) for airports with low visibility conditions.&lt;/dc:description&gt;</p>	literal or ref
<b>dc:format</b>	<p>Use for the terms generally applied to indicate the format of the cultural heritage object or the file format of a born digital object. Use the value "3D-PDF" if appropriate.</p> <p>&lt;dc:format&gt;paper&lt;/dc:format&gt;</p>	literal or ref
<b>dc:identifier</b>	<p>An identifier of the original CHO.</p> <p>&lt;dc:identifier&gt;urn:isbn:9780387097466&lt;/dc:identifier&gt;</p>	literal
<b>dc:language</b>	<p>The language of text CHOs and also for other types of CHO if there is a language aspect. <b>Mandatory</b> for TEXT objects, strongly recommended for other object types with a language element. Best practice is to use ISO 639 two- or three-letter primary language tags<sup>3</sup>. Repeat for multiple languages.</p> <p>&lt;dc:language&gt;it&lt;/dc:language&gt;</p>	literal
<b>dc:publisher</b>	<p>The name of the publisher of the CHO. If possible supply the identifier of the publisher from an authority source.</p> <p>&lt;dc:publisher&gt;Oxford University Press&lt;/dc:publisher&gt; or create a reference to an instance of the Agent class &lt;dc:publisher rdf:resource="http://www.oup.com/"&gt;</p>	literal or ref
<b>dc:relation</b>	<p>The name or identifier of a related resource, generally used for other related CHOs. Cf edm:isRelatedTo.</p> <p>&lt;dc:relation&gt;maps.crace.1/33&lt;/dc:relation&gt; (Shelf mark)</p> <p>Or to provide a link to another object:</p> <p>&lt;dc:relation rdf:resource="http://www.identifier/relatedObject"/&gt;</p>	literal or ref
<b>dc:rights</b>	<p>Use to give the name of the rights holder of the CHO if possible or for more general rights information. (Note that the controlled edm:rights property relates to the digital objects and applies to the edm:WebResource and/or edm:Aggregation).</p> <p>&lt;dc:rights&gt;Copyright © British Library Board&lt;/dc:rights&gt;</p>	literal or ref
<b>dc:source</b>	<p>A related resource from which the described resource is derived in whole or in part i.e. the source of the original CHO. (Not the name of the content holder: for this see edm:dataProvider .)</p> <p>&lt;dc:source&gt;Security Magazine pp 3-12&lt;/dc:source&gt;</p>	literal or ref
<b>dc:subject</b>	<p>The subject of the CHO. One of dc:subject or dc:coverage or dc:type or dcterms:spatial must be provided</p> <p>&lt;dc:subject&gt;trombone&lt;/dc:subject&gt;</p> <p>or create a reference to an instance of the Concept class</p> <p>&lt;dc :subject rdf:resource="http://id.loc.gov/authorities/subjects/sh85137992"/&gt;</p>	literal or ref

	>	
<b>dc:title</b>	The title of the CHO. Either dc:title or dc:description must be provided. Exact translations of the title can be provided using appropriate xml language attributes <dc:title xml:lang="en">Eight Weeks</dc:title> <dc:title xml:lang="it">Ocho semanas</ dc:title>	literal
<b>dc:type</b>	The nature or genre of the CHO. Ideally the term(s) will be taken from a controlled vocabulary. One of dc:type or dc:subject or dc:coverage or dcterms:spatial must be provided. <dc:type>Book</dc:type> or <dc:type>trombone</dc:type> or create a reference to an instance of the Concept class <dc:type rdf:about="http://www.mimo-db.eu/HornbostelAndSachs/356/">	literal or ref
<b>dcterms:alternative</b>	Any alternative title of the CHO including abbreviations or translations that may not be exact. <dcterms:alternativexml:lang="en">Eight weeks: a novel</dcterms:alternative>	literal
<b>dcterms:conformsTo</b>	An established standard to which the CHO conforms. <dcterms:conformsTo>W3C WCAG 2.0</dcterms:conformsTo> (conforms to web content accessibility guidelines). Or link to the resource <dcterms:conformsTo rdf:resource="http://www.w3.org/TR/WCAG/" />	literal or ref
<b>dcterms:created</b>	The date of creation of the CHO. <dcterms:created>Mid 16th century</dcterms:created> or <dcterms:created>1584</dcterms:created> or create a reference to an instance of the TimeSpan class <dcterms:created rdf:resource="http://semium.org/time/15xx_3_third"/>	literal or ref
<b>dcterms:extent</b>	The size or duration of the CHO. <dcterms:extent>13 cm</dcterms:extent> (the width of an original object). <dcterms:extent>34 minutes</dcterms:extent> (the duration of an audio file).	literal or ref
<b>dcterms:hasFormat</b>	A resource related to the CHO that is substantially the same as the CHO but in another format. <dcterms:hasFormat>http://upload.wikimedia.org/wikipedia/en/f/f3/Europeana_logo.png</dcterms:hasFormat> for a png image file of the described tiff resource Or as a link to a resource <dcterms:hasFormat	literal or ref

	<p>rdf:resource="http://upload.wikimedia.org/wikipedia/en/f/f3/Europeana_logo.png"/&gt;</p>	
<b>dcterms:hasPart</b>	<p>A resource that is included either physically or logically in the CHO.</p> <p>&lt;dcterms:hasPart&gt;Vol.2. Issue 1&lt;/dcterms:hasPart&gt;.</p>	literal or ref
<b>dcterms:hasVersion</b>	<p>Another, later resource that is a version, edition or adaptation of the CHO demonstrating substantive changes in content rather than format.</p> <p>&lt;dcterms:hasVersion&gt;The Sorcerer's Apprentice (translation by Edwin Zeydel, 1955)&lt;/dcterms:hasVersion&gt;. In this example the 1955 translation is a version of the described resource.</p>	literal or ref
<b>dcterms:isFormatOf</b>	<p>Another resource that is substantially the same as the CHO but in another format.</p> <p>&lt;dcterms:isFormatOf&gt;Europeana_logo.tiff&lt;/dcterms:isFormatOf&gt; where the resource being described is a png image file.</p>	literal or ref
<b>dcterms:isPartOf</b>	<p>A resource in which the CHO is physically or logically included. This property can be used for objects that are part of a hierarchy and will be used to support an appropriate display in the portal. For that purpose it will be necessary to supply a reference as the value..</p> <p>&lt;dcterms:isPartOf&gt;Crace Collection of Maps of London&lt;/dcterms:isPartOf&gt;</p> <p>Or link to parent object if part of a hierarchy of CHOs</p> <p>&lt;dcterms:isPartOf rdf:resource="http://data.europeana.eu/item/08701/1B0BACAA44D5A807E43D9B411C9781AAD2F96E65"/&gt;</p>	literal or ref
<b>dcterms:isReferenced By</b>	<p>Another resource that references, cites or otherwise points to the CHO.</p> <p>&lt;dcterms:isReferencedBy&gt;Till, Nicholas (1994) Mozart and the Enlightenment: Truth, Virtue and Beauty in Mozart's Operas, W. W. Norton &amp; Company &lt;/dcterms:isReferencedBy&gt;</p>	literal or ref
<b>dcterms:isReplacedBy</b>	<p>Another resource that supplants , displaces, or supersedes the CHO.</p> <p>&lt;dcterms:isReplacedBy&gt;http://dublincore.org/about/2009/01/05/bylaws/&lt;/dcterms:isReplacedBy&gt; where the resource described is an older version (http://dublincore.org/about/2006/01/01/bylaws/)</p> <p>or link</p> <p>&lt;dcterms:isReplacedBy rdf:resource="http://dublincore.org/about/2009/01/05/bylaws"/&gt;</p>	literal or ref
<b>dcterms:isRequiredBy</b>	<p>Another related resource that requires the CHO to support its function, delivery or coherence</p> <p>&lt;isRequiredBy&gt;http://www.myslides.com/myshow.ppt&lt;/isRequiredBy&gt; where the image being described is required for an</p>	literal or ref

	online slideshow.	
<b>dcterms:issued</b>	Date of formal issuance or publication of the CHO. <dcterms:issued>1993</dcterms:issued> or create a reference to an instance of the TimeSpan class <dcterms:issued rdf:resource="http://semium.org/time/17xx_3_third"/> (late 18th century)	literal or ref
<b>dcterms:isVersionOf</b>	Another, earlier resource of which the CHO is a version, edition or adaptation, demonstrating substantive changes in content rather than format. <dcterms:isVersionOf>The Sorcerer's Apprentice<dcterms:isVersionOf>. In this example The Sorcerer's Apprentice (translation by Edwin Zeydel, 1955) is the resource being described.	literal or ref
<b>dcterms:medium</b>	The material or physical carrier of the CHO. <dcterms:medium>metal</dcterms:medium>	literal or ref
<b>dcterms:provenance</b>	A statement of changes in ownership and custody of the CHO since its creation. Significant for authenticity, integrity and interpretation. <dcterms:provenance>Donated to The National Library in 1965</dcterms:provenance>	literal or ref
<b>dcterms:references</b>	Other resources referenced, cited or otherwise pointed to by the CHO. <dcterms:references>Honderd jaar Noorse schilderkunst</dcterms:references>	literal or ref
<b>dcterms:replaces</b>	A related resource that is supplanted, displaced, or superseded by the CHO. <dcterms:replaces>http://dublincore.org/about/2006/01/01/bylaws/</dcterms:replaces> where the resource described is a newer version (http://dublincore.org/about/2009/01/05/bylaws/) or link to resource <dcterms:replaces rdf:resource="http://dublincore.org/about/2006/01/01/bylaws"/>	literal or ref
<b>dcterms:requires</b>	Another resource that is required by the described resource to support its function, delivery or coherence. <dcterms:requires>http://ads.ahds.ac.uk/project/userinfo/css/old browsers.css</dcterms:requires> where the resource described is an HTML file at http://ads.ahds.ac.uk/project/userinfo/digitalTextArchiving.html	literal or ref
<b>dcterms:spatial</b>	Spatial characteristics of the CHO. i.e. what the CHO represents or depicts in terms of space (e.g. a location, co-ordinate or place). Either dcterms:spatial or dc:type or dc:subject or dc:coverage must be provided.	literal or ref

	<p>&lt;dcterms:spatial&gt;Portugal&lt;/dcterms:spatial&gt; or create a reference to an instance of the Place class</p> <p>&lt;dcterms:spatial rdf:resource="http://sws.geonames.org/2264397"/&gt;</p>	
<b>dcterms:tableOfContents</b>	<p>A list of sub-units of the CHO.</p> <p>&lt;dcterms:tableOfContents&gt;Chapter 1. Introduction, Chapter 2. History &lt;/dcterms:tableOfContents&gt;</p>	literal
<b>dcterms:temporal</b>	<p>Temporal characteristics of the CHO. i.e. what the CHO is about or depicts in terms of time (e.g. a period, date or date range.)</p> <p>&lt;dcterms:temporal&gt;Roman Empire&lt;/dcterms:temporal&gt; or create a reference to an instance of the TimeSpan class</p> <p>&lt;dcterms:temporal rdf:resource="http://semium.org/time/roman_empire"/&gt;</p>	literal or ref
<b>edm:currentLocation</b>	<p>The geographic location whose boundaries presently include the CHO. If the name of a repository, building, site, or other entity is used then it should include an indication of its geographic location.</p> <p>&lt;edm:currentLocation rdf:resource="http://sws.geonames.org/2950159"/&gt; (Identifier for Berlin)</p>	ref
<b>edm:hasMet</b>	<p>The identifier of an agent, a place, a time period or any other identifiable entity that the CHO may have "met" in its life.</p> <p>&lt;edm:hasMet rdf:resource="http://viaf.org/viaf/96994048"/&gt; (Identifier for William Shakespeare)</p> <p>&lt;edm:hasMet rdf:resource="http://sws.geonames.org/6620265"/&gt;(location identifier for Shakespeare's Globe theatre.)</p>	ref
<b>edm:hasType</b>	<p>The identifier of a concept, or a word or phrase from a controlled vocabulary (thesaurus etc) giving the type of the CHO. E.g. Painting from the AAT thesaurus. This property can be seen as a super-property of e.g. dc:format or dc:type to support "What" questions.</p> <p>&lt;edm:hasType&gt;Painting&lt;/edm:hasType&gt;</p>	ref or literal
<b>edm:incorporates</b>	<p>The identifier of another resource that is incorporated in the described CHO. E.g. the movie "A Clockwork Orange" incorporates Rossini's La Gazza Ladra" in its soundtrack.</p> <p>&lt;edm:incorporates rdf:resource="http://www.identifier/IncorporatedResource"/&gt;</p>	ref
<b>edm:isDerivativeOf</b>	<p>The identifier of another resource from which the described CHO has been derived. E.g. the identifier of Moby Dick when the Italian translation is the described CHO.</p> <p>&lt;edm:isDerivativeOf rdf:resource="http://www.identifier/SourceResource"/&gt;</p>	ref

<b>edm:isNextInSequence</b>	<p>The identifier of the preceding object where both objects are part of the same overall resource. Use this for objects that are part of a hierarchy or sequence to ensure correct display in the portal.</p> <pre>&lt;edm:isNextInSequence rdf:resource="http://www.identifier/PrecedingResource"/&gt;</pre>	ref
<b>edm:isRelatedTo</b>	<p>The identifier or name of a concept or other resource to which the described CHO is related. E.g. Moby Dick is related to XIX Century literature. Cf dc:relation.</p> <pre>&lt;edm:isRelatedTo&gt;Literature&lt;/edm:isRelatedTo &gt;</pre> <p>Or link to resource</p> <pre>&lt;edm:isRelatedTo rdf:resource="http://www.eionet.europa.eu/gemet/concept?cp=4850"/&gt;</pre>	ref or literal
<b>isRepresentationOf</b>	<p>The identifier of another object of which the described CHO is a representation. E.g. the identifier of the statue when the CHO being described is a painting of that statue.</p> <pre>&lt;edm:isRepresentativeOf rdf:resource="http://www.identifier/RepresentedResource"/&gt;</pre>	ref
<b>edm:isSimilarTo</b>	<p>The identifier of another resource to which the described CHO is similar.</p> <pre>&lt;edm:isSimilarTo rdf:resource="http://www.identifier/SimilarResource"/&gt;</pre>	ref
<b>edm:isSuccessorOf</b>	<p>The identifier of a resource to which the described CHO is a successor. E.g. "The Two Towers" is a successor of "Fellowship of the Ring".</p> <pre>&lt;edm:isSuccessorOf rdf:resource="http://dbpedia.org/resource/The_Fellowship_of_the_Ring"/&gt;</pre>	ref
<b>edm:realizes</b>	<p>If the CHO described is of type edm:PhysicalThing it may realize an information object. E.g. a copy of the Gutenberg publication realizes the Bible.</p> <pre>&lt;edm:realizes rdf:resource="http://www.identifier/PhysicalThing"/&gt;</pre>	ref
<b>edm:type</b>	<p>The value must be one of the types accepted by Europeana as it will support portal functionality :</p> <p>TEXT, VIDEO, SOUND, IMAGE, 3D. (For 3D, when applicable, use the value "3D-PDF" in dc:format )</p> <pre>&lt;edm:type&gt;IMAGE&lt;/edm:type&gt; (upper-case)</pre> <pre>&lt;edm:type&gt;3D&lt;/edm:type&gt; (upper-case)</pre>	literal
<b>owl:sameAs</b>	<p>Use to point to your own (linked data) representation of the object, if you have already minted a URI identifier for it. It is also possible to provide URIs minted by third-parties for the object.<sup>5</sup></p> <pre>&lt;owl:sameAs rdf:resource="http://www.identifier/SameResourceElsewhere"/&gt;</pre>	ref

### Propriedades da classe Web Resource (edm:WebResource)

Estas são as propriedades que podem ser utilizadas para a classe EDM:WebResource. Isso significa que eles são atributos de uma representação digital de um objeto do patrimônio cultural e não o próprio objeto.

Pode haver mais de um edm:WebResource para cada edm:ProvidedCHO. Eles serão linkados através do ore:Aggregation utilizando edm:hasView ou uma outra propriedade. Um deles deve ser uma propriedade obrigatória em edm:isShownAt ou edm:isShownBy.

Property	Note	Value type
<b>dc:creator</b>	For the creator of the web resource. If possible supply the identifier of the creator from an authority source. Repeat for multiple creators.  <dc:creator xml:lang="es">Biblioteca Nacional de España</dc:creator>  or create a reference to an instance of the Agent class <dc:creator rdf:resource="http://viaf.org/viaf/147143794"/>	literal or ref
<b>dc:description</b>	Use for an account or description of this digital representation <dc:description>Performance with Buccin trombone</dc:description>	literal or ref
<b>dc:format</b>	Use for the format of this digital representation. (Use the value "3D-PDF" if appropriate.) <dc:format>image/jpeg</dc:format>	literal or ref
<b>dc:rights</b>	Use for the name of the rights holder of this digital representation if possible or for more general rights information. Note the difference between this property and the mandatory, controlled edm:rights property below. <dc:rights> Copyright © British Library Board</dc:rights>	literal or ref
<b>dc:source</b>	A related resource from which the Web resource is derived in whole or in part. <dc:source>The name of the source video tape <dc:source>	literal or ref
<b>dcterms:conformsTo</b>	An established standard to which the web resource conforms. <dcterms:conformsTo>W3C WCAG 2.0</dcterms:conformsTo> (web content accessibility guidelines).	literal or ref
<b>dcterms:created</b>	Date of creation of the Web resource. <dcterms:created>2010</dcterms:created>  or create a reference to an instance of the TimeSpan class <dc:date rdf:resource="http://semium.org/time/2010"/>	literal or ref
<b>dcterms:extent</b>	The size or duration of the digital resource. <dcterms:extent>1h 26 min 41 sec</dcterms:extent>	literal or ref

<b>dcterms:hasPart</b>	<p>A resource that is included either physically or logically in the web resource.</p> <p>&lt;dcterms:hasPart rdf:resource="http://www.identifier/Part"/&gt;</p>	ref
<b>dcterms:isFormatOf</b>	<p>Another resource that is substantially the same as the web resource but in another format.</p> <p>&lt;dcterms:isFormatOf&gt;http://upload.wikimedia.org/wikipedia/en/f/f3/Europeana_logo.png&lt;/dcterms:isFormatOf&gt; for a png image file of the described tiff web resource.</p> <p>Or as a link to a resource</p> <p>&lt;dcterms:isFormatOf rdf:resource="http://upload.wikimedia.org/wikipedia/en/f/f3/Europeana_logo.png"/&gt;</p>	literal or ref
<b>dcterms:isPartOf</b>	<p>A resource in which the Web Resource is physically or logically included. This property can be used for web resources that are part of a hierarchy. Hierarchies can be represented as hierarchies of ProvidedCHOs or hierarchies of web resources but not both at the same time. See the Task Force report on representing hierarchical entities.</p> <p>&lt;dcterms:isPartOf rdf:resource="http://data.europeana.eu/item/08701/1B0BACAA44D5A807E43D9B411C9781AAD2F96E65"/&gt;</p>	ref
<b>dcterms:issued</b>	<p>Date of formal issuance or publication of the web resource.</p> <p>&lt;dcterms:issued&gt;1999&lt;/dcterms:issued&gt;</p> <p>or create a reference to an instance of the TimeSpan class</p> <p>&lt;dcterms:issued rdf:resource="http://semium.org/time/2010"/&gt;</p>	literal or ref
<b>edm:isNextInSequence</b>	<p>Where one CHO has several web resources, shown by multiple instances of the edm:hasView property on the ore:Aggregation this property can be used to show the sequence of the objects. Each web resource (apart from the first in the sequence) should use this property to give the URI of the preceding resource in the sequence.</p> <p>&lt;edm:isNextInSequence rdf:resource="http://data.europeana.eu/item/2020601/9A3907CB46B651DE91621933ECC31EC1DC52B33C"/&gt;links to the web resource for page 2 of a digitised diary from the WebResource for page 3.</p>	ref
<b>edm:rights</b>	<p>The value in this element will indicate the copyright, usage and access rights that apply to this digital representation. It is strongly recommended that a value is supplied for this property for each instance of a web resource. The rights statement specified at the level of the web resource will 'override' the statement specified at the level of the aggregation.</p> <p>The value in this element is a URI taken from the set of those defined for use in Europeana. A list of these can be found at <a href="http://pro.europeana.eu/web/available-rights-statements">http://pro.europeana.eu/web/available-rights-statements</a></p> <p>&lt;edm:rights</p>	ref

	<p>rdf:resource="http://creativecommons.org/publicdomain/mark/1.0"/&gt;</p> <p>&lt;edm:rights rdf:resource="http://www.europeana.eu/rights/rr-f"/&gt;</p> <p>Or create a reference to an instance of the cc:License class where additional details of the rights can be provided (such as an expiry date for the restrictions):</p> <p>http://www.europeana.eu/rights/out-of-copyright-non-commercial/2029-06-01</p> <p>or</p> <p>&lt;edm:rights rdf:resource="#statement_3000095353971"/&gt;</p>	
<b>owl:sameAs</b>	<p>Provide the URI of another web representation of the same resource.</p> <p>&lt;owl:sameAs rdf:resource="urn:soundcloud:150424305"&gt;</p>	ref

### Propriedades para a classe aggregation (ore:Aggregation)

Estas são as propriedades que podem ser utilizadas na classe **ore:Aggregation**. São atributos aplicados para todo um conjunto de recursos relacionados sobre um provedor CHO.

Property	Note	Value type
<b>edm:aggregatedCHO</b>	<p>The identifier of the source object e.g. the Mona Lisa itself. This could be a full linked open data URI or an internal identifier.</p> <p>&lt;edm:aggregatedCHO rdf resource="#UEDIN:214"/&gt;</p>	ref
<b>edm:dataProvider</b>	<p>The name or identifier of the data provider of the object (i.e. the organisation providing data to an aggregator). Identifiers will not be available until Europeana has implemented its Organisation profile.</p> <p>&lt;edm:dataProvider&gt;Palais des Beaux Arts de Lille&lt;/edm:dataProvider&gt;</p> <p>Or as a link to a resource</p> <p>&lt;edm:dataProvider rdf:resource="http://www.pba-lille.fr"/&gt;</p>	literal or ref
<b>edm:hasView</b>	<p>The URL of a web resource which is a digital representation of the CHO. This may be the source object itself in the case of a born digital cultural heritage object.</p> <p>edm:hasView should only be used where there are several views of the CHO and one (or both) of the mandatory edm:isShownAt or edm:isShownBy properties have already been used. It is for cases where one CHO has several views of the same object. (e.g. a shoe and a detail of the label of the shoe)</p>	ref

	<p>&lt;edm:hasView rdf:resource="http://www.mimo-db.eu/media/UEDIN/VIDEO/0032195v.mpg"/&gt;</p> <p>&lt;edm:hasView rdf:resource="http://www.mimo-db.eu/media/UEDIN/AUDIO/0032195s.mp3"/&gt;</p>	
<b>edm:isShownAt</b>	<p>The URL of a web view of the object in full information context. Either edm:isShownAt or edm:isShownBy is mandatory.</p> <p>&lt;edm:isShownAt rdf:resource="http://www.mimo-db.eu/UEDIN/214"/&gt;</p>	ref
<b>edm:isShownBy</b>	<p>The URL of a web view of the object. Either edm:isShownAt or edm:isShownBy is mandatory.</p> <p>&lt;edm:isShownBy rdf:resource="http://www.mimo-db.eu/media/UEDIN/IMAGE/0032195c.jpg"/&gt;</p>	ref
<b>edm:object</b>	<p>The URL of a representation of the CHO which will be used for generating previews for use in the Europeana portal. This may be the same URL as edm:isShownBy. See Europeana Portal Image Guidelines (<a href="http://pro.europeana.eu/technical-requirements">http://pro.europeana.eu/technical-requirements</a>) for information regarding the specifications of previews. This must be an image, even if it is for a sound object.</p> <p>&lt;edm:object rdf:resource="http://www.mimo-db.eu/media/UEDIN/IMAGE/0032195c.jpg"/&gt;</p>	ref
<b>edm:provider</b>	<p>The name or identifier of the provider of the object (i.e. the organization providing data directly to Europeana). Identifiers will not be available until Europeana has implemented its Organization profile.</p> <p>&lt;edm:provider&gt;Geheugen van Nederland&lt;/edm:provider&gt;</p> <p>or link to a resource</p> <p>&lt;edm:provider rdf:resource="http://www.geheugenvannederland.nl"/&gt;</p>	literal or ref
<b>dc:rights</b>	<p>Ideally this should be applied to the edm:WebResource or the edm:ProvidedCHO. It is included here for the conversion of data from ESE where it is not known which object the rights apply to.</p>	literal
<b>edm:rights</b>	<p>This is a mandatory property and the value given here should be the rights statement that applies to the digital representation as given (for example) in edm:object or edm:isShownAt/By, when these resources are not provided with their own edm:rights (see edm:rights documentation). The value for the rights statement in this element is a URI taken from the set of those defined for use in Europeana at <a href="http://pro.europeana.eu/available-rights-statements">http://pro.europeana.eu/available-rights-statements</a></p> <p>The rights statement given in this property will also by default apply to the previews used in the portal and will support portal search and display functionality.</p> <p>Where there are several web resources attached to one edm:ProvidedCHO the rights statement given here will be regarded as the "reference" value for all the web resources.</p>	ref

	<p>Therefore a suitable value should be chosen with care if the rights statements vary between different resources. In fact in such cases Europeana encourages the provision of separate rights statements for each individual web resource. For example, a low-resolution of a JPEG file could be CC-BY, while the high resolution version or a video showing the object would be CC-BY-NC. In such cases the rights statements given for the individual web resources would 'override' the one specified at the ore:Aggregation level. Any other associated web resources would still be governed by the edm:rights of the ore:Aggregation.</p> <pre>&lt;edm:rights rdf:resource="http://creativecommons.org/publicdomain/mark/1.0"/&gt;  &lt;edm:rights rdf:resource="http://www.europeana.eu/rights/rr-f"/&gt;</pre> <p>Or create a reference to an instance of the cc:License class where additional details of the rights can be provided (such as an expiry date for the restrictions):</p> <pre>http://www.europeana.eu/rights/out-of-copyright-non-commercial/2029-06-01</pre> <p>or <code>&lt;edm:rights rdf:resource="#statement_3000095353971"/&gt;</code></p>	
edm:ugc	<p>This is a mandatory property for objects that are user generated or user created that have been collected by crowdsourcing or project activity. The property is used to identify such content and can only take the value "true" (lower case).</p> <pre>&lt;edm:ugc&gt;true&lt;/edm:ugc&gt;</pre>	literal

## As classes Contextuais

### Propriedades para a classe Agent (edm:Agent)

A classe edm:Agent compreende pessoas, tanto individualmente como em grupos, que possuem potencial para realizar ações o qual eles podem ser responsabilizados.

Property	Note	Value type
skos:prefLabel	<p>The preferred form of the name of the agent. Although the maximum number of occurrences is set at 1, it can be interpreted as 1 per language tag.</p> <pre>&lt;skos:prefLabel xml:lang="fr"&gt;Courtois neveu aîné&lt;/skos:prefLabel&gt;  &lt;skos:prefLabel xml:lang="en"&gt;Courtois'eldest nephew&lt;/skos:prefLabel&gt;</pre>	literal

<b>skos:altLabel</b>	Alternative forms of the name of the agent. <skos:altLabel xml:lang="en">Courtois</skos:altLabel> <skos:altLabel xml:lang="fr">Augte. Courtois aîné</skos:altLabel>	literal
<b>skos:note</b>	A note about the agent e.g. biographical notes. <skos:note> Courtois neveu aîné started a company of the same name manufacturing brass instruments in Paris in 1803</skos:note>	literal
<b>dc:date</b>	A significant date associated with the Agent. <dc:date>1803</dc:date/>	literal or ref
<b>dc:identifier</b>	An identifier of the agent. <dc:creator>http://viaf.org/viaf/96994048</dc:creator>	literal
<b>dcterms:hasPart</b>	Reference to an Agent that is part of the Agent being described (e.g. a part of a corporation). <dcterms:hasPart rdf:resource="http://identifier/partOfCorporation/">	ref
<b>dcterms:isPartOf</b>	Reference to an agent that the described agent is part of. <dcterms:isPartOf rdf:resource="http://identifier/parentCorporation/">	ref
<b>edm:begin*</b>	The date the agent was born/established. <edm:begin>1795</edm:begin>	literal
<b>edm:end*</b>	The date the agent died/terminated. <edm:end>1885</edm:end>	literal
<b>edm:hasMet</b>	Reference to another entity which the agent has "met" in a broad sense. For example a reference to a Place class <edm:hasMet rdf:resource="http://sws.geonames.org/6620265/">	ref
<b>edm:isRelatedTo</b>	Reference to other entities, particularly other agents, with whom the agent is related in a generic sense. <edm:isRelatedTo rdf:resource="http://identifier/relatedAgent/">	ref
<b>foaf:name</b>	The name of the agent as a simple textual string. <foaf:name>Auguste Courtois</foaf:name>	literal
<b>rdaGr2:biographicalInformation</b>	Information pertaining to the life or history of the agent. <rdaGr2:biographicalInformation>Courtois neveu aîné started a company of the same name manufacturing brass instruments in Paris in 1803</rdaGr2:biographicalInformation>	literal
<b>rdaGr2:dateOfBirth</b>	The date the agent (person) was born. <rdaGr2:dateOfBirth>1795</rdaGr2:dateOfBirth>	literal

<b>rdaGr2:dateOfDeath</b>	The date the agent (person) died. <rdaGr2:dateOfDeath>1895</rdaGr2:dateOfDeath>	literal
<b>rdaGr2:dateOfEstablishment</b>	The date on which the agent (corporate body) was established or founded. <rdaGr2:dateOfEstablishment>1795</rdaGr2:dateOfEstablishment>	literal
<b>rdaGr2:dateOfTermination</b>	The date on which the agent (corporate body) was terminated or dissolved. <rdaGr2:dateOfTermination>1895</rdaGr2:dateOfTermination>	literal
<b>rdaGr2:gender</b>	The gender with which the agent identifies. <rdaGr2:gender>Female</rdaGr2:gender>	literal
<b>rdaGr2:placeOfBirth</b>	The town, city, province, state, and/or country in which a person was born. <rdaGr2:placeOfBirth>Lusaka, Northern Rhodesia</rdaGr2:placeOfBirth> <rdaGr2:placeOfBirth rdf:resource="http://sws.geonames.org/909137"/>	literal or ref
<b>rdaGr2:placeOfDeath</b>	The town, city, province, state, and/or country in which a person died. <rdaGr2:placeOfDeath>London, United Kingdom</rdaGr2:placeOfDeath> <rdaGr2:placeOfDeath rdf:resource="http://sws.geonames.org/2635167"/>	literal or ref
<b>rdaGr2:professionOrOccupation</b>	The profession or occupation in which the agent works or has worked. <rdaGr2:professionOrOccupation>Instrument Maker</rdaGr2:professionOrOccupation>	literal or ref
<b>owl:sameAs</b>	Another URI of the same agent. <owl:sameAs rdf:resource="http://www.identifier/sameResourceElsewhere"/>	ref

### Propriedades para a classe **Place (edm:Place)**

Uma localização espacial identificada pelo provedor e nomeada de acordo com alguns vocabulários ou convenções locais.

Property	Note	Value type
<b>wgs84_pos:lat</b>	The latitude of a spatial thing (decimal degrees). <wgs84_pos:lat>51.5075</wgs84_pos:lat>	Floating point7
<b>wgs84_pos:long</b>	The longitude of a spatial thing (decimal degrees)	Floating

	<wgs84_pos:long>-0.1231</wgs84_pos:long>	point7
<b>wgs84_pos:alt</b>	The altitude of a spatial thing (decimal metres above the reference) <wgs84_pos:alt>21</wgs84_pos:alt>	Floating point 7
<b>skos:prefLabel</b>	The preferred form of the name of the place. Maximum is 1 per language tag. <skos:prefLabel xml:lang="en">London</skos:prefLabel>	literal
<b>skos:altLabel</b>	Alternative forms of the name of the place. <skos:altLabel xml:lang="en">Greater London</skos:altLabel>	literal
<b>skos:note</b>	Information relating to the place. <skos:note xml:lang="en">Pop. 21m</skos:note>	literal
<b>dcterms:hasPart</b>	Reference to a place that is part of the place being described. <dcterms:hasPart rdf:resource="http://sws.geonames.org/2643741/"> (City of London)	ref
<b>dcterms:isPartOf</b>	Reference to a place that the described place is part of. <dcterms:isPartOf rdf:resource="http://sws.geonames.org/2635167/"> (United Kingdom)	ref
<b>edm:isNextInSequence</b>	Can be used to represent a sequence of Place entities over time e.g. the historical layers of the city of Troy. Use this for objects that are part of a hierarchy or sequence to ensure correct display in the portal.	ref
<b>owl:sameAs</b>	URI of a Place <owl:sameAs rdf:resource="http://sws.geonames.org/2635167/">(London)	ref

### Propriedades para a classe Timespan ou Period (edm:TimeSpan)

Um período de tempo tendo um início, um final e uma duração.

Property	Note	Value type
<b>skos:prefLabel</b>	The preferred form of the name of the timespan or period. Although the maximum number of occurrences is set at 1, it can be interpreted as 1 per language tag. <skos:prefLabel xml:lang="en">Roman Empire</skos:prefLabel>	literal
<b>skos:altLabel</b>	Alternative forms of the name of the timespan or period. <skos:altLabel xml:lang="fr">Empire romain (27 avant J.-C.-476 après J.-C.)</skos:altLabel >	literal

<b>skos:note</b>	Information relating to the timespan or period. <skos:note>The <b>Roman Empire</b> (Latin: <i>Imperium Romanum</i> ) was the post-Republican period of the ancient Roman civilization, characterised by an autocratic form of government and large territorial holdings around the Mediterranean in Europe, Africa, and Asia.</skos:note>	literal
<b>dcterms:hasPart</b>	Reference to a timespan which is part of the described timespan .	ref
<b>dcterms:isPartOf</b>	Reference to a timespan of which the described timespan is a part.	ref
<b>edm:begin</b>	The date the timespan started. <edm:begin>27 BC</edm:begin>	literal
<b>edm:end</b>	The date the timespan finished. <edm:end>476 AD</edm:end>	literal
<b>edm:isNextInSequence</b>	Can be used to represent a sequence of Time periods. Use this for objects that are part of a hierarchy or sequence to ensure correct display in the portal. <edm:isNextInSequence rdf:resource="http://semium.org/time/roman_republic"/> (The Roman Empire was preceded by the Roman Republic)	ref
<b>owl:sameAs</b>	The URI of a timespan <owl:sameAs rdf:resource="http://semium.org/time/roman_empire"/>	ref

### Propriedades para a classe **Concept (skos:Concept)**

Um conceito é definido como uma unidade de pensamento ou significado que vem de uma base de conhecimento organizado (como termos de assuntos de um dicionário de sinônimos ou vocabulário controlado) onde URI's ou identificadores locais foram criados para representar cada conceito.

Property	Note	Value type
<b>skos:prefLabel</b>	The preferred form of the name of the concept. Maximum instance is 1 per language tag. <skos:prefLabel xml:lang="fr">Buccin</skos:prefLabel> <skos:prefLabel xml:lang="de">Buccin</skos:prefLabel> <skos:prefLabel xml:lang="nl">Buccin</skos:prefLabel>	literal
<b>skos:altLabel</b>	Alternative forms of the name of the concept. <skos:altLabel xml:lang="en">Buccin</skos:altLabel>	literal

<b>skos:broader</b>	The identifier of a broader concept in the same thesaurus or controlled vocabulary. <skos:broader rdf:resource="http://www.mimo-db.eu/InstrumentsKeywords/4369_1"/>	ref
<b>skos:narrower</b>	The identifier of a narrower concept. <skos:narrower rdf:resource="http://narrower.term"/>	ref
<b>skos:related</b>	The identifier of a related concept <skos:related rdf:resource="http://related.term"/>	
<b>skos:broadMatch</b> <b>skos:narrowMatch</b> <b>skos:relatedMatch</b>	The identifier of a broader, narrower or related matching concepts from other concept schemes. <skos:broadMatch rdf:resource="http://broadMatch.term"/> <skos:narrowMatch rdf:resource="http://narrowMatch.term"/> <skos:relatedMatch rdf:resource="http://relatedMatch.term"/>	ref
<b>skos:exactMatch</b> <b>skos:closeMatch</b>	The identifier of close or exactly matching concepts from other concept schemes. <skos:exactMatch rdf:resource="http://exactMatch.term"/> <skos:closeMatch rdf:resource="http://closeMatch.term"/>	ref
<b>skos:note</b>	Information relating to the concept. <skos:note>The buccin is a visually distinctive trombone popularized in military bands in France between 1810–1845 which subsequently faded into obscurity.</skos:note>	literal
<b>skos:notation</b>	The notation in which the concept is represented. This may not be words in natural language for some knowledge organisation systems e.g. algebra <skos:notation rdf:datatype="http://www.w3.org/2001/XMLSchema#int">123</skos:notation>	string (+ rdf: datatype attribute)
<b>skos:inScheme</b>	The URI of a concept scheme	ref

## APÊNDICE B

O quadro abaixo lista os vocabulários externos em potencial que contribuem ou contribuíram para o enriquecimento do material da Europeana.

Vocabulary	URL	Language	Licence	Comment
UDC summary edition	<a href="http://www.udcc.org/udcsummary/php/index.php">http://www.udcc.org/udcsummary/php/index.php</a>	51 languages	summary edition available as CC0	
Getty Thesauri: AAT, TGN, ULAN	<a href="http://vocab.getty.edu">http://vocab.getty.edu</a>	en, es, nl, zh and several others	ODC BY	TGN and ULAN scheduled for release in 2014
VIAF	<a href="http://viaf.org/">http://viaf.org/</a>	international	ODC BY	VIAF subsumes ULAN. ULAN has more relations, but there are more languages in VIAF and the scope is different
EuroVoc	<a href="http://eurovoc.europa.eu/">http://eurovoc.europa.eu/</a>	24 languages	<a href="http://eurovoc.europa.eu/drupal/?q=de/legalnotice">http://eurovoc.europa.eu/drupal/?q=de/legalnotice</a>	
British Museum		en		People, Places, Subjects, Object types, Materials. Not aligned to anything.  Available in SKOS RDF
Pleiades	<a href="http://pleiades.stoa.org">http://pleiades.stoa.org</a>	en + latin/greek names	CC BY	Ancient place names
Pleiades Time Periods	<a href="http://pleiades.stoa.org/vocabularies/time-periods">http://pleiades.stoa.org/vocabularies/time-periods</a>	en	CC BY	Time periods of antiquity
DBpedia	<a href="http://wiki.dbpedia.org/About">http://wiki.dbpedia.org/About</a>	many	from version 3.4 on is licensed under the terms of the Creative Commons Attribution-ShareAlike 3.0	Excellent source for people, sub-city places (eg stadium, museum). Challenge is how to select classes to be used for indexing  With a tool that takes context analysis into account, it can be used for a broad range of terms



French national library : authors, places (geographic names) subjects	<a href="http://data.bnf.fr/semanticweb-en">http://data.bnf.fr/semanticweb-en</a>	French and other languages, also transliterate	Open Licence (=CC-BY)	200 000 authors at this date, 1,5 million, in 2-3 years  170 000 topics 110 000 places
Thesaurus PICO	<a href="http://purl.org/pico/thesaurus_4.3.0.skos.xml">http://purl.org/pico/thesaurus_4.3.0.skos.xml</a>	it, en	CC BY	Contact: ICCU, Sara Di Giorgio Available in SKOS RDF. On persons, organisations, classifications, periods, actor roles, object names...
Thesaurus of "The Israel Museum, Jeruzalem	<a href="http://www.judaica-europeana.eu/Search_Europeana_Collections_in_Hebrew.html">http://www.judaica-europeana.eu/Search_Europeana_Collections_in_Hebrew.html</a>	he, en		Contact Dr. Allison Kupietzky SKOSified Was used for the Europeana Judaica project
Joconde thesauri	<a href="http://www.culture.gouv.fr/documentation/joconde/fr/pres.htm">http://www.culture.gouv.fr/documentation/joconde/fr/pres.htm</a>	fr		Terminologies of the Joconde database, managed by MCC, French ministry of Culture and Communication Not SKOSified? Contact: Jeannette IVAIN Different vocabularies on periods, styles, entities, materials and techniques...
AM-MovE thesaurus	<a href="http://www.museumminzicht.be/public/musea_werk/thesaurus/zoeken/index.cfm">http://www.museumminzicht.be/public/musea_werk/thesaurus/zoeken/index.cfm</a>	nl	Free of charge for non-profit organisations	MoVe thesaurus, often used by museum sector in Flanders. Same topics as Getty AAT. In the future they will try to align the thesaurus more with the AAT. Available as XML.
Library of congress subject headings, name authority file, classifications...	<a href="http://id.loc.gov/">http://id.loc.gov/</a>	en	public domain	Offered through Linked Data service and downloadable in different formats: SKOS RDF, JSON, etc. Also see OCLC FAST. LCSH problem: many subjects are pre-coordinated (e.g. "Italian poetry--16th century"), which makes them unsuitable for enrichment unless the collection used exactly LCSH



RMAH thesauri on: object names, materials and techniques, geography, musical instrument classification,	<a href="http://www.rmah.be">www.rmah.be</a> ; <a href="http://www.carmentis.be">www.carmentis.be</a>	fr, nl, en	freely available on request	Thesauri focussing on the museum collections ranging from prehistoric times until Art Deco. Completely available in French, Dutch and English. Some concepts have scope notes. Based on AAT, British Museum and other terminology reference thesauri. Main problem: only exportable in CSV
Europeana photography vocabulary	<a href="http://www.europeana-photography.eu/">http://www.europeana-photography.eu/</a>	Translated in the 12 languages of the project partners	CC0?	Thematic thesaurus on photographic keywords, materials and techniques, classifications...developed within the framework of the Europeana photography project. SKOSified Contact: <a href="mailto:n.vansteen@kmg-mrah.be">n.vansteen@kmg-mrah.be</a>
Europeana Fashion vocabulary	<a href="http://www.europeanafashion.eu/">http://www.europeanafashion.eu/</a>	DE, SE, RS, FR, PT, GR, NL, SE, ES, AT, IT, EN	CC0	Thematic thesaurus on Fashion keywords, materials and techniques, classifications...developed within the framework of the Europeana photography project. SKOSified Contact: <a href="mailto:n.vansteen@kmg-mrah.be">n.vansteen@kmg-mrah.be</a>
KOKO Ontology Cloud	<a href="http://kansalliskirjasto.onki.fi/">http://kansalliskirjasto.onki.fi/</a> - Includes ca. 15 linked ontologies of general concepts	FI, SE, EN	CC BY	Will be maintained by the National Library in 2014-; fairly comprehensive collection of general concepts (ca. 40 000) in SKOS hierarchies <a href="mailto:eero.hyvonen@aalto.fi">eero.hyvonen@aalto.fi</a> , <a href="mailto:matias.frosterus@kansalliskirjasto.fi">matias.frosterus@kansalliskirjasto.fi</a>
Finnish History Ontology HISTO (Events)	<a href="http://www.ldf.fi/dataset/history/index.html">http://www.ldf.fi/dataset/history/index.html</a>	FI	CC BY	Major historical events (ca 1100) in Finnish history. We are combining it e.g. with the national biography of 6500 short biographies. <a href="mailto:eero.hyvonen@aalto.fi">eero.hyvonen@aalto.fi</a>
Ontology of Historical Persons TOIMO	<a href="http://www.ldf.fi/dataset/agents/index.html">http://www.ldf.fi/dataset/agents/index.html</a>	FI	CC BY	Over 150 000 agents harvested from various sources including ULAN. Not yet of good quality because of e.g. duplicates. <a href="mailto:eero.hyvonen@aalto.fi">eero.hyvonen@aalto.fi</a>



Ontology of Historical Places in Finland SAPO	<a href="http://onki.fi/en/browser/overview/sapo">http://onki.fi/en/browser/overview/sapo</a>	FI	CC BY	Finnish county history since late 1800 and some earlier places from Sweden-Finland. Planned to be extended. <a href="mailto:eero.hyvonen@aalto.fi">eero.hyvonen@aalto.fi</a>
Various additional ontologies of the ONKI.fi service.	<a href="http://onki.fi/en/browser/">http://onki.fi/en/browser/</a>	FI, SE, EN, ...		E.g., lots of biological name lists are available, such as birds (10 000) or mammals (6 000) of the world. <a href="mailto:eero.hyvonen@aalto.fi">eero.hyvonen@aalto.fi</a>
MIMO vocabulary	<a href="http://www.mimo-db.eu/InstrumentsKeywords/2204">http://www.mimo-db.eu/InstrumentsKeywords/2204</a>	EN, FR, NL, IT, DE, SE	CC BY	<a href="mailto:rbailly@cite-musique.fr">rbailly@cite-musique.fr</a>
National Library of Spain: Persons, Organizations, Subject Heading, Titles	<a href="http://datos.bne.es">http://datos.bne.es</a> and <a href="http://datahub.io/dataset/datos-bne-es">http://datahub.io/dataset/datos-bne-es</a>	ES and many other non-tagged languages	CC0	More than four million authorities (including persons, orgs and works) many of them linked to VIAF and DBpedia. Rich subject headings linked to LCSH. Linked Data, many formats available. Contact: <a href="mailto:dvila@fi.upm.es">dvila@fi.upm.es</a>
Lista de Encabezamientos de Materia (Subject Headings) of the Spanish Public Library Network	<a href="http://datahub.io/dataset/lista-encabezamientos-materia">http://datahub.io/dataset/lista-encabezamientos-materia</a>	ES	CC0	Linked to LCSH and RAMEAU. Linked Data.
GND (Integrated Authority File)	General information <a href="http://www.dnb.de/EN/gnd">http://www.dnb.de/EN/gnd</a> RDF-Dumps <a href="http://datendienst.d-nb.de/cgi-bin/mabit.pl?user!D=opendata&amp;pass=opendata&amp;cmd=login">http://datendienst.d-nb.de/cgi-bin/mabit.pl?user!D=opendata&amp;pass=opendata&amp;cmd=login</a>	mainly de, some other non-tagged languages	CC0	Updates in January, May and September. Links to VIAF, LCSH, RAMEAU, STW and others coming up. Contains Persons, corporate bodies, subject headings, works, events and more.
STW (Standardthesaurus Wirtschaft; Thesaurus for Economics)	<a href="http://zbw.eu/stw/">http://zbw.eu/stw/</a>	de, en	CC BY-NC-SA	Linked to GND and DBpedia
TheSoz (Thesaurus Sozialwissenschaften; Thesaurus for the Social Sciences)	<a href="http://datahub.io/dataset/gesis-theso">http://datahub.io/dataset/gesis-theso</a>	de	CC BY-NC-ND	Links to DBpedia



DDC (Dewey Decimal Classification)	<a href="http://dewey.info/">http://dewey.info/</a>	several (depends on edition)	CC BY-NC-ND (depends on edition and language)	
ISNI (ISO certified global standard)	<a href="http://www.isni.org">http://www.isni.org</a>	Names with linguistic and transliteration variances	copyright	6.4 million individuals 400,000 organisations The core metadata is freely available for viewing via the web site. Access is also available via a search API. Each assigned ISNI is accessible by a persistent URI.
AgroVoc	<a href="http://aims.fao.org/standards/agrovoc/">http://aims.fao.org/standards/agrovoc/</a> , LOD at <a href="http://aims.fao.org/standards/agrovoc/linked-open-data">http://aims.fao.org/standards/agrovoc/linked-open-data</a>	Arabic, Chinese, Czech, English, French, German, Hindi, Hungarian, Italian, Japanese, Korean, Lao, Persian, Polish, Portuguese, Russian, Slovak, Spanish, Thai, Turkish (depending on state of translation)	CC BY-NC-SA (for English, French, Russian and Spanish), for other languages copyright rest with the responsible institution	32,000 concepts covering all areas of interest to FAO, including food, nutrition, agriculture, fisheries, forestry, environment etc.