

UNIVERSIDADE FUMEC  
Faculdade de Ciências Empresariais  
Mestrado Profissional em Sistemas de Informação e Gestão do  
Conhecimento

UMA PROPOSTA DE LINKED DATA DE ACERVOS EM UMA  
REDE DE MUSEUS

ANTÔNIO AUGUSTO PONTELO COSTA

Belo Horizonte - MG  
2016

ANTÔNIO AUGUSTO PONTELO COSTA

UMA PROPOSTA DE LINKED DATA DE ACERVOS EM UMA  
REDE DE MUSEUS

Dissertação apresentada ao Curso de Mestrado Profissional em Sistemas de Informação e Gestão do Conhecimento da Faculdade de Ciências Empresariais da Universidade FUMEC, como parte dos requisitos para a obtenção do título de Mestre em Sistemas de Informação e Gestão do Conhecimento. Área de concentração: Gestão de Sistemas de Informação e do Conhecimento. Linha de pesquisa: Tecnologia e Sistemas de Informação.

Prof. orientador: Luiz Cláudio Gomes Maia

Belo Horizonte - MG  
2016

## RESUMO

Os museus são instituições que desempenham importante papel para a sociedade, com seus acervos de grande valor cultural e científico. É seu dever promover o acesso aos acervos e realizar ações de comunicação para a divulgação dos bens culturais que compõem suas coleções. Para isso, eles vêm empregando a Tecnologia da Informação e Comunicação para apoiar suas atividades, ampliar o leque de serviços prestados à sociedade, promover a cultura, a ciência e o conhecimento e divulgar e disponibilizar seus acervos por meio da *Web*. Para tornar a navegação mais intuitiva e natural e possibilitar a troca de informações entre os museus, visando à recuperação da informação e ao reuso e interoperabilidade dos dados, é preciso adaptá-las ao formato da *Web Semântica*. Este estudo propõe uma solução capaz de integrar os dados de acervos da Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais e disponibilizá-los na *Web* utilizando conceitos de *Web Semântica* e de *Linked Data*. Para atingir tal objetivo, utilizou-se a metodologia experimental, com base na qual dados de acervos de museus integrantes dessa rede foram coletados, agrupados e organizados, seguindo um modelo de dados comum, convertendo-os e adaptando-os para os formatos de dados da *Web Semântica* e do *Linked Data*. Ao final, foi desenvolvido um protótipo de aplicação para validar o estudo, com base na prova de conceito. Demonstrou-se que, seguindo os passos descritos, é possível integrar dados de acervos de diferentes museus, em diferentes formatos de dados, e disponibilizá-los na *Web* aplicando conceitos de *Web Semântica* e *Linked Data*.

**Palavras-chave:** *Web Semântica*. *Linked Data*. Museu. Acervo.

## ABSTRACT

The museums are institutions that play important role for society, with their collections with great cultural and scientific value. They must promote access to collections and carry out communication activities for the dissemination of cultural assets that make up its collections. For this, they have been adopting the Information and Communication Technology to support their activities, expand the range of provided services to society, culture, science and knowledge promotion and disseminate and make available their collections through the Web. To make a more intuitive and natural navigation and enable the information exchange between museums, aiming the information retrieval and the data reuse and interoperability, we need to adapt them to the Semantic Web format. This study proposes a solution to integrate the collection data of Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais and make them available on Web using Semantic Web and Linked Data concepts. To achieve this goal, this study used the experimental methodology, in which collection data from members of this institution were collected, grouped and organized, according to a common data model, converted and adapted to the data formats of the Semantic Web and Linked Data. Finally, an application prototype was developed to validate the study, based in concept prove. It has been shown that following the steps described, it's possible to integrate data from different collections of museums, in different data formats, and make them available on the Web applying concepts of Semantic Web and Linked Data.

**Keywords:** Semantic Web. Linked Data. Museum. Collection.

## LISTA DE QUADROS

Quadro 1 - Exemplos de triplas em RDF .....	32
Quadro 2 - Dados selecionados dos acervos da Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais .....	49
Quadro 3 – Grupos de Informação e Categorias de Informação selecionados do <i>International Committee for Documentation</i> .....	52
Quadro 4 - Classe e propriedades reutilizadas para a classe Imagem .....	58
Quadro 5 - Classe e propriedades reutilizadas para a classe Instituição .....	58
Quadro 6 - Classe e propriedades reutilizadas para a classe Item.....	59
Quadro 7 - Termos criados para algumas propriedades da classe Item .....	59
Quadro 8 - Prefixos e nomes das ontologias reusadas .....	60
Quadro 9 - Totais de <i>links</i> RDF para recursos externos.....	69

## LISTA DE FIGURAS

Figura 1 - Análise dos sites dos integrantes da RMECC quanto à disponibilização dos seus acervos na Web.....	13
Figura 2 - Representação do processo de recuperação da informação.....	22
Figura 3 - Linha do tempo da Web.....	27
Figura 4 - Diferença entre <i>links</i> convencionais e <i>links</i> semânticos.....	29
Figura 5 - Diagrama em nuvem do <i>Linking Open Data</i> maio de 2007.....	31
Figura 6 - Diagrama em nuvem do <i>Linking Open Data</i> agosto de 2014.....	31
Figura 7 - Representação gráfica dos exemplos de triplas em RDF.....	33
Figura 8 - Um exemplo de grafo RDF simples representando o livro <i>Semantic Web For Dummies</i> ...	36
Figura 9 - Exemplo de consulta SPARQL que busca livros de autoria de Jeff Pollock.....	37
Figura 10 - Arquitetura conceitual do projeto.....	46
Figura 11 - Diagrama de Entidade Relacionamento para os acervos da Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais.....	53
Figura 12 - Arquitetura geral da plataforma D2RQ.....	56
Figura 13 - Trecho de código do arquivo de mapeamento <i>rededemuseus-ori.ttl</i> .....	57
Figura 14 - Representação da reutilização de ontologias e termos externos da <i>Web Semântica</i> .....	61
Figura 15 - Trecho de código do arquivo de mapeamento <i>rededemuseus-modif.ttl</i> .....	63
Figura 16 - Tela inicial do D2R Server.....	64
Figura 17 - Dados referente a um item de acervo visualizados no D2R Server.....	65
Figura 18 - Trecho de código do arquivo RDF <i>rededemuseus-ori.rdf</i> .....	66
Figura 19 - Visualização dos grafos RDF que correspondem ou tem relação com um item de acervo	67
Figura 20 - Legenda dos símbolos usados pelo RDF Gravity.....	68
Figura 21 - Porcentagem das propriedades que foram ligadas a recursos externos.....	70
Figura 22 - Página inicial do protótipo de aplicação.....	71
Figura 23 - Consulta SPARQL que busca todos os itens que pertencem a determinada instituição.....	72
Figura 24 - Resultado da consulta SPARQL que busca todos os itens que pertencem a determinada instituição.....	73
Figura 25 - Consulta SPARQL que busca os itens de todas as instituições que no seu nome contém o termo pesquisado.....	74
Figura 26 - Resultado da consulta SPARQL que busca os itens de todas as instituições em cujo nome contém o termo pesquisado.....	74
Figura 27 - Consulta SPARQL que busca todos os dados de determinado item.....	75
Figura 28 - Resultado da consulta SPARQL que busca todos os dados de determinado item.....	76
Figura 29 - Consulta SPARQL que busca todas as imagens de determinado item.....	76
Figura 30 - Resultado da consulta SPARQL que busca todas as imagens de determinado item.....	76
Figura 31 - Consulta SPARQL que busca todos os dados de determinada instituição.....	77
Figura 32 - Resultado da consulta SPARQL que busca todos os dados de determinada instituição.....	77
Figura 33 - Visualização dos grafos RDF de um item de acervo sem o formato <i>Linked Data</i> .....	79
Figura 34 - Visualização dos grafos RDF de um item de acervo no formato <i>Linked Data</i> .....	80
Figura 35 - Tela do protótipo de aplicativo: explorar os acervos por instituição.....	82
Figura 36 - Tela do protótipo de aplicativo: lista de itens de determinada instituição.....	83

Figura 37 - Tela do protótipo de aplicativo: detalhes sobre um item do acervo .....	84
Figura 38 - Tela do protótipo de aplicativo: dados sobre a instituição .....	85
Figura 39 - Galeria de imagens de um item do acervo.....	86
Figura 40 - Mapa dos locais de produção dos itens dos acervos.....	87
Figura 41 - Tela de informações do mapa dos locais de produção dos itens dos acervos.....	88

## LISTA DE SIGLAS

API	<i>Application Programming Interface</i>
CIDOC	<i>International Committee for Documentation</i>
CSV	<i>Comma-separated Values</i>
EDM	<i>Europeana Data Model</i>
HTML	<i>HyperText Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
ICOM	<i>International Council of Museums</i>
IDE	<i>Integrated Development Environment</i>
IRI	<i>International Resource Identifier</i>
Java EE	<i>Java Enterprise Edition</i>
LOD	<i>Linking Open Data</i>
OWL	<i>Web Ontology Language</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>Resource Description Framework Schema</i>
RMECC	Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais
SGBD	Sistema de Gerenciamento de Banco de Dados
SPARQL	<i>Simple Protocol and RDF Query Language</i>
SQL	<i>Structured Query Language</i>
TI	Tecnologia da Informação
TICs	Tecnologias da Informação e Comunicação
UFMG	Universidade Federal de Minas Gerais
URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web Consortium</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>10</b>
1.1	Problema de pesquisa .....	12
1.2	Justificativa.....	12
1.3	Objetivos .....	16
1.3.1	Objetivo geral .....	16
1.3.2	Objetivos específicos .....	16
1.4	Aderência ao programa e interdisciplinaridade .....	17
1.5	Trabalhos relacionados .....	17
<b>2</b>	<b>REFERENCIAL TEÓRICO .....</b>	<b>20</b>
2.1	Recuperação da informação e interoperabilidade nos museus.....	20
2.2	<i>Web Semântica</i> .....	25
2.3	<i>Linked Data</i> .....	28
2.4	<i>Resource Description Framework</i> .....	32
2.5	Ontologia.....	34
2.6	SPARQL .....	35
2.7	Acervos de museus na <i>Web Semântica</i> .....	37
<b>3</b>	<b>METODOLOGIA .....</b>	<b>41</b>
3.1	Caracterização da pesquisa.....	41
3.2	Procedimentos metodológicos .....	42
3.3	Objeto da pesquisa.....	43
3.4	Origem dos dados.....	44
3.5	Proposta de arcabouço conceitual .....	45
<b>4</b>	<b>IMPLEMENTAÇÃO .....</b>	<b>48</b>
4.1	Coleta, seleção e organização dos dados .....	48
4.2	Integração e armazenamento dos dados em um único repositório .....	50
4.3	Conversão dos dados para o formato RDF .....	54
4.4	Adaptação para o formato <i>Linked Data</i> e ligação com outros recursos externos da <i>Web Semântica</i> .....	68
4.5	Consultas e visualização dos dados integrados .....	70
<b>5</b>	<b>VALIDAÇÃO .....</b>	<b>78</b>

5.1	Integração, conversão e adaptação dos dados dos acervos para os formatos RDF e <i>Linked Data</i> .....	78
5.2	Consultas e visualização dos dados integrados .....	81
6	CONSIDERAÇÕES FINAIS .....	89
	REFERÊNCIAS .....	92
	APÊNDICES .....	99
	APÊNDICE A – Arquivo de mapeamento relacional-RDF rededemuseus-ori.ttl.....	99
	APÊNDICE B – Arquivo de mapeamento relacional-RDF rededemuseus-modif.ttl.	104

## 1 INTRODUÇÃO

Museus são espaços que, entre outras funções, promovem a cultura, a pesquisa e a aprendizagem. Segundo Marchi e Costa (2004), eles são responsáveis por criar, manter e disponibilizar uma variedade de informações capazes de suportar suas atividades, por exemplo: exposições, educação, programas públicos, pesquisa e conservação. Algumas de suas características peculiares, como estrutura e riqueza de seus acervos, fazem deles um espaço educativo informal, que promove o conhecimento e a aprendizagem para alunos e o público em geral.

A Lei Federal 11.904, de 2009, que instituiu o *Estatuto de Museus* e dá outras providências, conceitua os museus como instituições sem fins lucrativos, abertas ao público, a serviço da sociedade e de seu desenvolvimento, que desempenham atividades de conservação, investigação, comunicação, interpretação e exposição de conjuntos e coleções com valor histórico, artístico, científico, técnico ou de qualquer outra natureza cultural, com o objetivo de promover a preservação, estudo, pesquisa, educação, contemplação e turismo (BRASIL, 2009). Para Holanda e Braz (2012), o papel que os museus desempenham na preservação da memória é fundamental para as áreas social, cultural e econômica.

Dentre outros elementos e aspectos característicos dos museus, destacam-se os acervos, que são compostos por itens que agregam valor à sociedade, à cultura e à ciência, de acordo com a temática do museu. O § 2º do art. 5º da Lei 11.904 (BRASIL, 2009) declara que o acervo dos museus é de interesse público e que sua proteção, valorização, pesquisa e acesso à sociedade representam um valor cultural de grande importância para o Brasil, país de grande diversidade cultural, regional, étnica e linguística. A Subseção 3 desta lei, que trata da difusão cultural e do acesso aos museus, esclarece que estes devem realizar ações de comunicação para divulgar e propiciar o acesso público aos bens culturais que compõem seus acervos, bem como elaborar e implementar programas de exposições voltados para sua vocação e tipologia, com o objetivo de estimular a reflexão e o reconhecimento de seus valores para a sociedade.

O desenvolvimento das Tecnologias da Informação e Comunicação (TICs), evidente nas últimas décadas, vem proporcionando novos meios orientados para a divulgação, o acesso e a disseminação da informação. Holanda e Braz (2012) e Marcondes (2012) relatam que os museus são exemplos de instituições culturais que vêm empregando essas tecnologias para apoiar suas atividades, ampliar o leque de serviços prestados à sociedade, promover a cultura, a ciência e o conhecimento e divulgar e disponibilizar seus acervos por meio da *Web*.

“O uso das tecnologias por parte dos museus tenta repercutir para a sociedade projeções do acervo formado na instituição, empregando a *Web* como fonte valiosa de disponibilização de conteúdos” (HOLANDA; BRAZ, 2012, p. 43).

Os museus contribuem para o crescimento de novas fontes eletrônicas de informações, que, ao disponibilizá-las na *Web*, geram vantagens para eles, como: frequente atualização dos recursos disponíveis, suporte a múltiplos usuários, acesso a qualquer hora por parte dos internautas, contribuição para o aperfeiçoamento dos usuários e acesso e disseminação da informação. Porém, essas informações, muitas vezes, são geridas por sistemas próprios de cada museu, dificultando a recuperação e disseminação da informação e gerando redundância nos bancos de dados desses sistemas (MARCHI; COSTA, 2004).

Segundo Hyvönen *et al.* (2005), Marchi e Costa (2004), Marcondes (2012) e Skevakis *et al.* (2013), os sistemas adotados pelos museus são particulares e específicos, tornando seus registros fechados em si mesmos. Por terem semânticas locais e funcionamento dependente, esses sistemas dificultam a integração e a interação de seus conteúdos com outros museus ou, até mesmo, com outros tipos de recursos informacionais disponíveis na *Web*. Os bancos de dados que armazenam as informações dos acervos estão situados em diferentes localidades e adotam sistemas e esquemas de banco de dados diferentes, criando grande obstáculo para a recuperação da informação. Consequentemente, os acervos não exploram suas sinergias e seus potenciais culturais e informacionais que poderiam ter.

Para disponibilizar as informações dos acervos na *Web*, tornando a navegação mais intuitiva e natural, e promover a troca de informações entre os museus, visando ao reuso e à interoperabilidade dos dados, é preciso adaptá-las para o formato da *Web Semântica*. Para Mori e Carvalho (2004), a *Web Semântica* é uma evolução da *Web* atual, em que o significado da informação é bem definido, tornando os dados da *Web* legíveis por humanos e por computadores.

A *Web Semântica* promove o compartilhamento e o reuso de dados entre diferentes aplicações, empresas e comunidades. Suas tecnologias, ferramentas e padrões tornam os recursos da Internet legíveis semanticamente tanto para os humanos quanto para os computadores, que por isso são capazes de fazer inferências na navegação e desempenhar tarefas mais sofisticadas, agregando maior valor à experiência do internauta (BERNERS-LEE; HENDLER; LASSILA, 2001; MARCONDES, 2012).

De acordo com Berners-Lee (2006), a *Web Semântica* é algo mais do que publicar dados na *Web*. É essencial estabelecer *links* entre os recursos, para que a *Web Semântica* possa ser expandida, divulgada e explorada por uma pessoa ou uma máquina. Segundo Heath

(2016), essa é a proposta do *Linked Data*, uma coleção de dados inter-relacionados semanticamente na *Web* (W3C, 2015a), que antes não eram conectados ou que eram conectados de outra forma mais simples. No *Linked Data*, os relacionamentos semânticos entre os recursos enriquecem a navegação, pois é possível explorar outros dados relacionados a determinado recurso em que se tenha interesse.

### **1.1 Problema de pesquisa**

Nesse cenário, formula-se o seguinte problema de pesquisa: Quais são os conceitos e tecnologias da *Web Semântica* que permitem a disponibilização de forma integrada dos acervos da Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais?

### **1.2 Justificativa**

A Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais (RMECC) é um projeto da Universidade Federal de Minas Gerais (UFMG) que tem por objetivo propor e desenvolver ações nas áreas de Pesquisa, Extensão, Graduação e Pós-Graduação, relacionadas aos Museus e Espaços de Ciência e Cultura da UFMG, com a finalidade de apoiar e estimular o desenvolvimento desses espaços. Segundo o site oficial da RMECC (UNIVERSIDADE FEDERAL DE MINAS GERAIS, 2015), os objetivos desse projeto são:

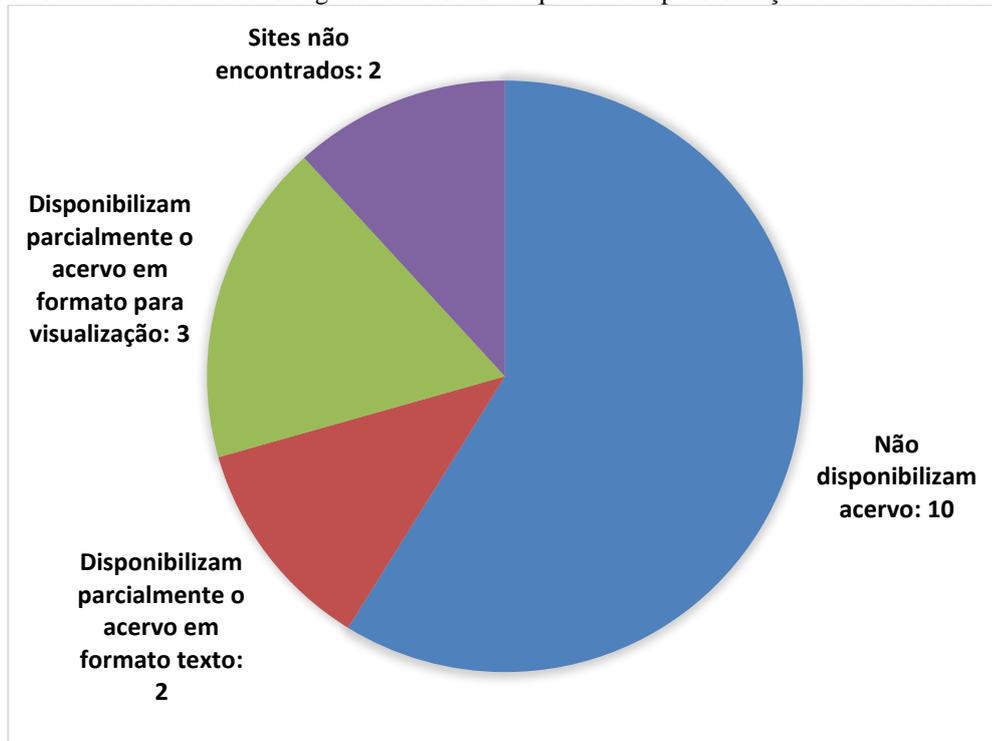
- Integrar os museus e os espaços que compõem a Rede, por meio de ações conjuntas;
- Promover a divulgação, intercâmbio institucional e parcerias interna e externamente à Rede;
- Viabilizar o intercâmbio científico, tecnológico e cultural entre os espaços integrantes da Rede e entre eles e as comunidades interna e externa à UFMG;

- Promover o gerenciamento, conservação e estratégias de uso e acesso às coleções dos acervos dos integrantes da Rede.

O regimento interno da RMECC (UNIVERSIDADE FEDERAL DE MINAS GERAIS, 2010), em seu art. 3º, estabelece que, para cumprir sua missão, a RMECC deve promover e/ou facilitar a comunicação entre a RMECC e os espaços integrantes, e destes entre si. A RMECC é composta atualmente por 20 museus e espaços de ciências e cultura. Destes, apenas três não são museus e/ou não são responsáveis pela guarda e disponibilização de acervos. Considerando então os 17 integrantes da RMECC que possuem acervo, procedeu-se à análise dos sites desses integrantes, buscando identificar como seus acervos são disponibilizados na *Web*.

A Figura 1 apresenta o resultado dessa análise.

Figura 1 - Análise dos sites dos integrantes da RMECC quanto à disponibilização dos seus acervos na *Web*



Fonte: Elaborada pelo autor, 2016.

A Figura 1 mostra que a maioria dos sites dos integrantes da RMECC não disponibiliza seus acervos na *Web*. Aqueles que disponibilizam parcialmente os acervos apresentam formatos de dados inadequados para compartilhamento, visualização e busca em conjunto com os outros sites dos integrantes da RMECC. Não há nenhuma comunicação ou troca de dados entre os poucos acervos que estão disponíveis na *Web* nessa Rede. Portanto, tem-se o mesmo cenário problemático descrito nos estudos de Hyvönen *et al.* (2005), Marchi e Costa (2004) e Marcondes (2012). Segundo esses autores, isso não favorece a recuperação

da informação nem a interoperabilidade dos dados dos acervos da RMECC. Por consequência, esses acervos não exploram suas sinergias nem os potenciais culturais e informacionais.

Tendo em vista os aspectos da Lei 11.904 (BRASIL, 2009), que caracterizam o papel importante do museu para a sociedade, o valor cultural e científico dos acervos de museus e o dever dos museus de promover acesso aos acervos e de realizar ações de comunicação para divulgar e propiciar o acesso público aos bens culturais que compõem seus acervos, assim como a atribuição da RMECC de integrar os museus e os espaços da Rede, de promover e facilitar a divulgação, comunicação e intercâmbio institucional, científico, tecnológico e cultural entre os integrantes e de desenvolver estratégias de uso e acesso às coleções dos acervos, justifica-se o desenvolvimento de uma solução que integre os dados de acervos da RMECC e disponibilize-os na *Web*.

A escolha da *Web Semântica* como a tecnologia norteadora deste estudo se deve ao fato de que a representação de acervos de museus nela viabiliza a recuperação da informação e a interoperabilidade dos dados, promovendo o conhecimento, a cultura e a ciência, com base nos trabalhos de Catarino (2014), Hyvönen *et al.* (2005), Isaac e Summers (2009), Marcondes (2012), Pastor-Sanchez, Mendez e Rodríguez-Muñoz (2009), Skevakis *et al.* (2013).

Outro aspecto importante, levantado por Hyvönen *et al.* (2005) e Skevakis *et al.* (2013), é que os acervos de museus têm a característica especial de se relacionarem semanticamente. Os itens que os compõem têm uma história e são relacionados semanticamente em vários aspectos com nosso ambiente e sociedade, bem como com outros itens de outros acervos. Essas associações semânticas podem ser estendidas a diferentes tipos de acervo em diferentes instituições, como no caso da RMECC, por meio da *Web Semântica*.

De acordo com Hyvönen *et al.* (2005), a tecnologia da *Web Semântica* permite novas possibilidades para a disponibilização de acervos de museus na *Web*, como a interoperabilidade dos dados dos acervos e o desenvolvimento de aplicações mais inteligentes, versáteis e amigáveis, na medida em que facilita a compreensão dos dados e de seus relacionamentos por parte dos computadores.

Além desses aspectos positivos gerais em relação à *Web Semântica*, Baker *et al.* (2011) cita alguns benefícios mais específicos para determinadas entidades que adotam essa tecnologia, dentro do contexto de bibliotecas e instituições culturais que, neste caso, se incluem os museus:

- Benefícios para pesquisadores, estudantes e clientes: como os recursos na *Web Semântica* e *Linked Data* são interligados semanticamente, a capacidade de descoberta e uso de dados pelos usuários são aprimoradas. A navegação entre os recursos das instituições e outros recursos externos da *Web* se torna mais sofisticada. Os usuários têm acesso aos recursos por meio de caminhos mais elaborados e diversificados. A *Web Semântica* conecta as coleções locais ao universo de informação representado na *Web*.
- Benefícios para organizações: as instituições culturais não possuem os recursos necessários para tornar suas informações em nível mais elevado de granularidade. Com o *Linked Data* é possível produzir diferentes tipos de dados sobre um mesmo recurso, de modo descentralizado, por diferentes atores, e posteriormente consolidado em uma mesma estrutura. A *Web Semântica* e o *Linked Data* podem ajudar as organizações a aprimorar seus processos internos de curadoria e as associações entre os objetos digitalizados e suas descrições. Conseqüentemente, isso melhora o processo de publicação dos dados da instituição na *Web*. O *Linked Data* possibilita uma presença maior na *Web* para as instituições culturais, onde os usuários e os pesquisadores podem encontrar suas informações mais facilmente.
- Benefícios para bibliotecários, arquivistas e curadores: com a *Web Semântica* e o *Linked Data*, esses profissionais podem criar um repositório de dados global, aberto e compartilhado que pode ser usado e reusado para descrever os recursos, com um nível de redundância de dados menor, se comparado com os processos de catalogação tradicionais.
- Benefícios para desenvolvedores e fornecedores: os desenvolvedores e fornecedores que lidam com essas tecnologias não ficam dependentes de formato de dados específico das instituições culturais. Portanto, eles têm maior liberdade de atuação no trabalho. A *Web Semântica* e o *Linked Data* suportam a recuperação e a combinação de dados compatível com todos os provedores de metadados. Os desenvolvedores não precisam trabalhar com formatos específicos de dados de bibliotecas ou arquivos, os quais requerem *softwares* e ferramentas personalizadas.

### 1.3 Objetivos

Este item do trabalho apresenta o objetivo geral do estudo, detalhado pelos objetivos específicos, a seguir.

#### 1.3.1 Objetivo geral

Integrar dados de acervos da Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais, utilizando conceitos da *Web Semântica*.

#### 1.3.2 Objetivos específicos

- Identificar e coletar os dados de acervos de museus da RMECC que irão compor a pesquisa;
- Agrupar e organizar os dados, utilizando um modelo de dados e um repositório comum para todos os acervos;
- Converter os dados dos acervos para o formato *Resource Description Framework* (RDF);
- Adaptar os dados em RDF para o formato *Linked Data* e interligá-los com outros recursos externos da *Web Semântica*;
- Propor um arcabouço conceitual;
- Testar o arcabouço conceitual proposto, por meio de implementação de um protótipo de aplicação capaz de visualizar os dados dos acervos de forma integrada, interativa e amigável.

## 1.4 Aderência ao programa e interdisciplinaridade

Este estudo trabalha questões sobre tecnologia e sistemas de informação, na medida em que aborda o tratamento e a disponibilização da informação sobre os acervos de museus na *Web*, contribuindo para a disseminação da informação e do conhecimento, com base na tecnologia da informação. Envolve, também, questões sobre sistemas de recuperação da informação, pois a *Web Semântica* é, em sua essência, um projeto de criação e estabelecimento de padrões de tecnologias que permitem o compartilhamento e o reuso de dados entre sistemas, empresas e comunidades, por meio de metadados descritivos que possibilitam a interligação, classificação e recuperação da informação disponível na *Web*.

Quanto à interdisciplinaridade, característica do programa de mestrado em Sistemas de Informação e Gestão do Conhecimento da Universidade FUMEC, este estudo aborda conceitos relacionados a Sistemas de Informação, Ciência da Informação e Tecnologia da Informação. Tendo em vista que esse programa de mestrado tem caráter profissional, com foco na realização de pesquisas aplicadas e práticas na área tecnológica, este estudo visa desenvolver um estudo prático aplicado, envolvendo os acervos dos integrantes da RMECC.

## 1.5 Trabalhos relacionados

Foram selecionados três trabalhos relacionados ao estudo aqui proposto, detalhados a seguir:

- Hyvönen *et al.* (2005) — envolve a disponibilização de acervos de museus na *Web Semântica* por meio de um portal.
- Skevakis *et al.* (2013) — relata a integração de dados de museus de História Natural e a disponibilização dos acervos na *Web Semântica* e no projeto Europeana.
- Azevedo (2014) — envolve o desenvolvimento de uma solução em *Linked Data* para integrar e disponibilizar dados sobre enchentes na Bacia do Rio Doce.

O trabalho “MuseumFinland – Finnish museums on the semantic web”, de autoria de Hyvönen *et al.* (2005), apresenta o portal semântico MuseumFinland, que

disponibiliza acervos de diferentes museus da Finlândia na *Web Semântica*. Hyvönen *et al.* (2005) demonstram como os museus, com seus acervos semanticamente ricos e inter-relacionados, podem criar um poderoso portal semântico juntos na *Web*. Os objetivos do desenvolvimento desse portal semântico foram: prover uma visão global dos acervos distribuídos; viabilizar a recuperação da informação, com base em conteúdo; disponibilizar conteúdos interligados semanticamente; e facilitar a publicação de conteúdo local por parte dos museus finlandeses.

O trabalho de Hyvönen *et al.* (2005) está relacionado com este estudo, na medida em que disponibiliza acervos de diferentes museus usando ferramentas e tecnologias da *Web Semântica*, integrando os dados dos acervos e desenvolvendo um meio para visualizá-los. Entretanto, o que difere os dois trabalhos prende-se às metodologias. No caso do portal MuseumFinland, todos os dados dos acervos eram baseados em ontologias específicas, o que facilitou o trabalho de conversão dos dados para o RDF. Já neste estudo o que se propõe é agrupar, organizar e adaptar os dados dos acervos, provenientes de diferentes formatos e bancos de dados.

O segundo trabalho tem o título de “Elevating Natural History Museums’ Cultural Collections to the Linked Data Cloud”, de autoria de Skevakis *et al.* (2013). Eles apresentam o projeto *Natural Europe Project*, que oferece uma solução organizada para viabilizar a disponibilidade, divulgação, exploração e valorização do conteúdo cultural, com a finalidade de promover a educação e a aprendizagem multilíngue e multicultural, usando tecnologias e práticas da *Web Semântica*. A origem dos dados foram os conteúdos dos acervos de seis museus de História Natural europeus, organizados pela federação *Natural Europe Cultural Digital Libraries Federation*.

O estudo de Skevakis *et al.* (2013) se relaciona com este estudo também pelo fato de disponibilizar acervos de diferentes museus usando ferramentas e tecnologias da *Web Semântica*. No entanto, este estudo propõe integrar dados de museus com diferentes temáticas, enquanto os museus envolvidos no trabalho de Skevakis *et al.* (2013) possuem uma mesma temática, ou seja, somente museus de História Natural. Outro aspecto que difere os dois trabalhos é que o de Skevakis *et al.* (2013) não descreve uma solução para a visualização dos dados integrados na *Web Semântica*, pois seus dados são integrados com o projeto Europeana, que, por sua vez, permite visualizá-los.

O terceiro trabalho, de autoria de Azevedo (2014), intitulado “Uma proposta para visualização de linked data sobre enchentes na Bacia do Rio Doce”, desenvolve um protótipo de aplicação capaz de obter dados heterogêneos relacionados a inundações na bacia do rio

Doce de diversas organizações públicas, integrá-los e disponibilizá-los para visualização em um Sistema de Informação Geográfica.

O trabalho de Azevedo (2014) está relacionado com este estudo principalmente nos aspectos metodológicos, pois propõe um protótipo de aplicação para integrar e disponibilizar visualmente dados de diferentes fontes, usando ferramentas e tecnologias da *Web Semântica* e do *Linked Data*. Porém, há diferenças nas origens dos dados, pois foram usados dados governamentais e de empresas. Há diferenças também nos tipos de dados e nas unidades de análise, visto que se tratou de dados sobre enchentes na bacia do rio Doce.

Todos esses três trabalhos relacionados serviram de alguma maneira para a concepção e construção deste estudo, seja no aspecto de abordagem, ferramental ou metodológico. Foram importantes referências também para a construção do arcabouço conceitual proposto, mas que não dispensam uma base teórica mais consistente que apoie o estudo. No próximo item, desenvolve-se o referencial teórico resultante de uma pesquisa bibliográfica na literatura e em trabalhos relacionados ao tema.

## 2 REFERENCIAL TEÓRICO

O referencial teórico levantado neste estudo aborda os conceitos de recuperação da informação e interoperabilidade nos museus, *Web Semântica*, *Linked Data*, *Resource Description Framework*, ontologia, SPARQL e acervos de museus disponibilizados na *Web Semântica*.

### 2.1 Recuperação da informação e interoperabilidade nos museus

O avanço das TICs, evidente nas últimas décadas, significou profundas mudanças na sociedade. A informação e o conhecimento ganharam grande importância, tornando-se o foco de um novo paradigma. Esse cenário é percebido por Alves (2005) quando relata que “a informação é considerada matéria-prima para o desenvolvimento social, econômico e cultural e aliada ao uso de tecnologia de informação passou a ser fator importante na construção de uma nova sociedade...” (ALVES, 2005, p. 11).

A Internet e a *Web* atingiram um volume de informação ainda não alcançado por qualquer outro meio de comunicação (HOLANDA; BRAZ, 2012), revolucionando todas as áreas do conhecimento e tornando-se a maior fonte de informação, em consequência da facilidade de disponibilização de informação e do volume de recursos, que cresce a cada dia (ALVES, 2005). A *Web* cresce cada vez mais com o passar do tempo, tanto em número de usuários quanto em quantidade de recursos disponíveis.

Holanda e Braz (2012), Marcondes (2012) e Santos Neto *et al.* (2013) incluem os museus como instituições culturais que adotaram as TICs para apoiar suas atividades, aumentar a gama de serviços prestados à sociedade, promover a cultura, a ciência e o conhecimento e disponibilizar e permitir acesso aos seus acervos por meio da *Web*.

Assim, é evidente que as tecnologias da informação e comunicação são cada vez mais empregadas para diversas atividades e setores da sociedade, inclusive pela comunidade científica e pelas instituições culturais, como por exemplo, os museus, que são um exemplo do uso dessas tecnologias, pois desempenham papel fundamental na preservação da memória em diversos seguimentos sociais, culturais e econômicos. O uso das tecnologias por parte dos museus tenta repercutir para a sociedade projeções do acervo formado na instituição, empregando a *Web* como fonte valorosa de disponibilização de conteúdos (HOLANDA; BRAZ, 2012, p. 43).

Para Souza e Alvarenga (2004), embora um dos objetivos da *Web* fosse facilitar o acesso, o intercâmbio e a recuperação de informações, ela foi desenvolvida sem muita organização e de modo descentralizado e anarquicamente. A *Web* cresceu exponencialmente e o volume de informações tomou imensas proporções, tornando-se um imenso repositório de documentos, no qual o processo de recuperação daquilo de que se necessita é complicado e impreciso.

A *Web* surgiu com a intenção de prover uma interface mais amigável, intuitiva e voltada para a apresentação e acesso ao crescente número e tipos de recursos da Internet. Mas o volume, extensão e variedade de recursos e a demanda dessa rede cresceram muito além das expectativas, tornando-se necessário o desenvolvimento de uma nova filosofia e de tecnologias e ferramentas para apoiá-la, além da ampliação da infraestrutura tecnológica (SOUZA; ALVARENGA, 2004).

Sayão e Marcondes (2008, p. 134) também relatam a mesma problematização: “A Internet é o retrato vivo da ‘explosão informacional’, do ‘caos informacional’, mas o acúmulo imenso e desordenado de informação é, ao mesmo tempo, a sua fortaleza e a sua fragilidade, sendo necessário, portanto, algum grau de ordenamento e intermediação”. Silva, Santos e Ferneda (2013, p. 28) concordam quando afirmam que “A *Web* se configura hoje como a maior base de dados existente e, conseqüentemente, a que necessita de maiores cuidados no tratamento descritivo de suas informações”.

A organização da informação é hoje uma necessidade no mundo globalizado, que valoriza a informação e o conhecimento, em que a *Web* representa uma importante fonte de enorme quantidade de conteúdos, que se multiplicam de variadas formas e continuamente. Para facilitar a compreensão dos recursos disponíveis na *Web*, é preciso estruturar os dados que os compõem. Vários *sites* na *Web* classificam seus recursos com o mínimo de interesse e dedicação que deveriam ter, sem considerar que os usuários que os visitam seriam os principais beneficiados de uma melhor estruturação dos dados, recuperação da informação e agregação de valor e, conseqüentemente, voltariam a visitá-los (ABARCA; MENA; NOVOA, 2010).

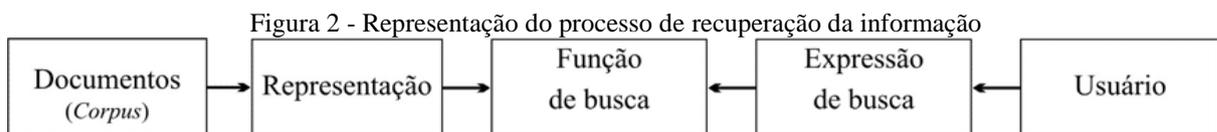
Ferneda (2003, p. 14) compila alguns conceitos sobre recuperação da informação encontrados na literatura: “a operação pela qual se seleciona documentos, a partir do acervo, em função da demanda do usuário; [...] fornecimento, a partir de uma demanda definida pelo usuário, dos elementos de informação documentária correspondentes; [...] a operação que fornece uma resposta mais ou menos elaborada a uma demanda, e esta resposta é convertida

num produto cujo formato é acordado com o usuário; [...] o tratamento da informação (catalogação, indexação, classificação)”.

O processo de recuperação da informação descrito por Ferneda (2003) corresponde à solicitação de uma informação por parte do usuário e à busca e identificação das informações correspondentes no conjunto de documentos, também chamado de *corpus*, que atendam à necessidade de informação do usuário. Para Ferneda (2003, p. 14-15), o usuário de um sistema de recuperação da informação está “interessado em recuperar ‘informação’ sobre um determinado assunto e não em recuperar dados que satisfazem sua expressão de busca, nem tampouco documentos, embora seja nestes que a informação estará registrada”.

Silva, Santos e Ferneda (2013) consideram a recuperação da informação um campo da Ciência da Computação que aborda a eficiência e a eficácia dos processos de busca em um sistema computacional, considerando a relevância dos resultados das buscas ao usuário do sistema e a coerência destes resultados com a expressão de busca fornecida. Os sistemas de recuperação da informação são responsáveis por apresentar os dados que atendem à solicitação, em um meio que seja possível ao usuário selecionar aqueles que realmente são de seu interesse (FERNEDA, 2003).

A Figura 2 representa o processo de recuperação da informação segundo Ferneda (2003).



Fonte: Ferneda, 2003.

Além da organização e estruturação dos dados, para a melhor recuperação da informação pelos sistemas de informação é também importante a característica de interoperabilidade. Para Souza e Alvarenga (2004), interoperabilidade é a capacidade de diferentes sistemas “conversarem” entre si. Sayão e Marcondes (2008) afirmam que dados e conteúdos diferentes de sistemas diferentes só podem ser integrados e reusados mediante a exploração de suas sinergias se estes forem geridos por sistemas altamente interoperáveis.

Segundo Payette *et al.* (1999), o objetivo geral da interoperabilidade é permitir que comunidades diferentes, dotadas de diferentes tipos de informações geridas por tecnologias distintas possam compartilhar suas informações e recursos por meio de processos de agregação e computação, gerando, assim, novos tipos de informação mais poderosos. Para

Miller (2000), uma organização interoperável é aquela que adota processos que asseguram que seus sistemas, procedimentos e cultura sejam geridos visando às oportunidades de compartilhamento e reuso de informação, tanto interna quanto externamente.

Sayão e Marcondes (2008) relatam que a interoperabilidade tem sido a questão mais priorizada pelos desenvolvedores de sistemas distribuídos de repositórios e de bibliotecas digitais em rede. Eles citam também outras áreas que focam e valorizam a interoperabilidade, como: indústria de conteúdos, instituições governamentais, educação e ensino a distância, arquivos e museus.

Miller (2000) explica os motivos que levam uma organização a se tornar interoperável. Nas últimas décadas, as organizações espalhadas pelo mundo acumularam grande volume de dados a respeito de vários assuntos. O resultado foi a construção de inúmeros repositórios de dados, desde pessoais até acervos de um grande museu. O acesso a esses dados tornou-se restrito a poucos sistemas e pessoas, com uma barreira de papelada e burocracia a impedir a recuperação e o acesso a eles por aqueles que precisariam tê-los.

Para qualquer entidade, seja ela pessoa, sistema ou organização, que queira integrar dados desses diferentes repositórios, geralmente, não há outra alternativa a não ser traduzir, redigitar ou adaptar dados de sistemas incompatíveis. Na “Era da Informação”, existe o consenso de que esses repositórios e o conhecimento obtido por meio da mineração dos dados neles representam um imenso valor. Para viabilizar o acesso a esses repositórios de dados, os sistemas que os gerem devem ser capazes de interoperar entre si (MILLER, 2000).

Os sistemas e as metodologias particulares e específicos adotados pelos museus tornam seus acervos isolados, com semânticas locais, impossibilitando a integração de seus conteúdos com outros acervos de museus ou, mesmo, com outros recursos disponíveis na *Web*. São sistemas com semântica própria, que gerenciam dados com formatos e padrões que impedem que os museus usufruam do potencial de integração e acesso que a *Web* oferece. Portanto, toda a sinergia e potencial culturais e informacionais que teriam os acervos dos museus, interligados entre eles e outros recursos da *Web* são desperdiçados (MARCONDES, 2012; SANTOS NETO *et al.*, 2013; SKEVAKIS *et al.*, 2013).

Marcondes (2012, p. 182) aborda, também, o acesso dos usuários ao conteúdo dos sistemas de arquivos, bibliotecas e museus. “Ao acessar esses catálogos na *Web*, o usuário fica como que prisioneiro desses contextos sistêmicos e institucionais específicos, praticamente sem possibilidades de navegar de fora para dentro ou de dentro para fora destes”.

Skevakis *et al.* (2013) acrescentam que dados sobre patrimônio cultural e biodiversidade são, sintática e semanticamente, heterogêneos, multilíngue e inter-relacionados e possuem semântica relevante. Instituições como museus, bibliotecas e arquivos gerenciam importantes informações científicas seguindo seus próprios padrões e boas práticas. Isso resulta em ineficiência na gestão dessas informações, impedindo que uma enorme quantidade de informação científica de qualidade seja recuperada e explorada. Gerenciar informações ricas e variadas semanticamente na *Web* de modo que os conteúdos sejam distribuídos e interoperáveis impõe requisitos que a *Web* tradicional não consegue resolver.

Bermes (2011, p. 1, tradução nossa) relata a inexistência da integração dos dados dos sistemas das instituições culturais. “Instituições de patrimônio cultural mantêm catálogos que são como silos, isolados uns dos outros, e isolados do amplo ecossistema da *Web*”.<sup>1</sup> Hyvönen *et al.* (2005) e Skevakis *et al.* (2013) concordam que, geralmente, os bancos de dados e os sistemas dos museus ficam situados em diferentes lugares e adotam modelos, padrões e esquemas distintos. Segundo Marchi e Costa (2004), os sistemas dos museus gerenciam informações importantes que apoiam suas atividades fins, porém esses sistemas são próprios, inacessíveis por outros sistemas de museus ou outras instituições. Esse problema também gera redundância nos bancos de dados.

Para esclarecer a riqueza semântica típica de informações de acervos culturais, como os de museus, e demonstrar como itens de acervos têm uma história e possuem relações entre eles, Hyvönen *et al.* (2005, p. 224) exemplificam usando uma cadeira.

Por exemplo, uma cadeira pode ser feita de madeira de carvalho e couro, pode ser de um certo estilo, foi desenhada por um designer famoso, foi fabricada por determinada empresa durante um período de tempo, foi usada em determinado edifício, juntamente com outras peças de mobiliário, e assim por diante. Outros itens de acervo, locais, períodos de tempo, profissionais, empresas, etc. podem estar relacionados com a cadeira, por meio de suas propriedades, constituindo implicitamente uma rede semântica complexa de associações. Essa rede semântica não se limita a um único acervo, mas se estende a outros acervos relacionados, em outros museus. A rede de associações semânticas pode ser estendida também para conteúdos de outros tipos em outras organizações (HYVÖNEN *et al.*, 2005, p. 224, tradução nossa).<sup>2</sup>

Para disponibilizar os acervos de museus na *Web*, proporcionando melhor recuperação da informação e promovendo o compartilhamento de dados entre museus, e

<sup>1</sup> “Cultural heritage institutions maintain catalogues that exist like silos, isolated from one another, and isolated from the wider ecosystem of the web” (BERMES, 2011, p. 1).

<sup>2</sup> “For example, a chair may be made of oak and leather, may be of a certain style, was designed by a famous designer, was manufactured by a certain company during a time period, was used in a certain building together with other pieces of furniture, and so on. Other collection items, locations, time periods, designers, companies, etc. can be related to the chair through their properties and implicitly constitute a complicated semantic network of associations. This semantic network is not limited to a single collection but spans over other related collections in other museums. The network of semantic associations can be extended to contents of other types in other organizations, as well” (HYVÖNEN *et al.*, 2005, p. 224).

destes com outros recursos disponíveis na *Web*, visando ao reuso e à interoperabilidade dos dados, e para usufruir de toda a sinergia e potencial cultural que existe nos acervos dos museus, eles devem ser adaptados para o formato da *Web Semântica* (CATARINO, 2014; HYVÖNEN *et al.*, 2005; MARCONDES, 2012; SANTOS NETO *et al.*, 2013; SKEVAKIS *et al.*, 2013; SOUZA, ALVARENGA, 2004).

## 2.2 *Web Semântica*

A *Web* armazena e disponibiliza grande quantidade de informações, que estão espalhadas pelo mundo. O volume e a diversidade dessas informações tendem a aumentar cada vez mais. Muitas vezes, essa situação dificulta a recuperação da informação de algo específico que se queira saber. Segundo Mori e Carvalho (2004), os dados disponíveis na Internet são estruturados para apresentar informações aos usuários, tornando esta rede um meio de comunicação direcionado também às pessoas, e não apenas às máquinas e aos computadores.

Ao disponibilizar os dados em formato mais apresentável na Internet, o contexto ou o significado desses dados podem ser perdidos ou não compreendidos pelos computadores. “Esta formatação dos dados, direcionada para a sua apresentação, dificulta a sua manipulação automatizada por meio de máquinas (computadores), uma vez que estas nem sempre são capazes de identificar a semântica associada a estes dados” (MORI; CARVALHO, 2004, p. 2). Para Mori e Carvalho (2004), um dos motivos de um resultado não satisfatório de uma busca na Internet é a incapacidade de se identificar o contexto ou o significado dos dados acessíveis. Os buscadores da Internet são mais eficientes quando se têm informações semânticas disponíveis.

De acordo com Berners-Lee, Hendler e Lassila (2001), a *Web Semântica* permite estruturar o significado dos conteúdos dos sites, tornando-os legíveis por computadores, que assim conseguem inferir ações mais sofisticadas e oferecer uma navegação mais interessante para o internauta. A *Web Semântica* é uma extensão da *Web* tradicional, na qual o significado das informações é bem definido, promovendo uma melhor cooperação entre computadores e usuários. Silva, Santos e Ferneda (2013, p. 34) explicam que a *Web Semântica* “se caracteriza por ser uma nova fase da *Web* na qual as informações dispersas na internet são

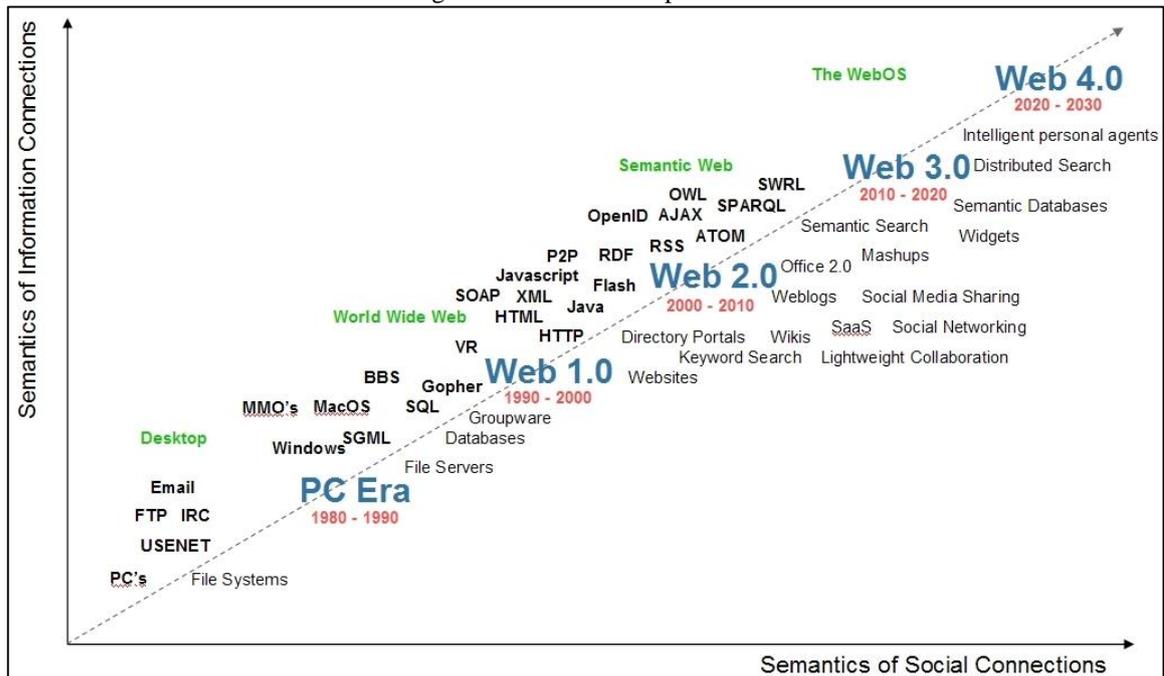
semanticamente descritas de modo a serem recuperadas com maior eficiência e relevância por motores de busca”.

Segundo W3C (2013), a *Web Semântica* provê ferramentas, tecnologias e conceitos comuns que permitem o compartilhamento e o reuso de dados entre aplicações, empresas e comunidades diversas. Trata-se de um projeto colaborativo coordenado pela *World Wide Web Consortium (W3C)*, que conta com a participação de vários pesquisadores e parceiros na área da indústria. Para Mori e Carvalho (2004), ela é considerada uma extensão da *Web* tradicional, na qual os significados das informações são bem definidos, permitindo que computadores capturem o significado do conteúdo de cada recurso e interliguem e classifiquem semanticamente os recursos disponíveis na *Web*.

De acordo com a W3C Brasil (2011), a *Web* já conhecida e tradicional é chamada de “*Web* de documentos”. A *Web* de dados é uma evolução da *Web* de documentos, baseada na *Web Semântica*. Esta *Web* de dados tem por objetivo expandir as possibilidades de uso dos computadores e possibilitar o desenvolvimento de sistemas que suportam interações na rede. Segundo Catarino (2014), a *Web Semântica* tem por objetivo tornar todos os dados da *Web* legíveis por humanos e por computadores, por meio do desenvolvimento de tecnologias, linguagens, padrões e recomendações.

Spivack (2007) apresenta um modelo de linha do tempo do passado, presente e futuro da *Web*, representado na Figura 3. De acordo com Pollock (2009), essa linha do tempo demonstra o potencial dos dados conectados por meio da tecnologia e o poder social das conexões que as pessoas podem fazer usando uma mesma tecnologia. As conexões sociais e as conexões de dados avançam proporcionalmente ao longo do tempo. Segundo Spivack (2007), a *Web Semântica* marca a transição da *Web 2.0* para a *Web 3.0*. Os nomes abaixo da linha tracejada representam as ferramentas que foram e serão desenvolvidas com a evolução da *Web*. Os nomes que aparecem acima dessa linha são tecnologias que suportaram essa evolução.

Figura 3 - Linha do tempo da Web



Fonte: Spivack, 2007.

A W3C é responsável por desenvolver e difundir recomendações e tecnologias para viabilizar a *Web Semântica*. Ela busca maximizar e enriquecer, cada vez mais, as possibilidades de uso da *Web*, por meio do desenvolvimento de protocolos, tecnologias e diretrizes. Os trabalhos da W3C são orientados por dois princípios: *Web* para todos e *Web* em todas as coisas. O primeiro princípio define que todos, indiscriminadamente, devem ter acesso ao conteúdo da *Web* e o segundo significa tornar o conteúdo da *Web* disponível em qualquer tipo de dispositivo, seja televisão, celular, *tablet*, etc. (CATARINO, 2014).

Marcondes (2012) descreve a *Web Semântica* como uma *Web* qualitativamente diferente da tradicional. Além de serem estruturados para sua exibição e leitura por usuários, os recursos possuem uma semântica legível por computadores. “As tecnologias da *Web Semântica* são a evolução da *Web* atual, em direção a mais amplas possibilidades de interação e de recuperação semântica da informação” (MARCONDES, 2012, p. 176).

Com a evolução da *Web* para uma rede mais participativa, colaborativa e social, conhecida como *Web 2.0*, a *Web Semântica* tornou-se mais importante. Ela propõe mudanças para a *Web*, em que a informação possui uma semântica bem definida, capaz de permitir melhor interação entre computadores e pessoas. Catarino (2014) afirma que para isso é necessário que os computadores tenham acesso a coleções de informações estruturadas, formadas por dados, metadados e seus relacionamentos, que auxiliam no processo de dedução automática para a aplicação de raciocínio automatizado. O conceito de *Linked Data* permite

que os dados na *Web* sejam interligados semanticamente, desde que estejam em um formato padrão acessível e gerenciável por tecnologias apropriadas, relacionando-se uns aos outros.

### 2.3 *Linked Data*

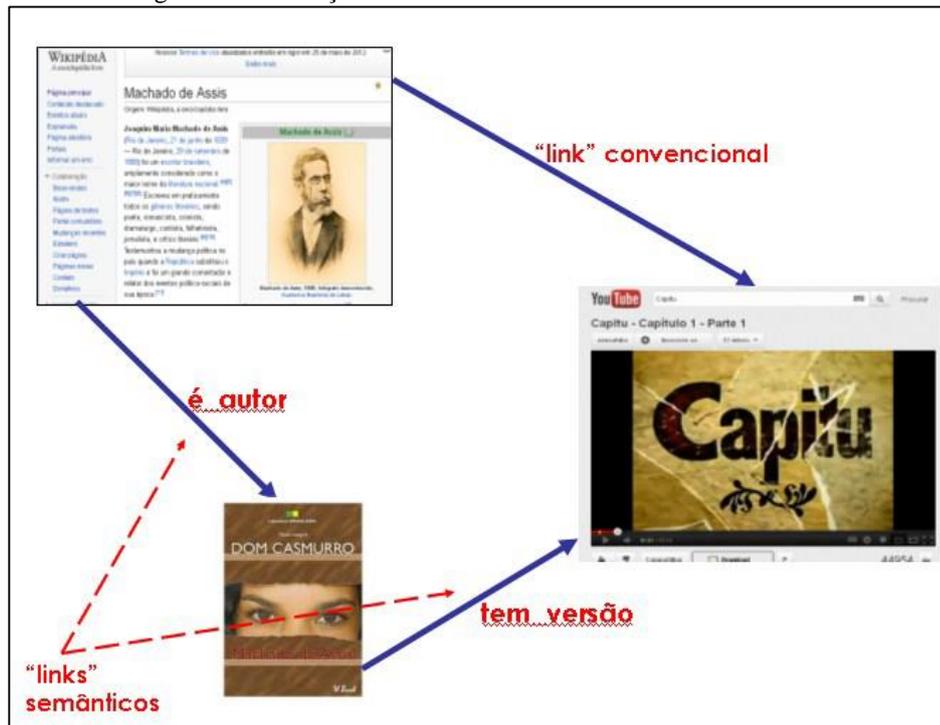
A *Web Semântica* é uma *Web* de dados. As tecnologias e os padrões da *Web Semântica* criam um ambiente em que os computadores conseguem buscar os dados e explorar seus relacionamentos semânticos. Entretanto, para viabilizar o funcionamento da *Web Semântica*, os dados devem ser disponibilizados e seus formatos devem seguir alguns padrões, além de serem passíveis de pesquisa e de gerenciamento por parte das ferramentas da *Web Semântica*. Para completar, a *Web Semântica* precisa ter acesso também aos relacionamentos semânticos que envolvem os dados disponíveis. Essa coleção de dados inter-relacionados na *Web* tem o nome de *Linked Data*, ou dados interligados, ou, ainda, dados linkados (W3C, 2015a).

Segundo Marcondes (2012), os dados interligados agregam valor e enriquecem a navegação, pois os recursos são ligados por *links* semânticos. Os *links* convencionais são etiquetas textuais, significativas apenas para os usuários, que servem de ponto de ligação para que programas navegadores acessem um recurso a partir de outro recurso, sem explicitar o significado dessa ligação. Já os *links* semânticos são significativos para os programas navegadores. Eles são processados de forma o mais completa possível, proporcionando melhor exploração e enriquecimento cognitivo a respeito da ligação entre ambos os recursos da *Web Semântica*.

Para mostrar a diferença entre *links* convencionais e *links* semânticos, Marcondes (2012) exemplifica a relação entre Machado de Assis, escritor brasileiro, e sua obra *Dom Casmurro*, que, por sua vez, é ligada a sua versão em vídeo chamada “Capitu”, uma minissérie brasileira de televisão. O *link* convencional apenas relaciona uma página sobre Machado de Assis diretamente ao vídeo Capitu, sem explicitar a semântica da relação. Para quem já sabe dessa relação é fácil deduzir, mas para outros que não saibam é importante e enriquecedor, além de o computador poder interpretar a semântica e ser capaz de fazer inferências e desempenhar tarefas mais sofisticadas.

A Figura 4 mostra esse exemplo de Marcondes (2012).

Figura 4 - Diferença entre *links* convencionais e *links* semânticos



Fonte: Marcondes, 2012.

Os dados interligados, que têm como base a funcionalidade da *Web Semântica*, agregam valor à navegação de um internauta, pois interligam os recursos informacionais na Internet por meio de *links* semânticos, dando um aspecto mais natural e intuitivo à navegação dos internautas. Para Marcondes (2012), os *links* na *Web* convencional servem para interligar páginas. Na proposta de *Linked Data*, os *links* são responsáveis por interligar recursos — por exemplo, o registro de um livro em um catálogo de biblioteca ou um verbete em uma enciclopédia sobre tal recurso —, além de possuírem significado e expressarem a relação semântica entre os recursos ligados.

A proposta de *Linked Data* é usar a *Web* como meio para inter-relacionar semanticamente dados que antes não eram conectados ou que eram conectados de outra forma mais simples (HEATH, 2016). Para Berners-Lee (2006), a *Web Semântica* é mais do que publicar dados na *Web*; é estabelecer *links* entre eles, para que a *Web* de dados possa ser explorada por uma pessoa ou máquina. Associada ao *Linked Data*, permite encontrar outros dados relacionados.

De acordo com Berners-Lee (2006), assim como a *Web* tradicional, a *Web* de dados também é formada por documentos. Porém, enquanto os *links* na *Web* tradicional são âncoras de relacionamento escritos em *HyperText Markup Language* (HTML), na *Web*

Semântica os *links* entre os dados são expressos em RDF. Os recursos, conteúdos ou conceitos são identificados unicamente pelo *Uniform Resource Identifier* (URI).

Berners-Lee (2006) estabelece quatro regras básicas para publicar dados no *Linked Data*: a) usar URIs para identificar qualquer recurso na *Web*, como se fosse um nome; b) os URIs devem implementar o protocolo *Hypertext Transfer Protocol* (HTTP), para que seja possível procurar por esses nomes; c) quando alguém procura por um URI, informações úteis e relevantes devem ser apresentadas, usando padrões da *Web Semântica*, por exemplo RDF, *Simple Protocol and RDF Query Language* (SPARQL); e d) um URI deve conter *links* para outros URIs na *Web*, para permitir a descoberta e exploração de outros recursos relacionados semanticamente.

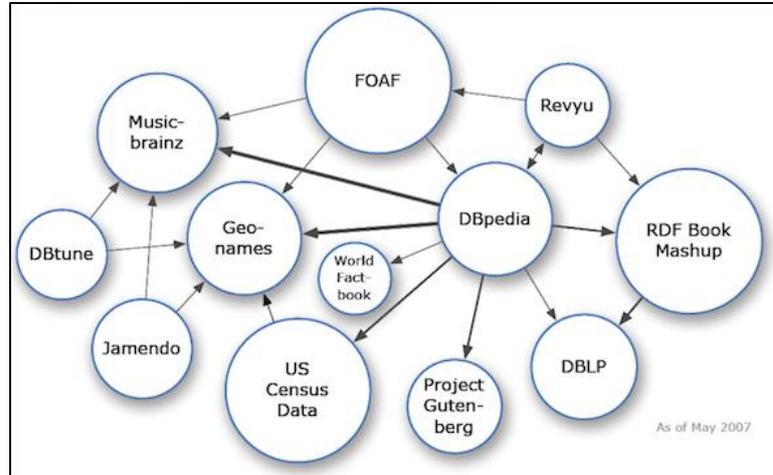
A W3C (2015a) dá um exemplo típico de uso do conceito de *Linked Data*, o DBpedia. Trata-se de um site que, resumidamente, torna os conteúdos da Wikipedia interligados semanticamente. O DBpedia não é importante somente por conter os dados da Wikipedia, mas também por criar relações desses dados com outros conjuntos de dados na *Web*. Por meio dessas relações extras, as aplicações são capazes de explorar conhecimentos adicionais e, provavelmente, mais precisos vindos de outros conjuntos de dados. Isso contribui para uma experiência muito mais enriquecedora e interessante para o usuário na *Web*.

Heath e Bizer (2011) relatam que muito da origem da *Web* de dados se deve aos esforços da comunidade de pesquisa da *Web Semântica*, em especial do projeto da W3C *Linking Open Data* (LOD), fundado em 2007, cujos objetivos eram: inicializar a *Web* de Dados com dados abertos já existentes e livres de licença; adaptá-los para o formato RDF, seguindo os princípios de *Linked Data*; e publicá-los na *Web*. A filosofia seguida sempre por esse projeto é que ele é aberto a qualquer entidade que queira publicar seus dados, bastando implementar os princípios de *Linked Data*. Esse foi o fator chave para o sucesso que se tornou o LOD.

A Figura 5 e a Figura 6 representam o diagrama em nuvem do LOD, no qual estão representados os conjuntos de dados publicados no formato *Linked Data*, tanto pela comunidade do LOD quanto por outras entidades e organizações (HEATH; BIZER, 2011). A Figura 5 apresenta o diagrama na sua versão de maio de 2007. A versão do diagrama da Figura 6 é de agosto de 2014. Comparando a Figura 5 com a Figura 6, são visíveis a evolução e a expansão do projeto LOD e do *Linked Data* desde sua criação. Cada círculo no diagrama é considerado um nó, que corresponde a um conjunto de dados publicado no *Linked Data*. As setas indicam a existência de *links* entre itens de nós distintos. Setas mais escuras indicam um

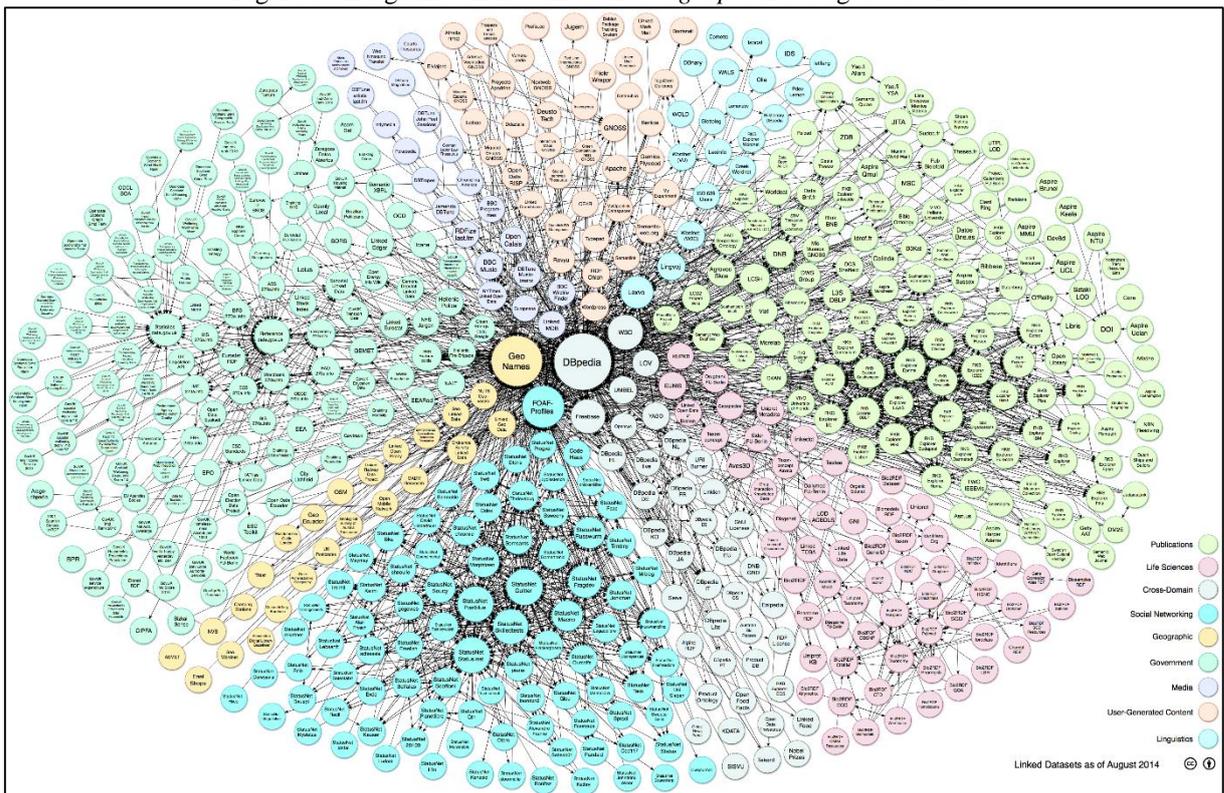
número maior de *links* no relacionamento e setas bidirecionais indicam uma relação mútua entre os nós (HEATH; BIZER, 2011).

Figura 5 - Diagrama em nuvem do *Linking Open Data* maio de 2007



Fonte: Schmachtenberg *et al.*, 2014.

Figura 6 - Diagrama em nuvem do *Linking Open Data* agosto de 2014



Fonte: Schmachtenberg *et al.*, 2014.

Para publicar os dados em *Linked Data*, é preciso usar tecnologia e formato comuns — nesse caso, o RDF — para que conjuntos de dados possam ser interligados, independentemente do seu formato de origem (W3C, 2015a).

## 2.4 Resource Description Framework

Para criar uma relação semântica entre os dados na *Web*, é preciso adotar o *Resource Description Framework* (RDF), modelo padrão para intercâmbio de dados na *Web*, considerado a base para a concepção da *Web Semântica*, que pode ser definido como um sistema de modelagem de dados ou, ainda, como uma forma de representação dos metadados dos recursos na *Web*. Ele descreve os recursos na *Web* por meio de triplas: *Subject*, *Predicate* e *Object*. *Subject*, ou sujeito, é o próprio recurso, identificado pelo seu *International Resource Identifier* (IRI), que é um identificador único na *Web*. *Predicate*, ou predicado, compreende as propriedades, ou os atributos, que descrevem o recurso. *Object*, ou objeto, envolve os valores correspondentes às propriedades dos recursos (CATARINO, 2014).

O guia do usuário do RDF, apresentado por Schreiber e Raimond (2014), relata que o RDF permite fazer assertivas em relação a recursos. Recursos podem ser qualquer coisa; por exemplo, documentos, pessoas, objetos e conceitos abstratos. Essas assertivas têm um formato simples. Elas sempre são compostas por: sujeito, predicado e objeto. Uma assertiva em RDF expressa uma relação entre dois recursos. O sujeito e o objeto representam os dois recursos envolvidos. O predicado representa a natureza do relacionamento entre os dois recursos. O relacionamento é representado sempre em um sentido (do sujeito para o objeto), sendo chamado no RDF de “propriedade”. Pelo fato de serem compostas por três elementos, as assertivas são também chamadas de “triplas”.

Schreiber e Raimond (2014) mostram exemplos de triplas em RDF, expressos informalmente em pseudocódigo, adaptado no Quadro 1.

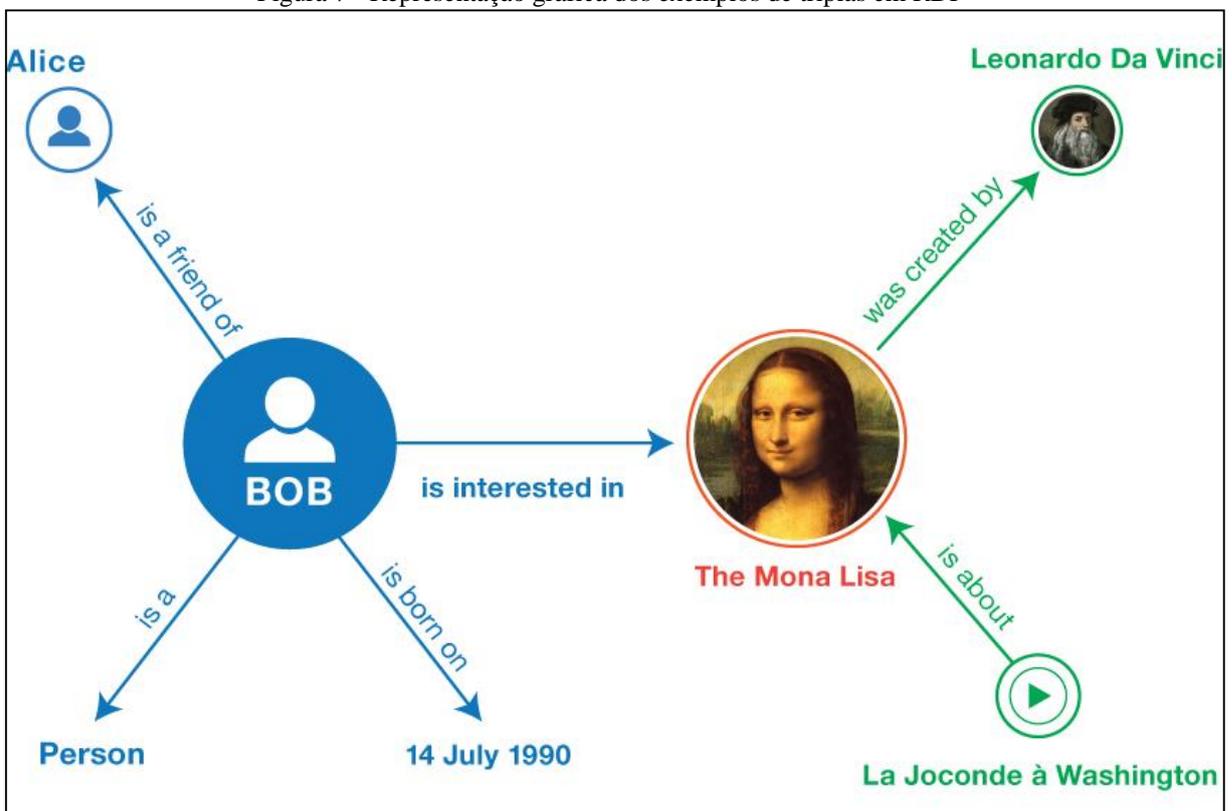
Quadro 1 - Exemplos de triplas em RDF

Sujeito	Predicado	Objeto
<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: Adaptado de Schreiber e Raimond, 2014.

Em RDF, o mesmo recurso pode ser referenciado em várias triplas. Nesse exemplo, Bob é o sujeito de quatro triplas, Mona Lisa é sujeito de uma tripla e objeto de duas triplas. É justamente essa possibilidade de se ter um mesmo recurso fazendo o papel de sujeito em uma tripla e de objeto em outra tripla que cria as conexões entre as triplas, uma das grandes vantagens do RDF. A representação do mesmo exemplo em Schreiber e Raimond (2014), apresentada na Figura 7, mostra as triplas conectados graficamente. Os grafos são compostos por nós e arcos, os sujeitos e os objetos das triplas são os nós e os predicados são os arcos.

Figura 7 - Representação gráfica dos exemplos de triplas em RDF



Fonte: Schreiber e Raimond, 2014.

Segundo Mori e Carvalho (2004), RDF é uma tecnologia desenvolvida pela W3C que tem uma arquitetura genérica de metadados capaz de descrever os recursos na *Web* por meio da adoção de padrões de metadados, com o objetivo de promover o intercâmbio de dados. Não se trata de uma linguagem, mas sim de um modelo de dados para descrever recursos semanticamente na *Web* por meio da implementação de metadados. O RDF usa o conceito de sentença (que é formada por um par de elementos: propriedade-valor) e um recurso ao qual essa propriedade está relacionada.

O padrão RDF é o primeiro passo em direção à *Web Semântica*. Ele expressa o significado dos documentos e dos recursos. É o padrão básico da proposta de dados interligados, que descreve os recursos por meio de assertivas sobre eles. Os conteúdos são publicados de forma estruturada, interconectados por meio de *links* semânticos, que expressam a natureza ou a semântica da ligação entre eles. A semântica no RDF é inteligível por computadores, porque é composta pela tripla sujeito, predicado e objeto de uma assertiva, que são os elementos do metamodelo do RDF. Em uma assertiva RDF sempre é possível identificar sobre qual recurso é feita a assertiva (sujeito), o tipo de assertiva que está envolvida (predicado) e o que especificamente é dito sobre o recurso (objeto) (MARCONDES, 2012).

Para completar os dados em RDF e prepará-los para serem interpretados semanticamente por computadores, é preciso definir termos e relações, para que a máquina consiga ler e extrair as informações de um documento. Esse papel é desempenhado pelas ontologias.

## 2.5 Ontologia

Explicam Souza e Alvarenga (2004, p. 136): “A palavra ‘ontologia’ deriva do grego *onto* (ser) e *logia* (discurso escrito ou falado)”. No campo de Tecnologia da Informação (TI), ontologia corresponde a um documento ou arquivo que formaliza os termos, os conceitos e suas relações, em uma área de conhecimento. Pode ser entendida também como uma especificação de relacionamento de entidades e suas interações em um domínio específico de conhecimento ou atividade. A ontologia possibilita construir um vocabulário comum que permite o compartilhamento e o reuso de informações entre membros de uma comunidade, sejam eles humanos ou computadores (SOUZA; ALVARENGA, 2004).

Carrasco, Thaller e Vidotti (2015) afirmam que o mapeamento de metadados para alguma ontologia é um modo de torná-los interoperáveis. Esses autores consideram a ontologia em seu trabalho como sendo:

[...] uma descrição formal explícita dos conceitos, em um determinado domínio, propriedades que descrevem as características e atributos de cada conceito, e restrições sobre os atributos. Uma ontologia, juntamente com todas as instâncias individuais de suas classes, constitui uma base de conhecimento (CARRASCO; THALLER; VIDOTTI, 2015, p. 212).

Ontologia é considerada também uma descrição direta de diferentes características e atributos de conceitos sobre determinado assunto. Uma ontologia, em uma área de conhecimento qualquer, consiste em um grupo de instâncias de classes do seu domínio compostas por propriedades que descrevem várias características e atributos dos conceitos (DANDAGI; SIDNAL, 2016).

No âmbito da *Web Semântica*, a W3C (2015b) explica que as ontologias definem os conceitos e as relações, também chamados de “termos”, que descrevem e representam determinada área de conhecimento. Ontologias são usadas para classificar os termos em uma aplicação particular, declarar relações e definir restrições de uso desses termos. De acordo com esse autor, na *Web Semântica* não há uma divisão muito bem definida entre o conceito de ontologia e o de vocabulário. Geralmente, refere-se a ontologia quando a coleção de termos é mais formal e complexa e o conceito vocabulário quando não há muita necessidade de formalismo.

Oliveira (2011) também apresenta o conceito de ontologia na área da *Web Semântica*. Segundo ele, a ontologia é responsável por definir termos e suas relações em determinada área de conhecimento, para que agentes de software possam interpretar e extrair o máximo de informação possível de um documento ou recurso na *Web*. A ontologia é considerada um fator chave para o desenvolvimento da *Web Semântica*, pois fornece um entendimento comum e compartilhado de um domínio, inteligível por pessoas e máquinas.

A ontologia desempenha papel muito importante na *Web Semântica*, promovendo o acesso, a interoperabilidade e a recuperação da informação, com base em conteúdo. As ontologias “permitem expressar regras possibilitando a um programa deduzir significados da informação guardados no documento, ou seja, permitem manipular os termos de uma maneira mais útil e eficiente” (OLIVEIRA, 2011, p. 9).

## 2.6 SPARQL

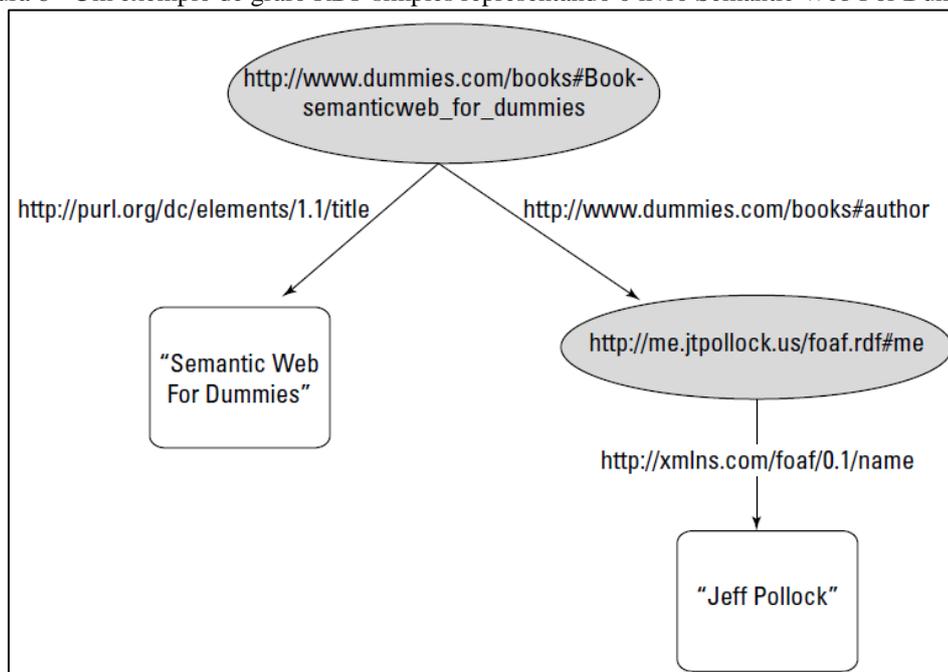
De acordo com Pollock (2009), o SPARQL, ou *Simple Protocol and RDF Query Language*, é uma linguagem de consulta e protocolo de acesso a dados no formato RDF. É considerada a linguagem de consulta primária da *Web Semântica*. Análogo à linguagem de consultas *Structured Query Language* (SQL) nos bancos de dados relacionais, é a linguagem de consulta padrão para dados em RDF. Desde 2008, tem sido a linguagem de consulta para

RDF oficialmente recomendada pela W3C. É capaz de consultar não só grafos RDF, como também outras fontes de dados que possam ser mapeadas para RDF. Como na maioria das linguagens de consultas declarativas, a consulta em SPARQL especifica um padrão nos dados que devem ser combinados em um conjunto de resultados. Dada uma tripla padrão específica como parâmetro em uma consulta, o processador SPARQL considera conjuntos de triplas no modelo RDF que correspondem ao padrão buscado (POLLOCK, 2009).

O SPARQL contém recursos para consultar padrões de grafos obrigatórios e opcionais, juntamente com suas conjunções e disjunções. Ele também suporta teste de valor extensível e consultas restritas por meio de grafo RDF. Os resultados retornados pela consulta podem ser conjuntos de resultados ou grafos RDF. A maioria das formas de consulta SPARQL é composta por um conjunto de padrões de triplas, chamados de “padrão de grafo básico”, semelhantes a triplas RDF, porém o sujeito, predicado e objeto podem ser representados por variáveis. Um padrão de grafo básico corresponde a um subgrafo dos dados em RDF quando os termos RDF desse subgrafo podem ser substituídos pelas variáveis. O resultado é um grafo RDF equivalente ao subgrafo (PRUD'HOMMEAUX; SEABORNE, 2008).

Um exemplo de consulta pode ser encontrado em Pollock (2009). A Figura 8 mostra o grafo RDF do livro *Semantic Web For Dummies*. O livro tem o título de *Semantic Web For Dummies*, cuja autoria é de Jeff Pollock.

Figura 8 - Um exemplo de grafo RDF simples representando o livro *Semantic Web For Dummies*



Fonte: Pollock, 2009.

A Figura 9 mostra um exemplo de uma consulta SPARQL que busca por livros de autoria de Jeff Pollock e ordena os resultados por título do livro. Na cláusula WHERE, são especificados os padrões de triplas. O primeiro padrão equivale às instâncias RDF que são do tipo livro (rdf:type Book). O segundo padrão equivale às instâncias que têm um relacionamento de autoria (book:author) com Jeff Pollock. Como esses dois padrões estão envolvidos pelos símbolos de chaves, “{ }”, na cláusula WHERE, trata-se de uma conjunção. O símbolo de interrogação na frente da palavra *livro* indica que se trata de uma variável que, neste caso, é aquilo que a consulta procura. Os símbolos de ponto final na cláusula WHERE indicam o final de um padrão de tripla. Essa consulta da Figura 9 retornaria o resultado “*Semantic Web For Dummies*” (POLLOCK, 2009).

Figura 9 - Exemplo de consulta SPARQL que busca livros de autoria de Jeff Pollock

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX books:<http://www.dummies.com/books#>

SELECT ?title
WHERE {
    ?book rdf:type books:Books .
    ?book books:author http://me.jtpollock.us/foaf.rdf#me .
    ?book dc:title ?title .
}
ORDER BY ?title
```

Fonte: Pollock, 2009.

## 2.7 Acervos de museus na Web Semântica

Dentre os trabalhos que envolvem a disponibilização de acervos de museus na Web Semântica, destaca-se o projeto Europeana que, segundo Doerr *et al.* (2010), consiste não somente de um portal que provê acesso a milhões de itens de acervos relacionados com diversos tipos de instituições e patrimônios culturais, mas também de um projeto que possibilita criar uma rica contextualização semântica na representação de acervos de bibliotecas, arquivos, museus e outras instituições culturais da Europa, e que também oferece funções semânticas mais complexas e elaboradas de tal maneira que não seria possível implementá-las por meio de sistemas tradicionais. Para isso, os dados dos acervos

representados na Europeana são sistematicamente interligados ao LOD e outros recursos da *Web* utilizando-se boas práticas e recomendações da *Web Semântica*.

Doerr *et al.* (2010, p. 2, tradução nossa) explica que “A estratégia técnica da Europeana foi concebida tanto para contribuir quanto para se beneficiar do crescimento do paradigma do *Linked Data*”.<sup>1</sup> A Europeana utiliza o *Europeana Data Model* (EDM), um modelo de dados desenvolvido para estruturar os dados dos acervos das instituições, independentemente do formato, padrão ou tipo de instituição. O EDM foi desenvolvido por meio de padrões e técnicas de domínio público, com base na *Web Semântica* (ISAAC, 2013).

Segundo Patrício (2012), antes do modelo EDM, a Europeana já adotava o *Europeana Semantic Elements*, um esquema de metadados comum a todos os dados de seus acervos, capaz de mapear conjuntos de dados heterogêneos para uma representação uniforme dos dados e promover a interoperabilidade entre os sistemas. O EDM não foi concebido para substituir o *Europeana Semantic Elements*, mas sim para aprimorá-lo e atender à demanda de adaptar os dados da Europeana para os formatos de *Web Semântica* e *Linked Data*.

Com o objetivo de encontrar outros trabalhos que envolvam museus na *Web Semântica*, procedeu-se à busca literal pelos termos *museu*, *museum*, *museo*, *acervo*, *collection* e *colección* no diagrama em nuvem do *Linking Open Data*, em sua versão de agosto de 2014, apresentado no item 2.3 deste trabalho, e incluindo também o nó que corresponde ao conjunto de dados da Europeana, por se tratar de dados de bibliotecas, arquivos, museus e outras instituições culturais da Europa, foram encontrados cinco conjuntos de dados publicados no LOD contendo algum dos termos de pesquisa. São eles:

- Amsterdam Museum em *Europeana Data Model* (EDM) — conjunto de dados do Museu de Amsterdã, que contém mais de 70.000 itens relacionados com a história da cidade. Seus dados foram convertidos para RDF seguindo o EDM, modelo de dados desenvolvido pela Europeana (DE BOER *et al.*, 2013). Foi publicado no LOD em 2011.
- British Museum Collection — contém os mesmos dados, porém em formato RDF, do acervo do Museu Britânico disponibilizado na *Web* tradicional. A versão atual dos dados está em fase de teste e o trabalho de desenvolvimento está em andamento. Foi publicado no LOD em 2011.

---

<sup>1</sup> “The Europeana technical strategy thus was conceived to both contribute to and benefit from the growing Linked Data paradigm” (Doerr *et al.*, 2010, p. 2).

- Europeana — conjunto de dados de acervos de bibliotecas, arquivos, museus e outras instituições culturais da Europa, representados no modelo de dados EDM, desenvolvido para este fim. Foi publicado no LOD em 2015.
- Mis Museos — conjunto de dados que oferece um índice semântico de museus, artistas e obras de arte. Os dados provêm de itens dos acervos da Rede Digital de Coleções de Museus da Espanha e de museus públicos da Espanha que não fazem parte dessa rede. Foi publicado no LOD em 2012.
- Museos de España — conjunto de dados de acervos de mais de 1.500 museus públicos e privados da Espanha. Publicado no LOD em 2012.

Baker, Noy e Swick (2012) apresentam uma seção do site oficial da W3C que descreve estudos de casos e casos de uso que envolvem a *Web Semântica*. Estudos de casos são descrições de sistemas desenvolvidos por organizações e que estão em produção atualmente. Casos de uso são exemplos de protótipos de sistemas desenvolvidos por organizações que ainda não usam oficialmente esses sistemas. Filtrando a lista de estudos de caso e casos de uso por área de atividade igual a “museum” e “cultural heritage”, os seguintes estudos de caso foram retornados:

- Aquaring — portal semântico europeu que reúne recursos do mundo aquático. A coleção digital do mundo aquático é de um domínio do conhecimento muito complexo, devido ao dinamismo e à heterogeneidade dos dados. O objetivo deste projeto foi reunir os dados de diferentes formatos e terminologias, armazenados em locais diferentes em um portal multilíngue na *Web Semântica* (RODRÍGUEZ, 2009).
- Europeana — disponibiliza milhões de objetos de acervos de bibliotecas, arquivos, museus e outras instituições culturais da Europa. Para isso, o projeto coleta metadados descritivos e *links* para recursos na *Web* de todas essas instituições, sendo, assim, um conjunto de dados e metadados muito heterogêneos (ISAAC, 2012).
- Portal do Patrimônio Cultural de Cantabria — portal semântico que integra informações relevantes sobre o patrimônio cultural da Cantabria, região do norte da Espanha. O objetivo deste portal foi reunir as informações, antes dispersas, em formatos diferentes, heterogêneas, fragmentadas e pouco detalhadas, deste patrimônio cultural (HERNÁNDEZ, 2007).

Neste trabalho, também foram citados dois estudos que não apareceram no diagrama em nuvem do LOD nem na lista de estudos de caso e casos de uso de *Web Semântica* da W3C, mas que também utilizam tecnologias e ferramentas da *Web Semântica* para integrar dados de acervos de museus, solucionar problemas de recuperação da informação e torná-los interoperáveis: o de Hyvönen *et al.* (2005) e o de Skevakis *et al.* (2013).

É possível perceber que, seguindo os passos descritos neste item do estudo, não foram encontrados museus brasileiros que utilizaram a *Web Semântica* para integrar e disponibilizar seus acervos na *Web*.

### 3 METODOLOGIA

Este item aborda os aspectos metodológicos deste estudo, contemplando a caracterização da pesquisa, os procedimentos metodológicos executados, o objeto da pesquisa, a origem dos dados dos acervos e a proposta do arcabouço conceitual.

#### 3.1 Caracterização da pesquisa

Quanto ao objetivo, este estudo é caracterizado como uma pesquisa exploratória; que, segundo Casarin e Casarin (2012), proporciona maior conhecimento sobre determinado problema ou fenômeno. Na maioria das vezes, configura-se como uma pesquisa preparatória, por abordar um tema ainda pouco explorado ou já conhecido, porém visto sob uma nova perspectiva. Este estudo pretende explorar os conceitos e as tecnologias sobre *Web Semântica* e *Linked Data* que permitem a integração de dados de acervos da RMECC e a disponibilização na *Web*.

Quanto ao método de pesquisa, este estudo trata-se de uma pesquisa experimental, com base em prova de conceito. Gil (2002, p. 47) apresenta conceitos sobre o método de pesquisa experimental: “Essencialmente, a pesquisa experimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto”. Para Cervo, Bervian e Silva (2007), a manipulação das variáveis na pesquisa experimental permite compreender a relação entre as causas e os efeitos de determinado fenômeno.

De acordo com Casarin e Casarin (2012), a pesquisa experimental se aplica quando se faz necessário intervir em um grupo estudado, para apurar os efeitos dessa intervenção. Este estudo propõe uma intervenção nos dados de acervos da RMECC, que atualmente não se encontram integrados e interoperáveis, formatando-os para o padrão da *Web Semântica*, aplicando os princípios de *Linked Data*. Depois dessa intervenção, os efeitos são verificados, para averiguar se a solução proposta satisfaz os objetivos deste trabalho, caracterizando, assim, uma pesquisa experimental.

Quanto à natureza, este estudo compreende uma pesquisa aplicada, cujo objetivo é a aplicação prática para solucionar problemas específicos. Para Gil (2008), a pesquisa

aplicada é aquela que investiga a aplicação e utilização do conhecimento na prática. “Sua preocupação está menos voltada para o desenvolvimento de teorias de valor universal que para a aplicação imediata numa realidade circunstancial” (GIL, 2008, p. 27). Mascarenhas (2012) acrescenta que a pesquisa aplicada estuda o problema em um contexto e busca soluções para os desafios enfrentados nesse ambiente específico.

Este estudo busca testar um arcabouço conceitual por meio da implementação de um protótipo de *software* que será aplicado na prática, no contexto da RMECC, usando parte dos acervos dos museus e dos espaços que a compõe, com o objetivo de desenvolver uma solução para integração dos dados dessa rede.

### 3.2 Procedimentos metodológicos

Como procedimentos metodológicos, os seguintes passos foram seguidos:

- a) Pesquisa bibliográfica na literatura existente e em trabalhos relacionados sobre tecnologias e ferramentas da *Web Semântica* e *Linked Data* que apoiem o estudo e a construção do arcabouço conceitual.
- b) Levantamento e coleta de dados de acervos da RMECC de diferentes bancos de dados, em diferentes formatos de dados, que fazem parte deste trabalho. Foram consultados os responsáveis pelos sistemas de banco de dados de acervos de cada integrante da RMECC e coletados os dados das coleções para fazerem parte deste estudo.
- c) Agrupamento e organização dos dados dos acervos em um modelo de dados comum e um único repositório. De acordo com Hyvönen *et al.* (2005), antes de adaptá-los para a *Web Semântica* é necessário consolidar os dados dos diferentes acervos em um repositório global.
- d) Converter os dados consolidados para o formato de dados da *Web Semântica*, o RDF. Segundo Catarino (2014), o RDF é o formato de dados padrão da *Web Semântica*, o qual permite criar relações semânticas entre os recursos na *Web*.
- e) Adaptação dos dados em RDF para o formato *Linked Data* e ligação dos dados com outros recursos externos da *Web Semântica*, por exemplo DBpedia e GeoNames. Segundo Berners-Lee (2006), Bizer, Cyganiak e

Heath (2007), Heath (2016) e Skevakis *et al.* (2013), o *Linked Data* envolve também a ligação dos dados em RDF com outros recursos externos já existentes na *Web*.

- f) Desenvolvimento de um protótipo de aplicação que permite integrar e visualizar os dados no formato *Web*. Para que o usuário final pudesse visualizar as interconexões entre os itens dos acervos dos museus, Hyvönen *et al.* (2005) representaram os itens como páginas *Web* interligadas.
- g) Prova de conceito para validar o experimento realizado.

### 3.3 Objeto da pesquisa

A unidade de análise desta pesquisa é a Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais, que, conforme consta em seu Regimento (UNIVERSIDADE FEDERAL DE MINAS GERAIS, 2010), constitui coordenadoria vinculada à Pró-Reitoria de Extensão da UFMG e tem por missão formular a proposição e o desenvolvimento de ações nas áreas de Pesquisa, Extensão, Graduação e Pós-Graduação, relacionadas aos Museus e Espaços de Ciência e Cultura da UFMG, para apoiar e estimular o desenvolvimento desses espaços. Integram a RMECC os museus e espaços de Ciência e Cultura da UFMG. São eles:

- Casa Padre Toledo
- Centro Cultural UFMG
- Centro de Estudos Literários e Culturais — Acervo de Escritores Mineiros
- Centro de Memória da Educação Física, do Esporte e do Lazer
- Centro de Memória da Enfermagem
- Centro de Memória da Engenharia
- Centro de Memória da Farmácia
- Centro de Memória da Medicina
- Centro de Memória da Odontologia
- Centro de Memória da Veterinária
- Centro de Pesquisa, Memória e Documentação da Faculdade de Educação

- Centro de Referência da Música de Minas — Museu Clube da Esquina
- Centro de Referência em Cartografia Histórica
- Centro de Referência em Patrimônio Geológico
- Espaço do Conhecimento UFMG
- Espaço Memória do Cinema
- Estação Ecológica UFMG
- Museu da Escola de Arquitetura
- Museu de Ciências Morfológicas
- Museu de História Natural e Jardim Botânico da UFMG

A RMECC, criada em 2000, tem os objetivos: integrar os museus e espaços da Rede; promover e facilitar a divulgação, comunicação e intercâmbio institucional, científico, tecnológico e cultural entre os integrantes; e desenvolver estratégias de uso e acesso às coleções dos acervos. Segundo o site oficial da RMECC, a estruturação deste órgão em rede favorece a qualificação das ações e a ampliação do atendimento do seu público-alvo, formado pela comunidade universitária, estudantes da educação básica, pesquisadores e a população em geral (UNIVERSIDADE FEDERAL DE MINAS GERAIS, 2015).

### 3.4 Origem dos dados

Como o objetivo deste trabalho é integrar dados de acervos da RMECC no formato de *Linked Data*, é necessário coletar dados de diferentes museus e espaços que compõem a RMECC. Dos 20 integrantes dessa rede, três não são museus e/ou não são responsáveis pela guarda e disponibilização de acervos. Dos 17 integrantes que possuem acervo, somente cinco disponibilizam seus acervos de alguma forma na *Web*. Por isso, foi preciso solicitar dados dos acervos não disponibilizados aos integrantes da RMECC que os guardam.

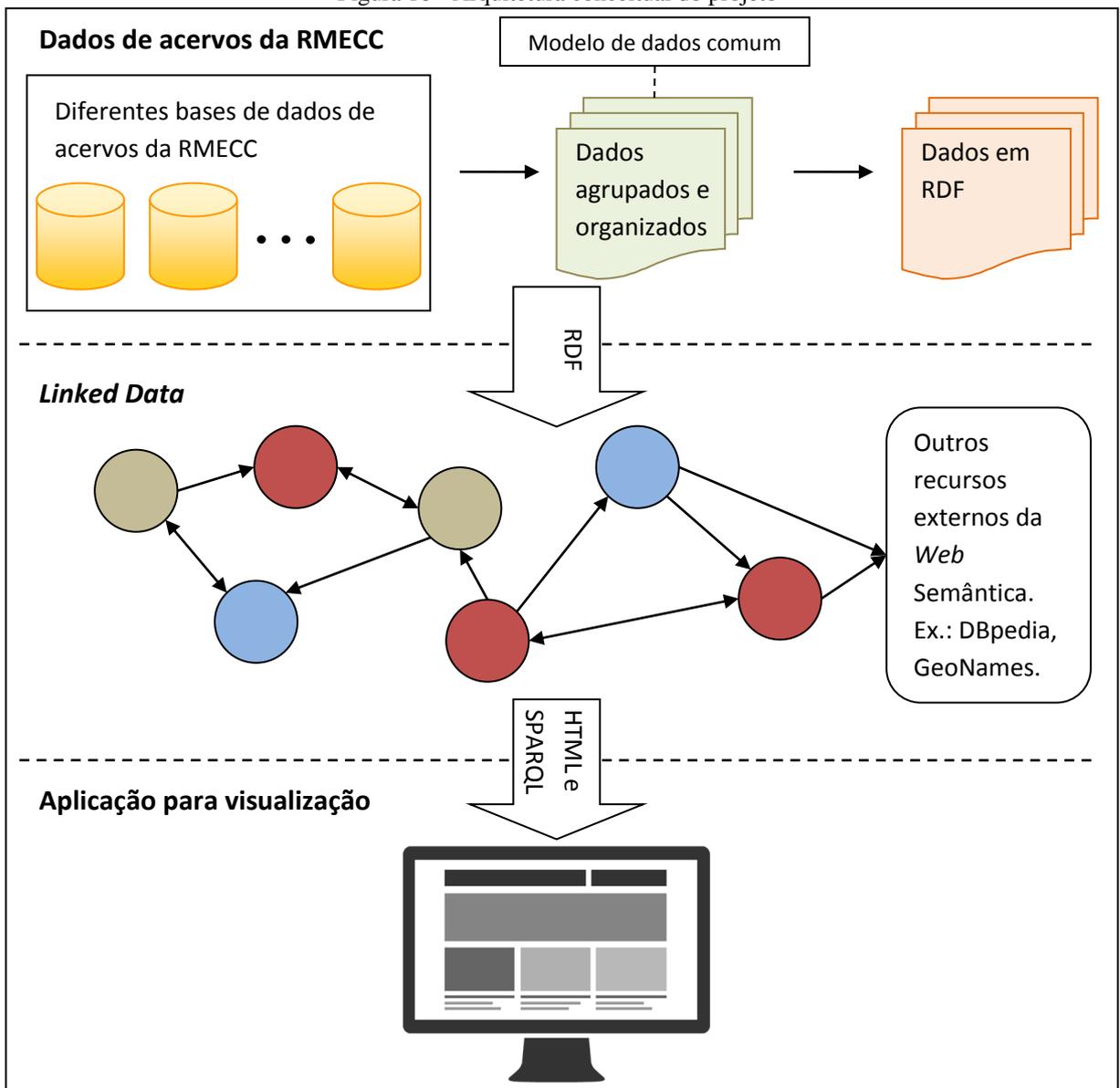
Como se pretende neste estudo propor uma solução capaz de integrar dados de acervos da RMECC utilizando conceitos da *Web Semântica* e *Linked Data*, não foram utilizados todos os dados de todos os acervos dos integrantes da rede. Foram selecionados dados de diferentes formatos e de diferentes acervos que fossem representativos e suficientes para testar o protótipo de aplicação.

Tanto os dados dos acervos da RMECC já disponibilizados de alguma forma na *Web* quanto aqueles dos acervos que foram levantados com os integrantes da Rede apresentam formatos e padrões diferentes e estão armazenados em bancos de dados distintos. Procedeu-se ao agrupamento e à organização dos dados dos acervos levantados. Após essa estruturação de todos esses dados dos acervos, foi aplicado o modelo RDF para representar as informações, já que esse é o padrão utilizado para *Linked Data* na *Web Semântica*.

### **3.5 Proposta de arcabouço conceitual**

A arquitetura conceitual, representada na Figura 10, segue um pouco do modelo de arcabouço conceitual apresentado por Azevedo (2014), trabalho relacionado a este e que possui semelhanças na metodologia de pesquisa. O arcabouço aqui proposto tem por objetivo identificar os principais componentes que irão fazer parte da solução proposta por este estudo, suas interconexões e responsabilidades, sem aprofundar nos detalhes de especificação do sistema. Essa arquitetura conceitual esboça os objetivos, as metodologias, as tecnologias e os conceitos levantados no referencial teórico deste trabalho.

Figura 10 - Arquitetura conceitual do projeto



Fonte: Elaborada pelo autor, 2016.

A arquitetura conceitual da Figura 10 é dividida em três camadas, sendo a sequência de leitura da mais acima para baixo. Na primeira camada, têm-se os dados de acervos dos museus e dos espaços integrantes da RMECC em formatos heterogêneos, armazenados em diferentes bancos de dados. Posteriormente, na mesma camada, os dados foram agrupados e organizados em um único repositório, aplicando um modelo de dados comum para todos os acervos. Ainda na primeira camada, os dados foram convertidos para o formato de dados da *Web Semântica*, chamado RDF.

Os dados em RDF permitem a transição da primeira para segunda camada, na qual os dados são adaptados para o formato de *Linked Data* e interligados a outros recursos externos da *Web Semântica*, por exemplo, DBpedia e GeoNames. Por meio das tecnologias

HTML e SPARQL, é possível fazer a transição da segunda para a terceira camada, na qual os dados de acervos da RMECC foram disponibilizados para visualização em um protótipo de aplicação.

## 4 IMPLEMENTAÇÃO

### 4.1 Coleta, seleção e organização dos dados

Para obter acesso aos dados dos 17 integrantes da RMECC que possuem acervo, foi enviada uma solicitação a cada um deles, na qual se explicavam os motivos e se especificavam as informações requisitadas. Esse processo teve a anuência e o apoio da RMECC, que forneceu informações sobre os contatos dos espaços integrantes. De acordo com a resposta obtida de três integrantes, seus acervos não foram inventariados e não possuíam registro deles em formato digital, o que inviabilizou sua inclusão neste estudo, restando, então, 14 integrantes.

Dos 14 integrantes que poderiam fazer parte deste estudo, foram coletados dados de acervos que pertencem a nove espaços. Trata-se de acervos distintos em relação ao formato dos dados adotado, no âmbito de temáticas variadas, como Artes, Arqueologia, Arquitetura, Biblioteca, Botânica, Cinema, Documentação e Preservação, Educação Física, Esporte e Lazer, Enfermagem, Engenharia, Farmácia, Geologia, História, Literatura, Medicina, Mineralogia, Paleontologia e Zoologia. Devido à variedade de informação dos itens que compõem esses nove acervos e dos formatos de arquivos que os descrevem, características necessárias para a composição deste trabalho, os dados obtidos foram considerados representativos e suficientes para a realização do experimento.

Os dados coletados apresentam formatos diferentes, como: arquivos de texto (extensões de arquivo .doc, .docx, .pdf), planilhas eletrônicas (extensões de arquivo .xls, .xlsx), páginas HTML, sistemas de bancos de dados e imagens (extensões de arquivo .jpg, .png).

No caso da RMECC, vários itens pertencentes ao mesmo acervo tratam de um mesmo assunto ou se assemelham quanto às características e aos dados. Para evitar repetição, procedeu-se à seleção de itens para compor a amostra. Dentre os dados de acervos dos nove espaços, foram selecionados dados de 280 itens, dando sempre prioridade a itens com informações as mais completas e heterogêneas possíveis, classificados dentre as temáticas citadas anteriormente. Foi dada preferência a itens representados por imagens, para enriquecer o protótipo de aplicação no aspecto visualização. Os dados selecionados foram organizados

por espaço aos quais pertencem e distribuídos em planilhas eletrônicas, para facilitar a manipulação.

O Quadro 2 relaciona resumidamente a quantidade e o tipo de dados dos acervos da RMECC selecionados para este estudo, por instituição. A coluna “Formatos de dados” apresenta os formatos diversos em que os dados foram enviados ou encontrados. A coluna “Disponibilização” indica se os dados foram coletados de arquivos enviados pela instituição, retirados diretamente de site ou de arquivos hospedados nele. A última coluna apresenta a quantidade de itens de cada acervo da RMECC selecionados.

Quadro 2 - Dados selecionados dos acervos da Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais

<b>Nome da instituição</b>	<b>Formatos de dados</b>	<b>Disponibilização</b>	<b>Quantidade de itens selecionados</b>
Centro de Estudos Literários e Culturais - Acervo de Escritores Mineiros	Arquivos de texto .doc	Arquivos enviados	44
Centro de Memória da Educação Física, do Esporte e do Lazer	Arquivos de texto .pdf	Arquivos no site	24
Centro de Memória da Enfermagem	Arquivos de texto .doc, planilhas eletrônicas .xls	Arquivos no site e arquivos enviados	20
Centro de Memória da Engenharia	Planilhas eletrônicas .xlsx	Arquivos enviados	35
Centro de Memória da Farmácia	Arquivos de texto .doc e .docx, planilhas eletrônicas .xlsx	Arquivos enviados	20
Centro de Memória da Medicina	Páginas HTML	Dados no site	15
Espaço Memória do Cinema	Planilhas eletrônicas .xlsx	Arquivos enviados	30
Museu da Escola de Arquitetura	Páginas HTML	Dados no site	21
Museu de História Natural e Jardim Botânico da UFMG	Arquivos de texto .doc e .docx, planilhas eletrônicas .xls e .xlsx, páginas HTML, sistemas de banco de dados	Arquivos enviados e dados no site	71
<b>Total</b>			<b>280</b>

Fonte: Elaborado pelo autor, 2016.

## 4.2 Integração e armazenamento dos dados em um único repositório

Para integrar os dados selecionados e armazená-los em um repositório comum, foi preciso, inicialmente, desenvolver um modelo de dados que fosse capaz de abranger as diferentes informações e temáticas dos itens dos acervos. Para isso, foram usadas como referência as Diretrizes Internacionais de Informação sobre Objetos de Museus: Categorias de Informação do *International Committee for Documentation* (CIDOC), encontradas em CIDOC - ICOM (2014), primeiro volume da coleção “Gestão e Documentação de Acervos: textos de referência”, que tem por objetivo publicar textos de referência estrangeiros traduzidos para a língua portuguesa que apoiam as atividades de documentação em museus e instituições culturais similares.

Segundo seu site oficial (CIDOC – ICOM, 2010), o CIDOC é um comitê que pertence ao *International Council of Museums* (ICOM). Visa organizar e publicar boas práticas e desenvolvimentos na área de Documentação de Museus, para apoiar as atividades das comunidades dos museus em geral. Segundo CIDOC – ICOM (2014), é referência internacional no campo da Documentação em Museus e instituições semelhantes. É composto por mais de 650 membros, distribuídos em 60 países, compreendendo profissionais especialistas em documentação, catalogação, gerentes de TI, desenvolvedores de sistemas, consultores e profissionais de formação.

De acordo com CIDOC – ICOM (2014), as Diretrizes Internacionais de Informação sobre Objetos de Museus: Categorias de Informação do CIDOC é um documento fundamental por apresentar conceitos, métodos e procedimentos orientados para as atividades de documentação em acervos museológicos. É uma referência internacional, fruto de estudos realizados ao longo do tempo, por diferentes profissionais da área, em diversos países.

Essas *Diretrizes* não são consideradas como normas e seu uso não é obrigatório. Elas servem como uma orientação geral e inicial sobre o que se deve considerar no momento da identificação e documentação dos itens que compõem um acervo. “Portanto, as Diretrizes apresentam uma proposta, que não é a única e tampouco definitiva, para orientar os profissionais a respeito de quais informações devem ser registradas quando há necessidade de se documentar um acervo, e como devem fazê-lo” (CIDOC – ICOM, 2014, p. 12).

As Diretrizes Internacionais de Informação sobre Objetos de Museus: Categorias de Informação do CIDOC reúnem categorias de informação que orientam o processo de registro dos objetos em coleções museológicas. Servem como guia para museus

individualmente, organizações de documentação ou ainda desenvolvedores de sistemas de documentação para museus. Vale destacar, para os propósitos deste estudo, que uma das atribuições das *Diretrizes* é servir de base para o compartilhamento de informações em um museu ou entre museus diferentes (CIDOC – ICOM, 2014).

As *Diretrizes* podem ser usadas como referência tanto para sistemas de inventário como para sistemas de catalogação de acervos. Sistemas de inventário dizem respeito às informações básicas de gestão do acervo de cada item, com informações sobre a definição de responsabilidades e segurança. Sistemas de catalogação envolvem informações mais detalhadas sobre a importância histórica dos itens do acervo (CIDOC – ICOM, 2014).

Para CIDOC – ICOM (2014), essas *Diretrizes* foram formadas para atender os museus e seus diversos tipos de temáticas de acervos, incluindo: Arqueologia, História da Cultura, Arte, Ciência e Tecnologia e Ciências Naturais. Esse foi também um fator importante para a escolha das *Diretrizes* como referência para este trabalho. Elas não contemplam todas as necessidades de detalhamento dos itens. Muitas vezes, é necessário incluir informações adicionais mais específicas sobre determinado acervo, de acordo com a necessidade de cada museu.

O CIDOC – ICOM (2014) apresenta 22 Grupos de Informação, que contêm uma ou mais Categorias de Informação. Conforme recomendado nesse documento, foram selecionados, de acordo com as necessidades deste estudo e a disponibilidade de informações, os seguintes Grupos de Informação e as respectivas Categorias de Informação para documentar os itens dos acervos, listados no Quadro 3.

Quadro 3 – Grupos de Informação e Categorias de Informação selecionados do *International Committee for Documentation*

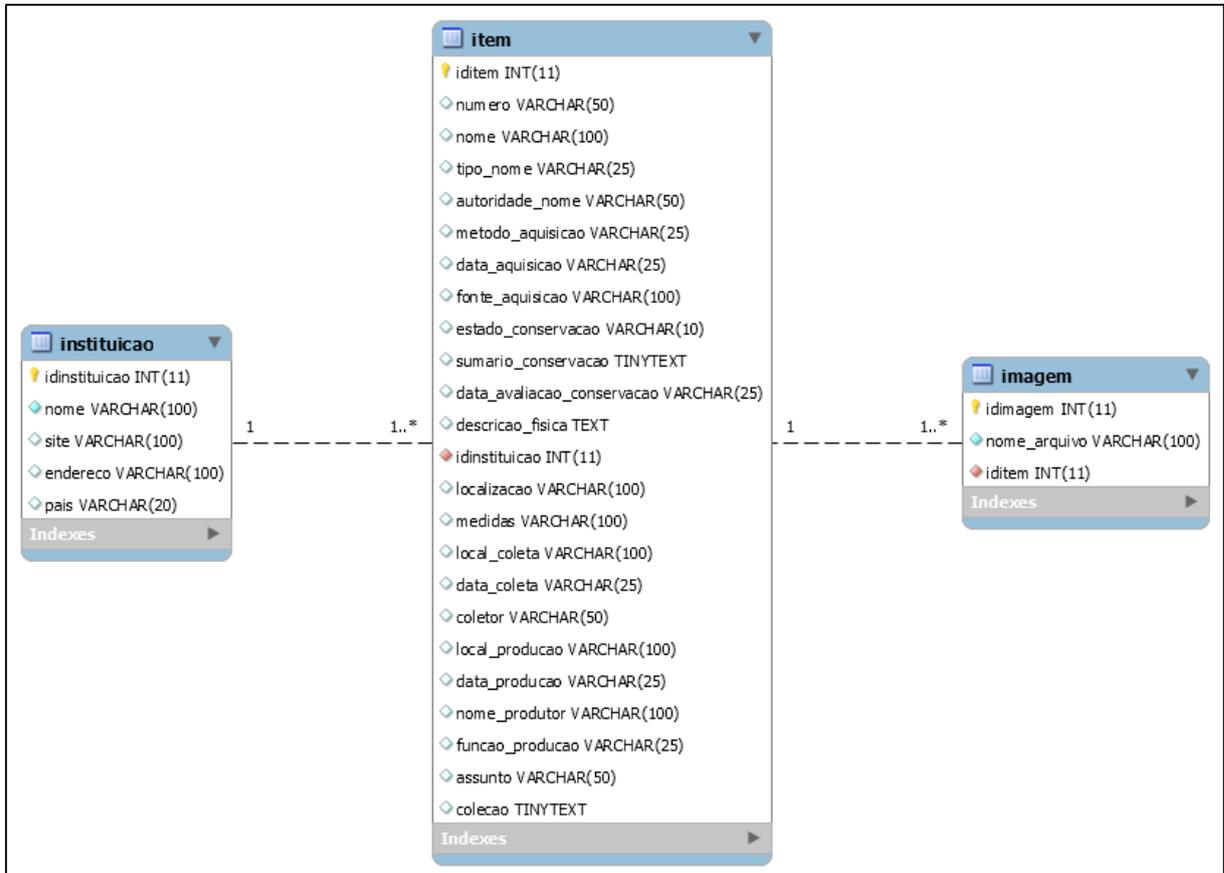
<b>Grupo de Informação</b>	<b>Categoria de Informação</b>
Informação sobre aquisição	Método de aquisição
	Data de aquisição
	Fonte da aquisição
Informação sobre estado de conservação	Estado de conservação
	Sumário do estado de conservação
	Data de avaliação do estado de conservação
Informação sobre descrição	Descrição física
Informação sobre imagem	Número de referência da imagem
Informação sobre instituição	Nome da instituição
	Endereço da instituição
	País da instituição
Informação sobre localização	Localização atual
Informação sobre medição	Medidas
Informação sobre coleta de objeto	Local da coleta
	Data da coleta
	Coletor
Informação sobre nome de objeto	Nome do objeto
	Tipo de nome do objeto
	Autoridade de nome do objeto
Informação sobre número de objeto	Número do objeto
Informação sobre produção de objeto	Local de produção
	Data de produção
	Nome do produtor
	Função da produção
Informação sobre assunto representado	Assunto representado

Fonte: Elaborado pelo autor, 2016.

Após esse levantamento das informações que seriam armazenadas sobre os itens dos acervos da RMECC, foi desenvolvido o modelo de dados, que corresponde ao diagrama de Entidade Relacionamento, representado na Figura 11. Para implementar o modelo de dados, foi utilizado o Sistema de Gerenciamento de Banco de Dados (SGBD) MySQL, por ser

um *software* livre, que permite importar dados no formato *Comma-separated Values* (CSV), compatível com a ferramenta utilizada neste estudo para manipular os dados em RDF, compatível também com a linguagem usada para o desenvolvimento de consultas SPARQL, ambos detalhados mais adiante.

Figura 11 - Diagrama de Entidade Relacionamento para os acervos da Rede de Museus e Espaços de Ciências e Cultura da Universidade Federal de Minas Gerais



Fonte: Elaborada pelo autor, 2016.

Conforme reproduzido na Figura 11, foram criadas três tabelas. A tabela “instituicao” armazena dados sobre cada um dos nove espaços da RMECC selecionados para este estudo. A esta tabela foram adicionados o campo chave primária “idinstituicao” e o campo “site”, que contém a *Uniform Resource Locator* (URL) do *Web* site da instituição. A tabela “item” armazena os dados dos itens dos acervos. A ela foram adicionados o campo chave primária “iditem” e o campo “colecao”, que descreve a coleção à qual um determinado item pertence. A tabela “imagem” é responsável por armazenar informações a respeito das imagens que representam determinado item de um acervo. Sua chave primária é o campo “idimagem”.

A Figura 11 mostra também os relacionamentos entre as tabelas. A tabela “instituicao” tem o relacionamento 1-N com a tabela “item”, o que significa que cada item

pertence a uma instituição, que, por sua vez, possui N itens. Para isso, o campo chave estrangeira “idinstituicao” foi criado na tabela “item”. A tabela “item” tem o relacionamento 1-N com a tabela “imagem”. Ou seja, um item possui N imagens e cada imagem representa um item. O campo chave estrangeira “iditem” foi criado na tabela “imagem” para fazer o relacionamento.

Após a criação do banco de dados, seguindo as especificações acima, foi necessário importar os dados dos acervos para o banco de dados. Com esse objetivo, foram criadas três planilhas eletrônicas, cada uma correspondendo a uma tabela do banco de dados e cada coluna das planilhas correspondendo a um campo de uma tabela. Os dados dos acervos, antes organizados em planilhas eletrônicas por instituição, foram copiados para estas três novas planilhas, que posteriormente foram salvas no formato CSV. Usando a interface administrativa do MySQL, procedeu-se à importação dos dados em cada tabela, por meio de cada arquivo CSV, integrando os dados de diferentes acervos em um repositório comum.

### **4.3 Conversão dos dados para o formato RDF**

Para fazer a conversão dos dados dos acervos em um banco de dados relacional para o formato de dados da *Web Semântica*, o RDF, foi utilizada a plataforma D2RQ. Tal procedimento foi o mesmo adotado nos estudos de Azevedo (2014), Pinto (2014) e Silva (2014), que também envolvem a conversão de dados relacionais para RDF. Segundo Cyganiak (2012), o D2RQ é uma plataforma capaz de acessar dados em banco de dados relacionais virtualmente como se fossem dados no formato RDF. Ele oferece acesso em formato RDF ao conteúdo do banco de dados relacional sem a necessidade de replicar todo o banco de dados em um armazém RDF.

O D2RQ é uma ferramenta gratuita que permite consultar dados usando a linguagem SPARQL, mesmo que esses dados não estejam formatados em RDF, acessar o conteúdo do banco de dados no formato *Linked Data* como se estivessem publicados na *Web*, gerar um arquivo RDF personalizado que corresponde a todo o banco de dados relacional e acessar as informações não formatadas em RDF do banco de dados relacional usando a funcionalidade *Apache Jena Application Programming Interface (API)* (CYGANIAK, 2012).

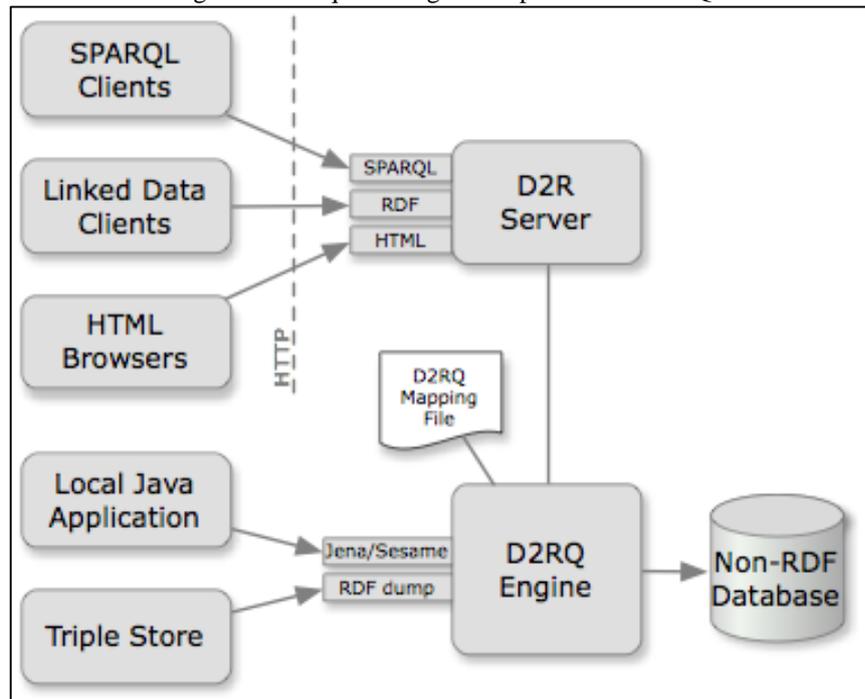
De acordo com Cyganiak (2012), os principais componentes do D2RQ são:

- *D2RQ Mapping Language*, uma linguagem declarativa de mapeamento que descreve as relações entre o modelo de dados relacional e os vocabulários RDF ou ontologias *Web Ontology Language* (OWL). O mapeamento gera um grafo RDF virtual, que contém informações provenientes do banco de dados relacional. O grafo RDF virtual pode ser acessado de formas diferentes, por meio do D2RQ, dependendo da implementação e do objetivo. O D2RQ oferece acesso ao banco de dados mapeado via SPARQL, servidor *Linked Data*, gerador de arquivo RDF, interface HTML simplificada e Apache Jena API (CYGANIAK *et al.*, 2012).
- *D2RQ Engine*, um *plug-in* para o conjunto de ferramentas Jena *Semantic Web* que usa os arquivos de mapeamento relacional-RDF para reescrever solicitações vindas do Jena API para a linguagem de consulta SQL, executa a consulta e retorna os resultados da consulta para as camadas mais externas de apresentação.
- *D2R Server*, servidor HTTP que disponibiliza uma visualização em *Linked Data*, uma visualização em HTML para a função de *debug* e um protocolo *endpoint* SPARQL do banco de dados relacional. Usa o arquivo de mapeamento na linguagem *D2RQ Mapping Language* para mapear o conteúdo do banco de dados relacional para o formato RDF e permite que os dados em RDF sejam consultados e explorados.

Cyganiak (2012) destaca que a ferramenta D2RQ segue os princípios de *Linked Data* e as boas práticas recomendadas para a *Web Semântica*. Por exemplo, o *D2R Server* atribui um URI para cada entidade que pertence ao banco de dados relacional e torna esse URI capaz de ser resolvido. Ou seja, um recurso RDF pode ser recuperado simplesmente acessando o URI da entidade na *Web*. Outro exemplo: a interface *Web* simples e os grafos RDF navegáveis compartilham o mesmo URI.

A Figura 12 apresenta a arquitetura geral da plataforma D2RQ.

Figura 12 - Arquitetura geral da plataforma D2RQ



Fonte: Cyganiak, 2012.

O primeiro passo usando a plataforma D2RQ constituiu em executar a ferramenta *generate-mapping*. Segundo Cyganiak (2012), esta ferramenta gera um arquivo de mapeamento na linguagem *D2RQ Mapping Language*, por meio da análise da estrutura do banco de dados relacional, na qual estão armazenados os dados. Esse arquivo de mapeamento, também chamado “mapeamento padrão”, mapeia cada tabela para uma nova classe *Resource Description Framework Schema* (RDFS), dando a ela o mesmo nome da tabela de origem, e cada coluna da tabela para uma propriedade da classe RDFS, que recebe o mesmo nome da coluna correspondente. O arquivo de mapeamento pode ser usado do jeito que foi gerado ou pode ser customizado de acordo com a necessidade.

O arquivo de mapeamento é escrito na linguagem *Terse RDF Triple Language*, mais conhecida como *Turtle*. Segundo Beckett *et al.* (2014), trata-se de uma sintaxe textual para RDF que permite escrever grafos RDF, por meio de uma linguagem mais compacta e natural, com abreviações para uso comum de padrões e tipos de dados.

A Figura 13 mostra trechos do arquivo de mapeamento “rededemuseus-ori.ttl”, gerado pela ferramenta *generate-mapping* do D2RQ, sem modificações. Este arquivo também foi reproduzido na íntegra no Apêndice A.

Figura 13 - Trecho de código do arquivo de mapeamento rededemuseus-ori.ttl

```

...
094 # Table item
095 map:item a d2rq:ClassMap;
096     d2rq:dataStorage map:database;
097     d2rq:uriPattern "item/@@item.iditem@";
098     d2rq:class vocab:item;
099     d2rq:classDefinitionLabel "item";
100     .
101 map:item__label a d2rq:PropertyBridge;
102     d2rq:belongsToClassMap map:item;
103     d2rq:property rdfs:label;
104     d2rq:pattern "item #@@item.iditem@";
105     .
106 map:item_iditem a d2rq:PropertyBridge;
107     d2rq:belongsToClassMap map:item;
108     d2rq:property vocab:item_iditem;
109     d2rq:propertyDefinitionLabel "item iditem";
110     d2rq:column "item.iditem";
111     d2rq:datatype xsd:integer;
112     .
113 map:item_numero a d2rq:PropertyBridge;
114     d2rq:belongsToClassMap map:item;
115     d2rq:property vocab:item_numero;
116     d2rq:propertyDefinitionLabel "item numero";
117     d2rq:column "item.numero";
118     .
119 map:item_nome a d2rq:PropertyBridge;
120     d2rq:belongsToClassMap map:item;
121     d2rq:property vocab:item_nome;
122     d2rq:propertyDefinitionLabel "item nome";
123     d2rq:column "item.nome";
...

```

Fonte: Elaborada pelo autor, 2016.

Na Figura 13, as linhas 094 a 099 declaram a classe mapeada do tipo “vocab:item”, tipo padrão gerado pelo *generate-mapping*, que corresponde à tabela “item”. As linhas 106 a 123 definem as propriedades da classe “item”, que correspondem aos campos “iditem”, “numero” e “nome”, com seus tipos, gerados por padrão, “vocab:item\_iditem”, “vocab:item\_numero” e “vocab:item\_nome”, respectivamente.

De acordo com Bizer, Cyganiak e Heath (2007) e Hyland, Atezing e Villazón-Terrazas (2014), o ideal é reusar ontologias padronizadas, difundidas e já bem conhecidas pela comunidade da *Web Semântica* sempre que for possível, e não simplesmente usar os vocabulários gerados automaticamente por ferramentas como o *generate-mapping* do D2RQ. Antes de definir qualquer novo termo, é preciso pesquisar por outros termos que expressem o mesmo significado para os dados e reusá-los, para, assim, contribuir com a inclusão e expansão da *Web Semântica*.

Cyganiak (2012) relata que o *generate-mapping* gera uma nova ontologia para cada banco de dados, associando os nomes das tabelas como nomes das classes e os nomes

das colunas como nomes das propriedades. Aplicações de *Web Semântica* são capazes de interpretar mais facilmente os dados se o arquivo de mapeamento for customizado, substituindo os termos gerados automaticamente por termos de ontologias mais usados e de livre acesso. Para isso, pode-se usar qualquer ferramenta de edição de texto.

O próximo passo consistiu em customizar o arquivo de mapeamento, reutilizando ontologias e termos mais difundidos e aceitos pela comunidade da *Web Semântica*. Foram feitas buscas na *Web* para cada classe e propriedade que representassem melhor a entidade em questão.

O Quadro 4, o Quadro 5 e o Quadro 6 mostram as classes e propriedades reutilizadas com sucesso para as classes Imagem, Instituição e Item, respectivamente. Já o Quadro 7 mostra os termos criados para algumas propriedades da classe Item, por não ter encontrado outros que melhor representassem tais entidades.

Quadro 4 - Classe e propriedades reutilizadas para a classe Imagem

<b>Termo</b>	<b>Tipo</b>	<b>Nome da tabela ou campo</b>
crm:E38_Image	classe	imagem
crm:P1_is_identified_by	propriedade	idimagem
crm:P48_has_preferred_identifier	propriedade	nome_arquivo
crm:P138_represents	propriedade	iditem

Fonte: Elaborado pelo autor, 2016.

Quadro 5 - Classe e propriedades reutilizadas para a classe Instituição

<b>Termo</b>	<b>Tipo</b>	<b>Nome da tabela ou campo</b>
foaf:Organization	classe	instituição
crm:P1_is_identified_by	propriedade	idinstituicao
skos:prefLabel	propriedade	nome
foaf:homepage	propriedade	site
dbo:address	propriedade	endereco
dbo:country	propriedade	pais

Fonte: Elaborado pelo autor, 2016.

Quadro 6 - Classe e propriedades reutilizadas para a classe Item

<b>Termo</b>	<b>Tipo</b>	<b>Nome da tabela ou campo</b>
crm:E19_Physical_Object	classe	item
crm:P1_is_identified_by	propriedade	iditem
crm:P48_has_preferred_identifier	propriedade	numero
crm:P102_has_title	propriedade	nome
crm:P30i_custody_transferred_through	propriedade	metodo_aquisicao
dbp:acquisitionDate	propriedade	data_aquisicao
crm:P29i_received_custody_through	propriedade	fonte_aquisicao
crm:P3_has_note	propriedade	descricao_fisica
crm:P55_has_current_location	propriedade	localizacao
crm:P43_has_dimension	propriedade	medidas
dcterms:spatial	propriedade	local_producao
dcterms:created	propriedade	data_producao
crm:P108i_was_produced_by	propriedade	nome_producao
dcterms:subject	propriedade	assunto
crm:P46i_forms_part_of	propriedade	colecacao
crm:P52_has_current_owner	propriedade	idinstituicao

Fonte: Elaborado pelo autor, 2016.

Quadro 7 - Termos criados para algumas propriedades da classe Item

<b>Termo</b>	<b>Tipo</b>	<b>Nome da tabela ou campo</b>
vocab:name_type	propriedade	tipo_nome
vocab:named_by	propriedade	autoridade_nome
vocab:conservation_status	propriedade	estado_conservacao
vocab:conservation_description	propriedade	sumario_conservacao
vocab:conservation_evaluation_date	propriedade	data_avaliacao_conservacao
vocab:gathering_place	propriedade	local_coleta
vocab:gathering_date	propriedade	data_coleta
vocab:gatherer	propriedade	coletor
vocab:production_function	propriedade	funcao_producao

Fonte: Elaborado pelo autor, 2016.

No arquivo de mapeamento, os termos são precedidos por prefixos, que indicam a quais ontologias eles pertencem. Foram reutilizados termos de seis ontologias diferentes, todas bem conhecidas e aceitas na *Web Semântica*.

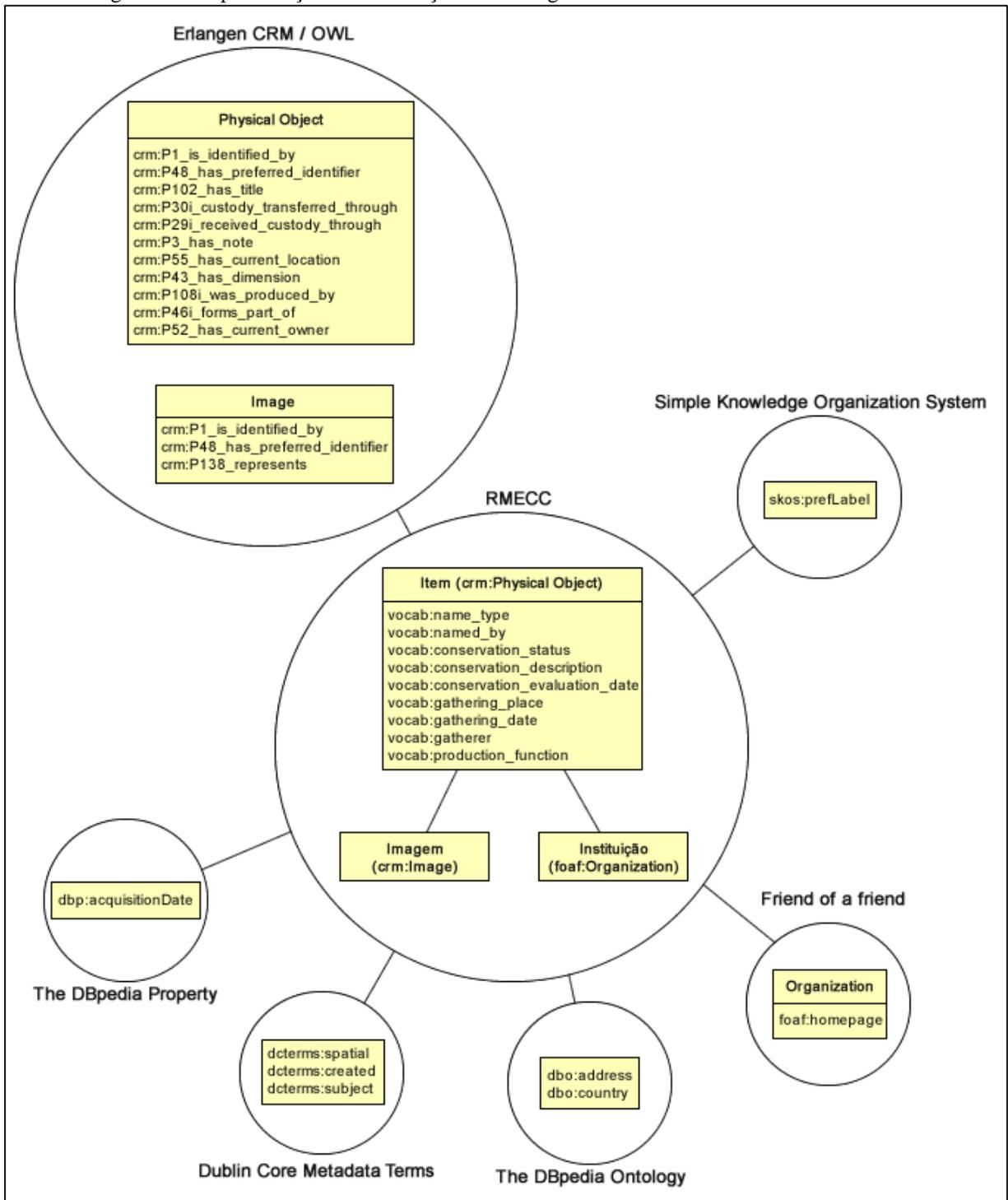
O Quadro 8 resume os prefixos e os nomes dessas ontologias.

Quadro 8 - Prefixos e nomes das ontologias reusadas

<b>Prefixo utilizado</b>	<b>Nome da ontologia reusada</b>
crm	<i>Erlangen CRM / OWL</i>
foaf	<i>Friend of a friend</i>
skos	<i>Simple Knowledge Organization System</i>
dbo	<i>The DBpedia Ontology</i>
dbp	<i>The DBpedia Property</i>
dcterms	<i>Dublin Core Metadata Terms</i>

Fonte: Elaborado pelo autor, 2016.

A Figura 14 representa graficamente a reutilização de ontologias e termos externos da *Web Semântica*, por meio de ligações entre eles, internas e externas ao conjunto de dados dos acervos. Os círculos brancos são os vocabulários ou ontologias, com exceção da RMECC, que representa o conjunto de dados dos acervos. Nos retângulos estão representadas as classes, texto em negrito e centralizado, e as propriedades reutilizadas ou criadas para representar as propriedades no conjunto de dados deste trabalho. À frente dos nomes das classes dentro do conjunto de dados da RMECC, entre parênteses, indicam as classes externas reutilizadas.

Figura 14 - Representação da reutilização de ontologias e termos externos da *Web Semântica*

Fonte: Elaborada pelo autor, 2016.

É importante destacar que a ontologia mais usada neste estudo e a que teve maior prioridade na escolha foi o *Erlangen CRM / OWL*. Trata-se de uma interpretação e

implementação em *Web Ontology Language* (OWL)<sup>1</sup> da ontologia CIDOC *Conceptual Reference Model* (CIDOC CRM), que, por sua vez, foi desenvolvida pelo CIDOC, mesma instituição que publicou as Diretrizes Internacionais de Informação sobre Objetos de Museus: Categorias de Informação, citadas no item 4.2 deste trabalho.

Segundo o ICOM (2016), o CIDOC CRM define uma estrutura formal por meio da qual é possível descrever conceitos e relacionamentos, implícitos e explícitos, usados para a documentação de patrimônios culturais. De acordo com Carrasco, Thaller e Vidotti (2015, p. 211), o CIDOC CRM “é uma ontologia do domínio do patrimônio cultural, criada como uma ferramenta para integração da informação”. O *Erlangen CRM / OWL* foi adotado por alguns projetos na *Web Semântica* que envolvem museus, por exemplo, o *British Museum Collection*, citado no item 2.7 deste trabalho.

A Figura 15 mostra um trecho do arquivo “rededemuseus-modif.ttl”, versão modificada do arquivo “rededemuseus-ori.ttl”, que reusa os termos das ontologias listadas no Quadro 8. O arquivo “rededemuseus-modif.ttl” também se encontra na íntegra no Apêndice B. Nesse arquivo de mapeamento customizado, as linhas 099 a 104 declaram a classe “item” sendo do tipo “crm:E19\_Physical\_Object”, reusado da ontologia *Erlangen CRM / OWL*. As linhas 111 a 128 definem as propriedades “iditem”, “numero” e “nome” dessa mesma classe como sendo do tipo “crm:P1\_is\_identified\_by”, “crm:P48\_has\_preferred\_identifier” e “crm:P102\_has\_title”, respectivamente.

---

<sup>1</sup> OWL é uma linguagem usada para definir e instanciar ontologias próprias para a *Web*. É voltada especialmente para representar informações contidas em documentos ou recursos na *Web* que precisam ser processadas por aplicações. A OWL define o significado de termos em um vocabulário e suas relações capazes de serem interpretados semanticamente por computadores na *Web* (MCGUINNESS; HARMELEN, 2004).

Figura 15 - Trecho de código do arquivo de mapeamento rededemuseus-modif.ttl

```

...
099 # Table item
100 map:item a d2rq:ClassMap;
101     d2rq:dataStorage map:database;
102     d2rq:uriPattern "item/@@item.iditem@";
103     d2rq:class crm:E19_Physical_Object;
104     d2rq:classDefinitionLabel "item";
105     .
106 map:item__label a d2rq:PropertyBridge;
107     d2rq:belongsToClassMap map:item;
108     d2rq:property rdfs:label;
109     d2rq:pattern "item #@@item.iditem@";
110     .
111 map:item_iditem a d2rq:PropertyBridge;
112     d2rq:belongsToClassMap map:item;
113     d2rq:property crm:P1_is_identified_by;
114     d2rq:propertyDefinitionLabel "item iditem";
115     d2rq:column "item.iditem";
116     d2rq:datatype xsd:integer;
117     .
118 map:item_numero a d2rq:PropertyBridge;
119     d2rq:belongsToClassMap map:item;
120     d2rq:property crm:P48_has_preferred_identifier;
121     d2rq:propertyDefinitionLabel "item numero";
122     d2rq:column "item.numero";
123     .
124 map:item_nome a d2rq:PropertyBridge;
125     d2rq:belongsToClassMap map:item;
126     d2rq:property crm:P102_has_title;
127     d2rq:propertyDefinitionLabel "item nome";
128     d2rq:column "item.nome";
...

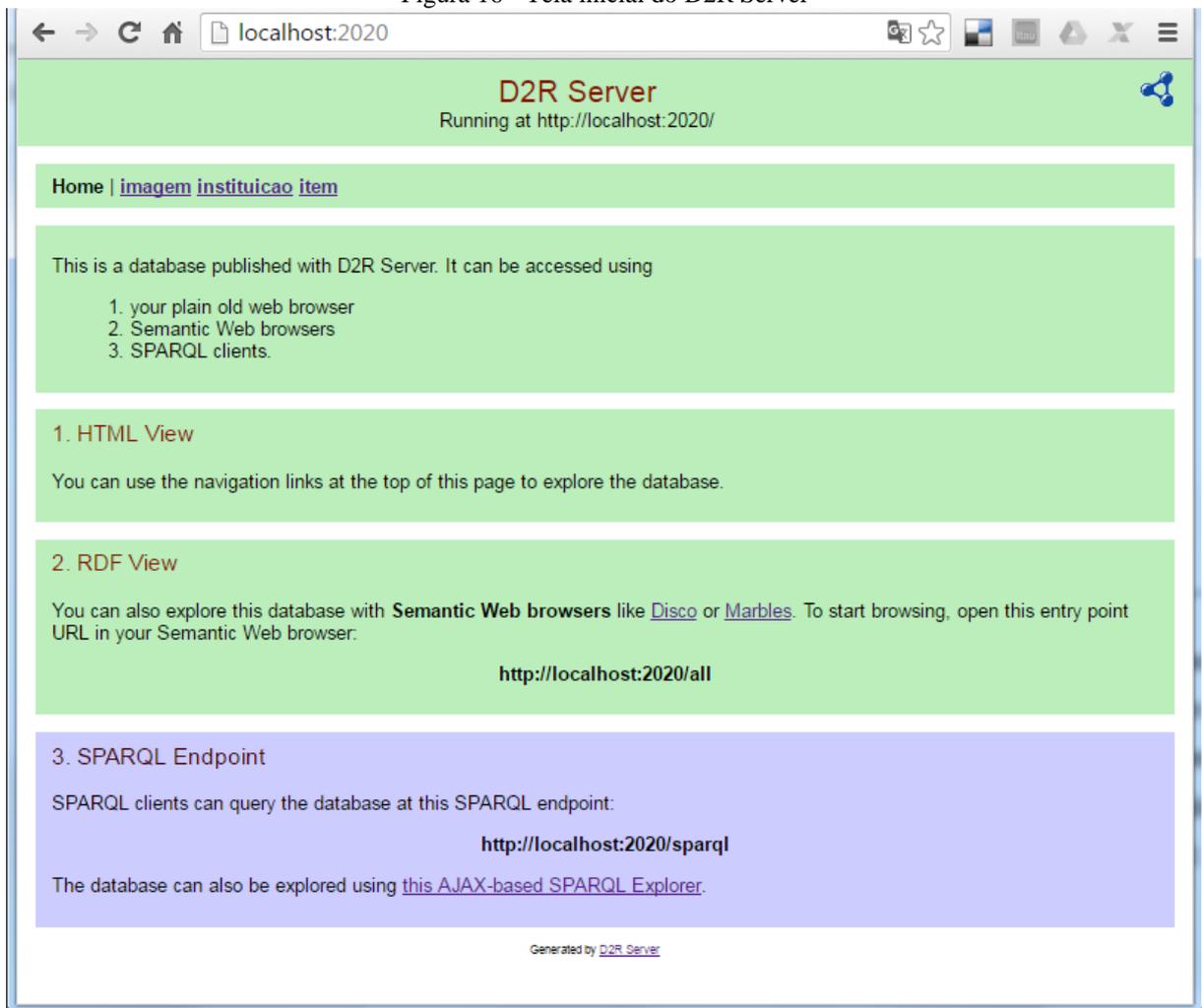
```

Fonte: Elaborada pelo autor, 2016.

Após a personalização do arquivo de mapeamento, utilizou-se a ferramenta do D2RQ chamada *D2R Server*, passando como parâmetro o arquivo “rededemuseus-modif.ttl”. Esta ferramenta serviu para teste e conferência de como os dados foram mapeados para o formato RDF — neste caso, virtualmente, porque os dados não foram convertidos para RDF ainda de fato.

A Figura 16 mostra a tela inicial do *D2R Server* já configurado com o arquivo de mapeamento.

Figura 16 - Tela inicial do D2R Server



Fonte: Elaborada pelo autor, 2016.

O D2R *Server* oferece uma visão HTML simples, que foi útil para visualizar os dados por tabela, na visão relacional, ou por classe, na visão RDF. *Links* foram disponibilizados no canto esquerdo superior, um para cada classe do banco de dados.

A Figura 17 mostra os dados referente ao item com “iditem” igual a 126. O item tem o nome de “Pesos de precisão”. É representado por duas imagens, que são facilmente acessadas por meio de *links*. Pertence ao acervo da instituição identificada pelo “idinstituicao” igual a 5, Centro de Memória da Farmácia, acessível via *link* também.

Figura 17 - Dados referente a um item de acervo visualizados no D2R Server



localhost:2020/page/item/126

**item #126**  
Resource URI: <http://localhost:2020/resource/item/126>

[Home](#) | [All item](#)

Property	Value
crm:P102_has_title	Pesos de precisão
crm:P108i_was_produced_by	
is crm:P138_represents of	< <a href="http://localhost:2020/resource/imagem/29">http://localhost:2020/resource/imagem/29</a> >
is crm:P138_represents of	< <a href="http://localhost:2020/resource/imagem/30">http://localhost:2020/resource/imagem/30</a> >
crm:P1_is_identified_by	126 (xsd:integer)
crm:P29i_received_custody_through	
crm:P30i_custody_transferred_through	
crm:P3_has_note	Os pesos têm formas cilíndricas com botão e pesos variáveis, fabricados em latão. O suporte é de plástico, com divisões diferenciadas de acordo com o tamanho dos pesos, o suporte ainda tem uma subdivisão interna com pequenas folhas de alumínio em mg.
crm:P43_has_dimension	Alt.: 3cm Comp.:15,5cm Prof.:7,2cm
crm:P46i_forms_part_of	
crm:P48_has_preferred_identifier	CMF-000.010
crm:P52_has_current_owner	< <a href="http://localhost:2020/resource/instituicao/5">http://localhost:2020/resource/instituicao/5</a> >
crm:P55_has_current_location	Em exposição (A.5/P.4)
dbp:acquisitionDate	
vocab:conservation_description	Arranhados, desgaste natural, desgaste por uso, sujidades, adesivo de papel inapropriado, presença de ferrugem
vocab:conservation_evaluation_date	
vocab:conservation_status	
dcterms:created	
vocab:gatherer	
vocab:gathering_date	
vocab:gathering_place	
rdfs:label	item #126
vocab:name_type	Nome comum
vocab:named_by	
vocab:production_function	
dcterms:spatial	
dcterms:subject	
rdf:type	crm:E19_Physical_Object

Fonte: Elaborada pelo autor, 2016.

Para converter os dados dos acervos para o formato RDF, foi utilizada a ferramenta *dump-rdf* da plataforma D2RQ, passando como parâmetro o arquivo “rededemuseus-modif.ttl”. Segundo Cyganiak (2012), esta ferramenta é responsável por replicar todo o conteúdo do banco de dados relacional em um único arquivo RDF, usando como referência o arquivo de mapeamento indicado.

A Figura 18 mostra um trecho do arquivo RDF “rededemuseus-ori.rdf”, gerado pelo *dump-rdf*. As linhas 8707 a 8715 representam em RDF os dados referentes à instituição Museu de História Natural e Jardim Botânico da UFMG.

Figura 18 - Trecho de código do arquivo RDF rededemuseus-ori.rdf

```

...
8707 <rdf:Description rdf:about="#instituicao/9">
8708 <skos:prefLabel>Museu de História Natural e Jardim Botânico da UFMG</skos:prefLabel>
8709 <crm:P1_is_identified_by
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">9</crm:P1_is_identified_by>
8710 <dbo:country>Brasil</dbo:country>
8711 <rdfs:label>instituicao #9</rdfs:label>
8712 <dbo:address>Rua Gustavo da Silveira, 1035 - Santa Inês - Belo Horizonte - MG</dbo:address>
8713 <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
8714 <foaf:homepage rdf:resource="https://www.ufmg.br/mhnpjb/">
8715 </rdf:Description>
...

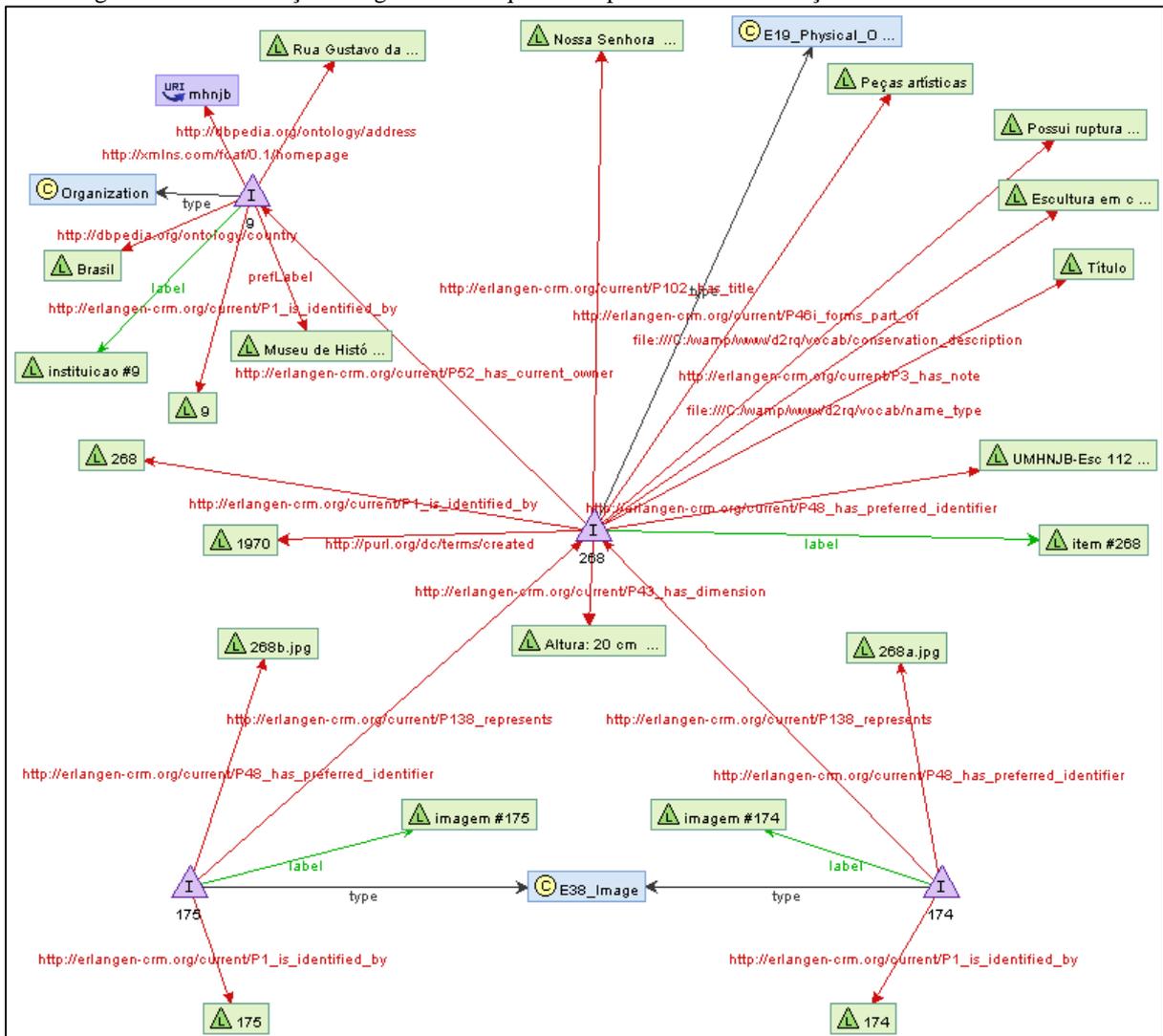
```

Fonte: Elaborada pelo autor, 2016.

Um exemplo de ferramenta usada para visualizar como ficaram os dados interligados dos acervos da RMECC em RDF é o RDF *Gravity*, que, segundo Goyal e Westenthaler (2016), é uma ferramenta utilizada para visualizar grafos em RDF e OWL. Este *software* oferece uma visualização simples, porém muito útil da estrutura dos grafos RDF, com opção de filtro de dados, busca textual e visualização de partes específicas de um conjunto de dados RDF.

A Figura 19, gerada por meio do RDF *Gravity*, é uma visualização dos grafos RDF que correspondem ou têm relação com o item representado pelo “iditem” igual a 268, de nome igual à “Nossa Senhora das Graças”, que pertence ao acervo da instituição Museu de História Natural e Jardim Botânico da UFMG.

Figura 19 - Visualização dos grafos RDF que correspondem ou tem relação com um item de acervo



Fonte: Elaborada pelo autor, 2016.

A Figura 19 mostra os três tipos de classes que representam os dados dos acervos. Existe uma instância da classe Instituição, que é o Museu de História Natural e Jardim Botânico da UFMG, a qual tem a guarda do item de nome “Nossa Senhora das Graças”, que é uma instância da classe Item. Existem duas instâncias da classe Imagem, que representam este mesmo item do acervo. Os relacionamentos semânticos, ou predicados, são representados por setas, que unem os sujeitos e os objetos. Os símbolos usados no RDF *Gravity* e compõem a Figura 19 são explicados em forma de legenda pela Figura 20.

Figura 20 - Legenda dos símbolos usados pelo RDF Gravity

 Organization	Conceitos ou classes.
	Instâncias de classes.
 Peças artísticas	Valor literal (string, inteiro, etc.)
 mhnb	Recurso URI

Fonte: Elaborada pelo autor, 2016.

#### 4.4 Adaptação para o formato *Linked Data* e ligação com outros recursos externos da *Web Semântica*

Para Bizer, Cyganiak e Heath (2007), um dos propósitos básicos de *Linked Data* é interligar dados de diferentes conjuntos de dados por meio de *links* RDF. Existem duas maneiras de estabelecer *links* RDF entre diferentes conjuntos de dados: manualmente ou automaticamente. A configuração automática de *links* RDF é recomendada apenas para os conjuntos de dados maiores e é realizada por meio da execução de algoritmos de ligação automatizados. A configuração manual de *links* RDF é indicada para conjuntos de dados de pequeno porte. Pode ser feita por meio de qualquer *software* editor de texto. Esta foi a abordagem escolhida para este estudo, por se tratar de um conjunto de dados menor.

Segundo Bizer, Cyganiak e Heath (2007), antes de estabelecer *links* RDF manualmente é preciso conhecer um pouco sobre os conjuntos de dados aos quais se pretende interligar os recursos. Para obter uma visão geral dos diferentes conjuntos de dados disponíveis que podem ser usados para interligar recursos com o conjunto de dados, o LOD é uma boa referência, citado no item 2.3. Seguindo as orientações desses autores, com base na disponibilidade e nas características, foram eleitos seis conjuntos de dados do LOD diferentes para interligar os dados RDF dos acervos: DBpedia, Freebase, GeoNames, *Linked Geo Data*, *Linked MDB* e Yago.

O DBpedia é o conjunto de dados que reúne as informações estruturadas da enciclopédia comunitária Wikipedia no formato da *Web Semântica*. O Freebase é um banco de dados aberto, com informações diversas, coordenado pelo Google. O GeoNames e o

*Linked Geo Data* são conjuntos de dados que contêm informações semânticas geoespaciais de vários lugares do mundo. O *Linked MDB*, também conhecido por *Linked Movie DataBase*, é um projeto que busca publicar o primeiro conjunto de dados abertos sobre filmes na *Web Semântica*. O Yago reúne informações gerais derivadas de outros conjuntos de dados.

Seguindo as orientações de Bizer, Cyganiak e Heath (2007), depois de definidos os conjuntos de dados que serão interligados com os dados dos acervos, foram feitas buscas nesses conjuntos de dados por recursos que se relacionam semanticamente com algum item de algum acervo da RMECC. No total, foram estabelecidos 1.095 *links* RDF para recursos externos do DBpedia, Freebase, GeoNames, *Linked Geo Data*, *Linked MDB* e Yago, conforme especificado no Quadro 9.

Quadro 9 - Totais de *links* RDF para recursos externos

<b>Conjunto de dados externo</b>	<b>Quantidade de <i>links</i> externos</b>
DBpedia	368
Freebase	290
GeoNames	133
<i>Linked Geo Data</i>	90
<i>Linked MDB</i>	38
YAGO	176
<b>Total</b>	<b>1095</b>

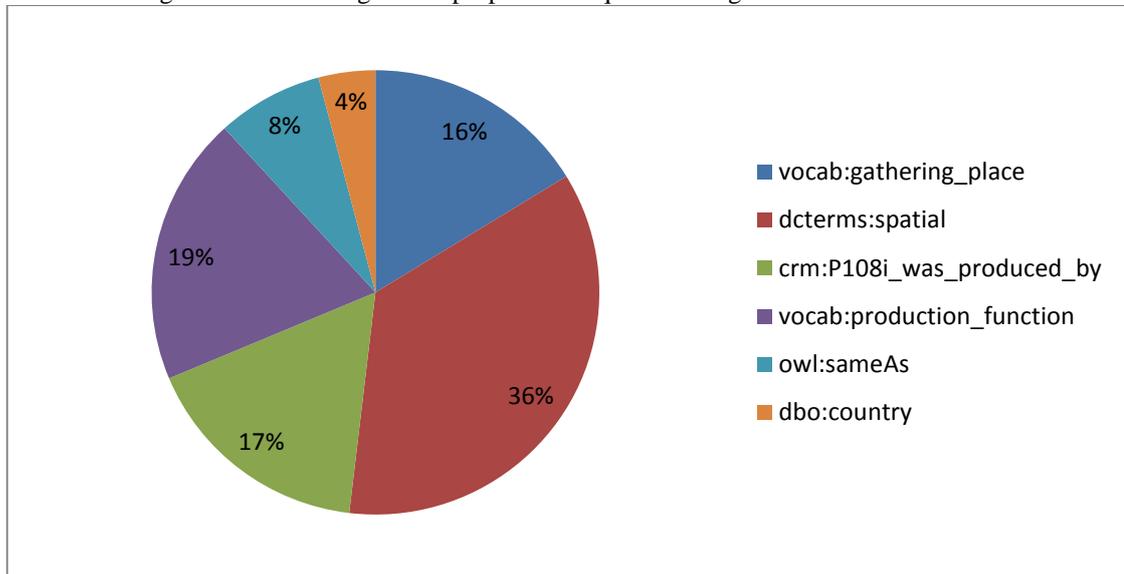
Fonte: Elaborado pelo autor, 2016.

Para estabelecer *links* RDF para recursos de outros conjuntos de dados, a propriedade, ou recurso, em questão precisa ser algo representado na *Web*. Algumas propriedades em particular do conjunto de dados desenvolvido neste estudo não são passíveis de ligação. As propriedades que envolvem identificação e nomeação, como chaves primárias ou números de identificação, são restritas para os dados dos acervos. As propriedades a respeito de datas, apesar de serem importantes para os acervos, representam números que não foram encontrados nos conjuntos de dados externos. As informações sobre descrição física, aquisição, estado de conservação e localização dizem respeito aos recursos internos do conjunto de dados, ou seja, dos itens dos acervos da RMECC. Portanto, não existem recursos que os representam em conjuntos de dados externos.

O gráfico da Figura 21 mostra quais propriedades do conjunto de dados dos acervos foram ligadas a recursos externos e a proporção de cada uma referente aos 1.095 *links* RDF. Um detalhe importante é que foi acrescentada a propriedade “sameAs” da OWL, a qual indica que determinado recurso do conjunto de dados refere-se à mesma coisa de um conjunto

de dados externo. Esta propriedade é muito útil para a descoberta e ligação semântica entre recursos de conjuntos de dados diferentes.

Figura 21 - Porcentagem das propriedades que foram ligadas a recursos externos



Fonte: Elaborada pelo autor, 2016.

#### 4.5 Consultas e visualização dos dados integrados

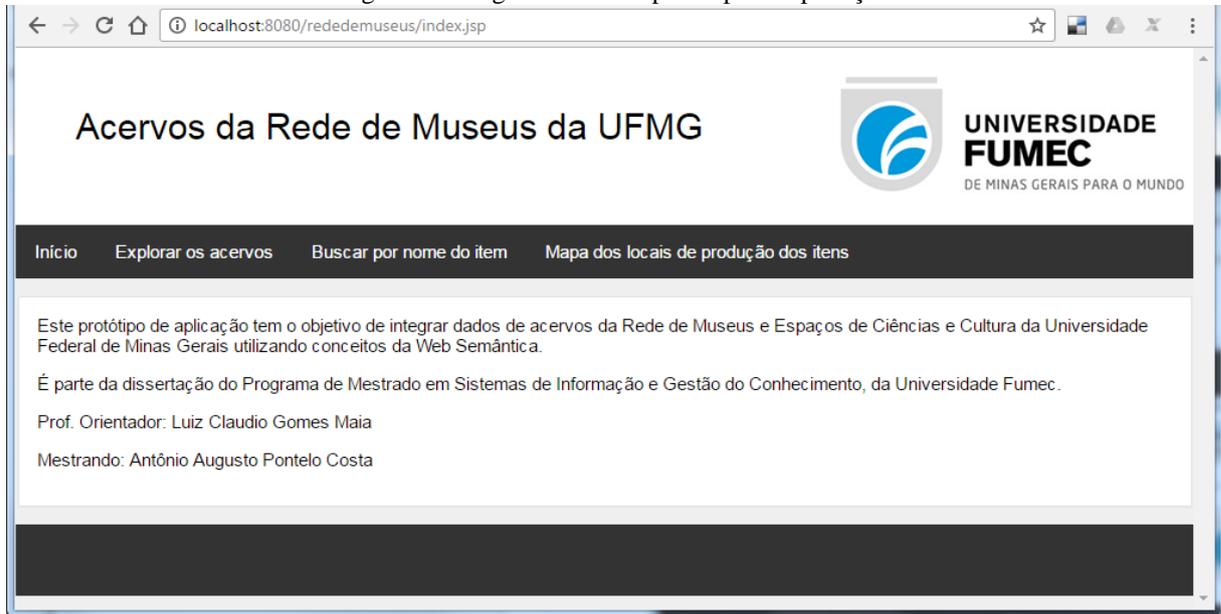
Para realizar as consultas e visualizar os dados dos acervos integrados no formato RDF, utilizou-se a ferramenta *D2RQ Engine*, componente da plataforma *D2RQ*, citado no item 4.3. Segundo Cyganiak (2012), por meio de arquivos de mapeamento relacional-RDF na linguagem *D2RQ Mapping Language*, o *D2RQ Engine* é capaz de reescrever solicitações vindas do Jena API para a linguagem de consulta SQL, executar a consulta solicitada e retornar os resultados da consulta para as camadas mais externas de apresentação.

O *D2RQ Engine* oferece bibliotecas na linguagem de programação Java. Por isso, o protótipo de aplicação proposto por este trabalho foi implementado nesta linguagem. Por se tratar de um protótipo de aplicação *Web*, foi utilizada a plataforma Java *Enterprise Edition* (Java EE). O ambiente de desenvolvimento utilizado foi o Eclipse *Integrated Development Environment* (IDE), por ser compatível com as ferramentas necessárias e por ser de conhecimento prévio deste autor.

A página inicial do protótipo de aplicação desenvolvido pode ser vista na Figura 22. O menu superior oferece três opções: explorar os acervos por museu ou espaço; buscar

itens pelo nome, digitando um termo para a pesquisa; e mapa dos locais de produção dos itens.

Figura 22 - Página inicial do protótipo de aplicação



Fonte: Elaborada pelo autor, 2016.

Para a implementação das funcionalidades do protótipo, foram desenvolvidas consultas na linguagem SPARQL, linguagem padrão para consulta de dados no formato RDF. A seguir, apresentam-se as consultas e seus respectivos resultados, obtidos no D2R Server.

A Figura 23 apresenta a consulta SPARQL que busca todos os itens que pertencem a determinada instituição — no caso, a instituição com “idinstituicao” igual a 3. A opção de explorar os acervos por museu ou espaço possibilita ao usuário ter uma visão geral sobre os itens de um determinado acervo, sem a necessidade de fornecer algum dado como parâmetro. Uma lista dos itens do acervo escolhido é apresentada para que o usuário possa clicar em um desses itens. É possível obter maiores detalhes tanto sobre um determinado item quanto sobre o museu responsável pela sua guarda. Os mesmos prefixos vistos na consulta da Figura 23 são utilizados nas demais consultas implementadas. Foram ocultadas nas consultas seguintes para evitar repetição.

Figura 23 - Consulta SPARQL que busca todos os itens que pertencem a determinada instituição

```

01 PREFIX owl: <http://www.w3.org/2002/07/owl#>
02 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
03 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
04 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
05 PREFIX dbo: <http://dbpedia.org/ontology/>
06 PREFIX dbp: <http://dbpedia.org/property/>
07 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
08 PREFIX vocab: <http://localhost:2020/resource/vocab/>
09 PREFIX dcterms: <http://purl.org/dc/terms/>
10 PREFIX map: <file://wamp/www/d2rq/rededemuseus-modif.ttl#>
11 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
12 PREFIX crm: <http://erlangen-crm.org/current/>
13 SELECT ?nomeItem ?nomeInstituicao ?idItem WHERE {
14     ?instituicao crm:P1_is_identified_by 3 .
15     ?instituicao skos:prefLabel ?nomeInstituicao .
16     ?item crm:P52_has_current_owner ?instituicao .
17     ?item crm:P102_has_title ?nomeItem .
18     ?item crm:P1_is_identified_by ?idItem .
19 }

```

Fonte: Elaborada pelo autor, 2016.

Figura 24 - Resultado da consulta SPARQL que busca todos os itens que pertencem a determinada instituição

SPARQL results:		
nomeltem	nomInstituicao	idItem
"Programa Educativo de Escolas de Enfermagem"	"Centro de Memória da Enfermagem"	69
"Pontos de Dúvida na Ética Hospitalar para Enfermeiras, Médicos, Padres e Religiosos"	"Centro de Memória da Enfermagem"	70
"O Viver Ético"	"Centro de Memória da Enfermagem"	71
"Páginas de História da Enfermagem"	"Centro de Memória da Enfermagem"	72
"A Soja nas Anemias Infantis de Origem Alimentar"	"Centro de Memória da Enfermagem"	73
"Inspectoria de Demographia e Educação Sanitária"	"Centro de Memória da Enfermagem"	74
"Penicilino-terapia da Sífilis"	"Centro de Memória da Enfermagem"	75
"The Center for the Study of the History of Nursing "	"Centro de Memória da Enfermagem"	76
"Précis d'assistance sociale- Des infirmières vositeuses et hospitalières"	"Centro de Memória da Enfermagem"	77
"Memórias da Fundação Ezequiel Dias"	"Centro de Memória da Enfermagem"	78
"Recipes for Institutions"	"Centro de Memória da Enfermagem"	79
"Enfermagem ? Leis, decretos e portarias"	"Centro de Memória da Enfermagem"	80
"A Infância e a Recreação"	"Centro de Memória da Enfermagem"	81
"Postos de Puericultura e Associações de Proteção à Maternidade e à Infância"	"Centro de Memória da Enfermagem"	82
"Assistência de Enfermos"	"Centro de Memória da Enfermagem"	83
"A Escola Pitoresca e Outros Trabalhos"	"Centro de Memória da Enfermagem"	84
"Influência da Flora sobre a Evolução Humana"	"Centro de Memória da Enfermagem"	85
"Initiation a la Méthode des testes"	"Centro de Memória da Enfermagem"	86
"Laboratory Manual of Chemistry"	"Centro de Memória da Enfermagem"	87
"Manual de Bioquímica"	"Centro de Memória da Enfermagem"	88

Fonte: Elaborada pelo autor, 2016.

A Figura 25 mostra a consulta SPARQL que busca todos os itens de todas as instituições que no seu nome contém o termo pesquisado, recebido via parâmetro. Para ilustrar, foi usado o termo "microfotografia". A opção de buscar itens pelo nome é importante para os usuários que buscam informações específicas sobre um determinado assunto ou item de acervo. Para isso, basta o usuário digitar um termo de pesquisa e uma lista dos itens de todos os museus que correspondem à pesquisa é retornada. Trata-se de uma pesquisa integrada, pois é realizada em todos os itens de acervos, de todos os museus.

Figura 25 - Consulta SPARQL que busca os itens de todas as instituições que no seu nome contém o termo pesquisado

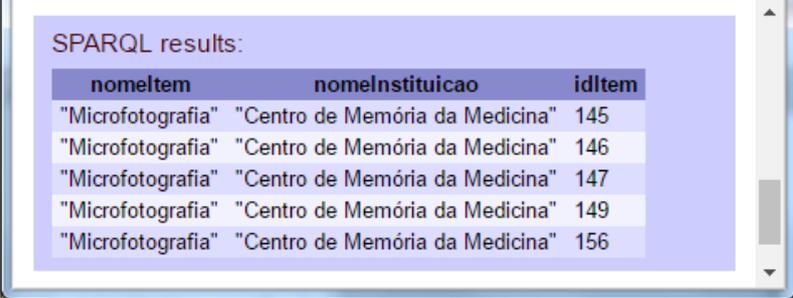
```

...
11 SELECT ?nomeItem ?nomeInstituicao ?idItem WHERE {
12     ?instituicao crm:P1_is_identified_by ?idInstituicao .
13     ?instituicao skos:prefLabel ?nomeInstituicao .
14     ?item crm:P52_has_current_owner ?instituicao .
15     ?item crm:P102_has_title ?nomeItem .
16     ?item crm:P1_is_identified_by ?idItem .
17     FILTER regex(?nomeItem, "microfotografia", "i" ) .
18 }

```

Fonte: Elaborada pelo autor, 2016.

Figura 26 - Resultado da consulta SPARQL que busca os itens de todas as instituições em cujo nome contém o termo pesquisado



nomeItem	nomeInstituicao	idItem
"Microfotografia"	"Centro de Memória da Medicina"	145
"Microfotografia"	"Centro de Memória da Medicina"	146
"Microfotografia"	"Centro de Memória da Medicina"	147
"Microfotografia"	"Centro de Memória da Medicina"	149
"Microfotografia"	"Centro de Memória da Medicina"	156

Fonte: Elaborada pelo autor, 2016.

A Figura 27 mostra a consulta que retorna todos os dados de determinado item, para exibição dos detalhes sobre ele. Neste caso, foi passado como parâmetro o “iditem” igual a 1. Por meio desta consulta, é possível obter todos os dados referentes a um item específico de um acervo, independente do formato desses dados. Assim, o usuário pode aprofundar seus conhecimentos sobre o item e sobre outros recursos internos e externos à RMECC que por ventura se relacionam semanticamente com este item de acervo, por meio de *links*.

Figura 27 - Consulta SPARQL que busca todos os dados de determinado item

```

...
13 SELECT ?idItem ?numero ?nome ?tipoNome ?autoridadeNome ?metodoAquisicao ?dataAquisicao
14     ?fonteAquisicao ?estadoConservacao ?sumarioConservacao ?dataAvaliacaoConservacao
15     ?descricaoFisica ?nomeInstituicao ?localizacao ?medidas ?localColeta ?dataColeta
16     ?coletor ?localProducao ?dataProducao ?nomeProdutor ?funcaoProducao ?assunto
17     ?colecacao WHERE {
18         ?item crm:P1_is_identified_by ?idItem .
19         ?item crm:P48_has_preferred_identifier ?numero .
20         ?item crm:P102_has_title ?nome .
21         ?item vocab:name_type ?tipoNome .
22         ?item vocab:named_by ?autoridadeNome .
23         ?item crm:P30i_custody_transferred_through ?metodoAquisicao .
24         ?item dbp:acquisitionDate ?dataAquisicao .
25         ?item crm:P29i_received_custody_through ?fonteAquisicao .
26         ?item vocab:conservation_status ?estadoConservacao .
27         ?item vocab:conservation_description ?sumarioConservacao .
28         ?item vocab:conservation_evaluation_date ?dataAvaliacaoConservacao .
29         ?item crm:P3_has_note ?descricaoFisica .
30         ?item crm:P52_has_current_owner ?instituicao .
31         ?item crm:P55_has_current_location ?localizacao .
32         ?item crm:P43_has_dimension ?medidas .
33         ?item vocab:gathering_place ?localColeta .
34         ?item vocab:gathering_date ?dataColeta .
35         ?item vocab:gatherer ?coletor .
36         ?item dcterms:spatial ?localProducao .
37         ?item dcterms:created ?dataProducao .
38         ?item crm:P108i_was_produced_by ?nomeProdutor .
39         ?item vocab:production_function ?funcaoProducao .
40         ?item dcterms:subject ?assunto .
41         ?item crm:P46i_forms_part_of ?colecacao .
42         ?item crm:P1_is_identified_by 1 .
43         ?instituicao skos:prefLabel ?nomeInstituicao .
44 }

```

Fonte: Elaborada pelo autor, 2016.

Figura 28 - Resultado da consulta SPARQL que busca todos os dados de determinado item

idItem	numero	nome	tipoNome	autoridadeNome	metodoAquisicao	dataAquisicao	fonteAquisicao	estadoConservacao
1	"	"O mundo dos senhores de escravos: dois ensaios de interpretação"	"Título"	"	"Doação"	"2009"	"Luana Diana dos Santos"	"Bom"

Fonte: Elaborada pelo autor, 2016.

A Figura 29 apresenta a consulta que retorna todas as imagens de determinado item, para exibição no protótipo de aplicação. Foi utilizado o parâmetro “iditem” igual a 140. Além de apresentar os dados sobre determinado item de acervo, são apresentadas as imagens que o representam. Elas podem ser visualizadas digitalmente no protótipo de aplicação em formato galeria e em tamanho original. Dessa forma, o usuário pode obter um contato maior com um determinado item de seu interesse, sem a necessidade de visitar o museu e conhecê-lo pessoalmente.

Figura 29 - Consulta SPARQL que busca todas as imagens de determinado item

```

...
13 SELECT ?idImagem ?nomeArquivo WHERE {
14     ?item crm:P1_is_identified_by 140 .
15     ?imagem crm:P138_represents ?item .
16     ?imagem crm:P1_is_identified_by ?idImagem .
17     ?imagem crm:P48_has_preferred_identifier ?nomeArquivo .
18 }

```

Fonte: Elaborada pelo autor, 2016.

Figura 30 - Resultado da consulta SPARQL que busca todas as imagens de determinado item

idImagem	nomeArquivo
56	"140a.jpg"
57	"140b.jpg"

Fonte: Elaborada pelo autor, 2016.

A Figura 31 apresenta a consulta que busca todos os dados de determinada instituição. Foi passado como parâmetro o “idinstituicao” igual a 9. Ao visualizar os dados e as imagens sobre um determinado item de acervo, o usuário pode se interessar em saber maiores detalhes sobre o museu responsável pela sua guarda. Por meio desta consulta, os

dados sobre a instituição são retornados e apresentados para o usuário. Um *link* para o site do museu também é disponibilizado, no qual o usuário pode clicar e visitá-lo.

Figura 31 - Consulta SPARQL que busca todos os dados de determinada instituição

```

...
13 SELECT ?idInstituicao ?nome ?site ?endereco ?pais WHERE {
14     ?instituicao crm:P1_is_identified_by ?idInstituicao .
15     ?instituicao skos:prefLabel ?nome .
16     ?instituicao foaf:homepage ?site .
17     ?instituicao dbo:address ?endereco .
18     ?instituicao dbo:country ?pais .
19     ?instituicao crm:P1_is_identified_by 9 .
20 }

```

Fonte: Elaborada pelo autor, 2016.

Figura 32 - Resultado da consulta SPARQL que busca todos os dados de determinada instituição



idInstituicao	nome	site	endereco	pais
9	"Museu de História Natural e Jardim Botânico da UFMG"	< <a href="https://www.ufmg.br/mhnb/">https://www.ufmg.br/mhnb/</a> >	"Rua Gustavo da Silveira, 1035 - Santa Inês - Belo Horizonte - MG"	"Brasil"

Fonte: Elaborada pelo autor, 2016.

Com o auxílio da ferramenta Google Maps API<sup>1</sup>, foi implementada outra funcionalidade no protótipo de aplicação, que consiste em um mapa interativo, no qual foram adicionadas marcas nos locais de produção dos itens de acervos da RMECC. É possível deslocar a área visível e aproximar ou afastar a visualização do mapa. Este mapa apresenta para o usuário uma visão geral das localidades aonde os itens dos acervos foram produzidos. Ao clicar em uma das marcas no mapa, é apresentada ao usuário uma tela de informações da cidade, do estado ou do país em questão, com *links* semânticos correspondentes à essa localidade nos conjuntos de dados externos do DBPedia e do Geonames. Ao clicar nos *links* externos, o usuário pode explorar os dados sobre o local de produção dos itens dos acervos como, por exemplo, população, coordenadas geográficas, pontos turísticos, etc. Na mesma tela de informações, há também uma lista com os nomes e respectivos *links* de acesso às informações dos itens dos acervos que foram produzidos em determinada localidade.

<sup>1</sup> <https://developers.google.com/maps/?hl=pt-br>

## 5 VALIDAÇÃO

Para validar o arcabouço conceitual proposto neste estudo, esta seção descreve os resultados alcançados e revela como as funcionalidades de integração, conversão e adaptação dos dados dos acervos para os formatos RDF e *Linked Data*, bem como as consultas e a visualização desses dados integrados, foram contempladas pelo arcabouço.

### 5.1 Integração, conversão e adaptação dos dados dos acervos para os formatos RDF e *Linked Data*

Com base no desenvolvimento e aplicação do modelo de dados que corresponde ao Diagrama de Entidade Relacionamento, representado na Figura 11, baseado nas Diretrizes Internacionais de Informação sobre Objetos de Museus: Categorias de Informação do CIDOC, foi possível integrar e armazenar os dados dos diferentes acervos da RMECC em um único repositório: um banco de dados relacional.

Os dados foram mapeados para o formato RDF, formato de dados adotado na *Web Semântica*, com o auxílio da ferramenta *generate-mapping*, da plataforma D2RQ, que analisou a estrutura do banco de dados dos acervos e gerou um arquivo de mapeamento relacional-RDF. Ontologias e termos mais conhecidos pela comunidade da *Web Semântica* foram reutilizados, substituindo alguns termos criados por padrão pelo *generate-mapping*, seguindo as orientações descritas em Bizer, Cyganiak e Heath (2007) e Hyland, Atemezing e Villazón-Terrazas (2014), e, ainda, colaborando para o reuso e compartilhamento de dados e expansão da *Web Semântica*.

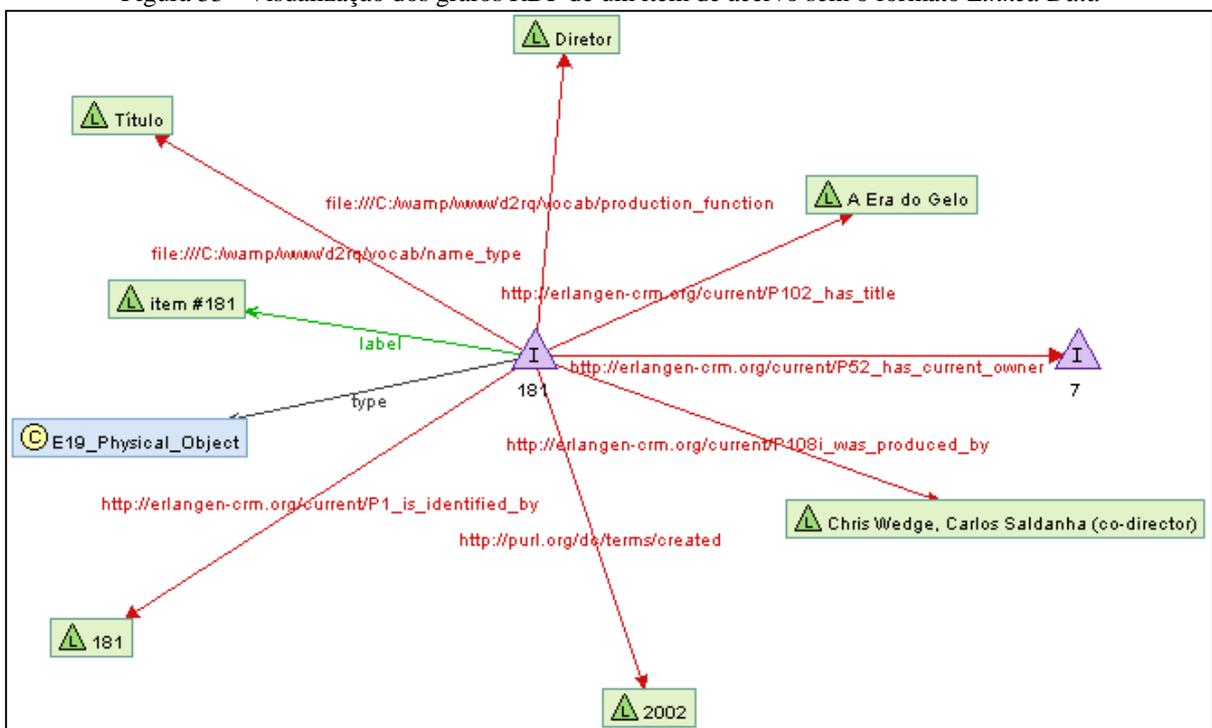
O arquivo de mapeamento personalizado serviu de parâmetro para a ferramenta *dump-rdf* da plataforma D2RQ, usada para gerar um arquivo RDF, que corresponde a todo o banco de dados dos acervos. Apesar de ser gerado no formato de dados da *Web Semântica*, o arquivo RDF ainda não estava no formato de *Linked Data* — ou seja, interligado a outros recursos de outros conjuntos de dados publicados na *Web*. De acordo com Berners-Lee (2006), Bizer, Cyganiak e Heath (2007), Heath (2016) e Skevakis *et al.* (2013), *Linked Data* não é apenas a publicação dos dados no formato RDF, mas também a interligação semântica desses dados com outros recursos externos já publicados na *Web*.

Para adaptar o arquivo RDF para o formato *Linked Data*, foi utilizada a configuração manual de *links* RDF, por meio de um *software* editor de texto que, segundo Bizer, Cyganiak e Heath (2007), é um procedimento indicado para conjunto de dados de menor volume.

A Figura 33 e a Figura 34 mostram um exemplo da diferença dos dados dos acervos em RDF antes e depois da adaptação para o *Linked Data* e da interligação a outros recursos externos. Ambas as figuras foram geradas pela ferramenta de visualização de grafos RDF *Gravity*, citada no item 4.3. Portanto, aplica-se a mesma legenda apresentada pela Figura 20.

Na Figura 33, o item do acervo, identificado por 181, e a instituição, identificada por 7, à qual pertence este item, são representados por símbolos de instâncias. Ou seja, não são recursos que possuem um URI específico que lhes permita serem interligados a outros recursos externos. A propriedade “*crm:P108i\_was\_produced\_by*”, que especifica quem produziu tal item, tem valor do tipo literal; ou seja, um texto que simplesmente descreve por quem foi produzido.

Figura 33 - Visualização dos grafos RDF de um item de acervo sem o formato *Linked Data*



Fonte: Elaborada pelo autor, 2016.

Na Figura 34, o mesmo item com título “A Era do Gelo” e a instituição à qual pertence, de nome “Espaço Memória do Cinema”, são representados pelo símbolo de recurso URI. Isso significa que foram atribuídos URIs para estas entidades. Dessa forma, eles estão



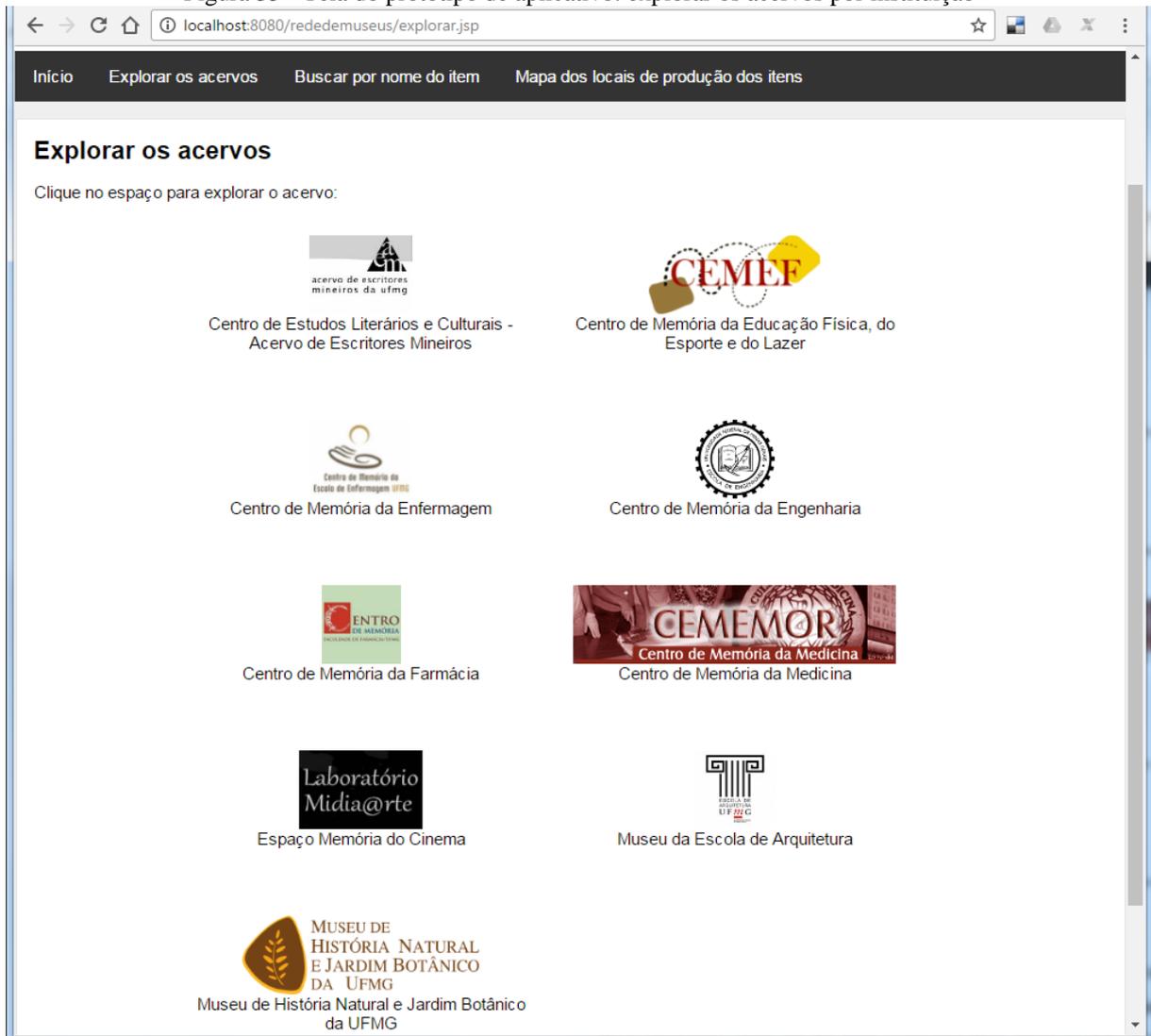
## 5.2 Consultas e visualização dos dados integrados

Para permitir que um internauta possa explorar os acervos dos espaços que fazem parte da RMECC e conhecer os itens desses acervos com maior detalhe, em uma única aplicação que permita integrar e visualizar esses dados no formato *Web*, mesmo esses dados estando no formato RDF da *Web Semântica*, foi desenvolvido um protótipo de aplicação na linguagem *Java Web* e consultas SPARQL que utiliza a ferramenta *D2RQ Engine* da plataforma D2RQ e as bibliotecas do Jena API.

O protótipo de aplicação desenvolvido permite visualizar os dados em RDF dos acervos da RMECC de forma integrada, interativa e amigável na *Web*. É possível explorar os itens que compõem determinado acervo por instituição, pesquisar por determinado termo nos nomes dos itens de todos os acervos, ver com mais detalhes os dados de um item específico, visualizar os dados sobre a instituição que o detém, visualizar as imagens que o representam e ainda interagir com um mapa que mostra os locais de produção de cada item de acervo.

A Figura 35 mostra a tela do protótipo de aplicativo na opção de explorar os acervos por instituição. Os espaços e os museus da RMECC que fazem parte deste estudo foram representados por imagens de seus respectivos logotipos. Nesta tela, é possível escolher uma instituição que se tenha interesse em conhecer seu acervo.

Figura 35 - Tela do protótipo de aplicativo: explorar os acervos por instituição



Fonte: Elaborada pelo autor, 2016.

Ao clicar em um espaço da RMECC, é apresentada uma lista dos itens deste acervo, como a da Figura 36. Esta lista mostra o nome do item e a instituição a qual pertence. Para saber mais sobre um dos itens, o usuário deve clicar no *link* correspondente.

Figura 36 - Tela do protótipo de aplicativo: lista de itens de determinada instituição

Nome do item	Instituição	ID
Pesos de latão em cepos de madeira	Centro de Memória da Farmácia	124
Pesos de latão em cepo de madeira	Centro de Memória da Farmácia	125
Pesos de precisão	Centro de Memória da Farmácia	126
Balança granatária	Centro de Memória da Farmácia	127
Capsulador	Centro de Memória da Farmácia	128
Enferizador de madeira	Centro de Memória da Farmácia	129
Cabos de Kolle em estante	Centro de Memória da Farmácia	130
Balança tríplice escala de mesa	Centro de Memória da Farmácia	131
Capsuleiro	Centro de Memória da Farmácia	132
Suporte para fabricação de capsulas amiláceas	Centro de Memória da Farmácia	133
Oculares de microscópio	Centro de Memória da Farmácia	134
Balança de Roberval	Centro de Memória da Farmácia	135
Comparador de Hellige	Centro de Memória da Farmácia	136
Esterilizador de seringas	Centro de Memória da Farmácia	137
Capsulas amiláceas	Centro de Memória da Farmácia	138
Bico de Bunsen	Centro de Memória da Farmácia	139
Balança de precisão	Centro de Memória da Farmácia	140
Nefelômetro	Centro de Memória da Farmácia	141
Micropipeta de volume variável	Centro de Memória da Farmácia	142
Balança pesa cartas	Centro de Memória da Farmácia	143

Fonte: Elaborada pelo autor, 2016.

Clicando em um dos itens da lista, é possível visualizar mais detalhes sobre o item e uma pré-visualização das imagens correspondentes, como na Figura 37. Os campos são listados no canto esquerdo da tela, enquanto seus valores são listados na parte central da tela. As imagens são reproduzidas no canto direito da tela, em formato de galeria de imagens. Para saber mais sobre a instituição que guarda o item, o usuário pode clicar no *link* dessa instituição. O usuário também pode clicar em uma das imagens para visualizá-la em tamanho original.

Figura 37 - Tela do protótipo de aplicativo: detalhes sobre um item do acervo

localhost:8080/rededemuseus/buscaitem?idItem=140

Início Explorar os acervos Buscar por nome do item Mapa dos locais de produção dos itens

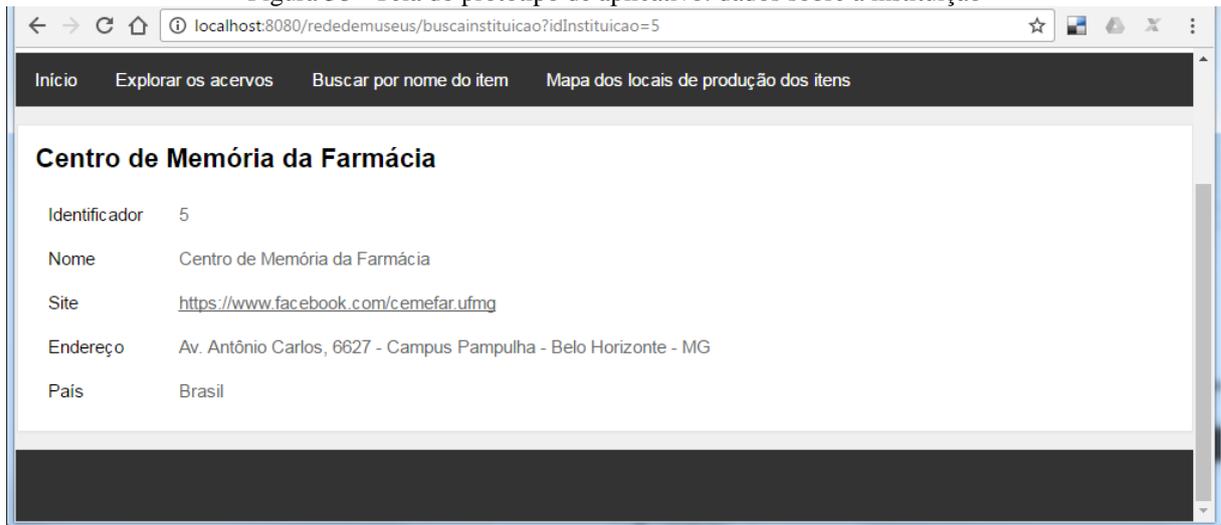
### Balança de precisão

Identificador	140
Número	CMF-000.083
Nome	Balança de precisão
Tipo de nome	Nome comum
Sumário do estado de conservação	Sujidades, madeira danificada, presença de ferrugem.
Descrição física	A balança de precisão possui um travessão feito de uma liga metálica leve e rígida, apoiado em um pivô, que por sua vez é sustentado por uma chapa fixada no topo central da base. Os pratos são pendurados em ganchos igualmente apoiados em pivôs por meio de chapas. No ponto central do travessão, uma agulha se desloca ao longo de uma escala, indicando os movimentos do conjunto. O instrumento é mantido em um envoltório de vidro que o protege contra poeira, corrosão ou acidente, e impede que as correntes de ar provoquem oscilações.
Instituição	<a href="#">Centro de Memória da Farmácia</a>
Localização	Em exposição (A.5/P.6)
Medidas	Alt.: 41cm Comp.: 36cm Prof.: 20,5cm
Data de produção	1985

Fonte: Elaborada pelo autor, 2016.

Clicando no nome da instituição, os dados sobre ela são apresentados como na Figura 38. Assim como na tela de detalhes sobre o item, nesta tela os campos são listados no canto esquerdo e seus dados na parte central. O usuário pode acessar o site do museu ou do espaço da RMECC por meio de *link*.

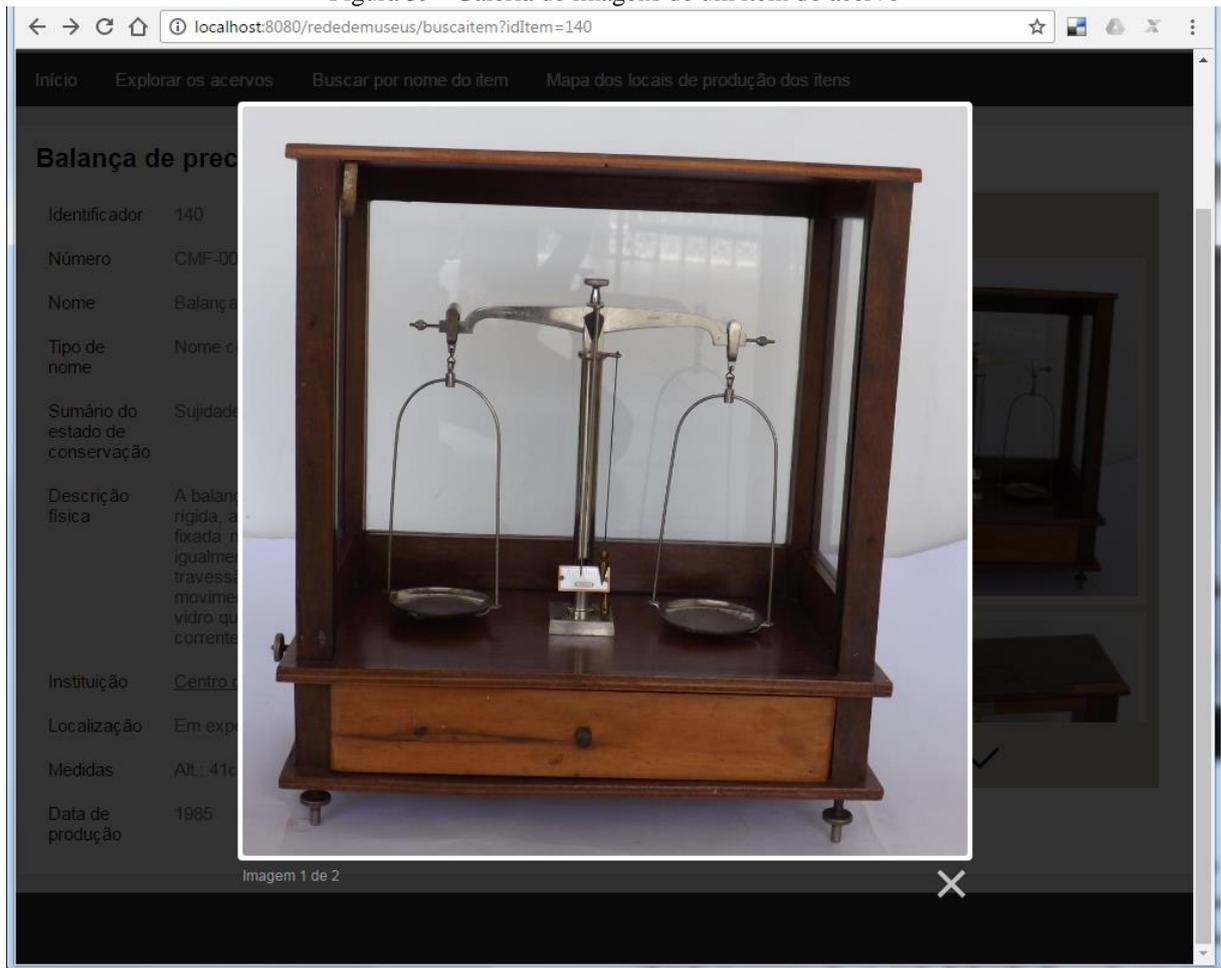
Figura 38 - Tela do protótipo de aplicativo: dados sobre a instituição



Fonte: Elaborada pelo autor, 2016.

Ao clicar em uma das imagens do item do acervo na tela de detalhes do item, ela é expandida em forma de galeria de imagens, sendo possível alterná-las, como na Figura 39. Nesta tela, o usuário pode visualizar as imagens que representam determinado item, em tamanho original.

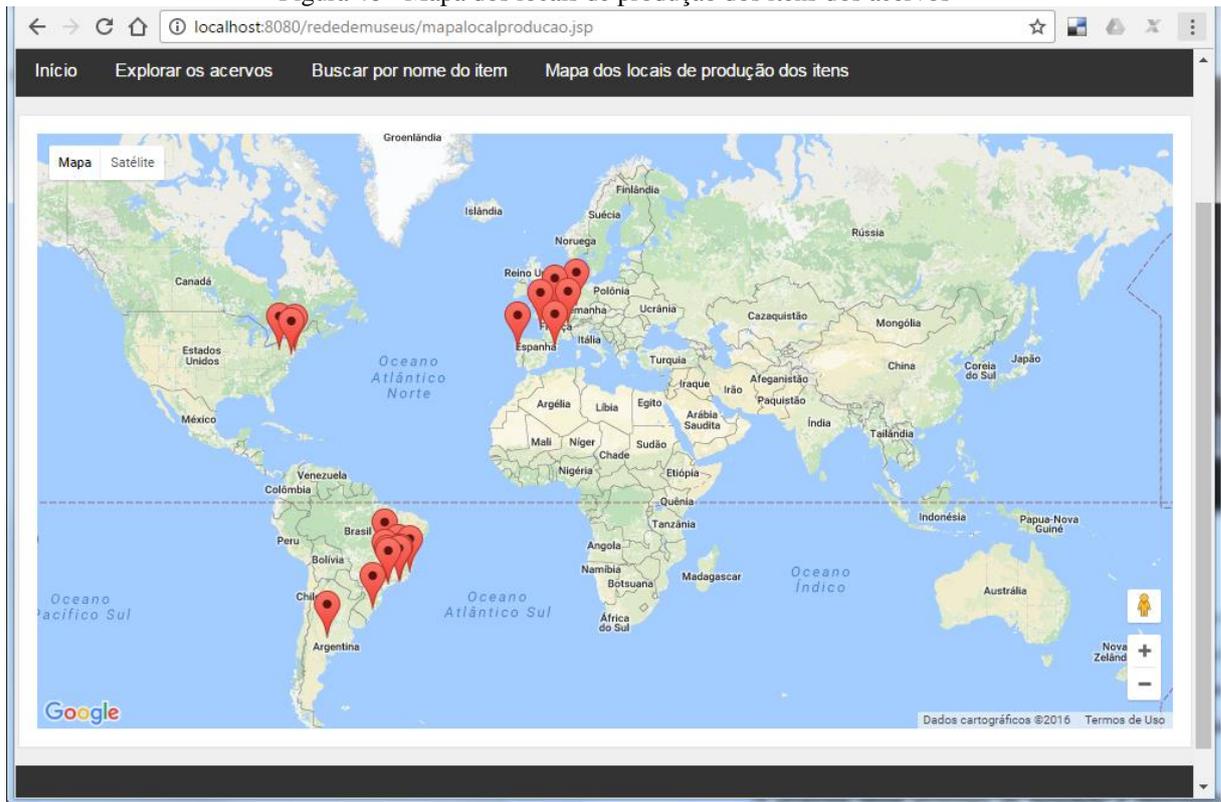
Figura 39 - Galeria de imagens de um item do acervo



Fonte: Elaborada pelo autor, 2016.

A Figura 40 apresenta a tela do mapa interativo, com as marcas nos locais de produção dos itens de acervos da RMECC. Nesta figura, a visualização do mapa é mais ampla, com todas as marcas dos locais de produção. A visualização do mapa pode ser mais afastada ou aproximada, ou ainda deslocada, de acordo com a interação do usuário com o mapa.

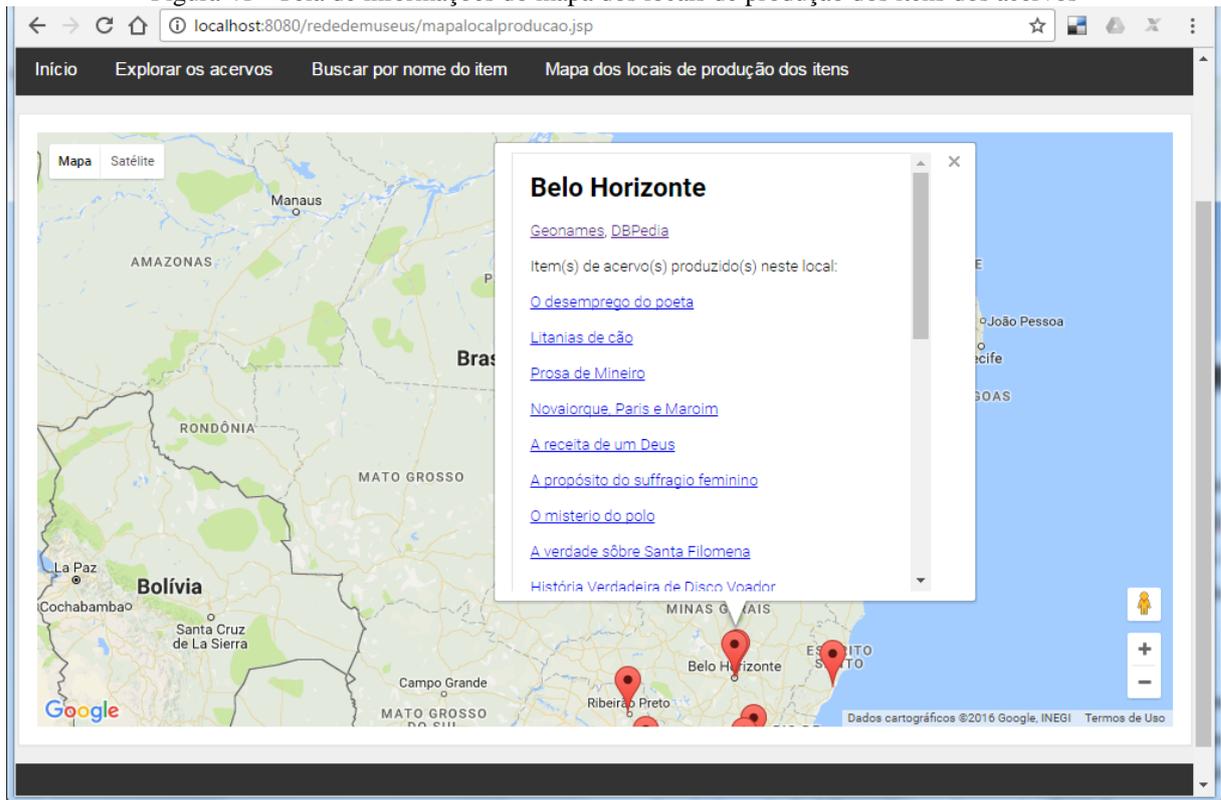
Figura 40 - Mapa dos locais de produção dos itens dos acervos



Fonte: Elaborada pelo autor, 2016.

Ao clicar em uma das marcas, é apresentada uma tela de informações sobre a localidade correspondente, com *links* para recursos do DBPedia e do Geonames, como representado na Figura 41. Nesta mesma tela de informações, é apresentada também uma lista de *links* para os itens que foram produzidos no local escolhido, que remetem o usuário para as páginas sobre os itens, semelhantes à Figura 37.

Figura 41 - Tela de informações do mapa dos locais de produção dos itens dos acervos



Fonte: Elaborada pelo autor, 2016.

A proposta desse protótipo de aplicação é demonstrar que, seguindo os passos descritos na implementação deste trabalho, é possível construir uma solução capaz de integrar dados de diferentes acervos de museus e disponibilizá-los na *Web* usando conceitos da *Web Semântica* e *Linked Data*. Apesar de sua simplicidade, este protótipo de aplicação comprova que é possível construir aplicações mais elaboradas e consistentes utilizando todo o arcabouço proposto por este trabalho.

## 6 CONSIDERAÇÕES FINAIS

Com base nos resultados obtidos por meio deste experimento, conclui-se que é possível integrar dados de acervos de diferentes museus, independentemente do formato de dados adotado, e disponibilizá-los na *Web* aplicando conceitos de *Web Semântica* e *Linked Data*. O arcabouço conceitual cumpriu os objetivos propostos neste trabalho.

No item **1.3.2**, foram listados seis objetivos específicos. O primeiro objetivo, sobre identificação e coleta dos dados de acervos de museus da RMECC que fizeram parte da pesquisa, foi alcançado no item **4.1**, no qual foram selecionados dados de acervos de nove espaços diferentes, conforme o Quadro 2.

O segundo objetivo, referente ao agrupamento e à organização dos dados utilizando um modelo de dados e um repositório comum para todos os acervos, foi consolidado no item **4.2**, no qual se apresentou um modelo de dados baseado nas Diretrizes Internacionais de Informação sobre Objetos de Museus: Categorias de Informação do CIDOC, que corresponde ao diagrama de Entidade Relacionamento, representado na Figura 11. Este modelo foi implementado em um banco de dados no SGBD MySQL, no qual foram importados os dados dos acervos selecionados.

O terceiro objetivo, sobre a conversão dos dados dos acervos para o formato RDF, foi alcançado no item **4.3**, no qual foi apresentada a plataforma D2RQ, ferramenta utilizada para gerar o arquivo de mapeamento relacional-RDF. Este arquivo de mapeamento foi posteriormente customizado, reutilizando-se ontologias e termos mais conhecidos na comunidade da *Web Semântica*, conforme recomendações e boas práticas da W3C. Novamente, foi usada a plataforma D2RQ para gerar o arquivo RDF, correspondente a todo o banco de dados dos acervos.

O quarto objetivo proposto, quanto à adaptação dos dados em RDF para o formato *Linked Data* e à interligação deles com outros recursos externos da *Web Semântica*, foi alcançado no item **4.4**, no qual se demonstrou que foram criados 1.095 *links* RDF para recursos externos de seis conjuntos de dados diferentes, publicados na *Web* e no projeto LOD.

Foi proposto um arcabouço conceitual, representado no item **3.5**, para integrar os dados dos acervos da RMECC utilizando conceitos da *Web Semântica* e *Linked Data*, atendendo ao quinto objetivo deste trabalho.

O sexto objetivo específico, sobre o teste do arcabouço conceitual proposto por meio de implementação de um protótipo de aplicação que permite visualizar os dados dos

acervos de forma integrada, interativa e amigável, foi cumprido no item **4.5**, que descreve o protótipo de aplicação criado para tal fim. Foram apresentadas imagens da tela do protótipo e das consultas SPARQL e seus resultados, capazes de trabalhar os dados em RDF dos acervos. O item **5.2** valida o protótipo de aplicação, mostrando algumas telas mais específicas e suas funcionalidades.

Tendo em vista o tema abordado e o envolvimento de diferentes tipos de acervos, este estudo pode contribuir para outros museus que tenham interesse de integrar e publicar os dados de seus acervos nos formatos da *Web Semântica* e do *Linked Data*.

Com alguns ajustes, conforme a necessidade, o arcabouço conceitual pode ser útil a outros tipos de abordagens, dados ou instituições que queiram adaptar e integrar seus dados utilizando recursos da *Web Semântica* e *Linked Data* e disponibilizá-los na *Web*. As ferramentas utilizadas foram satisfatórias e podem ser aplicadas em outros estudos similares.

A estrutura de dados implementada neste estudo, especificada no item **4.2**, que teve como referência as Diretrizes Internacionais de Informação sobre Objetos de Museus: Categorias de Informação do CIDOC, mostrou ser uma solução considerável para bancos de dados de diferentes acervos, inclusive para os espaços da RMECC. Mesmo que o objetivo de um museu não seja disponibilizar os dados de seu acervo no formato RDF, este modelo de dados pode ser aplicado em seus sistemas de inventário e catalogação de acervos. Para isso, deve-se personalizar a estrutura de dados, adequando-a de acordo com a demanda da instituição.

Quanto maior a quantidade de informações sobre os itens de um acervo, maior o número de interligações RDF possíveis entre esse conjunto de dados e outros disponíveis na *Web Semântica*. Consequentemente, maior seu valor e expressividade na coleção de dados do *Linked Data*. Portanto, a completude dos dados dos acervos mostrou ser um ponto crucial para este tipo de conjunto de dados.

Como sugestões para a realização de trabalhos futuros, indicam-se:

- Detalhamento do modelo de dados proposto: o modelo de dados aplicado no banco de dados dos acervos pode ser mais detalhado, para tornar-se abrangente e atender melhor às necessidades de diferentes museus e outras instituições.
- Melhorias no protótipo de aplicação: o protótipo de aplicação implementado serviu para validar este estudo, mas aplicando-se melhorias em suas funcionalidades e aparência visual é possível torná-lo uma aplicação mais evoluída e funcional, podendo ser adotada em qualquer instituição.

- Aplicação do arcabouço conceitual em outros tipos de dados: o arcabouço conceitual proposto neste estudo pode ser adotado em outros trabalhos que envolvam outros tipos de dados e instituições. Algumas características são particulares, como o modelo de dados e as ontologias e termos reutilizados no RDF, mas com algumas alterações, a aplicação do arcabouço conceitual é viável. Também é viável a utilização das mesmas ferramentas adotadas nos procedimentos seguidos neste experimento em outros estudos e conjuntos de dados.

A riqueza dos relacionamentos semânticos entre os itens de acervos de museus e entre estes e outros recursos disponíveis na *Web* é capaz de contribuir para a expansão e valorização da *Web Semântica* e do *Linked Data*. Para tanto, os conjuntos de dados desses acervos devem ser adequados para o formato de dados RDF, reutilizando ontologias e termos difundidos na comunidade da *Web Semântica*, e interligados a outros recursos e conjuntos de dados expressivos. Assim, os dados dos acervos podem ser mais bem explorados e divulgados para a sociedade, tornando-se uma navegação na *Web* mais interessante e enriquecedora.

## REFERÊNCIAS

ABARCA, Alonso Cavieres; MENA, Sergio Fredes; NOVOA, Arturo Ramírez. Tesauros y Web Semántica: Diseño metodológico para estructurar contenidos Web mediante SKOS-Core. **Serie Bibliotecología y Gestión de Información**, n. 57, p. 1-64, maio 2010. Disponível em:

<[http://eprints.rclis.org/14420/1/Serie\\_N%C2%BA\\_57\\_Mayo\\_2010\\_Tesauros\\_y\\_Web\\_Sem%C3%A1ntica.pdf](http://eprints.rclis.org/14420/1/Serie_N%C2%BA_57_Mayo_2010_Tesauros_y_Web_Sem%C3%A1ntica.pdf)>. Acesso em: 06 abr. 2016.

ALVES, Rachel Cristina Vesú. **Web Semântica: uma análise focada no uso de metadados**. 2005. 180 f. Dissertação (Mestrado em Ciência da Informação da Faculdade de Filosofia e Ciências) – Universidade Estadual Paulista – UNESP, Marília, 2005.

AZEVEDO, Patrícia Carolina Neves. **Uma proposta para visualização de linked data sobre enchentes na Bacia do Rio Doce**. 2014. 104 f. Dissertação (Mestrado Profissional em Sistemas de Informação e Gestão do Conhecimento da Faculdade de Ciências Empresariais – FACE) - Universidade FUMEC, Belo Horizonte, 2014.

BAKER, Thomas *et al.* **Library Linked Data Incubator Group Final Report**. W3C Incubator Group Report, 2011. Disponível em:

<<https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>. Acesso em: 23 out. 2016.

BAKER, Thomas; NOY, Natasha; SWICK, Ralph. **Semantic Web Case Studies and Use Cases**, 2012. Disponível em: <<https://www.w3.org/2001/sw/sweo/public/UseCases/>>. Acesso em: 06 abr. 2016.

BECKETT, David *et al.* **Terse RDF Triple Language**. W3C Recommendation, 2014. Disponível em: <<https://www.w3.org/TR/turtle/>>. Acesso em: 01 ago. 2016.

BERMES, Emmanuelle. Convergence and Interoperability: a Linked Data perspective. In: IFLA WORLD LIBRARY AND INFORMATION CONGRESS, 77., 2011, Porto Rico. **Anais...** Porto Rico, 2011, p. 1-12.

BERNERS-LEE, Tim. **Linked Data**. Design Issues, 2006. Disponível em:

<<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 09 abr. 2016.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web. **Scientific American**, v. 284, n. 5, maio 2001. Disponível em:

<[http://iscl2918929391.googlecode.com/svn-history/r347/trunk/RPC/Slides/p01\\_theSemanticWeb.pdf](http://iscl2918929391.googlecode.com/svn-history/r347/trunk/RPC/Slides/p01_theSemanticWeb.pdf)>. Acesso em: 07 abr. 2016.

BIZER, Chris; CYGANIAK, Richard; HEATH, Tom. **How to Publish Linked Data on the Web**, 2007. Disponível em <<http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/>>. Acesso em: 01 ago. 2016.

BRASIL. **Lei nº 11.904**, de 14 de janeiro de 2009. Institui o Estatuto de Museus e dá outras providências. Brasília. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2007-2010/2009/Lei/L11904.htm](http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2009/Lei/L11904.htm)>. Acesso em: 20 mar. 2016.

CARRASCO, Laís Barbudo; THALLER, Manfred; VIDOTTI, Silvana Aparecida Borsetti Gregório. Ontologia Cidoc CRM no contexto dos ambientes digitais de patrimônios culturais. **Liinc em Revista**, Rio de Janeiro, v. 11, n. 1, p. 208-222, maio 2015. Disponível em: <<http://liinc.revista.ibict.br/index.php/liinc/article/view/789>>. Acesso em: 01 ago. 2016.

CASARIN, Helen de Castro Silva; CASARIN, Samuel José. **Pesquisa científica: da teoria à prática**. 1. ed. Curitiba: InterSaberes, 2012. 200 p.

CATARINO, Maria Elisabete. *Simple Knowledge Organization System: construindo sistemas de organização do conhecimento no contexto da Web Semântica*. **Informação & Tecnologia (ITEC)**, João Pessoa, v. 1, n. 1, p. 17-28, jan./jun. 2014. Disponível em: <<http://www.okara.ufpb.br/ojs2/index.php/itec/article/view/19307>>. Acesso em: 22 jun. 2015.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino; SILVA, Roberto da. **Metodologia Científica**. 6. ed. São Paulo: Pearson Prentice Hall, 2007. 162 p.

CIDOC - ICOM. **Declaração de Princípios de Documentação em Museus e Diretrizes Internacionais de Informação sobre Objetos de Museus**: Categorias de Informação do Comitê Internacional de Documentação (CIDOC-ICOM). São Paulo: Secretaria de Estado de Cultura de São Paulo; Associação de Amigos do Museu do Café; Pinacoteca do Estado de São Paulo, 2014. 76 p.

CIDOC – ICOM. **ICOM International Committee For Documentation**. 2010. Disponível em: <<http://network.icom.museum/cidoc/>>. Acesso em: 28 jul. 2016.

CYGANIAK, Richard. **D2RQ: Accessing Relational Databases as Virtual RDF Graphs**, 2012. Disponível em: <<http://d2rq.org/>>. Acesso em: 01 ago. 2016.

CYGANIAK, Richard *et al.* **The D2RQ Mapping Language**, 2012. Disponível em: <<http://d2rq.org/d2rq-language>>. Acesso em: 01 ago. 2016.

DANDAGI, S. Vidya; SIDNAL, Nandini. Semantic Web: creating and querying. **International Journal of Web & Semantic Technology (IJWesT)**, v. 7, n. 1, p. 17-27, jan. 2016. Disponível em: <<http://aircconline.com/ijwest/V7N1/7116ijwest02.pdf>>. Acesso em: 01 ago. 2016.

DE BOER, Victor *et al.* Amsterdam Museum Linked Open Data. **Semantic Web**, v. 4, n. 3, p. 237-243, 2013. Disponível em: <[http://www.semantic-web-journal.net/sites/default/files/swj293\\_2.pdf](http://www.semantic-web-journal.net/sites/default/files/swj293_2.pdf)>. Acesso em: 06 abr. 2016.

DOERR, Martin *et al.* The europeana data model (edm). In: IFLA WORLD LIBRARY AND INFORMATION CONGRESS, 76., 2010, Gothenburg. **Anais...** Gothenburg, 2010, p. 1-12.

FERNEDA, Edberto. **Recuperação de Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação**. 2003. 137 f. Teste (Doutorado em Ciências da Comunicação da Escola de Comunicação e Artes) - Universidade de São Paulo, São Paulo, 2003.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 175 p.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008. 200 p.

HEATH, Tom. **Linked Data**. Linked Data Community, 2016. Disponível em: <<http://linkeddata.org/>>. Acesso em: 02 abr. 2016.

HEATH, Tom; BIZER, Christian. Linked data: Evolving the web into a global data space. **Synthesis lectures on the semantic web: theory and technology**, v. 1, n. 1, p. 1-136, fev. 2011. Disponível em: <<http://linkeddatatbook.com/editions/1.0/>>. Acesso em: 07 abr. 2016.

HERNÁNDEZ, Francisca. **An Ontology of Cantabria's Cultural Heritage**. Semantic Web Use Cases and Case Studies, 2007. Disponível em: <<https://www.w3.org/2001/sw/sweo/public/UseCases/FoundationBotin/>>. Acesso em: 06 abr. 2016.

HOLANDA, Cíntia; BRAZ, Márcia Ivo. Indexação automática de conteúdos na web: análise de sites de museus. **Biblionline**, João Pessoa, v. 8, n. 1, p. 42-59, jan./jun. 2012. Disponível em: <<http://www.okara.ufpb.br/ojs2/index.php/biblio/article/view/11766/7509>>. Acesso em: 11 maio 2015.

HYLAND, Bernadette; ATEMEZING, Ghislain; VILLAZÓN-TERRAZAS, Boris. **Best Practices for Publishing Linked Data**. W3C Working Group Note, 2014. Disponível em: <<https://www.w3.org/TR/ld-bp/>>. Acesso em: 01 ago. 2016.

HYVÖNEN, Eero *et al.* MuseumFinland: Finnish museums on the semantic web. **Web Semantics: Science, Services and Agents on the World Wide Web**, v. 3, n. 2, maio 2005. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S157082680500017X>>. Acesso em: 18 set. 2015.

ICOM. **The CIDOC Conceptual Reference Model**, 2016. Disponível em: <<http://www.cidoc-crm.org/>>. Acesso em: 01 ago. 2016.

ISAAC, Antoine. **Enriching and sharing cultural heritage data in Europeana**. Semantic Web Use Cases and Case Studies, 2012. Disponível em: <<https://www.w3.org/2001/sw/sweo/public/UseCases/Europeana/>>. Acesso em: 04 abr. 2016.

ISAAC, Antoine. **Europeana Data Model Primer**. Europeana Data Model Documentation, 2013. Disponível em: <[http://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf)>. Acesso em: 20 out. 2016.

ISAAC, Antoine; SUMMERS, Ed. **SKOS Simple Knowledge Organization System Primer**. W3C Working Group Note, 2009. Disponível em: <<http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>>. Acesso em: 16 jul. 2015.

MARCHI, Ana Carolina Bertoletti de; COSTA, Antônio Carlos da Rocha. Uma proposta de padrão de metadados para objetos de aprendizagem de museus de ciências e tecnologia. **Renote**, Porto Alegre, v. 2, n. 1, mar. 2004. Disponível em: <<http://seer.ufrgs.br/index.php/renote/article/view/13660/7946>>. Acesso em: 18 maio 2015.

MARCONDES, Carlos Henrique. “*Linked data*” – dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na *web*. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 17, n. 34, p. 171-192, maio/ago. 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/24410>>. Acesso em: 16 jun. 2015.

MASCARENHAS, Sidnei Augusto. **Metodologia científica**. 1. ed. São Paulo: Pearson Education do Brasil, 2012. 125 p.

MCGUINNESS, Deborah L.; HARMELEN, Frank van. **OWL Web Ontology Language Overview**. W3C Recommendation, 2004. Disponível em: <<https://www.w3.org/TR/owl-features/>>. Acesso em: 01 ago. 2016.

MILLER, Paul. Interoperability: What Is It and Why Should I Want It? **Ariadne**, n. 24, 2000. Disponível em: <<http://www.ariadne.ac.uk/issue24/interoperability?ref=SevSevil.Com>>. Acesso em: 07 abr. 2016.

MORI, Alexandre; CARVALHO, Cedric Luiz de. **Metadados no Contexto da Web Semântica**. Goiânia: Universidade Federal de Goiás, 2004. 18 p. Relatório Técnico RT-INF\_002-04. Disponível em: <[http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_002-04.pdf](http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_002-04.pdf)>. Acesso em: 13 jul. 2015.

OLIVEIRA, Alysson Vicuña de. Introdução a Web Semântica, Ontologia e Máquinas de Busca. **Revista Tecnologias em Projeção**, v. 2, n.1, p. 7-10, jun. 2011. Disponível em: <<http://revista.faculdadeprojecao.edu.br/index.php/Projecao4/article/view/74/62>>. Acesso em: 01 ago. 2016.

PASTOR-SANCHEZ, Juan-Antonio; MENDEZ, Francisco Javier Martínez; RODRÍGUEZ-MUÑOZ, José Vicente. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. **Information Research**, Lund, v. 14, n. 4, dez. 2009. Disponível em: <<http://InformationR.net/ir/14-4/paper422.html>>. Acesso em: 04 maio 2015.

PATRÍCIO, Helena Simões. A Europeana 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: Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas, 11., 2012, Lisboa. **Anais...** Lisboa, 2012.

PAYETTE, Sandra *et al.* Interoperability for Digital Objects and Repositories. **D-Lib Magazine**, v. 5, n. 5, maio 1999. Disponível em: <<http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/may99/payette/05payette.html>>. Acesso em: 06 abr. 2016.

PINTO, Vitor Afonso. **Supporting Competitive Intelligence With Linked Enterprise Data**. 2014. 102 f. Dissertação (Mestrado Profissional em Sistemas de Informação e Gestão do Conhecimento da Faculdade de Ciências Empresariais – FACE) - Universidade FUMEC, Belo Horizonte, 2014.

POLLOCK, Jeffrey T. **Semantic Web for Dummies**. 1. ed. Hoboken: John Wiley & Sons, 2009. 384 p.

PRUD'HOMMEAUX, Eric; SEABORNE, Andy. **SPARQL Query Language for RDF**. W3C Recommendation, 2008. Disponível em: <<https://www.w3.org/TR/rdf-sparql-query/>>. Acesso em: 06 abr. 2016.

RODRÍGUEZ, Marta González. **Online Resource for Information on Aquatic Sciences**. Semantic Web Use Cases and Case Studies, 2009. Disponível em: <<https://www.w3.org/2001/sw/sweo/public/UseCases/Aquaring/>>. Acesso em: 06 abr. 2016.

SANTOS NETO, Antonio Laurindo dos *et al.* Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. **TransInformação**, Campinas, v. 25, n. 1, p. 81-87, jan./abr. 2013. Disponível em: <<http://www.scielo.br/pdf/tinf/v25n1/a08v25n1.pdf>>. Acesso em: 02 abr. 2016.

SAYÃO, Luís Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **TransInformação**, Campinas, v. 20, n. 2, p. 133-148, maio/ago. 2008. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/530/510>>. Acesso em: 02 abr. 2016.

SCHMACHTENBERG, Max et al. **The Linking Open Data cloud diagram**. 2014. Disponível em: <<http://lod-cloud.net/>>. Acesso em: 06 abr. 2016.

SCHREIBER, Guus; RAIMOND, Yves. **RDF 1.1 Primer**. W3C Working Group Note, 2014. Disponível em: <<http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>>. Acesso em: 16 jul. 2015.

SILVA, Márcio Martins da. **Desenvolvimento de um observatório de empresas de software no Brasil com recursos da Web Semântica**. 2014. 188 f. Dissertação (Mestrado Profissional em Sistemas de Informação e Gestão do Conhecimento da Faculdade de Ciências Empresariais – FACE) - Universidade FUMEC, Belo Horizonte, 2014.

SILVA, Renata Eleutério da; SANTOS, Plácida Leopoldina Ventura Amorim da Costa; FERNEDA, Edberto. Modelos de recuperação de informação e web semântica: a questão da relevância. **Informação & Informação**, Londrina, v. 18, n. 3, p. 27-44, set./dez. 2013. Disponível em: <<http://repositorio.unesp.br/handle/11449/114705>>. Acesso em: 20 out. 2016.

SKEVAKIS, Giannis *et al.* Elevating Natural History Museums' Cultural Collections to the Linked Data Cloud. **SDA**, v. 13, p. 40-51, 2013. Disponível em: <<http://ceur-ws.org/Vol-1091/paper4.pdf>>. Acesso em: 10 abr. 2016.

SOUZA, Renato Rocha; ALVARENGA, Lídia. 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. Disponível em: <<http://www.scielo.br/pdf/ci/v33n1/v33n1a16>>. Acesso em: 06 abr. 2016.

SPIVACK, Nova. **How the WebOS Evolves?**. 2007. Disponível em: <<http://www.novaspivack.com/technology/how-the-webos-evolves>>. Acesso em: 12 abr. 2016.

UNIVERSIDADE FEDERAL DE MINAS GERAIS. Câmara de Extensão. **Resolução nº 02/2010**, de 09 de julho de 2010. Institui a Rede de Museus e Espaços de Ciência e Cultura da UFMG, e aprova seu regimento interno. Disponível em: <<https://www.ufmg.br/rededemuseus/index.php/a-rede/marco-regulatorio>>. Acesso em: 06 abr. 2016.

UNIVERSIDADE FEDERAL DE MINAS GERAIS. **Rede de Museus e Espaços de Ciências e Cultura da UFMG**. 2015. Disponível em: <<https://www.ufmg.br/rededemuseus/index.php/a-rede/quem-somos>>. Acesso em: 06 abr. 2016

W3C. **Linked Data**, 2015a. Disponível em: <<http://www.w3.org/standards/semanticweb/data>>. Acesso em: 13 jul. 2015.

W3C. **Vocabularies**, 2015b. Disponível em: <<http://www.w3.org/standards/semanticweb/data>>. Acesso em: 01 ago. 2016.

W3C. **W3C Semantic Web Activity**, 2013. Disponível em: <<http://www.w3.org/2001/sw/>>. Acesso em: 13 jul. 2015.

W3C Brasil. **Web Semântica**, 2011. Disponível em: <<http://www.w3c.br/Padroes/WebSemantica>>. Acesso em: 13 jul. 2015.

## APÊNDICES

### APÊNDICE A – Arquivo de mapeamento relacional-RDF rededemuseus-ori.ttl

```

@prefix map: <#> .
@prefix db: <> .
@prefix vocab: <vocab/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix d2rq: <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#> .
@prefix jdbc: <http://d2rq.org/terms/jdbc/> .

```

```

map:database a d2rq:Database;
    d2rq:jdbcDriver "com.mysql.jdbc.Driver";
    d2rq:jdbcDSN "jdbc:mysql:///rededemuseus";
    d2rq:username "root";
    jdbc:autoReconnect "true";
    jdbc:zeroDateTimeBehavior "convertToNull";
.

```

# Table imagem

```

map:imagem a d2rq:ClassMap;
    d2rq:dataStorage map:database;
    d2rq:uriPattern "imagem/@ @imagem.idimagem@@";
    d2rq:class vocab:imagem;
    d2rq:classDefinitionLabel "imagem";
.

```

```

map:imagem__label a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:imagem;
    d2rq:property rdfs:label;
    d2rq:pattern "imagem#@ @imagem.idimagem@";
.

```

```

map:imagem_idimagem a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:imagem;
    d2rq:property vocab:imagem_idimagem;
    d2rq:propertyDefinitionLabel "imagem idimagem";
    d2rq:column "imagem.idimagem";
    d2rq:datatype xsd:integer;
.

```

```

map:imagem_nome_arquivo a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:imagem;
    d2rq:property vocab:imagem_nome_arquivo;
    d2rq:propertyDefinitionLabel "imagem nome_arquivo";
    d2rq:column "imagem.nome_arquivo";
.

```

```

map:imagem_iditem__ref a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:imagem;
    d2rq:property vocab:imagem_iditem;
    d2rq:refersToClassMap map:item;
    d2rq:join "imagem.iditem => item.iditem";
.

```

# Table instituicao

```

map:instituicao a d2rq:ClassMap;
    d2rq:dataStorage map:database;
    d2rq:uriPattern "instituicao/@ @instituicao.idinstituicao @@";
    d2rq:class vocab:instituicao;
    d2rq:classDefinitionLabel "instituicao";
.

map:instituicao__label a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property rdfs:label;
    d2rq:pattern "instituicao # @ @instituicao.idinstituicao @@";
.

map:instituicao_idinstituicao a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property vocab:instituicao_idinstituicao;
    d2rq:propertyDefinitionLabel "instituicao idinstituicao";
    d2rq:column "instituicao.idinstituicao";
    d2rq:datatype xsd:integer;
.

map:instituicao_nome a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property vocab:instituicao_nome;
    d2rq:propertyDefinitionLabel "instituicao nome";
    d2rq:column "instituicao.nome";
.

map:instituicao_site a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property vocab:instituicao_site;
    d2rq:propertyDefinitionLabel "instituicao site";
    d2rq:column "instituicao.site";
.

map:instituicao_endereco a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property vocab:instituicao_endereco;
    d2rq:propertyDefinitionLabel "instituicao endereco";
    d2rq:column "instituicao.endereco";
.

map:instituicao_pais a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property vocab:instituicao_pais;
    d2rq:propertyDefinitionLabel "instituicao pais";
    d2rq:column "instituicao.pais";
.

# Table item
map:item a d2rq:ClassMap;
    d2rq:dataStorage map:database;
    d2rq:uriPattern "item/@ @item.iditem @@";
    d2rq:class vocab:item;
    d2rq:classDefinitionLabel "item";
.

map:item__label a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:item;
    d2rq:property rdfs:label;
    d2rq:pattern "item # @ @item.iditem @@";
.

map:item_iditem a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:item;
    d2rq:property vocab:item_iditem;
    d2rq:propertyDefinitionLabel "item iditem";
    d2rq:column "item.iditem";

```

```

        d2rq:datatype xsd:integer;
    .
    map:item_numero a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_numero;
        d2rq:propertyDefinitionLabel "item numero";
        d2rq:column "item.numero";
    .
    map:item_nome a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_nome;
        d2rq:propertyDefinitionLabel "item nome";
        d2rq:column "item.nome";
    .
    map:item_tipo_nome a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_tipo_nome;
        d2rq:propertyDefinitionLabel "item tipo_nome";
        d2rq:column "item.tipo_nome";
    .
    map:item_autoridade_nome a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_autoridade_nome;
        d2rq:propertyDefinitionLabel "item autoridade_nome";
        d2rq:column "item.autoridade_nome";
    .
    map:item_metodo_aquisicao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_metodo_aquisicao;
        d2rq:propertyDefinitionLabel "item metodo_aquisicao";
        d2rq:column "item.metodo_aquisicao";
    .
    map:item_data_aquisicao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_data_aquisicao;
        d2rq:propertyDefinitionLabel "item data_aquisicao";
        d2rq:column "item.data_aquisicao";
    .
    map:item_fonte_aquisicao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_fonte_aquisicao;
        d2rq:propertyDefinitionLabel "item fonte_aquisicao";
        d2rq:column "item.fonte_aquisicao";
    .
    map:item_estado_conservacao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_estado_conservacao;
        d2rq:propertyDefinitionLabel "item estado_conservacao";
        d2rq:column "item.estado_conservacao";
    .
    map:item_sumario_conservacao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_sumario_conservacao;
        d2rq:propertyDefinitionLabel "item sumario_conservacao";
        d2rq:column "item.sumario_conservacao";
    .
    map:item_data_avaliacao_conservacao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_data_avaliacao_conservacao;
        d2rq:propertyDefinitionLabel "item data_avaliacao_conservacao";

```

```

        d2rq:column "item.data_avaliacao_conservacao";
    .
    map:item_descricao_fisica a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_descricao_fisica;
        d2rq:propertyDefinitionLabel "item descricao_fisica";
        d2rq:column "item.descricao_fisica";
    .
    map:item_localizacao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_localizacao;
        d2rq:propertyDefinitionLabel "item localizacao";
        d2rq:column "item.localizacao";
    .
    map:item_medidas a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_medidas;
        d2rq:propertyDefinitionLabel "item medidas";
        d2rq:column "item.medidas";
    .
    map:item_local_coleta a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_local_coleta;
        d2rq:propertyDefinitionLabel "item local_coleta";
        d2rq:column "item.local_coleta";
    .
    map:item_data_coleta a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_data_coleta;
        d2rq:propertyDefinitionLabel "item data_coleta";
        d2rq:column "item.data_coleta";
    .
    map:item_coletor a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_coletor;
        d2rq:propertyDefinitionLabel "item coletor";
        d2rq:column "item.coletor";
    .
    map:item_local_producao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_local_producao;
        d2rq:propertyDefinitionLabel "item local_producao";
        d2rq:column "item.local_producao";
    .
    map:item_data_producao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_data_producao;
        d2rq:propertyDefinitionLabel "item data_producao";
        d2rq:column "item.data_producao";
    .
    map:item_nome_produzor a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_nome_produzor;
        d2rq:propertyDefinitionLabel "item nome_produzor";
        d2rq:column "item.nome_produzor";
    .
    map:item_funcao_producao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_funcao_producao;
        d2rq:propertyDefinitionLabel "item funcao_producao";

```

```
        d2rq:column "item.funcao_producao";
    .
    map:item_assunto a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_assunto;
        d2rq:propertyDefinitionLabel "item assunto";
        d2rq:column "item.assunto";
    .
    map:item_colecao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_colecao;
        d2rq:propertyDefinitionLabel "item colecao";
        d2rq:column "item.colecao";
    .
    map:item_idinstituicao__ref a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:item_idinstituicao;
        d2rq:refersToClassMap map:instituicao;
        d2rq:join "item.idinstituicao => instituicao.idinstituicao";
    .
```

## APÊNDICE B – Arquivo de mapeamento relacional-RDF rededemuseus-modif.ttl

```

@prefix map: <file:/wamp/www/d2rq/rededemuseus-modif.ttl#> .
@prefix vocab: <vocab/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix d2rq: <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#> .
@prefix jdbc: <http://d2rq.org/terms/jdbc/> .
@prefix crm: <http://erlangen-crm.org/current/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix dbo: <http://dbpedia.org/ontology/>.
@prefix dbp: <http://dbpedia.org/property/>.
@prefix dcterms: <http://purl.org/dc/terms/>.

map:database a d2rq:Database;
    d2rq:jdbcDriver "com.mysql.jdbc.Driver";
    d2rq:jdbcDSN "jdbc:mysql:///rededemuseus";
    d2rq:username "root";
    jdbc:autoReconnect "true";
    jdbc:zeroDateTimeBehavior "convertToNull";
    .

# Table imagem
map:imagem a d2rq:ClassMap;
    d2rq:dataStorage map:database;
    d2rq:uriPattern "imagem/@ @imagem.idimagem@@";
    d2rq:class crm:E38_Image;
    d2rq:classDefinitionLabel "imagem";
    .

map:imagem__label a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:imagem;
    d2rq:property rdfs:label;
    d2rq:pattern "imagem # @ @imagem.idimagem@";
    .

map:imagem_idimagem a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:imagem;
    d2rq:property crm:P1_is_identified_by;
    d2rq:propertyDefinitionLabel "imagem idimagem";
    d2rq:column "imagem.idimagem";
    d2rq:datatype xsd:integer;
    .

map:imagem_nome_arquivo a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:imagem;
    d2rq:property crm:P48_has_preferred_identifier;
    d2rq:propertyDefinitionLabel "imagem nome_arquivo";
    d2rq:column "imagem.nome_arquivo";
    .

map:imagem_iditem__ref a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:imagem;
    d2rq:property crm:P138_represents;
    d2rq:refersToClassMap map:item;
    d2rq:join "imagem.iditem => item.iditem";
    .

# Table instituicao

```

```

map:instituicao a d2rq:ClassMap;
    d2rq:dataStorage map:database;
    d2rq:uriPattern "instituicao/@ @instituicao.idinstituicao @@";
    d2rq:class foaf:Organization;
    d2rq:classDefinitionLabel "instituicao";
.

map:instituicao__label a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property rdfs:label;
    d2rq:pattern "instituicao # @ @instituicao.idinstituicao @@";
.

map:instituicao_idinstituicao a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property crm:P1_is_identified_by;
    d2rq:propertyDefinitionLabel "instituicao idinstituicao";
    d2rq:column "instituicao.idinstituicao";
    d2rq:datatype xsd:integer;
.

map:instituicao_nome a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property skos:prefLabel;
    d2rq:propertyDefinitionLabel "instituicao nome";
    d2rq:column "instituicao.nome";
.

map:instituicao_site a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property foaf:homepage;
    d2rq:propertyDefinitionLabel "instituicao site";
    d2rq:uriColumn "instituicao.site";
.

map:instituicao_endereco a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property dbo:address;
    d2rq:propertyDefinitionLabel "instituicao endereco";
    d2rq:column "instituicao.endereco";
.

map:instituicao_pais a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:instituicao;
    d2rq:property dbo:country;
    d2rq:propertyDefinitionLabel "instituicao pais";
    d2rq:column "instituicao.pais";
.

# Table item
map:item a d2rq:ClassMap;
    d2rq:dataStorage map:database;
    d2rq:uriPattern "item/@ @item.iditem @@";
    d2rq:class crm:E19_Physical_Object;
    d2rq:classDefinitionLabel "item";
.

map:item__label a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:item;
    d2rq:property rdfs:label;
    d2rq:pattern "item # @ @item.iditem @@";
.

map:item_iditem a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:item;
    d2rq:property crm:P1_is_identified_by;
    d2rq:propertyDefinitionLabel "item iditem";
    d2rq:column "item.iditem";

```

```

        d2rq:datatype xsd:integer;
    .
    map:item_numero a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P48_has_preferred_identifier;
        d2rq:propertyDefinitionLabel "item numero";
        d2rq:column "item.numero";
    .
    map:item_nome a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P102_has_title;
        d2rq:propertyDefinitionLabel "item nome";
        d2rq:column "item.nome";
    .
    map:item_tipo_nome a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:name_type;
        d2rq:propertyDefinitionLabel "item tipo_nome";
        d2rq:column "item.tipo_nome";
    .
    map:item_autoridade_nome a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:named_by;
        d2rq:propertyDefinitionLabel "item autoridade_nome";
        d2rq:column "item.autoridade_nome";
    .
    map:item_metodo_aquisicao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P30i_custody_transferred_through;
        d2rq:propertyDefinitionLabel "item metodo_aquisicao";
        d2rq:column "item.metodo_aquisicao";
    .
    map:item_data_aquisicao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property dbp:acquisitionDate;
        d2rq:propertyDefinitionLabel "item data_aquisicao";
        d2rq:column "item.data_aquisicao";
    .
    map:item_fonte_aquisicao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P29i_received_custody_through;
        d2rq:propertyDefinitionLabel "item fonte_aquisicao";
        d2rq:column "item.fonte_aquisicao";
    .
    map:item_estado_conservacao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:conservation_status;
        d2rq:propertyDefinitionLabel "item estado_conservacao";
        d2rq:column "item.estado_conservacao";
    .
    map:item_sumario_conservacao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:conservation_description;
        d2rq:propertyDefinitionLabel "item sumario_conservacao";
        d2rq:column "item.sumario_conservacao";
    .
    map:item_data_avaliacao_conservacao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:conservation_evaluation_date;
        d2rq:propertyDefinitionLabel "item data_avaliacao_conservacao";

```

```

        d2rq:column "item.data_avaliacao_conservacao";
    .
    map:item_descricao_fisica a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P3_has_note;
        d2rq:propertyDefinitionLabel "item descricao_fisica";
        d2rq:column "item.descricao_fisica";
    .
    map:item_localizacao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P55_has_current_location;
        d2rq:propertyDefinitionLabel "item localizacao";
        d2rq:column "item.localizacao";
    .
    map:item_medidas a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P43_has_dimension;
        d2rq:propertyDefinitionLabel "item medidas";
        d2rq:column "item.medidas";
    .
    map:item_local_coleta a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:gathering_place;
        d2rq:propertyDefinitionLabel "item local_coleta";
        d2rq:column "item.local_coleta";
    .
    map:item_data_coleta a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:gathering_date;
        d2rq:propertyDefinitionLabel "item data_coleta";
        d2rq:column "item.data_coleta";
    .
    map:item_coletor a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:gatherer;
        d2rq:propertyDefinitionLabel "item coletor";
        d2rq:column "item.coletor";
    .
    map:item_local_producao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property dcterms:spatial;
        d2rq:propertyDefinitionLabel "item local_producao";
        d2rq:column "item.local_producao";
    .
    map:item_data_producao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property dcterms:created;
        d2rq:propertyDefinitionLabel "item data_producao";
        d2rq:column "item.data_producao";
    .
    map:item_nome_produutor a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P108i_was_produced_by;
        d2rq:propertyDefinitionLabel "item nome_produutor";
        d2rq:column "item.nome_produutor";
    .
    map:item_funcao_producao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property vocab:production_function;
        d2rq:propertyDefinitionLabel "item funcao_producao";

```

```
        d2rq:column "item.funcao_producao";
    .
    map:item_assunto a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property dcterms:subject;
        d2rq:propertyDefinitionLabel "item assunto";
        d2rq:column "item.assunto";
    .
    map:item_colecao a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P46i_forms_part_of;
        d2rq:propertyDefinitionLabel "item colecao";
        d2rq:column "item.colecao";
    .
    map:item_idinstituicao__ref a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:item;
        d2rq:property crm:P52_has_current_owner;
        d2rq:refersToClassMap map:instituicao;
        d2rq:join "item.idinstituicao => instituicao.idinstituicao";
    .
```