

Dados abertos interligados:

publicação, recuperação e
integração de acervos de arquivos,
bibliotecas e museus na web

Carlos Henrique Marcondes



**CULTURA
ACADÊMICA**
Editora

As tecnologias de dados abertos vinculados fazem parte da Web Semântica e são uma nova forma de representar conteúdos para que os programas, além de publicá-los, possam compreender seus significados e processá-los de forma mais inteligente.

Neste novo paradigma tecnológico os conteúdos são independentes de programas específicos; podem ser publicados diretamente na web e ser enriquecidos, integrados e contextualizados por seus curadores originais e/ou por terceiros. Da mesma forma, eles são estruturados em formatos que permitem seu uso por usuários humanos e aplicações informáticas.

Este livro é dirigido a profissionais de informação, gestores de acervos de instituições arquivísticas, bibliotecas e museus ou profissionais de tecnologia da informação de instituições que estejam desenvolvendo ou planejem desenvolver projetos de digitalização e publicação de seus acervos segundo as tecnologias de dados abertos interligados.

“Agradeço também à equipe das editoras Oficina Acadêmica, Marília e Cultura Acadêmica, São Paulo, pela presteza no apoio técnico.”



**CULTURA
ACADÊMICA**
Editora

Títulos de la colección:

1. Cibermetría. Midiendo el espacio red
Enrique Orduña-Malea, Isidro F. Aguillo
2. La web social como nuevo medio de comunicación y evaluación científica
Amalia Mas-Bleda, Isidro F. Aguillo
3. SEO. Introducción a la disciplina del posicionamiento en buscadores
Jorge Serrano-Cobos
4. Métricas de la web social para bibliotecas
Nieves González-Fernández-Villavicencio
5. Calidad en sitios web. Método de análisis general, ecommerce, imágenes, hemerotecas y turismo
Rafael Pedraza-Jiménez, Lluís Codina y Javier Guallar (coord.)
6. Gestión de datos de investigación
Fabiano Couto Corrêa
7. Desafío a la investigación estándar en comunicación. Crítica y alternativas
Manuel Goyanes
8. Bibliotecas integradas
Pablo Parra-Valero
9. Gestión de información personal. Integración mediante el correo electrónico
Jorge Franganillo

Podéis consultar nuestro catálogo em
www.editorialuoc.com

Arquivos, bibliotecas e museus têm sido guardiões da memória e cultura desde o surgimento destas instituições na Antiguidade. Catálogos têm sido os instrumentos através dos quais instituições de memória e cultura como arquivos, bibliotecas e museus divulgam seus acervos. Os catálogos passaram de listas de títulos, de fichários desdobrados por títulos, autores e assuntos até, com o advento da tecnologia dos computadores (a partir da década de 1960), em catálogos em linha.

Com o surgimento da Web os catálogos de arquivos, bibliotecas e museus tiveram um grande avanço, passando a estar disponíveis ai e podendo então ser consultados a partir de qualquer lugar e a qualquer hora. Também os acervos veem sendo digitalizados e disponibilizados em linha. No entanto a tecnologia atual dos sistemas de catálogos na Web apresenta como grande limitação restringir o acesso e a interligação dos conteúdos ao escopo do sistema de catálogo, transformando este em um “silo” que aprisiona a informação e limitando assim as possibilidades de integração de acervos digitais de arquivos, bibliotecas e museus entre si e aos demais conteúdos existentes na Web.

Web vêm se desenvolvendo segundo a visão da Web Semântica. As tecnologias de dados abertos interligados são parte da visão da Web Semântica e uma nova forma de representar conteúdos que permitem a programas, mais que simplesmente publicarem estes conteúdos, também “compreenderem seus significados”, podendo assim processá-los de forma mais “inteligente”.

O livro apresenta as tecnologias de dados abertos interligados tanto do ponto de vista técnico quanto de sua aplicação à publicação de conteúdos de catálogos de arquivos, bibliotecas e museus. Um esquema geral de gestão de acervos digitais com vistas a sua publicação na Web utilizando as potencialidades dessas tecnologias é apresentado, discutindo as atribuições dos profissionais de informação na curadoria de acervos digitais e abrangendo temas críticos como: atribuição de identificadores persistentes, criação de “links” a partir de campos específicos dos catálogos, promoção de “links” para acervos publicados segundo estas tecnologias, integração de conteúdos de catálogos de arquivos, bibliotecas e museus, identificação e curadoria dos itens destacados de cada acervo, etc.



CARLOS HENRIQUE MARCONDES

Professor Titular da Universidade Federal Fluminense, Rio de Janeiro, Brasil, professor Visitante da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, pesquisador nível I do CNPq. Realizou estágio de pós-doutorado na Universidad Carlos III de Madrid sob a supervisão do Prof. José Antonio Moreiro González. É investigador Colaborador da Universidad Complutense de Madrid, onde parte do grupo de pesquisas “BIBLIOTECA Y SOCIEDAD”. Apresentou várias palestras seminários e oficinas sobre os temas desse livro.

ISBN 978-65-5954-039-6



**DADOS ABERTOS INTERLIGADOS:
PUBLICAÇÃO, RECUPERAÇÃO E INTEGRAÇÃO DE
ACERVOS DE ARQUIVOS,
BIBLIOTECAS E MUSEUS NA WEB**

CARLOS HENRIQUE MARCONDES

**DADOS ABERTOS INTERLIGADOS:
PUBLICAÇÃO, RECUPERAÇÃO E INTEGRAÇÃO DE
ACERVOS DE ARQUIVOS,
BIBLIOTECAS E MUSEUS NA WEB**

Marília/Oficina Universitária
São Paulo/Cultura Acadêmica

2021



**CULTURA
ACADÊMICA**
Editora

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS - FFC
UNESP - campus de Marília

Diretor

Prof. Dr. Marcelo Tavella Navega

Vice-Diretor

Dr. Pedro Geraldo Aparecido Novelli

Conselho Editorial

Mariângela Spotti Lopes Fujita (Presidente)

Adrián Oscar Dongo Montoya

Célia Maria Giacheti

Cláudia Regina Mosca Giroto

Marcelo Fernandes de Oliveira

Marcos Antonio Alves

Neusa Maria Dal Ri

Renato Geraldi (Assessor Técnico)

Rosane Michelli de Castro

Parecerista

Prof. Dr. Guilherme Ataíde Dias

Docente Associado no Departamento de Ciência da Informação da Universidade Federal da Paraíba.

Ficha catalográfica

Serviço de Biblioteca e Documentação - FFC

-
- M321d Marcondes, Carlos Henrique.
Dados abertos interligados : publicação, recuperação e integração de acervos de arquivos, bibliotecas e museus na web / Carlos Henrique Marcondes. – Marília : Oficina Universitária ; São Paulo : Cultura Acadêmica, 2021.
124 p. : il.
Inclui bibliografia
ISBN 978-65-5954-039-6 (Impresso)
ISBN 978-65-5954-040-2 (Digital)
DOI: <https://doi.org/10.36311/2021.978-65-5954-040-2>
1. Ciência da informação. 2. Dados ligados. 3. Web semântica. 4. Metadados. 5. Tecnologia da informação. 6. Recuperação da informação. I. Título.

CDD 020

Copyright © 2020, Faculdade de Filosofia e Ciências

Editora afiliada:



Associação Brasileira de
Editoras Universitárias

Cultura Acadêmica é selo editorial da Editora UNESP
Oficina Universitária é selo editorial da UNESP - campus de Marília

DADOS ABERTOS INTERLIGADOS: PUBLICAÇÃO, RECUPERAÇÃO E INTEGRAÇÃO DE ACERVOS DE ARQUIVOS, BIBLIOTECAS E MUSEUS NA WEB

OBJETIVOS

Apresentar as tecnologias de dados abertos interligados e sua aplicação na publicação, recuperação e interligação de conteúdos de catálogos de arquivos, bibliotecas e museus

RESUMO

Arquivos, bibliotecas e museus têm sido guardiões da memória e cultura desde o surgimento destas instituições na Antiguidade. Catálogos têm sido os instrumentos através dos quais instituições de memória e cultura como arquivos, bibliotecas e museus divulgam seus acervos. Os catálogos passaram de listas de títulos, de fichários desdobrados por títulos, autores e assuntos até, com o advento da tecnologia dos computadores (a partir da década de 1960), em catálogos em linha.

Com o surgimento da Web (cerca de 2000) os catálogos de arquivos, bibliotecas e museus tiveram um grande avanço, passando a estar disponíveis ai e podendo então ser consultados a partir de qualquer lugar e a qualquer hora. Também os acervos veem sendo digitalizados e disponibilizados em linha. No entanto a tecnologia atual dos sistemas de catálogos na Web apresenta como grande limitação restringir o acesso e a interligação dos conteúdos ao escopo do sistema de catálogo, transformando este em um “silo” que aprisiona a informação e limitando

assim as possibilidades de integração de acervos digitais de arquivos, bibliotecas e museus entre si e aos demais conteúdos existentes na Web.

A partir de 2000 a Web vêm se desenvolvendo segundo a visão da Web Semântica. As tecnologias de dados abertos interligados são parte da visão da Web Semântica e uma nova forma de representar conteúdos que permitem a programas, mais que simplesmente publicarem estes conteúdos, também “compreenderem seus significados”, podendo assim processá-los de forma mais “inteligente”.

O livro apresenta as tecnologias de dados abertos interligados tanto do ponto de vista técnico quanto de sua aplicação à publicação de conteúdos de catálogos de arquivos, bibliotecas e museus. Um esquema geral de gestão de acervos digitais com vistas a sua publicação na Web utilizando as potencialidades dessas tecnologias é apresentado, discutindo as atribuições dos profissionais de informação na curadoria de acervos digitais e abrangendo temas críticos como: atribuição de identificadores persistentes, criação de “links” a partir de campos específicos dos catálogos, promoção de “links” para acervos publicados segundo estas tecnologias, integração de conteúdos de catálogos de arquivos, bibliotecas e museus, identificação e curadoria dos itens destacados de cada acervo, etc.

Este livro é dirigido a profissionais de informação, gestores de acervos de instituições arquivísticas, bibliotecas e museus ou profissionais de tecnologia da informação de instituições que estejam desenvolvendo ou planejem desenvolver projetos de digitalização e publicação de seus acervos segundo as tecnologias de dados abertos interligados.

Este livro é uma versão revista e ampliada de “Datos abiertos enlazados de archivos, bibliotecas y museos en la Web” publicado pelo Editorial UOC de Barcelona em 2018, com ISBN 978-84-9180-307-2. O autor é Professor Titular da Universidade Federal Fluminense, Rio de Janeiro, Professor Visitante da ECI/UFMG, Brasil, pesquisador nível I do CNPq. Realizou estágio de pós-doutorado na Universidad Carlos III de Madrid sob a supervisão do Prof. José Antonio Moreira González. É investigador Colaborador da Universidad Complutense de Madrid, onde parte do grupo de pesquisas “BIBLIOTECA Y SOCIEDAD”. Apresentou seminário e oficina sobre os temas deste livro na UCM em abril de 2017.

AGRADECIMENTOS

Quem tem um amigo, tem um tesouro. Devo este livro inteiramente a minha amiga, professora Maria Antonia Garcia Moreno, da Universidade Complutense de Madrid, que me incentivou a escrevê-lo depois de ter apresentado uma oficina sobre este tema na UCM em 2017.

A professora Maria Antonia Garcia Moreno e o professor Tony Hernández Pérez, da Universidade Carlos III de Madrid, traduziram o texto da edição em espanhol deste livro. Sem o apoio e estímulo de ambos, esse livro não seria escrito.

Agradeço ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) seu apoio às minhas pesquisas sobre o tema.

Agradeço também, a professora Mariângela Spotti Lopes Fujita, da Editora Cultura Universitária, da Faculdade de Filosofia e Ciências da UNESP/Marília, pelo seu incentivo em publicar esta versão do livro em português.

SUMÁRIO

1. INTRODUÇÃO-----	13
1.1. A tecnologia atual de disponibilização de conteúdos de catálogos arquivos, bibliotecas e museus	
1.2. Evolução histórica dos registros bibliográficos e dos “software” de catálogos	
1.3. A Interoperabilidade e os acervos digitais de arquivos, bibliotecas e museus	
1.4. Surgimento da Web e as limitações atuais das tecnologias de registros de catálogos	
1.5. O que estamos interligando: objetos digitais de acervos de memória e cultura	
1.6. Considerações finais	
2. TECNOLOGIAS WEB SEMÂNTICA E LOD – DADOS ABERTOS INTERLIGADOS	31
2.1. Representação de conteúdos: RDF	
2.2. Identificadores persistentes, URLs X URIs, IRIs	
2.3. Licenças de uso	
2.4. Consulta a bases de dados: linguagem/protocolo SPARQL	
2.5. Representação de conteúdos processáveis por programas e semântica	
2.6. Projetos de dados abertos interligados em arquivos, bibliotecas e museus	

3.1. Vocabulários usados frequentemente com as tecnologias LOD

3.1.1 Dublin Core Metadata Elements

3.1.2. Vocabulários da Getty Foundation

3.1.3. ICONCLASS

3.1.4. Vocabulários da Library of Congress

3.1.5. VIAF – Virtual International Authority File

3.1.6. Iniciativa SPAR - Semantic Publishing and Referencing
Ontologies

3.1.7. Vocabulários da Wikipedia, DBpedia e Wikidata

3.1.8. GEONAMES

3.1.9. Schema.org

3.1.10. VRA – Visual Resources Association

3.1.11. AGROVOC

3.1.12. SKOS - Simple Knowledge Organization System.

3.1.13. Registros de vocabulários compatíveis com as tecnologias
de dados abertos interligados

3.2. Modelos conceituais na área de informação e documentação

3.2.1. O que são modelos conceituais e seu papel na representação
conteúdos de arquivos, bibliotecas e museus

3.2.2. Modelo FRBR (bibliotecas)

3.2.3. Modelo CIDOC CRM (museus)

3.2.4. EDM - Europeana Data Model

3.2.5. Modelo RiC-CM (arquivos)

3.2.6. Modelo BIBFRAME

3.2.7. Outros modelos conceituais na área de informação e
documentação

3.2.8. Considerações finais

4. PUBLICAÇÃO DE CATÁLOGOS NA WEB USANDO AS TECNOLOGIAS LOD ----	99
4.1. As novas tarefas do curador digital na publicação de acervos como LOD	
4.3. Convertendo campos de catálogos para dados abertos interligados	
4.4. Promoção e enriquecimento de relacionamentos entre de catálogos publicados como LOD e outros recursos	
4.5. Ferramentas	
4.5.1. Navegadores de dados abertos	
4.5.2. Editores RDF	
4.5.3. Validadores	
4.5.4. Conversores de registros de catálogos	
4.5.5. Bancos de dados	
4.5.6. Ferramentas para descobrir “links”	
4.5.7. Outras ferramentas	
5. CONSIDERAÇÕES FINAIS -----	113
REFERÊNCIAS -----	115

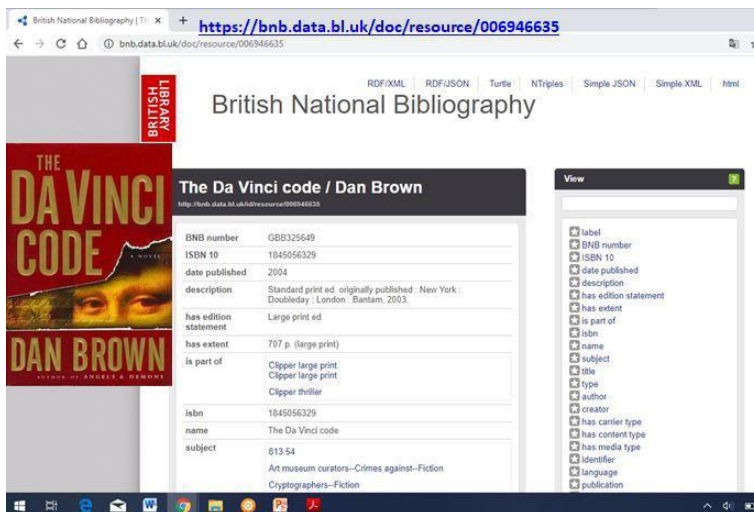
1. INTRODUÇÃO

O que queremos dizer com publicar acervos digitais com as tecnologias de dados abertos interligados?

Vamos considerar o seguinte “link”: <https://bnb.data.bl.uk/doc/resource/006946635>.

Se “colarmos” esse “link” janela de um navegador como o Google Chrome e acionarmos “Enter” seremos levados para a seguinte página.

Figura 1 - Um registro da BNB publicado como dados abertos interligados¹



The screenshot shows a web browser displaying the British National Bibliography (BNB) record for 'The Da Vinci code' by Dan Brown. The URL in the address bar is <https://bnb.data.bl.uk/doc/resource/006946635>. The page features a 'BRITISH ADVISORY' logo and a 'View' panel on the right. The main content area displays the following metadata:

The Da Vinci code / Dan Brown	
BNB number	GBB325649
ISBN 10	1845056329
date published	2004
description	Standard print ed. originally published New York, Doubleday ; London : Bantam, 2003.
has edition statement	Large print ed.
has extent	707 p. (large print)
is part of	Clipper large print. Clipper thriller.
isbn	1845056329
name	The Da Vinci code
subject	813.54 Art museum curators--Crimes against--Fiction. Cryptographers--Fiction.

The 'View' panel on the right lists various metadata fields with checkboxes, including: label, BNB number, ISBN 10, date published, description, has edition statement, has extent, is part of, isbn, name, subject, title, type, author, creator, has carrier type, has content type, has media type, identifier, language, and publication.

Fonte: elaborado pelo autor

¹ Disponível em: <https://bnb.data.bl.uk/doc/resource/006946635>. Acesso em: 21 de jun. de 2020.

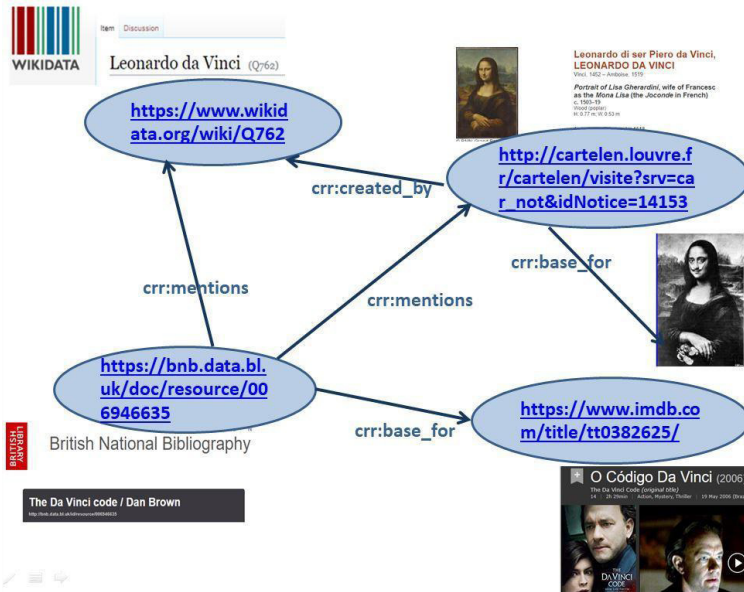
O “link” conduz à página do registro do livro “O Código Da Vinci”, do autor Dan Brown, na British National Bibliography; a página se parece com uma página html comum, amigável e legível para nós. No entanto esta página não está em formato html, está formatada segundo as tecnologias de dados abertos interligados. A página é formada em um conjunto de triplas, descrevendo o recurso identificado pelo “link” acima, cada tripla contendo o “link” que identifica o recurso, uma de suas propriedades e o valor dessa propriedade, como por exemplo:

```
<https://bnb.data.bl.uk/doc/resource/006946635> <dc:title>  
<The da Vinci Code>.
```

Pensemos agora em outros recursos que possam haver na Web relacionados com o Código Da Vinci; o mais óbvio seria a Mona Lisa, pintura que pertence ao acervo do Museu do Louvre em Paris, França, de autoria de Leonardo Da Vinci (que tem o “link” <https://www.louvre.fr/en/oeuvre-notices/mona-lisa-portrait-lisa-gherardini-wife-francesco-del-giocondo>); outro recurso seria o artigo sobre o pintor Leonardo Da Vinci da Wikidata, a versão da Wikipedia em dados abertos interligados (que tem o “link” <https://www.wikidata.org/wiki/Q762>); outro recurso seria o registro do filme “The da Vinci Code”, estrelado por Tom Hanks, no Internet Movie Database (que tem o “link” <https://www.imdb.com/title/tt0382625/>); outro seria a versão autoretrato de Salvador Dali como Mona Lisa, feita pelo pintor em colaboração com o fotógrafo Philippe Hasman (que tem o “link” <http://www.studiolo.org/Mona/MONASV12.htm>).

Os quatro recursos têm “links” válidos, estão portanto publicados diretamente na Web e estão relacionados entre si. Poderíamos dizer que os três recursos se relacionam assim: o livro de Dan Brown “menciona” a Mona Lisa, que esta foi “criada por” de Leonardo Da Vinci, que o livro de Dan Brown foi a “base para” o filme estrelado por Tom Hanks e que Dali fez uma paródia da Mona Lisa. Com “links” específicos para os significados dos relacionamentos “menciona”, “criado por” e “base para” (MARCONDES, 2020) poderíamos interligar os quatro recursos inicialmente isolados e assim criar uma narrativa, ou um novo recurso, formado por estes recursos mais as suas interrelações, como na figura seguinte.

Figura 2 - Diferentes recursos publicados na Web como dados abertos interligados



Fonte: elaborado pelo autor

Assim funcionam os dados abertos interligados. Os quatro registros estão publicados, isto é, são acessíveis diretamente na Web através de seus “links” e estão interligados através, não de “link” comuns como os usados para navegação entre páginas html, mas por “links” semânticos que possuem um significado específico, de modo a compor uma narrativa.

Este livro é sobre como interligar registros de acervos digitais de arquivos, bibliotecas e museus, publicados na Web com as tecnologias de dados abertos interligados, formando novos recursos, como exposições ou aulas virtuais; esses novos recursos são, compostos por registros digitais interligados de acervos diferentes instituições. Assim são potencializadas sinergias, ampliada a disseminação destes acervos e as experiências dos usuários em navegar por eles.

1.1. A TECNOLOGIA ATUAL DE DISPONIBILIZAÇÃO DE CONTEÚDOS DE CATÁLOGOS ARQUIVOS, BIBLIOTECAS E MUSEUS

Desde as décadas de 1980-1990 surgiram vários projetos de digitalização e acesso em linha de acervos de cultura e memória em todo o mundo. Nos EUA o projeto pioneiro foi o American Memory², no período 1990-1995, que digitalizou conteúdos do acervo da Library of Congress e os distribuía em mídias como CD-ROM e videodiscos para escolas e outras instituições de ensino. O programa Digital Library Initiative³, desde 1994, desenvolveu tecnologias e implementou projetos-piloto, recebendo apoio de agências federais americanas como a NSF, DARPA e NASA no montante U\$68 milhões (FOX, 1999). Na sua primeira fase o projeto aportou recursos a 6 iniciativas, que juntavam criação de projetos-piloto e desenvolvimento de tecnologias específicas (processamento de textos, imagens e reconhecimento de fala): Universidade da Califórnia em Berkeley e Santa Barbara, Universidade de Stanford, Califórnia, Universidade de Illinois em Urbana-Champaign, Universidade de Michigan e Universidade Carnegie-Mellon. O programa se desdobrou em uma segunda fase.

Na Europa o programa eContent - European digital content on the global networks -, cuja primeira fase desenvolveu-se de 2001 a 2004⁴ e a segunda⁵, de 2005-2008, deu o impulso inicial das políticas públicas europeias de digitalização de conteúdos. Esse programa é um resultado direto das recomendações da reunião do Conselho Europeu em Lisboa, no ano de 2000 (MENDO CARMONA; TEJADA ARTIGAS, 2014). Dessas políticas públicas resultaram também o desenvolvimento do Portal Europeu de Arquivos e a Biblioteca Europeia, esta, a antecessora da Biblioteca Europeia.

Nestas iniciativas a digitalização, acesso e reuso de conteúdos de instituições de memória e cultura é encarada como uma prioridade e também como um investimento, com impactos para além da área cultural, em setores econômicos como educação, turismo, indústrias criativas.

Como relata Kapsalis (2016), comentando estudo feito em instituições que adotaram o livre acesso a seus acervos digitais na Web,

² Ver <https://memory.loc.gov/ammem/dli2/html/lcndlpl.html>.

³ Ver https://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=100660.

⁴ Ver http://cordis.europa.eu/pub/econtent/docs/call1_brochure.pdf.

⁵ Ver http://cordis.europa.eu/econtent/customisation/cust_projects.htm.

essas instituições têm muito a lucrar com a publicação dos seus acervos em formatos abertos e de livre acesso. Esta decisão institucional joga um papel fundamental para potencializar o reuso, aumentar seus públicos, tanto no seu sítio Web quanto presenciais, aumentar sua exposição na mídia e facilitar seu acesso a recursos de fomento.

As tecnologias LOD vêm contribuir para ampliar o acesso e reuso dos acervos digitais de arquivos, bibliotecas e museus. Elas se constituem em um passo adiante em relação às tecnologias correntes de acesso a estes conteúdos através da Web, os catálogos.

Catálogos em linha são hoje a ferramenta básica para a descoberta de recursos em arquivos, bibliotecas e museus. Um catálogo em linha, ou OPAC -“online public access catalog” - como é conhecido pela sigla em inglês, é um programa de computador que permite a usuários em linha em uma rede, ou através da Web, consultar uma base de dados de registros catalográficos. Um catálogo em linha é um programa *reativo*, isto é, funciona ou reage quando um usuário submete ao sistema suas necessidades de informação, especificando uma consulta através de um conjunto de palavras-chaves, termos controlados, nomes de autores, etc. O programa busca e retorna registros da sua base de dados que atendam aos critérios especificados.

Desde o surgimento dos sistemas computadorizados em linha e da Internet que catálogos em linha vêm sendo usados por arquivos, bibliotecas e museus para disseminarem seus acervos. Com o surgimento da Web a partir da década de 1990 os catálogos de arquivos, bibliotecas e museus tiveram um grande avanço, passando a estar disponíveis na Web, podendo então ser consultados a partir de qualquer lugar e à qualquer hora. Simultaneamente os itens mais significativos de diversos os acervos veem sendo digitalizados e disponibilizados em catálogos em linha. Os catálogos em linha constituem a tecnologia atual de disseminação de acervos de arquivos, bibliotecas e museus. Exemplos de catálogos em linha de arquivos⁶, bibliotecas⁷ e museus⁸ podem ser encontrados nos “links” indicados.

⁶ Catálogo do Archives Hub, Reino Unido, <https://archiveshub.jisc.ac.uk>.

⁷ Catálogo da Biblioteca Ambrosiana, Milão, Itália, <http://ambrosiana.comperio.it/%20>.

⁸ Catálogo Online do Museu do Louvre, França, http://cartelen.louvre.fr/cartelen/visite?srv=crt_frm_rs&langue=en&initCritere=true.

1.2. EVOLUÇÃO HISTÓRICA DOS REGISTROS BIBLIOGRÁFICOS E DOS “SOFTWARE” DE CATÁLOGOS

As primeiras aplicações computacionais aplicadas à documentação foram desenvolvidas por Peter Luhn, um pesquisador da IBM, na década de 1960. Eram aplicações que geravam índices de palavras a partir do título de publicações.

O antecessor dos atuais programas de catálogos em linha disponíveis na Web foi o sistema MEDLARS (Medical Literature Analysis and Retrieval System) operado pela Biblioteca Nacional de Medicina dos EUA, lançado em 1964. O MEDLARS permitia consultar uma base de dados de registros de artigos biomédicos. Funcionava por processamento “em lotes”, modalidade característica dos primeiros computadores das décadas de 1950 a 1970. O processamento em lotes era o oposto dos atuais sistemas de recuperação de informações em linha interativos, onde um usuário submete uma consulta e os registros são recuperados imediatamente; o usuário pode examinar os resultados recuperados pelo sistema e reformular sua estratégia de pesquisa. No MEDLARS uma consulta à base de dados era submetida ao sistema através de cartões perfurados e os itens da base de dados recuperados eram impressos em uma listagem, sem qualquer interação com o usuário.

A partir de 1971 foi lançada uma versão em linha do MEDLARS, o MEDLINE – MEDLARS On Line). Apesar dos avanços da versão em linha, o MEDLINE ainda era de acesso bastante restrito, operado por bibliotecários e pesquisadores que tinham acesso direto à NLM e não pelo público. Essa situação muda a partir de 1990, com o surgimento da Web e dos navegadores, quando é lançada uma versão pública, disponível na Web, do MEDLINE, o PubMed⁹.

A Library of Congress - LC - dos EUA, que publicava seu catálogo de fichas bibliográficas desde o início do século XX, permitindo que outras bibliotecas reutilizassem as fichas em seus catálogos, também desenvolve projetos de publicar esse catálogo em meio magnético usando computadores. Esse projeto, o MARC – Machine Readable Cataloging – catalogação Legível por computadores, tem seus primeiros resultados

⁹ Dados históricos sobre a NLM podem ser encontrados em MILESTONES IN NLM HISTORY. Disponível em: <https://apps.nlm.nih.gov/175/milestones.cfm>. Acesso em: 13 jan. 2017.

em 1969, com a geração do catálogo da LC em fitas magnéticas, a mídia portátil da época. No formato MARC todos os itens de uma ficha catalográfica como título, autor, editor, cabeçalhos de assunto, etc., eram gravados em registros computacionais que reproduziam as fichas de um catálogo. Fitas magnéticas com os registros catalográficos eram geradas pelo sistema e podiam ser adquiridas por outras bibliotecas, que por sua vez geravam seus próprios catálogos imprimindo as fichas dos itens de seu acervo previamente catalogados pela LC.

A década de 1980 vê surgirem muitos fornecedores comerciais de ILS – “Integrated Library Systems” – Sistemas Integrados de Gestão de Bibliotecas (ROWLEY, 2002) e de sistemas equivalentes para arquivos e museus, que tinham entre suas funções os catálogos em linha.

1.3. A INTEROPERABILIDADE E OS ACERVOS DIGITAIS DE ARQUIVOS, BIBLIOTECAS E MUSEUS

Ao lado das tecnologias de catálogos em linha, desenvolveram-se também as tecnologias que visavam integrar conteúdos de diversos catálogos distribuídos de diferentes instituições, denominadas tecnologias de interoperabilidade. Esta questão é bastante crítica quando se trata de conteúdos de acervos de instituições como arquivos, bibliotecas e museus. Embora tenham em comum o fato de serem instituições de memória e cultura, suas tradições e metodologias de curadoria e tratamento de acervos visando sua disseminação são muito diversas, evoluíram separadamente ao longo dos anos. É com o surgimento da Web e, em especial, da Web Semântica, que se abrem as potencialidades para integração desses acervos. O objeto deste livro, a integração de acervos digitais de arquivos, bibliotecas e museus, é uma preocupação e um desafio conceitual, metodológico e tecnológico há já bastante tempo (FOX; MARCHIONINI, 1998). Essa questão é conhecida como *interoperabilidade*.

Pode-se compreender interoperabilidade como a propriedade de sistemas diferentes (por ex. catálogos de bibliotecas digitais, instrumentos de pesquisa arquivísticos automatizados, sistemas de gestão de acervos museológicos), através de padrões tecnológicos, instrumentos semânticos, acordos ou propostas, serem capazes de operarem em conjunto visando a execução de uma tarefa, intercambiarem seus conteúdos mantendo o

significado desses conteúdos no sistema original. Quando se fala em conteúdos de diferentes catálogos distribuídos, a interoperabilidade diz respeito fundamentalmente a recuperar esses conteúdos de forma significativa para os usuários que consultam um catálogo.

A interoperabilidade consiste em diferentes dimensões, também chamadas por alguns autores de níveis (CUENCA, 2014; MOEN, 2001; PAYETTE et al., 1999). Essas dimensões seriam:

- *infraestrutura tecnológica*, formada por protocolos como o HTTP¹⁰ e linguagens de marcação como XML¹¹;
- a dimensão *sintática*, formado por conjuntos de metadados como MARC, Dublin Core, o conjunto de metadados de objetos museológicos LIDO¹², os campos previstos na norma ISAD(G);
- a dimensão *semântica*, formado por instrumentos semânticos como vocabulários controlados, terminologias, modelos conceituais como FRBR, CIDOC CRM, EDM, RiC-CM, etc, que fixam o significado dos elementos descritivos; e
- a dimensão *administrativa/política*, formado por fóruns comuns, padrões e normas, acordos e convênios entre instituições.

Cuenca (2014) sugere o acréscimo de uma dimensão *jurídica*; uma vez que esta se falando de disponibilização e intercâmbio de conteúdos, questões relativas à direitos autorais, direitos de acesso, licenças de uso (reuso) teriam que ser endereçadas nessa dimensão.

As diferentes soluções de interoperabilidade utilizadas até hoje enfatizam o aspecto da troca ordenada de conteúdos de forma significativa; formatos de metadados como MARC e Dublin Core têm um papel fundamental nestas soluções de interoperabilidade. Essas soluções podem ser classificadas com relação ao momento em que se dá esta troca de

¹⁰ HTTP Hypertext Transport Protocol, ver <https://www.w3.org/Protocols/>.

¹¹ XML Extended Markup Language, um padrão do W3C, ver em <https://www.w3.org/XML/>.

¹² What is LIDO, Lightweight Information Describing Objects, ver em <http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>.

metadados. Metadados podem ser trocados *antes de serem utilizados*, isto é, podem ser primeiro agregados em uma base de dados comum como pré-condição para seu uso; ou podem ser trocados *simultaneamente ao momento de sua consulta e utilização*. A primeira solução é a dos formatos MARC ou LILACS¹³, utilizado para a montagem de um catálogo coletivo/base de dados comum, ou o uso do padrão de metadados Dublin Core – DC - para a coleta automática de metadados no protocolo OAI-PMH.

A segunda solução é a utilizada pelo protocolo Z39.50 ou, no mais recente protocolo SRU/SRW (MORGAN, 2004), que permitem, a partir de um programa cliente, a consulta e recuperação de registros MARC ou MODS¹⁴ ou DC, de vários sistemas de catálogos diferentes simultaneamente, como se fossem um único catálogo virtual.

Nestas soluções a interoperabilidade sintática, isto é, a conformidade com um padrão, é garantida por um específico formato de codificação de conteúdos, como a norma ISO2709¹⁵ ou a linguagem XML¹⁶ na qual são codificados os metadados Dublin Core¹⁷ – DC - no protocolo OAI-PMH. Quanto à interoperabilidade semântica, isto é, os conteúdos originários de um sistema A fazem sentido quando importados ou integrado em um outro sistema B, no caso de formatos como o MARC, ela é garantida por padrões como o uso associado do MARC com o CCAA2 - Código de Catalogação Anglo-Americano, 2ª. edição, que especificaria não só a semântica de cada campo como também as regras de como transcrevê-lo num registro bibliográfico. Este esquema vem sendo usado por bibliotecas do mundo inteiro há décadas. Deve ser lembrado que os sistemas a serem interoperados neste caso têm seus conteúdos completamente *uniformes*, formados por registros bibliográficos em formatos como MARC ou LILACS, descritos segundo o CCAA2.

No caso de do uso do padrão de metadados DC pode-se dizer que a interoperabilidade semântica é informal, já que o conteúdo de cada elemento é descrito informalmente; seria o caso de Autor no sistema 1

¹³ Formato LILACS. Disponível em: <http://metodologia.lilacs.bvsalud.org/download/P/LILACS-2-ManualDescricao-pt.pdf>.

¹⁴ MODS. Disponível em: <http://www.loc.gov/standards/mods/>.

¹⁵ INTERNATIONAL STANDARD ORGANIZATION. ISO2709 Information and documentation: format for information exchange. 2008. Disponível em: http://www.iso.org/iso/catalogue_detail.htm?csnumber=41319.

¹⁶ Linguagem XML. Disponível em: <http://www.w3.org/XML/>.

¹⁷ Ver <http://dublincore.org/documents/dces/>.

e Artista no sistema 2, ambos correspondendo ao elemento dc:creator. Isto acarreta problemas como no caso de se buscar a interoperabilidade entre acervos arquivísticos, bibliográficos e museológicos; ao contrário dos acervos bibliográficos tornados interoperáveis através de formatos como o MARC, este acervos são pouco uniformes. Aspectos como a granularidade – registros bibliográficos representam uma obra, que pode ter várias edições e exemplares, registros arquivísticos representam agrupamentos como fundos ou séries, registros museológicos representam objetos únicos -, são difíceis de serem contornados e unificados através de um padrão de metadados tão genérico como o DC.

A adoção dos chamados Perfis de Aplicação – Application Profiles¹⁸ -, tem sido uma das formas de tratar o problema de aspectos específicos da descrição de acervos digitais de arquivos não contemplados pelo padrão DC. No entanto, o desdobramento do padrão DC em elementos específicos de um Perfil de Aplicação, como foi o caso dos objetos museológicos (MILLER, 2000; PERKINS; SPINAZZE, 1999) não resolve o problema específico da interoperabilidade entre acervos de arquivos, bibliotecas e museus – bastante heterogêneos entre si -, antes o agrava. Apesar dessas questões, uma experiência interessante do uso do DC, estendido com qualificadores, é o conjunto de metadados ESE - Europeia Semantic Elements, usado para descrever os objetos digitais recebidos pelos diferentes provedores – arquivos, bibliotecas e museus – da Biblioteca Europeia¹⁹.

Recentemente novas arquiteturas com relação às bibliotecas digitais, o setor pioneiro na interoperabilidade, têm sido propostas. Estas novas propostas têm influência, naturalmente, na questão da interoperabilidade. A partir do Digital Library Manifesto, foi criado o DELOS Digital Library Reference Model (CANDELA et al., 2008). As novas tecnologias disponíveis, como a Web Semântica e a computação em grid²⁰ juntamente como modelo de referência citado trouxeram novas arquitetura para as bibliotecas digitais, baseadas não mais em um sistema único, mas na modularização de componentes de software e na distribuição de recursos informacionais e computacionais. Estas novas propostas são

¹⁸ Dublin Core Application profiles. Disponível em: <http://dublincore.org/documents/profileguidelines/>.

¹⁹ Ver <https://pro.europeana.eu/page/ese-documentation>.

²⁰ Grid computing. https://en.wikipedia.org/wiki/Grid_computing.

conhecidas como arquiteturas ponto a ponto (P2P), arquiteturas orientadas a serviços (SoA), funcionando sobre uma infraestrutura tecnológica de computação em grid (CANDELA et al. 2005). Com relação a questão da interoperabilidade as novas arquiteturas não mais centralizam e agregam metadados e sim têm como base uma proposta descentralizada.

A arquitetura ponto a ponto funciona com diversas bibliotecas digitais que se organizam em uma rede para fornecerem seus serviços. A arquitetura se baseia na existência de índices comuns para os recursos disponíveis nas bibliotecas da rede. Quando uma consulta é feita a um dos nós da rede os índices são consultados e se este nó não dispõe do recurso, a consulta é repassada ao nó correspondente e o recurso recuperado e entregue ao usuário. A arquitetura ponto a ponto funciona como o serviço Napster de fornecimento de cópias de músicas. Ao invés das cópias estarem em um único repositório, estão distribuídas em vários, e o Napster simplesmente direciona a requisição por uma música ao repositório correspondente.

Já a arquitetura orientada a serviços tira partido de um padrão para oferta e integração de serviços da Web Semântica, o Web Service²¹. Um web service é um tipo de serviço oferecido através da Web que pode ser acionado por uma interface padronizada, acionada por regras descritas segundo o padrão Web Service Description Language (WSDL)²². Existe também um diretório de localização de serviços específicos, consultado através do protocolo. Os “web services” poderiam integrar serviços oferecidos através da Web como a localização de um livro em uma biblioteca, sua aquisição em uma livraria, a verificação do seu cartão de crédito no seu banco e a entrega do livro através de uma transportadora. Web services são complementados através de um serviço de diretório e localização de serviços disponíveis que funcionam segundo o padrão Web Service, chamado Universal Description, Discovery and Integration (UDDI)²³.

1.4. SURGIMENTO DA WEB E AS LIMITAÇÕES ATUAIS DAS TECNOLOGIAS DE REGISTROS DE CATÁLOGOS

²¹ Web Services. Disponível em: https://en.wikipedia.org/wiki/Web_service.

²² WSDL. Disponível em: <http://www.w3.org/TR/wsdl>.

²³ UDDI. Disponível em: <http://pt.wikipedia.org/wiki/UDDI>.

A Web, surgida no início da década de 1990 (BERNERS-LEE; CAILLIAU, 1990), promoveu um grande avanço na tecnologia dos catálogos em linha, permitindo que catálogos de arquivos, bibliotecas e museus se tornassem realmente de acesso público, passando a estar disponíveis e podendo ser consultados a partir de qualquer lugar e a qualquer hora.

The Vatican Library is an extraordinary repository of rare books and manuscripts. Among its 150,000 manuscripts are early copies of works by Aristotle, Dante, Euclid, Homer, and Virgil. Yet today access to the Library is limited. Because of the time and cost required to travel to Rome, only some 2000 scholars can afford to visit the Library each year. (MINTZER et al., 1996).

Nesse artigo sobre a digitalização da Biblioteca do Vaticano, os autores chamam a atenção sobre as possibilidades abertas com a digitalização e disponibilização na Web dos mais de 150.000 livros e manuscritos raros, um tesouro cultural. Antes do projeto, a Biblioteca só era capaz de atender cerca de 2000 acadêmicos anualmente.

Esta questão ilustra as potencialidades dos acervos digitais em memória e cultura. Quando comparados aos acervos tradicionais – manuscritos, documentos ou obras raras impressas, objetos físico - os objetos digitais têm propriedades específicas: um *alcance* e uma *plasticidade* muito maiores. Com relação ao alcance, um objeto digital disponível na Web pode ser acessado por qualquer usuário, a qualquer hora e desde qualquer lugar; seu valor cultural se propaga em uma proporção impossível aos objetos físicos. Também o acesso de um usuário não impede o acesso de outros, já que, exceto por limitações de banda de Internet, o que são acessadas por diferentes usuários são cópias do objeto.

Com relação à plasticidade, além de apoiarem e serem material para pesquisas acadêmicas, prestando-se para finalidades inusitadas, podem ser recombinados e agregados a outros recursos formando novos recursos, usados como materiais educativos, comporem publicações, exposições ou aulas virtuais. Estes objetos também tem um potencial econômico, vêm

sendo considerados um insumo para as chamadas indústrias criativas, como pode ser constatado na página do projeto Europeana Creative²⁴:

The project aims to support and promote the re-use of cultural resources that are made available via Europeana - a website that provides access to digital resources of Europe's museums, libraries, archives and audio-visual collections. *Europeana Creative* stimulated the re-use of this wealth of material by creative industries active in Design, History Education, Natural History Education, Social Media and Tourism.

Ao mesmo tempo em que surge a Web seus criadores propuseram uma visão para o seu desenvolvimento, conhecida como Web Semântica. Desde então a Web vem se desenvolvendo segundo esta visão. As tecnologias da Web Semântica dizem respeito a publicação e disponibilização de conteúdos. Propõe uma nova forma de representar conteúdos que permite a programas, mais que simplesmente publicarem esses conteúdos para leitura e compreensão de seres humanos, possam também “compreenderem seus significados” e assim processá-los de forma mais “inteligente”. Esses conteúdos seriam publicados diretamente na Web, isto é, acessíveis por “links”, e estruturados, de modo que a Web Semântica possa ser consultada como uma base de dados.

Se comparada à proposta da Web Semântica a tecnologia atual dos sistemas de catálogos na Web apresenta grandes limitações, ao restringir o acesso e a possível interligação dos conteúdos mantidos no catálogo ao escopo do sistema, transformando este em um “silo” que aprisiona a informação. Além disso, os conteúdos dos acervos mantidos no catálogo em linha só fazem sentido, só têm “semântica”, no escopo do sistema de catálogo, seu significado está restrito a este escopo. Um usuário tem que acessar o sistema de catálogo, fazer “login” nele para acessar seus registros. É o sistema de catálogo que recupera, por exemplo, registros em MARC, com seus identificadores de conteúdo codificados, que só fazem sentido para catalogadores que conheçam o formato MARC, e os exibe em campos que fazem sentido a um usuário, como autor, título, editor, data de publicação, etc., exibindo-os na tela. Essa tecnologia limita assim as possibilidades de integração de acervos digitais de arquivos, bibliotecas e

²⁴ <https://euroclio.eu/projects/partner-europeana-creative/>

museus entre si e aos demais conteúdos existentes na Web, encerrando os conteúdos de catálogos de arquivos, bibliotecas e museus em um mundo próprio, fechado, isolado dos outros fluxos de informação da Web.

Os SRI atuais – neles incluído os catálogos de arquivos, bibliotecas e museus, mas também bases de dados, bibliotecas e repositórios digitais, sejam eles físicos ou digitais - trabalham desde sempre baseados no pressuposto, implícito, de que um usuário com uma questão a ser respondida é atendido pelo SRI *através da identificação e acesso a documentos/informações* nos quais se supõe que ele possa encontrar a resposta para sua questão. As políticas de desenvolvimento de acervos, de representação/descrição/catalogação/indexação que envolve os SRIs convencionais apoiam esse pressuposto.

As tecnologias da Web Semântica e dados abertos interligados trabalham em um pressuposto diferente. Elas permitem criar representações legíveis por programas das próprias *coisas* sobre as quais um usuário formula suas questões. Se um usuário interroga a Web Semântica com questões acerca da Pedra da Roseta²⁵ ele não vai recuperar documentos sobre a Pedra da Roseta, mas sim descrições da mesma, sua propriedades mais importantes, além de “links” para outras “coisas” associadas ao mesmo tema, como Jean-François Champollion²⁶, Napoleão Bonaparte²⁷, hieróglifos²⁸, Egito²⁹, etc., criadas por uma autoridade muito especial, o curador desse objeto de patrimônio, o Museu Britânico³⁰.

Outro exemplo, interrogando um SRI convencional sobre a população da cidade do Rio de Janeiro, ele recuperará documentos que trazem dados estatísticos sobre a cidade do Rio de Janeiro, ou séries históricas sobre a evolução da população da cidade do Rio de Janeiro. Ao consultar dados sobre a própria cidade do Rio de Janeiro³¹ representada segundo as tecnologias LOD (usando a linguagem SPARQL³²), ele encontrará o

²⁵ Ver https://pt.wikipedia.org/wiki/Pedra_de_Roseta; ver também <https://www.wikidata.org/wiki/Q48584>.

²⁶ Ver https://pt.wikipedia.org/wiki/Jean-François_Champollion.

²⁷ Ver https://pt.wikipedia.org/wiki/Napoleão_Bonaparte.

²⁸ Ver <https://pt.wikipedia.org/wiki/Hieróglifo>.

²⁹ Ver <https://pt.wikipedia.org/wiki/Egito>.

³⁰ Ver a Pedra da Roseta no Museu Britânico, através de seu URI, <http://collection.britishmuseum.org/id/object/YCA62958>.

³¹ Ver http://dbpedia.org/resource/Rio_de_Janeiro.

³² SPARQL Query language for RDF, <https://www.w3.org/TR/rdf-sparql-query/>.

número de habitantes da cidade; poderá também consultar diretamente pela população da cidade do Rio de Janeiro, e receberá como resposta o número de seus habitantes. A interligação entre essas “coisas” (na verdade, suas representações digitais) publicadas na Web fornece um rico contexto para o usuário. É a chamada “*Web das coisas*”³³.

SRI's automatizados típicos são todos baseados na tecnologia de arquivos invertidos e algoritmos para resolver operadores booleanos. Essa tecnologia remota aos primeiros SRI's da década de 1970. Ele resolve consultas do tipo:

- Dado um assunto ou um autor, recupere referências sobre esse assunto, ou desse autor;

Mas é incapaz de resolver consultas do tipo:

- Recupere referências sobre um dado assunto, de autores filiados a universidades e cursos de pós-graduação com conceito CAPES 7. Ou

- Recupere referências sobre um dado assunto, de autores que foram receberem apoio do Edital Universal CNPq 2016 e que são filiados à UFF.

A segunda e terceira consultas envolvem navegar por uma rede de fatos básicos inter-relacionados, as referências sobre um dado assunto, os autores dessas referências, os programas de pós-graduação a que esses autores são filiados e as notas dos programas segundo a avaliação da CAPES, quais dos autores receberam apoio do Edital Universal CNPq 2016 e quais são filiados à UFF.

Imaginemos também uma consulta em uma base de dados de artigos biomédicos na qual queremos recuperar referências sobre as possíveis relações entre uma substância e uma doença. A Lógica, ou Álgebra, Booleana³⁴, a linguagem de consulta na qual se baseia a tecnologia atual dos SRI, expressaria essa consulta através de: “substância .AND. doença”. No entanto, na área biomédica, relações entre substâncias e doenças tanto podem ser do tipo <substância> <causa> <doença> quanto <substância>

³³ Ver em https://en.wikipedia.org/wiki/Web_of_Things.

³⁴ Ver https://en.wikipedia.org/wiki/Boolean_algebra.

<previne> <doença>. A especificidade de relações como essas não consegue ser expressa através da Lógica Booleana.

1.5. O QUE ESTAMOS INTERLIGANDO: OBJETOS DIGITAIS DE ACERVOS DE MEMÓRIA E CULTURA

O acesso a acervos digitalizados de patrimônio cultural através da Web – livros e documentos raros, manuscritos, objetos arqueológicos, artefatos diversos com valor histórico e cultural - pressupõe sua representação através de metadados que lhes descrevam, forneçam pontos de acesso e assinalem contexto, acompanhadas de cópias ou imagens dos objetos físicos – livros raros, manuscritos e outros documentos digitalizados, fotografias ou vídeos dos objetos. Essas cópias ou imagens dos objetos físicos substituem os mesmos quando acessados através da Web.

Esta atividade digitalização de acervos se baseia em toda uma atividade prévia de valorização, documentação e curadoria desses acervos físicos, realizada por arquivos, bibliotecas e museus. Esta é a missão e o papel que as sociedades atribuem às instituições de memória e cultura.

O objeto digital que será publicado, interligado e acessado através da Web utilizando as tecnologias LOD, é um agregado de metadados de diversas naturezas, associado a cópias ou imagens digitais do objeto físico. Esse agregado é um *objeto digital complexo*, formado de metadados e identificadores que o associam a diversos arquivos. A esse agregado chamaremos neste livro de objeto digital de patrimônio – ODP. Um ODP tem, em geral, uma estrutura e componentes como são mostrados a seguir.

- OPD

- identificador persistente

- metadados do objeto

(esses metadados se referem ou ao objeto físico original, quando o ODP é uma representação de um objeto físico, ou ao objeto nativo digital, quando for o caso)

- metadados descritivos
- metadados temáticos

- metadados relativos aos direitos autorais e de reprodução do objeto em sí.
- metadados da cópia ou imagem do objeto digital (de uma ou mais cópias ou imagens)
- identificador persistente da cópia ou imagem digital do objeto
- metadados técnicos/tecnológicos (formato, tamanho, resolução, etc.)
- metadados relativos aos direitos autorais e de reprodução da cópia ou imagem digital do objeto (quando o ODP for uma representação de um objeto físico).

Diversos padrões endereçam a estrutura e os componentes de objetos digitais complexos, como METS³⁵, ORE³⁶ e ProvidedCHO, o padrão de objeto digital complexo da Biblioteca Europeia (ISAAC, 2013).

1.6. CONSIDERAÇÕES FINAIS

“With linked data, when you have some of it, you can find other, related, data”. (BERNERS-LEE, 2006).

Imagine um típico catálogo Web de biblioteca de onde foi recuperado um registro MARC do livro Don Quijote, de Cervantes. Esse registro esta associado a uma cópia digital da edição original do Quijote. O registro MARC exhibe os campos **100 Entrada Principal Nome Pessoal** e **245 Título e campos relacionados ao Título**, como se segue.

```
1001#$$aMiguel de Cervantes Saavedra$d1547-1616  
24510$aDon Quijote$cMiguel de Cervantes Saavedra
```

³⁵ Ver METS – Metadata Encoding & Transmission Standard, <http://www.loc.gov/standards/mets/>.

³⁶ Ver ORE - Object Reuse and Exchange, <https://www.openarchives.org/ore/>.

Certamente que este registro seria enriquecido se, ao invés de simplesmente informar que Miguel de Cervantes é o autor de *Don Quijote*, informasse também dados biográficos sobre Miguel de Cervantes, uma sinopse do *Don Quijote*, outros livros ou artigos que analisam a obra, obras de arte ou ilustrações que tem por tema o *Don Quijote*, autores que influenciaram e que foram influenciados por Cervantes, a inserção de Cervantes e do *Don Quijote* na história da literatura, etc. Também esse registro e, por conseguinte, a biblioteca que contém esta obra, seriam valorizados se um “link” pudesse ser feito do sítio Web de uma exposição ou de um seminário acadêmico sobre a obra de Cervantes para este registro, permitindo o acesso à cópia digital da obra.

Permitir todo este enriquecimento e valorização dos registros de catálogos, dos acervos e de instituições como arquivos, bibliotecas e museus é a promessa das tecnologias da Web Semântica e dos dados abertos interligados. Estas tecnologias propiciam a publicação de acervos digitais de arquivos, bibliotecas e museus diretamente na Web sem a intermediação e as limitações semânticas dos sistemas gerenciadores de catálogos. Além disso, essas tecnologias, ao viabilizam novos tipos de relações culturalmente significativas (MARCONDES, 2020) que podem ser estabelecidas entre objetos digitais pertencentes a estes acervos, ampliam suas sinergias além de integrarem esses acervos a outros conteúdos disponibilizados na Web, como filmes, lugares, sítios históricos, imagens estáticas e em movimento, enciclopédias, agências de turismo, universidades e centros de pesquisa, monumentos, etc. Acervos digitais publicados segundo as tecnologias LOD ganham novos curadores, a medida que qualquer um pode criar um “link” para um ODP. As tecnologias LOD transformam a Web em uma base de dados que pode ser consultada diretamente usando linguagens como SPARQL.

2. TECNOLOGIAS DA WEB SEMÂNTICA E LOD – “LINKED OPEN DATA” - DADOS ABERTOS INTERLIGADOS

Em um texto de 2001, dirigido a um público amplo e não especialista em computação, Tim Berners-Lee, o cientista da computação criador da Internet e da Web, delinea a visão da Web Semântica. Segundo esta proposta, a Web Semântica seria uma evolução da Web atual na qual os programas “comprenderiam” o conteúdo das páginas e assim poderiam nos auxiliar em tarefas muito mais complexas que simplesmente exibir esse conteúdo. Dados que compõe a Web seriam estruturados de forma a permitir que programas possam interagir de forma generalizada esses dados e uns com os outros, ao contrário da Web atual em que programas são especializados em determinados tipos e formatos de dados.

A Web atual é formada por dados em formato de páginas hipertextuais que, uma vez processados pelos programas navegadores, como Chrome, Internet Explorer, Firefox, etc, se tornam legíveis somente por humanos. essas páginas são interligadas através de hiperlinks que servem para navegar entre uma página e outra, acessíveis através de URLs e frequentemente sujeitos aos erros 404, “Página não encontrada”. A tarefa dos programas navegadores se limita a decodificar o texto da página em HTML exibir seu conteúdo em formato legível por pessoas e acessar outra

página quando um “link” é acionado, permitindo a navegação de página para página.

Na proposta da Web Semântica teríamos uma Web em que qualquer recurso aí disponibilizado teria seu conteúdo, sua utilidade, seu funcionamento, enfim, o que ele é e como pode ser operado, etc., descrito através de metadados, de modo que programas, além de pessoas, pudessem compreendê-lo e operar com ele. Recursos são qualquer “coisa” que exista na forma de dados na Web, tanto “coisas” digitais, que só têm existência na Web, como um documento em .pdf, uma imagem .jpg ou arquivo de áudio .mp3, quanto “coisas” físicas, que têm uma representação na Web, como uma pessoa, representada por sua página no Facebook, ou uma universidade, representada através de sua página institucional, ou um sítio histórico, representado através de um arquivo de imagem.

O modelo de dados da Web Semântica se baseia na ideia simples que recursos podem ser representados através de afirmações descrevendo suas propriedades. Cada afirmação é composta de sujeito (o recurso que está sendo descrito) uma propriedade do recurso e o valor dessa propriedade. Suponhamos que queremos representar uma pessoa P1, de nome Joana Lopes, que estuda na U1, a UFF – Universidade Federal Fluminense, tem 21 anos e mora no local L1, cujo endereço “Rua X, n. y...”. Ela poderia ser representada assim:

- Joana Lopes
- estuda na UFF
- tem 21 anos
- mora na Rua X, número y.

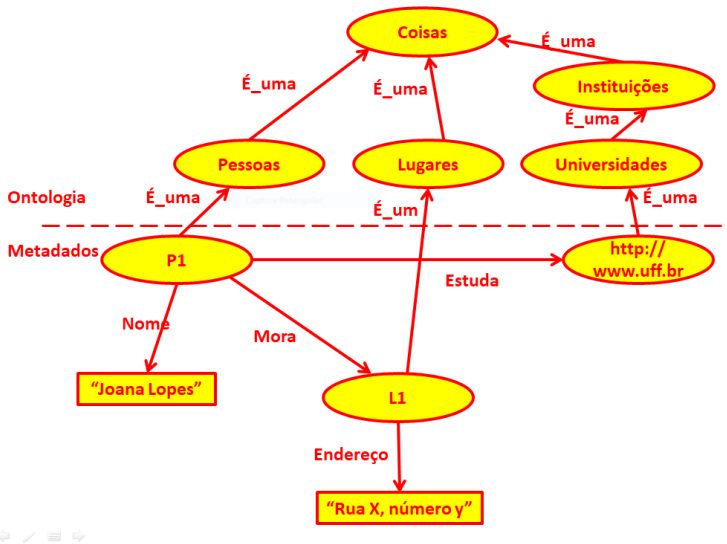
No exemplo o recurso Joana é representado através de afirmações descrevendo três de suas propriedades. Na primeira o sujeito é “Joana Lopes”, a propriedade é “estuda” e o valor dessa propriedade é “UFF”. Propriedades e valores são metadados assinalados ao recurso que está sendo descrito.

Como estamos no ambiente Web, também poderíamos substituir “UFF” pelo “link” para a página da UFF, assim: - estuda na <http://www.uff.br>

uff.br. Dessa maneira estaríamos *interligando* os dados do recurso Joana Lopes com os dados do recurso UFF.

As propriedades e seus valores - ou metadados - utilizadas para descrever os recursos também teriam um significado preciso, sendo referenciadas a ontologias computacionais. Nesse caso P1 é referenciado à uma ontologia como sendo (É_uma) uma Pessoa, L1 como sendo um Lugar e UFF como sendo uma universidade (que, por sua vez, é uma Instituição). Esse modelo de dados seria um grafo, ou rede semântica¹, como na figura a seguir; nele as coisas seriam ovais, os textos seriam retângulos e os “hiperlinks” semânticos seriam setas.

Figura 3 - Modelo de dados da Web Semântica



Fonte: elaborado pelo autor

Interligando um grafo a outro usando padrões tecnológicos e licenças abertos (dados abertos interligados) é formado um grafo global, a Web de dados. Ao estarem representados em formato inteligível por programas e interligados, a Web de dados pode ser consultada como uma base de dados e não somente ser navegável como a Web atual.

¹ Ver https://en.wikipedia.org/wiki/Semantic_network.

A partir deste exemplo intuitivo podemos compreender melhor o funcionamento das tecnologias LOD. Elas se sustentam em três bases.

- RDF - Resource Description Framework – modelo de dados codificado na linguagem XML para descrever recursos de modo que estas descrições sejam “compreensíveis” por programas, além de por pessoas. A descrição de um recurso em RDF se dá sempre através de afirmações de suas propriedades e dos valores dessas propriedades para esse recurso. O modelo de dados do RDF baseia-se em afirmações formadas por três elementos, o sujeito da afirmação, o recurso que está sendo descrito, uma das propriedades desse recurso, e o valor dessa propriedade. A seguir é mostrado um outro exemplo de afirmação que permitirá ilustrar outros aspectos fundamentais das tecnologias LOD:

“a página <http://www.uff.br> (o recurso a ser descrito) tem como autor (a propriedade) Carlos H. Marcondes (o valor da propriedade autor para esse recurso)”.

- uso de identificadores persistentes e válidos globalmente em toda a Web – IRIs ao invés dos URLs - como “links” para acessar recursos e navegar entre eles. Os atuais URLs são meros endereços ou mecanismos para acessar um recurso, não identificadores. Por serem endereços não são identificadores apropriados, pois se o endereço do recurso ou o mecanismo tecnológico para acessá-lo se modificarem, acontecem as “quebras de links” tão comuns na navegação pela Web atual. Numa Web baseada em conteúdos “interligados”, os hiperlinks ou identificadores dos recursos, têm que ter a máxima confiabilidade.

Os IRIs² (anteriormente chamados de URIs³, dos quais são uma generalização, permitindo o uso na sua constituição de caracteres de alfabetos como o grego, cirílico, chinês, coreano, japonês, árabe, ao contrário dos URIs que só permitem o uso de caracteres do alfabeto latino ocidental) são identificadores persistentes de recursos, que não mudam ao sabor da reorganização interna das pastas e nomes de um servidor, nem de mudanças na tecnologia e nos mecanismos para acessar um recurso. Os

² Ver https://en.wikipedia.org/wiki/Internationalized_Resource_Identifier.

³ Ver https://en.wikipedia.org/wiki/Uniform_Resource_Identifier.

IRIs são identificadores como os CPF pessoais, ou os ISBN, permanentes e imutáveis. Como não são um endereço nem um mecanismo de acesso, têm que, de alguma maneira, serem convertidos em endereços ou mecanismos de acesso para que o recurso possa ser acessado. IRIs (como também outros tipos de identificadores permanentes, como DOIs⁴, handles⁵, etc.) dependem de serviços de resolução, responsáveis por realizarem esta conversão, chamada de dereferenciação, os IRIs são dereferenciados em URLs para acesso ao recurso.

- Uso de vocabulários padronizados, reconhecidos e avaliados por comunidades de usuários em domínios específicos. Como o modelo de dados do RDF só possui como pressupostos semânticos a existência de três elementos distintos, o recurso, a propriedade e o valor da propriedade, propriedades e valores podem ser substituídos por IRIs referenciando vocabulários que especificam exatamente e inequivocamente o significado e restrições de uso de uma propriedade ou o valor dessa propriedade.

Na declaração RDF anteriormente citada, a propriedade autor poderia ser substituída por uma IRI referenciando um vocabulário específico. O exemplo ficaria assim:

```
<http://www.uff.br> <http://purl.org/dc/elements/1.1/creator> "Carlos H. Marcondes".
```

Neste caso o vocabulário é o Dublin Core Metadata Elements Set, um vocabulário de elementos que servem para descrever documentos digitais. Este vocabulário tem um elemento, "dc:creator", onde é especificado e definido o que é o autor de um documento.

Na mesma declaração RDF anterior, o valor da propriedade dc:creator também poderia ser substituído, de uma cadeia de caracteres, "Carlos H. Marcondes", por um "link" para a página do Prof. Marcondes na UFF, assim.

```
<http://www.uff.br> <http://purl.org/dc/elements/1.1/creator> <http://www.professores.uff.br/mrcondes>
```

⁴ Ver <http://www.doi.org/>.

⁵ Ver <https://www.handle.net/>.

Outra questão importante, o que são dados *abertos*? Tim Berners-Lee, ao propor uma Web de dados abertos interligados (BERNERS-LEE, 2006) sugeriu diferentes graus de abertura dos dados, por ele chamados de esquema das cinco estrelas. O esquema proposto é hoje um critério de qualidade do grau de abertura dos dados de um recurso. É descrito a seguir.

- 1 estrela: “torne seus recursos disponíveis na Web (tanto faz o formato) sob uma licença aberta” (vermos as licenças abertas na seção 2.3).
- 2 estrelas – “torne seus recursos disponíveis como dados estruturados (ex. excel no lugar de imagem escaneada”.
- 3 estrelas – “utilize formatos não-proprietários (ex. CSV e não excel)”.
- 4 estrelas – “utilize URIs para identificar recursos. Isso vai ajudar as pessoas a apontarem para eles”.
- 5 estrelas – “conecte seus dados com dados de outras pessoas para prover contexto (dados linkados)”⁶.

É essa combinação de atributos que tornam os dados realmente abertos, encontráveis e reutilizáveis. Em algumas áreas como pesquisas científicas, a questão do reuso dos dados de pesquisa é bastante sensível, tem um peso econômico e vem ganhando cada vez mais relevância, sendo incorporada a políticas dos órgãos de fomento em todo o mundo (NATIONAL INSTITUTE OF HEALTH, 2015; NATIONAL SCIENCE FOUNDATION, 2015?; OEC, 2007).

Nas seções seguintes as bases tecnológicas dos LOD são detalhadas, exceto no que diz respeito aos vocabulários, que são contemplados em um capítulo especial, o capítulo 3.

2.1. REPRESENTAÇÃO DE CONTEÚDOS: RDF

RDF é um modelo de dados para representar recursos e suas propriedades em formato legível por programas. Um recurso é qualquer

⁶ A explicação para as 5 estrelas foram tiradas de <http://5stardata.info/pt-BR/>.

coisa disponibilizada na Web, acessível através de um IRI. Existem recursos que só existem no ambiente Web, como um artigo científico em formato .pdf, um arquivo .mp3 com a 5ª. Sinfonia de Beethoven, um vídeo qualquer em um arquivo .mp4. Existem também recursos que são representações digitais de coisas que existem no mundo real, como a página do facebook de Joana, uma fotografia digital da Porta de Alcalá em Madrid ou o registro no catálogo Web do Museu Britânico da Pedra da Rosseta; essas representações digitais de coisas físicas, pessoas, objetos, monumentos, etc., constituem a chamada “web das coisas”⁷.

RDF usa uma construção específica na linguagem XML para representar um recurso. Dizer que RDF é um modelo de dados significa que, ao contrário de um documento em XML comum, um documento RDF segue uma sintaxe específica onde estão claramente especificados o sujeito, o predicado e o objeto de uma afirmação; afirmações são a base do modelo de dados RDF. Como foi visto anteriormente no início deste capítulo, um recurso é descrito (representado) através de uma ou mais afirmações sobre o mesmo. Vejamos no exemplo a seguir, na notação RDF/XML.

```
<?xml version="1.0">
<rdf:RDF
  <rdf:Description rdf:about "http://www.w3.org/
  DesignIssues/LinkedData">
    <autor>BERNERS-LEE, T.</autor>
    <ano>2006</ano>
  </rdf:Description>
</rdf:RDF>
```

Este é um documento em XML. Só que segue uma sintaxe específica. Existe uma marcação especial, **rdf:Description rdf:about** que contém o IRI do recurso que está sendo descrito, neste caso a página “**http://**

⁷ Ver <https://www.w3.org/WoT/>.

www.w3.org/DesignIssues/LinkedData". À declaração do recurso que está sendo descrito seguem-se as declarações de duas de suas propriedades, que são o autor do recurso e o ano de sua publicação, constituindo assim duas afirmações sobre o (mesmo) recurso.

Um programa que acessasse este documento RDF “compreenderia” minimamente que trata-se de um recurso, dado por seu IRI, com duas propriedades (que ele não “compreenderia”, mas saberia, pelo menos, que tratam-se de propriedades desse recurso), autor e ano. Mesmo com esta compreensão limitada e sem compreender o que significam autor e ano, este programa poderia responder a questões como 1- o que ou quem é o autor do recurso dado pela IRI? ou 2- o que é 2006? Estas respostas poderiam ser dadas simplesmente pelo sintaxe recurso-propriedade-valor de um documento RDF. Estas questões poderiam ser formuladas esquematicamente da seguinte maneira.

<recurso> <propriedade> <valor>

- <IRI> <autor> <? >, e a resposta seria “BERNERS-LEE, T.”;
- <IRI> <? > “2006”, e a resposta seria <ano>.

Vamos desenvolver agora uma segunda versão do mesmo documento RDF, só que agora introduziremos um elemento que vai agregar semântica compreensível por computadores ao documento RDF; trata-se do “namespaces” ou espaço de nomes. Um espaço de nomes indica, através de um IRI, um vocabulário público, aberto, disponível na Web, onde estão especificados elementos com significado específico, que serão usados como propriedades e valores de propriedades em um documento RDF. Vejamos a segunda versão do exemplo.

```
<?xml version="1.0">
```

```
<rdf:RDF
```

```
  xmlns:rdf=" http://www.w3.org/1999/02/22-rdf-syntax-ns"
```

```
  xmlns:dc="http://purl.org/dc/elements/1.1">
```



```
<rdf:Description rdf:about "http://www.w3.org/DesignIssues/
LinkedData">
  <rdf:type>document</rdf:type>
  <dc:creator>BERNERS-LEE, T.</dc: creator>
  <dc:date>2006</dc:date>
  </rdf:Description>
</rdf:RDF>
```

Nesta segunda versão há a declaração de dois espaços de nomes - “xmlns”, XML namespace -, indicando dois vocabulários, o primeiro que indica a sintaxe de um documento RDF (xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns) e o segundo, o vocabulário Dublin Core – DC - (xmlns:dc="http://purl.org/dc/elements/1.1"); ambos os vocabulários são indicados através de suas respectivas IRIs. O primeiro vocabulário serve para especificar a propriedade rdf:type, e o valor dessa propriedade, “document”; ambos, a propriedade e seu valor, o que significam type e document, são definidos claramente nos vocabulários RDF syntax e Dublin Core.

Sobre o vocabulário DC, voltaremos a discutir em maior profundidade adiante. Nesta segunda versão do exemplo o DC serve para especificar a semântica específica das propriedades do recurso dc:creator, o criador, o autor, de um documento, e dc:date, a data de publicação de um documento. O desenvolvedor de um programa que irá tratar de este documento RDF que utiliza o vocabulário DC poderá consultar a especificação dos elementos dc:creator e dc:date disponíveis na IRI <http://purl.org/dc/elements/1.1> e construir o programa para que trate esses conteúdos exatamente de acordo com sua semântica (padronizada, universal, consensada) conforme especificada nesta IRI.

Em um documento RDF <recurso> <propriedade> <valor> formam o que é chamado de uma tripla. Triplas são uma das formas mais comuns de representar documentos RDF. A segunda versão do documento do nosso exemplo poderia ser representada através das seguintes três triplas, na notação N-Triples:

```
<http://www.w3.org/DesignIssues/LinkedData> <http://www.w3.org/1999/02/22-rdf-syntax-ns/type> <"document">  
<http://www.w3.org/DesignIssues/LinkedData>  
  <http://purl.org/dc/elements/1.1/creator> <"BERNERS-LEE, T.">  
<http://www.w3.org/DesignIssues/LinkedData>  
  <http://purl.org/dc/elements/1.1/date> <"2006">
```

A representação em triplas favorece o armazenamento de documento RDF em bancos de dados que podem ser facilmente consultados, respondendo questões como as mostradas anteriormente. Para responder a questões em conjuntos de dados representados como RDF, em triplas ou não, existe uma linguagem de consulta, a SPARQL, que será vista adiante na seção 2.5.

Além das notações RDF/XML e N-triples, existe também a notação TURTLE – Terse RDF Triple Language⁸, uma representação textual de um documento RDF. Nesta notação o nosso exemplo ficaria assim:

```
PREFIX xmlns:rdf= http://www.w3.org/1999/02/22-rdf-syntax-ns  
PREFIX xmlns:dc=http://purl.org/dc/elements/1.1  
<http://www.w3.org/DesignIssues/LinkedData>  
  rdf:type "document";  
  dc:creator "BERNERS-LEE, T.";  
  dc:date "2006".
```

Outra questão a respeito de vocabulários semânticos indicados através de IRIs diz respeito a propriedades e valores que têm nomes idênticos. Como tratar propriedades e valores que têm nomes idênticos?

⁸ Ver <http://www.w3.org/TR/turtle/>.

Segundo estudo realizado por Koehler (2002), em uma amostra aleatória de páginas Web avaliadas por um período de quatro anos (dezembro de 1996 a fevereiro de 2001), somente cerca de 34% dos URL's permaneciam válidos e ativos. Um estudo mais recente mostra que o problema persiste (KLEIN et al., 2014). Esses últimos autores, embora se refiram à informação em ciência, tecnologia e medicina, mencionam um problema adicional, a perda de contexto, por eles chamada de “context drift”, bastante importante para a compreensão de qualquer conteúdo.

Isso acontece porque os URL, uma tecnologia dos primórdios da Web, são o endereço de uma página em um servidor. Qualquer mudança na estrutura de pastas ou de nomes de pastas desse servidor faz com que um URL se altere. Se esse URL havia sido guardado previamente em um “bookmark” ou “favoritos” e for acionada, ocorrerá erro 404. Usar um endereço em um servidor como localizador se mostrou bastante frágil. Posteriormente os URL foram usados também para acionarem programas a partir de páginas Web. A seguir estas questões são sistematizadas.

- URL como endereço de uma página:

https://elpais.com/ccaa/2017/10/16/catalunya/1508134587_768684.html

- URL para acionar programas e passar parâmetros para esses programas:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000100010&lng=en&nrm=iso&tlng=pt

Como já vimos, no primeiro caso qualquer alteração nos nomes ou estrutura de pastas no servidor vai alterar o URL provocando erro 404; no segundo caso, qualquer alteração na tecnologia, como por exemplo, alteração do programa que está sendo acionado ou nos parâmetros necessários ao seu acionamento, também provocará o erro 404.

Os problemas dos URLs como mecanismo para acesso aos recursos logo foram percebidos; ao longo do tempo surgiram várias alternativas, que passaremos a examinar.

Era necessário que os recursos na Web fossem acessados através de identificadores que não se alterassem nunca, assim como o CPF identifica permanentemente um cidadão, ou o ISBN identifica permanentemente um livro, ou o ISSN identifica permanentemente um periódico. Com o ISBN e o ISSN a área de documentação e informação já havia identificado a necessidade de identificadores persistentes. Com os recursos digitais existe a mesma necessidade. O uso de identificadores persistentes também está ligada a questão da preservação a longo prazo dos recursos digitais (DAVIDSON, 2006) disponibilizados na Web; estes devem ser acessíveis hoje ou daqui a duzentos anos, sem qualquer problema de quebra de “links”. Essa questão está naturalmente ligada ao compromisso das instituições de memória e cultura em proverem identificadores persistentes aos recursos que disponibilizam na Web e garantirem sua persistência, uma das atividades de curadoria digital mais importantes que essas instituições passam a assumir na gestão de seus acervos digitais. Várias recomendações enfatizam esse compromisso, dirigidas a museus (ICOM/CIDOC, 2013?), arquivos (NATIONAL ARCHIVES AND RECORDS ADMINISTRATION, 2004) e bibliotecas, (LARSON, 2005); NATIONAL LIBRARY OF AUSTRALIA, 2001).

O consenso geralmente aceito¹⁰, (MCKENNA; WYNS, 2010; KLEIN et al., 2014;; TONKIN, 2008) é que identificadores persistentes devem atender aos seguintes *requisitos*.

- identificadores persistentes devem ser únicos em todo o espaço da Web;
- identificadores persistentes devem ser válidos indefinidamente, não podem ser alterados ao longo do tempo;
- identificadores persistentes devem permitir acesso a recursos de forma independente de sua localização física em um servidor ou repositório; a localização física de um recurso poderá mudar sem que isto acarrete alteração no seu identificador persistente ou impeça o acesso direto ao recurso.

¹⁰ Ver <http://www.ands.org.au/guides/persistent-identifiers-expert>

Tim Berners-Lee discute e reforça esses requisitos em texto famoso (BERNERS-LEE, 1998), dedicado ao URI, com título “Cool URIs don’t change”. O URI – Uniform Resource Identifier – é identificador persistente usado com as tecnologias LOD. Uma versão generalizado do URI é o IRI – Internationalized Resource Identifier. Enquanto o URI usa somente os caracteres do alfabeto latino (correspondente ao código de caracteres ASCII¹¹), o IRI usa caracteres de diferentes alfabetos (correspondente ao código de caracteres UNICODE¹²).

Assim como o URL é acionado pelo protocolo HTTP, para acessar um recurso (como no exemplo, https://elpais.com/ccaa/2017/10/16/catalunya/1508134587_768684.html), os IRIs/URIs funcionam da mesma maneira, são HTTP IRIs/URIs.

Na Web das coisas os IRIs identificaram, como já foi visto, tanto coisas físicas quantos recursos nativos digitais. O protocolo HTTP permite que os programas navegadores¹³ negociem com os servidores que hospedam os recursos sobre o melhor formato segundo o qual um recurso deve ser enviado para um programa, seja ele um programa navegador que precisa receber os dados do recurso em um formato legível por pessoas, seja ele um programa agente que precisa receber os dados do recurso em formato RDF. Assim, um IRI pode identificar um recurso, sua representação em formato legível por pessoas e também sua representação em formato legível por programas. Um servidor que abrigasse um recurso dado pelo seu IRI:

- [http://dbpedia.org/resource/Rio_de_Janeiro_\(cidade\)](http://dbpedia.org/resource/Rio_de_Janeiro_(cidade)), poderia responder a uma solicitação para acessá-lo, de duas maneiras:

- dado o IRI <http://dbpedia.org/page/Corcovado>, solicitado por um programa navegador (que tem um usuário humano por traz), o servidor responderia com um conteúdo legível por pessoas, por exemplo, uma página em HTML, [http://dbpedia.org/resource/Rio_de_Janeiro_\(cidade\).html](http://dbpedia.org/resource/Rio_de_Janeiro_(cidade).html);

- dado o mesmo IRI, <http://dbpedia.org/data/Corcovado>, mas agora solicitado por um programa agente, que quer “compreender”

¹¹ Ver <https://en.wikipedia.org/wiki/ASCII>.

¹² Ver <https://en.wikipedia.org/wiki/Unicode>

¹³ Programas navegadores são o Internet Explorer, o Google Chrome, o Mozilla Firefox, o Opera, etc.

e processar o conteúdo do recurso, o servidor responderia [http://dbpedia.org/resource/Rio_de_Janeiro_\(cidade\).rdf](http://dbpedia.org/resource/Rio_de_Janeiro_(cidade).rdf).

A interação entre o servidor que abriga o recurso e os programas que querem acessá-lo é chamada “content negotiation” e faz parte do protocolo HTTP¹⁴. O envio dos dados apropriados ao programa que o solicita (.html para humanos, .rdf para programas agentes) através do seu IRI é denominado de *resolução* ou dereferenciação do IRI. Um IRI é resolvido localmente pelo servidor que abriga o recurso, usando o serviço DNS – Domain Name System¹⁵ um dos protocolos que fazem parte do protocolo da Internet, o TCP/IP¹⁶.

Os casos descritos acima não são os únicos. O servidor que abriga o recurso, juntamente com o DNS local, se encarregam de enviar ao programa cliente uma versão *adequada* do recurso, ou mesmo, outros meios de acessá-lo; por exemplo, o IRI da propriedade `edm:hasMet`, <http://www.europeana.eu/schemas/edm/hasMet>, é resolvido acessando uma página contendo toda a documentação do EDM; faça você mesmo uma experiência, copiando esse IRI e colando na janela de um programa navegador.

Existem outras formas de resolver identificadores persistentes que não diretamente pelo DNS servidor que abriga o recurso. Outros tipos de identificadores persistentes são resolvidos por sistemas centrais de resolução. Identificadores persistentes que são resolvidos dessa maneira são o DOI e o Handle.

O DOI – digital object identifier¹⁷ – é o identificador persistente de um consórcio de grandes editores de periódicos científicos. No funcionamento do DOI um editor recebe um conjunto de DOIs e, a cada artigo científico que publica, um DOI é assinalado. Um DOI tem a forma de um conjunto de números, assim:

- <http://dx.doi.org/10.1590/S1519-69842009000200030>.

¹⁴ Ver <http://www.ietf.org/rfc/rfc2616.txt>.

¹⁵ Ver https://en.wikipedia.org/wiki/Domain_Name_System.

¹⁶ Ver https://en.wikipedia.org/wiki/Internet_protocol_suite.

¹⁷ Ver <http://www.doi.org/>.

A primeira parte do DOI até a primeira “/” serve para acionar o serviço central de resolução do consórcio DOI. A segunda parte, até a segunda “/” identifica o editor e sua publicação. A última parte é o identificador do artigo.

Quando um “link” com um DOI, como o do exemplo, é acionado ou digitado na janela de um programa navegador, o serviço central de resolução do consórcio DOI é acionado. Este serviço mantém uma base de dados que permite associar cada DOI ao URL do recurso. O URL é então devolvido ao programa navegador para que este possa acessar o recurso. Esse procedimento permite que um editor altere as pastas e/ou seus nomes no servidor que abriga seus artigos quando queira; uma vez que o DOI de um artigo não foi alterado, o editor só terá que atualizar o novo URL na base de dados do serviço central de resolução do consórcio DOI. O Handle¹⁸ funciona da maneira semelhante.

Outros sistemas de identificadores persistentes são o ARK¹⁹ – Archival Resource Key – para a identificação de objetos digitais arquivísticos, o PURL – Persistent Uniform Resource Locator²⁰ – o ORCID²¹ – Open Researcher and Contributor ID – para a identificador de pesquisadores, para ser usado em citações bibliográficas e evitar formas diferentes de transcrever o nome do pesquisador, e o ISNI – International Standard Name Identifier, para identificar de artistas, produtores, cantores, etc.

2.3. LICENÇAS DE USO

Dados, para realmente abertos, têm que poder ser reusados. As licenças de uso jogam um papel fundamental para tornar dados realmente abertos e reusáveis. A Web, desde sua criação sempre foi um espaço aberto. As tecnologias de informação por seu lado também facilitam a cópia e redistribuição de diferentes recursos. Estas facilidades tornaram necessário que o reuso e compartilhamento desses recursos fosse minimamente disciplinado. As licenças padronizadas surgiram com esta finalidade.

¹⁸ Ver <https://www.handle.net/>.

¹⁹ Ver <https://confluence.ucop.edu/display/Curation/ARK>.

²⁰ Ver <http://bibpurl.oclc.org/>.

²¹ Ver <https://orcid.org/>.

Licenças padronizadas de uso são declarações, que foram amplamente discutidas e consensadas entre a comunidade da Web, que estabelecem grãos de possibilidades de reuso dos recursos disponibilizados; essas possibilidades são geralmente menos restritivas que os tradicionais “copyrights”. Em 2001 foi surgida a Creative Commons, uma organização não governamental criada com a finalidade de estabelecer as licenças padronizadas, hoje conhecidas como licenças Creative Commons. Estas licenças estabelecem as possibilidades de reuso, distribuição e compartilhamento de obras criativas – textos, música, fotos, vídeos, etc., que estejam disponíveis na Web. Um determinado recurso é disponibilizado segundo um dos tipos de licenças Creative Commons; este tipo de licença fica claramente definido ao ser acessado o recurso, geralmente através de um logotipo padronizado e/ou de um “link” para o texto da licença. Qualquer um que queira reusar, distribuir, compartilhar, alterar ou reformatar esse recurso deverá fazê-lo obedecendo aos termos da licença Creative Commons sob a qual o recurso é licenciado. Uma forma fácil de compreender esta questão é visitar a página da Wikipedia sobre as licenças Creative Commons, https://pt.wikipedia.org/wiki/Licenças_Creative_Commons. Essa entrada da Wikipedia, aliás como todas as outras, está liberada segundo a licença “Attribution-ShareAlike 3.0 Unported”²², um dos tipos de licença Creative Commons. Um usuário que deseje reusar, compartilhar ou distribuir o conteúdo desta entrada da Wikipedia tem o direito de: Compartilhar — copiar e redistribuir o material em qualquer suporte ou formato; Adaptar — remixar, transformar, e criar a partir do material para qualquer fim, mesmo que comercial. Pode fazer isso desde que atenda às seguintes restrições: Atribuição — deve ser dado o crédito apropriado, prover um “link” para a licença e indicar se mudanças foram feitas. Um usuário deve fazê-lo em qualquer circunstância razoável, mas de maneira alguma que sugira ao licenciante a apoiá-lo ou o uso que esse usuário está fazendo desse recurso; Compartilhar com igual licença — se o recurso for remixado, transformado, ou seja, se for feita outra criação a partir do material original, esta nova criação deve ser distribuída sob a mesma licença que o original.





As licenças Creative Commons são apresentadas em três versões: como um instrumento legal convencional, em linguagem técnica de

²² Ver https://pt.wikipedia.org/wiki/Wikipédia:Texto_da_licença_Creative_Commons_Attribution-ShareAlike_3.0_Unported.

Direito, como um texto em linguagem simplificada, para ser compreensível por leigos, por você ou por mim, e em um formato legível e identificável por programas, usando RDF²³.

Os diferentes tipos de licenças Creative Commons trabalham combinando de um a três dos quatro seguintes critérios: 1- *atribuição de créditos ao autor da obra original*, 2- *compartilhamento da obra original, sem alterar a licença original*, 3- *uso não comercial*, e 4- *permissão de criar obras derivadas ou remixar, isto é alterar, expandir, reescrever a obra original*. Esses critérios estão explicados no Quadro 1 a seguir.

Quadro 1 – Licenças Creative Commons

Ícone	Licença	Descrição
	Atribuição (BY)	Um usuário tem o direito de copiar, distribuir, exibir e executar a obra e fazer trabalhos derivados dela, conquanto que atribua créditos devidos ao autor ou licenciador, na maneira especificada por estes.
	Compartilha Igual (SA)	Um usuário devem distribuir obras derivadas somente sob uma licença idêntica àquela da obra original.
	Não Comercial (NC)	Um usuário pode copiar, distribuir, exibir e executar a obra e fazer trabalhos derivados dela, desde que seja para fins não-comerciais
	Sem Derivações (ND)	Um usuário pode copiar, distribuir, exibir e executar somente cópias idênticas da obra, não podendo criar derivações da mesma.








Fonte: elaborado pelo autor

Quando em um dos tipos de licença, algum dos critérios não é mencionado, isso significa que não há restrições relativas a esse critério; por exemplo, quando o critério (NC) não é mencionado na licença, significa que a obra pode ser distribuída inclusive com fins comerciais; quando o critério (SA) não é mencionado na licença, significa que a obra pode ser distribuída com licenças diferentes da licença da obra original; quando o critério (ND) não é mencionado na licença, significa que a obra pode ser alterada e remixada.

²³ Ver https://wiki.creativecommons.org/wiki/Pt:CC_REL.

Um resumo das licenças Creative Commons, combinando um ou até três dos quatro critérios, é mostrado Quadro 2 a seguir.

Quadro 2 – “Links” para as licenças Creative Commons

Logotipo da licença	Descrição	Sigla	“Link” para a licença
	Recurso liberado globalmente sem restrições	CC0	
	Atribuição de crédito ao autor da obra original	BY	https://creativecommons.org/licenses/by/4.0/legalcode
	Atribuição + Compartilha Igual	BY-SA	https://creativecommons.org/licenses/by-sa/4.0/legalcode
	Atribuição + Não Comercial	BY-NC	https://creativecommons.org/licenses/by-nc/4.0/legalcode
	Atribuição + Sem Derivações	BY-ND	https://creativecommons.org/licenses/by-nd/4.0/legalcode
	Atribuição + Não Comercial + Compartilha Igual	BY-ND-SA	https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode
	Atribuição + Não Comercial + Sem Derivações	BY-ND-NC	https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode

Fonte: elaborado pelo autor

Ao publicar coleções digitais na Web segundo as tecnologias LOD, especial atenção deve ser tomada com relação às licenças a serem atribuídas aos objetos digitais. Geralmente quando se trata de obras raras ou manuscritos, estas estão em domínio público e a instituição pode decidir sob que tipo de licença publicará as cópias digitais das obras; a princípio, a instituição que permitiu a cópia digital é a detentora dos direitos de licenciamento da cópia digital. Com documentos arquivísticos de valor históricos acontece o mesmo. Cuidado deve ser tomado com documentos de arquivos pessoais, em especial quando são doados pelas famílias; esses podem ter restrições especiais quanto à sua publicação. Nas fotografias de objetos museológicos, em princípio, o direito autoral é do fotógrafo; o museu deve se precaver a respeito, fazendo um contrato com o fotógrafo

que garante ao museu os direitos irrestritos sobre as fotografias das obras. Obras de arte podem ainda ter restrições especiais quanto à sua reprodução, que se adicionam às do fotógrafo que criou sua cópia digital.

Estas questões devem ser cuidadosamente equacionadas em todo projeto, anteriormente a qualquer publicação de acervos na Web. Um texto bastante abrangente, desde um ponto de vista jurídico, que sistematiza essas questões, é o relatório “Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment body” (BESEK, 2003), feito para o Council of Library and Information Resources, dos EUA.

2.4. CONSULTA A BASES DE DADOS: LINGUAGEM/PROTOCOLO SPARQL

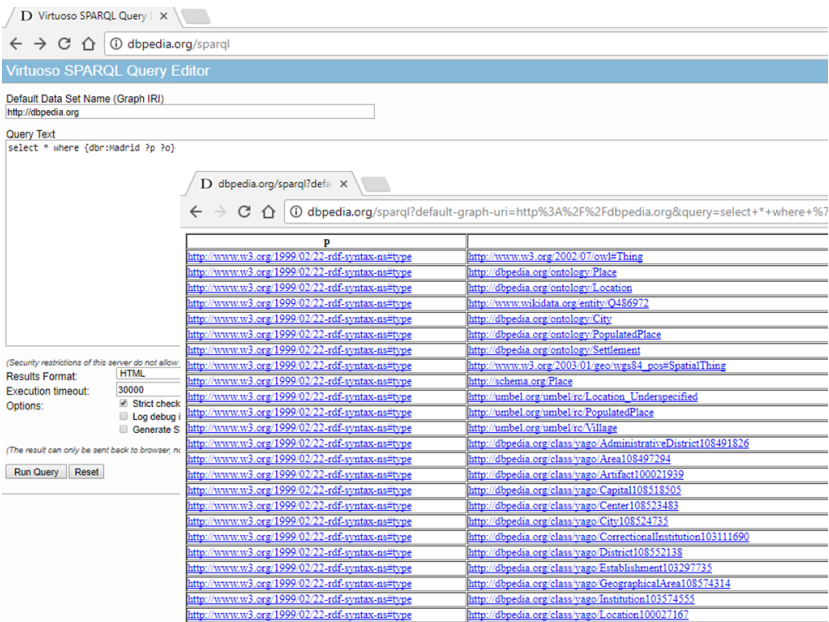
Conjuntos de triplas RDF podem ser disponibilizadas de duas maneiras, ou como arquivos (“dumps” de arquivos .rdf) ou armazenadas em ferramentas como programas gerenciadores de triplas, “triplestore” como o Openlink Virtuoso. Essa e outras ferramentas serão vistas com mais detalhes na seção 4.6.

SPARQL - SPARQL Protocol and RDF Query Language – é uma linguagem de consulta a dados em RDF, uma linguagem de consulta semântica. Um gerenciador de triplas RDF pode ser acessada na Web através de um “link” chamado de “end point”. Nesse “end point” podem ser feitas consultas ao conjunto de triplas através da linguagem SPARQL. SPARQL é tanto uma linguagem de consulta a conjuntos de triplas quanto um protocolo que permite a um usuário em um navegador cliente submeter consultas a vários “end points” hospedados em servidores distintos simultaneamente.

Toda tripla RDF é estruturada em <sujeito>, <predicado> e <objeto>. SPARQL opera restringindo o padrão das triplas. Assim, a consulta “<Madrid>, <? >, <? >” retornará todas as triplas disponíveis na base do “end point” especificado nas quais Madrid é o sujeito; um exemplo de duas dessas triplas seriam <Madrid>, <type>, <City> e <Madrid>, <populationTotal>, <3141991>, ou seja, Madrid é um recurso do tipo cidade e Madrid tem uma população total de 3141991 habitantes.

O “end point” SPARQL da DBpedia²⁴ permite um modo amigável de fazer consultas SPARQL, somente variando o padrão das triplas <sujeito>, <predicado> e <objeto>. Bastaria entrar no formulário de consulta com o padrão das triplas que se deseja recuperar. Veja na figura seguinte como poderia ser formulada a consulta anterior; nela, “Madrid” é o sujeito da tripla. Na figura também são mostrados alguns resultados.

Figura 4. Consulta em SPARQL no “end point” da DBpedia

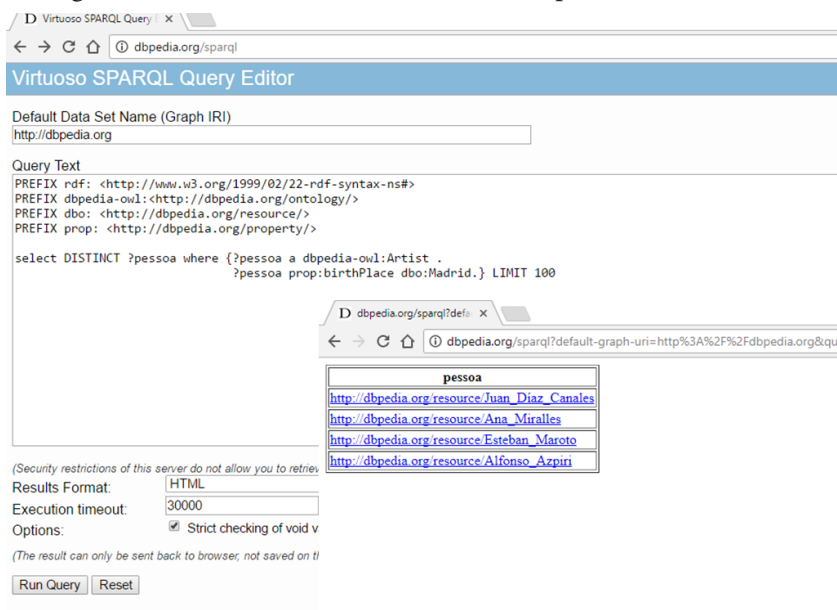


Fonte: elaborado pelo autor

Na próxima figura vemos uma outra consulta em feita ao “end point” da DBpedia, sobre triplas com duas restrições: 1- pessoas que tenham nascido em “Madrid”, e dentre elas 2- pessoas que sejam “artistas”. Nesta consulta a sintaxe completa SPARQL é mostrada; na mesma figura são mostrados os resultados.

²⁴ Ver em <http://dbpedia.org/sparql>.

Figura 5. Consulta em SPARQL ao “end point” da DBPedia.



Fonte: elaborado pelo autor

A sintaxe completa da consulta em SPARQL do exemplo anterior é a seguinte:

- **PREFIX declaration, para IRIs/URIs dos prefixos a serem usados como abreviatura dos vocabulários:**

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>

PREFIX dbo: <http://dbpedia.org/resource/>

PREFIX prop: <http://dbpedia.org/property/>

- **DATASET definition, para especificar o grafo ou “end point” que será consultado:**

FROM http://dbpedia.org

- **RESULT clause, para especificar a informação que será devolvida pela consulta:**

```
select DISTINCT ?pessoa
```

- **QUERY PATTERN, para especificar o padrão das triplas que serão devolvidas:**

```
where {?pessoa a dbpedia-owl:Artist . ?pessoa prop:birthPlace  
dbo:Madrid.}
```

- **QUERY MODIFIERS, para especificar ordenamentos, limites, etc.:**

```
LIMIT 100.
```

Os resultados também poderiam ser apresentados em RDF/XML, como se segue:

```
<rdf:RDF xmlns:res=http://www.w3.org/2005/sparql-results#  
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">  
<rdf:Description rdf:nodeID="rset">  
<rdf:type rdf:resource="http://www.w3.org/2005/sparql-  
results#ResultSet" />  
<res:resultVariable>pessoa</res:resultVariable>  
<res:solution rdf:nodeID="r0">  
<res:binding rdf:nodeID="r0c0">  
<res:variable>pessoa</res:variable>  
<res:value rdf:resource="http://dbpedia.org/resource/Juan_Díaz_  
Canales"/>  
</res:binding>  
</res:solution>  
<res:solution rdf:nodeID="r1">  
<res:binding rdf:nodeID="r1c0">  
<res:variable>pessoa</res:variable>
```

```
<res:value rdf:resource="http://dbpedia.org/resource/Ana_
Miralles"/></res:binding>
</res:solution>
<res:solution rdf:nodeID="r2">
<res:binding rdf:nodeID="r2c0">
<res:variable>peessoa</res:variable>
<res:value rdf:resource="http://dbpedia.org/resource/Esteban_
Maroto"/>
</res:binding>
</res:solution>
<res:solution rdf:nodeID="r3">
<res:binding rdf:nodeID="r3c0">
<res:variable>peessoa</res:variable>
<res:value rdf:resource="http://dbpedia.org/resource/Alfonso_
Azpiri"/>
</res:binding>
</res:solution>
</rdf:Description>
</rdf:RDF>.
```

Existem várias fontes²⁵ e materiais de referência (XQUERY/ SPARQL TUTORIAL, 2017) que permitem aprofundar o conteúdo desta breve introdução à SPARQL.

Embora permita consultas complexas, com diversos critérios simultaneamente, SPARQL não é uma linguagem simples para o usuário final. A interface SPARQL da DBpedia permite consultas simplificadas, somente variando o padrão ?s, ?p, ?o, como mostrado na Figura 4. Existem

²⁵ Ver <https://www.w3.org/2009/Talks/0615-qbe/>, <https://www.cambridgesemantics.com/blog/semantic-university/learn-sparql/sparql-by-example/>, <http://www.linkeddatatools.com/querying-semantic-data>.

vários desenvolvimentos que facilitam ao usuário final consultar bases de dados em RDF²⁶, (DIAZ; ARENAS; BENEDIKT, 2016).

2.5. PROJETOS DE DADOS ABERTOS INTERLIGADOS EM ARQUIVOS, BIBLIOTECAS E MUSEUS

Nesta seção estão relacionados projetos importantes que utilizaram as tecnologias LOD para publicação de seus dados. Os projetos declaram como objetivos aumentar a visibilidade, o aproveitamento e reuso dos dados desses acervos, além de integrar esses acervos com outros também disponíveis na Linked Data Cloud.

- O projeto Linking Lives do portal britânico Archives Hub oferece um serviço de dados abertos, o Archives Hub Linked Data²⁷, incluindo um “end point” SPARQL²⁸ e a possibilidade de fazer “download” de conjuntos de dados. O projeto tem como objetivos: “**Linking Lives** is exploring ways to present Linked Data. We aim to show that archives can benefit from being presented as a part of the diverse data sources on the Web to create full biographical pictures, enabling researchers to make connections between people and events”.

- O Alexandria Archives Institute, uma organização sem fins lucrativos, desenvolve, desde 2012, o projeto “Exploring Biogeography of Early Domestic Animals using Linked Open Data”²⁹, com dados arqueológicos.

- Park (2015) apresenta o projeto de LOD dos Arquivos Nacionais da Coreia.

- O projeto de dados abertos da Bibliothèque nationale de France³⁰ se torna público em 2011. Em junho de 2016, disponibilizava 8.000.000 de documentos interligados a cerca de 2.000.000 registros de autoridades³¹.

²⁶ Ver <http://uispin.org/ui.html>.

²⁷ Ver em <http://data.archiveshub.ac.uk/>.

²⁸ Ver em <http://data.archiveshub.ac.uk/sparql>.

²⁹ Ver <https://alexandriarchive.org/linked-data/>

³⁰ Ver <http://data.bnf.fr/>.

³¹ Ver <http://data.bnf.fr/about>.

Os objetivos do projeto são:

“make the data produced by the BnF more visible on the Web, federate the data produced by the BnF, both within and outside the catalogues, contribute to collaboration and metadata exchange by creating links between structured and trustable resources, facilitate reuse of metadata (under Open License) by third parties. The data.bnf.fr project endeavours to make the data produced by Bibliothèque nationale de France (French National Library) more useful on the Web”.

Os dados publicados seguem a “Licence ouverte française”, usada pelo governo francês. Essa licença é semelhante à Creative Commons, adaptada à legislação francesa de copyright.

- A Biblioteca Nacional de España, através do portal datos.bn.es³², publica seus catálogos como dados abertos. O portal e as tecnologias usadas foram desenvolvidos em conjunto com a universidade Politécnica de Madrid. O primeiro conjunto de dados foi disponibilizado em 2011; atualmente estão publicados dados de 3.000.000 de obras e 1.500.000 autoridades, organizados em 500 diferentes temas. Os dados são publicados segundo a licença Creative Commons CC0, sem restrições de uso (vista na seção 2.3). Os dados podem exibidos em um navegador convencional e também serem baixados em formato RDF/Turtle.

- A British Library, que publica a British National Bibliography³³ como dados abertos; existe também um “end point” SPARQL disponível. A Bibliothèque Nationale de France³⁴, desde 2011; em junho de 2016, disponibilizava 8.000.000 de documentos interligados a cerca de 2.000.000 de autoridades³⁵. Os dados são publicados segundo a licença Creative Commons CC0.

- A Deutsche Nationalbibliothek mantém um serviço de dados abertos³⁶ onde é possível fazer “download” de conjuntos de dados de

³² Ver <http://datos.bn.es>.

³³ Ver <http://bnb.data.bl.uk/>.

³⁴ Ver <http://data.bnf.fr/>.

³⁵ Ver <http://data.bnf.fr/about>.

³⁶ Ver http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkedata_node.html.

coleções ou temas específicos. Os dados são publicados segundo a licença Creative Commons CC0.

- A Library of Congress (EUA) mantém também um serviço de dados abertos³⁷, com ênfase nos seus arquivos de autoridades e vocabulários; é possível fazer “download” dos dados em diferentes formatos RDF.

- A biblioteca Europeia tem um projeto de dados abertos interligados desde 2012³⁸. O projeto se desenvolveu em duas fases, a primeira disponibilizando 20.000.000 de itens do acervo, e a segunda, 36.000.000. Existe um “end point” SPARQL³⁹ e foi desenvolvida uma API – Application Program Interface⁴⁰ - ⁴¹ com a qual foram criadas diversas aplicações para navegar, visualizar e georeferenciar dados. Os conteúdos publicados pela Europeia segundo as tecnologias LOD são liberados pela licença Creative Commons CC0. Detalhes sobre o projeto podem ser encontrados em (HASHOFER; ISAAC, 2011).

- o British Museum implementou seu projeto de dados abertos a partir de 2011. Seu serviço disponibiliza dados em um “end point” SPARQL⁴², em formatos como RDF/XML, N-Triples e TURTLE. Maiores detalhes podem ser encontrados no “press release”⁴³.

- O Museo del Prado anuncia o lançamento de seu novo sítio Web em um “press release” de 2015⁴⁴. Segundo o “press release”, o novo sítio utiliza as tecnologias LOD para criar um “knowledge graph”, que se baseia no modelo conceitual CIDOC CRM, e interliga o acervo do Prado com acervos de outras instituições que também usam essas tecnologias, como DBpedia, Wikidata, Freebase, Europeia, BBC, the New York Times, the British Museum e the Getty Foundation.

³⁷ Ver <http://id.loc.gov/>.

³⁸ Ver <https://pro.europeana.eu/resources/apis/linked-open-data>.

³⁹ Ver <https://pro.europeana.eu/resources/apis/linked-open-data#sparql-endpoint>.

⁴⁰ Interface de Programação de Aplicações, ver https://pt.wikipedia.org/wiki/Interface_de_programação_de_aplicações

⁴¹ Ver <https://pro.europeana.eu/resources/apis/linked-open-data#tools>.

⁴² Ver <http://collection.britishmuseum.org>.

⁴³ Ver http://www.britishmuseum.org/about_us/news_and_press/press_releases/2011/semantic_web_endpoint.aspx.

⁴⁴ Ver <https://www.museodelprado.es/en/whats-on/new/the-museo-del-prado-is-now-at-the-forefront-of/ac800655-62f6-41d9-ba1b-5907726bff07>.

- O projeto LOD do Rijksmuseum de Amsterdam foi inaugurado em 2016⁴⁵. O projeto foi desenvolvido em cooperação com a Biblioteca Europeia é o resultado de uma estratégia previamente formulada de abertura dos dados de suas coleções, que incluiu o desenvolvimento de uma API para que aplicações pudessem acessar os dados do museu. Os dados são modelados de acordo com o EDM e CIDOC CRM. Os dados abertos do catálogo usam como identificador persistente um sistema misto de persistente URL combinado com handle. Maiores detalhes sobre o projeto podem ser encontrados em DIJKSHOORN (2017).

- American Art Collaborative (AAC)⁴⁶, Linked Open Data Initiative, é um iniciativa de 14 museus de arte dos Estados Unidos, apoiados pela Mellon Foundation, para publicar seus acervos como dados abertos. O AAC promove o uso de vocabulários comuns para a representação das peças, como o CIDOC CRM, os vocabulários da Getty Foundation e o uso de identificadores persistentes para as peças das coleções, além de “linkar” as peças com recursos como a DBpedia. Promove também o uso de ferramentas comuns para navegação e integração de acervos.

O capítulo seguinte analisa maneiras para agregar semântica aos conteúdos disponibilizados na Web em geral e, em especial, aos conteúdos de acervos de arquivos, bibliotecas e museus, através do uso de vocabulários.

⁴⁵ Ver <https://www.slideshare.net/saschel/open-cultural-heritage-data-the-rijksmuseum>.

⁴⁶ Ver <http://americanartcollaborative.org/>.

3. EM BUSCA DE UMA REPRESENTAÇÃO MAIS RICA DOS ACERVOS EM ARQUIVOS, BIBLIOTECAS E MUSEUS: VOCABULÁRIOS

Como foi visto no capítulo anterior, vocabulários são um dos elementos chave das tecnologias de dados abertos interligados. São eles que padronizam como são descritos recursos ao agregarem significados precisos aos objetos, propriedades e valores das propriedades dos recursos descritos em RDF.

Vocabulários são conjuntos sistematizados de entidades semióticas - CONCEITOS -, triádicas (PEIRCE, 1994), cada um dos quais REPRESENTA alguma coisa (OBJETO ou REFERENTE), de alguma maneira (SIGNO: termo ou código), que gera um SIGNIFICADO para seus usuários. Servem para representar/modelar um domínio, fixando para a comunidade deste domínio o significado das coisas aí existentes.

A norma ISO/DIS 25964-1 (2011) estabelece que um tesauros é composto por conceitos estruturados sistemicamente. Tesauros são instrumentos para a indexação e recuperação de documentos e outros recursos em SRI; usam portanto a linguagem na interação com estes sistemas, formulando consultas que utilizam termos linguísticos.

Um conceito, ideia ou pensamento, é sempre o pensamento de alguma coisa, o objeto, que para ser comunicado a outros, necessita

ser externalizado através de um termo linguístico. Quando estivermos falando de termos que fazem parte de um vocabulário, teremos sempre subentendida essa relação triádica que inclui também o conceito e o objeto referido.

Instrumentos para controle e padronização semânticos como glossários, vocabulários especializados ou listas terminológicas de assuntos, listas de autoridades, etc., são utilizados há bastante tempo na área de documentação. Esses instrumentos tiveram que evoluir com a emergência das tecnologias da informação e, mais recentemente, das tecnologias da Web Semântica e Dados Abertos Interligados. Estas tecnologias trouxeram novos tipos de instrumentos para controle e padronização semânticos como os modelos conceituais e as ontologias computacionais.

Vocabulários são os portadores dos significados e como tal são um dos elementos chave nas tecnologias de dados abertos interligados. Conceitos podem ser identificados por identificadores únicos persistentes, IRIs, como já foi visto no capítulo anterior, e podem ser acompanhados pelos termos que os expressam, por definições, notas de escopo, origem e histórico de adoção dos termos, fontes de pesquisa sobre os mesmos, etc. Além dos conceitos, vocabulários são mais ou menos estruturados contendo diversos tipos de relacionamentos, como por exemplo:

- simples relacionamentos de ordem alfabética, como glossários e listas alfabéticas de termos;
- relacionamentos de classe-subclasse dos objetos referenciados, como em uma taxonomia;
- relacionamentos classe-subclasse e relacionamentos associativos como em um tesouro;
- relacionamento classe-subclasse e diversos tipos de relacionamentos associativos contendo identificadores únicos persistentes, escopo, domínio e cardinalidade destes relacionamentos, e regras de aplicação (axiomas) destes, como nos modelos conceituais e ontologias computacionais.

Vários autores (OBRST, 2006; SOUZA; TUDHOPE; ALMEIDA, 2012) propõe uma classificação dos vocabulários, conforme

eles sejam informais ou mais formalizados, ou conforme eles sejam adequados ao uso por pessoas ou por máquinas. Naturalmente a aplicação de vocabulários com as tecnologias LOD requer que esses vocabulários sejam o mais formalizados possíveis, para poderem ser utilizados por máquinas sem ambiguidades semânticas. A figura a seguir mostra um quadro conjunto dos principais tipos de vocabulários, desde os tradicionais até os mais avançados, segundo a sua “expressividade semântica, ou seja, sua capacidade de representar acuradamente as entidades em um domínio.

Figura 6 - Vocabulários quanto a complexidade e expressividade “semântica”.



Fonte: elaborado pelo autor

Existem dois tipos de vocabulários:

-1- Vocabulários compostos por um Elenco de Campos (propriedades) descritivos, como Autor, Título, Editor, Assunto, etc., chamados de “Property vocabularies por Zeng (2019), o que cada termo significa em si, o *nome do termo* – por exemplo, os 15 elementos do vocabulário Dublin Core tem nomes padronizados com uma semântica definida como dc:creator, dc:title, etc., ou os campos do formato MARC como 100 Entrada Principal por Nome Pessoal, 245 Título e indicação de responsabilidade, etc;

-2- **Vocabulários compostos por Valores** ou conteúdos possíveis de UM único campo (ou propriedade), chamados de “Data vocabularies por Zeng (2019).

- ex: propriedade ASSUNTO: em vocabulários como MESH, LCSH;

- ex: propriedade AUTORIDADES, autor: VIAF, listas de autoridades

- ex: propriedade LUGAR GEOGRÁFICO: Getty Thesaurus of Geographic Names (TGN)

Na próxima seção estaremos discutindo e mostrando exemplos de vários vocabulários operacionais, desde os mais simples como o Dublin Core até mais complexos, como ontologias computacionais. Na seção 3.2. estaremos discutindo os modelos conceituais mais usuais na área de documentação, como LRM/FRBR, CIDOC CRM, RiC-CM, etc.

3.1. VOCABULÁRIOS USADOS FREQUENTEMENTE COM AS TECNOLOGIAS LOD

Antes de começarmos a apresentar e discutir os diferentes vocabulários mais usados com as tecnologias de dados abertos interligados é importante sistematizarmos quais seriam os requisitos para que um vocabulário possa ser usado com estas tecnologias.

Estes requisitos, de acordo com a maioria de autores que trabalham com essa questão (CARACCILO et al., 2012; ZENG, 2019), seriam:

- Os conceitos do vocabulário devem ter escopo global, isto é, serem independentes de qualquer sistema local, terem semântica global e serem isentos, o mais possível de particularidades regionais;
- O vocabulário deve ser multilíngue, seus conceitos devem ter termos traduzidos em diversos idiomas;

- Cada conceito deve ter um identificador único, além de um identificador persistente, um IRI/URI que permita que ele possa ser acessado e “dereferenciado” através da Web e usado em triplas RDF;
- Cada conceito deve poder ser representado em linguagem de máquina segundo os padrões usados com as tecnologias de dados abertos interligados: RDF/XML, TURTLE, JASON;
- O vocabulário deve poder ser exportado, ou ser feito “download”, no todo ou em parte, em formato SKOS.

3.1.1 DUBLIN CORE METADATA ELEMENTS

O vocabulário DC - Dublin Core - é o vocabulário mais largamente utilizado na descrição de recursos na Web, devido a sua simplicidade e generalidade. Sua concepção minimalista permite que praticamente qualquer tipo de recurso Web – documentos vídeos, imagens, páginas, etc., possa ser descrito genericamente usando o DC. Existem inúmeras experiências do uso do DC em arquivos (HUVILA, 2008; MCKEMMISH, et al., 1999), bibliotecas e museus (ANDRESEN, 2006; DOERR, 2003; NEVILE; LISSONNET, 2005; WEIBEL, 1999). Por isso tem sido sempre o vocabulário usado quando o objetivo é a interoperabilidade entre sistemas distintos, uma questão chave para o tema deste livro; diferentes vocabulários podem ser mapeados para o DC. É o formato básico para intercâmbio de metadados no protocolo OAI-PMH, como foi visto no na seção 1.2. É um vocabulário simples, sem nenhuma estruturação entre seus 15 conceitos, chamados de elementos. Simples e genéricos, os 15 elementos DC formam um núcleo ou “core”, aplicada a uma grande variedade de objetos digitais. É a língua franca para representação de informação na Internet. Surgiu a partir de uma oficina sobre metadados da OCLC¹ e NCSA² ocorrida em 1995 na cidade de Dublin, Ohio, EUA, daí sua denominação. Seu objetivo original era definir um conjunto, de poucos elementos de significado óbvio e intuitivo, que pudesse ser usado pelos próprios autores para descreverem seus trabalhos, tendo em vista a proliferação de recursos eletrônicos e a impossibilidade dos profissionais de informação em catalogá-los.

¹ OCLC – On Line Computer Library Center

² NCSA – National Center of Supercomputing Applications.

Os 15 elementos do vocabulário de metadados Dublin Core são os seguintes:

- **Title** (Título) - Um Title é o nome pelo qual o recurso é formalmente conhecido;
- **Creator** (Autor, responsável pelo documento) - Exemplos de um Creator incluem uma pessoa, uma organização ou um serviço. O nome de um Creator deve ser usado para indicar esta entidade;
- **Subject** (Assunto) - o Subject será expresso por palavras ou termos-chave, ou códigos de classificação, que descrevam o assunto de um recurso. Boas práticas recomendadas seriam selecionar valores de um vocabulário controlado ou esquema de classificação;
- **Publisher** (Publicador, quem torna o documento disponível na Internet) - exemplos de um Publisher incluem uma pessoa, uma organização, um serviço ou, mais especificamente, um arquivo, biblioteca ou museus que está publicando seu acervo como dados abertos interligados. O nome de um Publisher deve ser usado para indicar a entidade responsável pela publicação/disponibilização do recurso;
- **Contributor** (outros Colaboradores - por ex. editores, tradutores, organizadores, etc) - exemplos de um Contributor incluem uma pessoa, uma *organização ou um serviço*;
- **Description** (Descrição) - um resumo/abstract, um sumário ou uma descrição livre do conteúdo;
- **Date** (Data) - Date é associado a criação ou de publicação do recurso; boas práticas recomendam codificar o conteúdo de Date são definidas pela Norma ISO 8601 e seguem o padrão AAAA-MM-DD;
- **Type** (Tipo de recurso - homepage, romance, poesia, software, dicionário, artigo de periódico, livro eletrônico, etc.) - Type inclui termos descrevendo categorias gerais, funções, gêneros ou níveis de agregação do conteúdo; boas práticas recomendam selecionar valores para Type de um vocabulário controlado (por exemplo, a

lista Dublin Core de Types); para descrever a manifestação física de um recurso, use o elemento Format;

- **Format** (Formato do arquivo que contém o documento eletrônico - texto, PDF, LaTeX, HTML, WORD ou outro) - tipicamente Format deve incluir o tipo de mídia ou dimensões do recurso. Format deve ser usado para determinar o software, hardware ou outro equipamento necessário para exibir ou operar o recurso; exemplos de dimensões incluem *tamanho* e *duração*; boas práticas recomendam selecionar valores de um vocabulário controlado (por exemplo, a lista de Internet Media Types [MIME]³ definindo formatos de mídia computacional);

- **Indetifier** (Identificador) – o “link” para acessar o recurso; boas práticas recomendam identificar um recurso por meio de um texto ou número em conformidade com um formato em um sistema de identificação padronizado e largamente aceito como Internationalized Resource Identifier (IRI), o Digital Object Identifier (DOI) e o International Standard Book Number (ISBN), etc;

- **Relation** (Relação) - relacionamentos com outros documentos, por ex. versões); boas práticas recomendam referenciar o recurso por meio de um texto em conformidade com sistema de identificação formal;

- **Source** – (Fonte) - o recurso a ser descrito pode ser derivado de um recurso fonte no todo ou em parte; boas práticas recomendam referenciar o recurso por meio de um texto ou número em conformidade com um sistema formal de identificação;

- **Language** (Linguagem) - boas práticas recomendam usar valores para o elemento Language definidos na RFC 1766 que incluem um código de duas letras (tirado da norma ISO 639, seguido opcionalmente de um código de duas letras do país (tirado da norma ISO 3166). Por exemplo, ‘en’ para Inglês, ‘fr’ para Francês, ou ‘es’ para Espanhol;

³ Ver <https://www.iana.org/assignments/media-types/media-types.xhtml>

- **Coverage** (Cobertura) - é um item temático onde é informada a localização espacial (um nome de lugar ou coordenadas geográficas), ou temporal, períodos temporais (um nome de período, data ou intervalo de datas) ou jurisdição (tal como um nome de uma entidade administrativa) do documento que está sendo descrito; boas práticas recomendam selecionar valores de um vocabulário controlado (por exemplo, do Getty Thesaurus de nomes Geográficos – TGN⁴) e, quando for o caso de nomes de lugares ou períodos serem usados ao invés de códigos numéricos, coordenadas ou intervalos de datas;

- **Rights** (Direitos) - Texto livre especificando qualquer restrição referente a direitos autorais; o elemento Rights deve conter uma declaração de direitos de uso de um recurso ou a referência a um serviço que preveja essa informação. A informação de Rights frequentemente incluem direitos de propriedade intelectual, de Copyright e vários direitos de propriedade; se o elemento Rights está ausente isto não significa que o recurso seja aberto ou não tenha nenhuma restrição quanto o uso; nenhuma suposição deverá ser feita em relação ao status ou a respeito dos direitos que regulam o uso do recurso.

O DC Pode ser usado com qualificadores para cada elemento, por exemplo, “**dc:contributor.orientador**” para especificar o orientador de uma tese. Há dois tipos de qualificadores; por refinamento, como já foi visto no exemplo anterior, “**orientador**” refina, especifica o significado de “**contributor**”; no outro tipo de qualificador um elemento é qualificado especificando o vocabulário ou “schema” de onde os valores do elemento são tirados; por exemplo, o elemento **dc:language** pode ser qualificado tirando seu conteúdo da tabela de idiomas ISO 6302, o elemento **dc:subject** tirando seus conteúdos do Tesaurus MeSH. Por exemplo, **dc:subject “schema=MeSH”:Carcinoma**. Os qualificadores servem para especificam mais precisamente o significado de um elemento DC.

Os elementos DC são implementados em sistemas informatizados de diferentes formas: como campos da base de dados de um sistema de

⁴ Ver <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>.

catálogo que correspondam aos elementos DC; como propriedades de um recurso descrito em RDF, como visto na seção 2.1; também podem ser codificados nas tags “meta” da linguagem HTML⁵, quando os metadados DC são assinalados a uma página .html. Veja o exemplo a seguir.

```
<meta name="dc.format" content="text/html">  
<meta name="dc.language" content="pt">  
<meta name="dc.creator" content="Marcondes, C. H.">  
<meta name="dc.title" content="Universidade Federal  
Fluminense">
```

O conjunto de elementos DC juntamente com os termos usados como qualificadores por refinamento formam um vocabulário mais amplo, o DC Terms⁶, também largamente utilizado descrição de recursos na Web e com as tecnologias LOD para especificar propriedades de recursos descritos em RDF, como visto no capítulo 2, seção 2.1. Os DC Terms têm sua semântica definida e são identificados por IRIs.

3.1.2. VOCABULÁRIOS DA GETTY FOUNDATION⁷

São 4, o **Art & Architecture Thesaurus (AAT)**, o **Cultural Name Objects Authority (CONA)** para nomes de sítios como o Pantheon de Atenas ou a torre Eiffel, o **Getty Thesaurus of Geographic Names (TGN)** para lugares, e o **Union List of Artists Names (ULAN)**. São muito usados na área de museologia e patrimônio cultural. Dos 4 o AAT, o TGC e o ULAN tem seus termos preparados para serem usados com as tecnologias LOD; todos possuem definições ou notas de escopo e identificadores únicos; os termos que são propriedades trazem também especificados seus domínios e escopos. O TGC e o ULAN permitem que seus termos sejam exportados em RDF/XML, TURTLE e JASON-LD.

⁵ As tags “meta” fazem parte do conteúdo de uma página html mas não são mostradas quando a página é exibida por um navegador, ver https://www.w3schools.com/Tags/tag_meta.asp.

⁶ Ver <http://dublincore.org/documents/dcmi-terms/>.

⁷ Ver <http://www.getty.edu/research/tools/vocabularies/cona/index.html>.

3.1.3. ICONCLASS⁸, A MULTILINGUAL CLASSIFICATION SYSTEM FOR CULTURAL CONTENT.

É um sistema de classificação com foco em artes e iconografia. É multilíngue, tem seus conceitos organizados em hierarquias a partir de 10 categorias. Todos os seus conceitos têm um código que permite a criação de IRIs únicos. ICONCLASS é licenciado segundo a Open Database License⁹. Seus conceitos podem ser exportados/baixados nos formatos SKOS/RDF e JASON.

3.1.4. VOCABULÁRIOS DA LIBRARY OF CONGRESS¹⁰

Tanto os LC Subject Headings quanto os LC Name Authority File estão disponíveis para serem usados com as tecnologias LOD, além de outros vocabulários mantidos pela LC.

3.1.5. VIAF – VIRTUAL INTERNATIONAL AUTHORITY FILE¹¹

É uma base de dados internacional e cooperativa de controle de autoridades. Além do nome padronizado de um autor ou instituição, dados históricos e/ou biográficos sobre o mesmo, cada autoridade tem também seu IRI, permitindo referenciá-lo como um “link” na descrição de um recurso em RDF.

3.1.6. INICIATIVA SPAR - SEMANTIC PUBLISHING AND REFERENCING ONTOLOGIES¹²

São um conjunto de vocabulários integrados relacionados a todos os aspectos domínio das publicações – documentos, suas partes, referências bibliográficas, citações, situação de um documento em cada estágio de publicação, etc. Todas as ontologias estão em formato RDFS ou OWL.

⁸ Ver <http://www.iconclass.nl/>.

⁹ Ver <http://opendatacommons.org/licenses/odbl/1.0/>.

¹⁰ Ver <http://id.loc.gov/>.

¹¹ Ver <http://viaf.org/>.

¹² Ver <http://www.sparontologies.net/ontologies>.

3.1.7. VOCABULÁRIOS DA WIKIPEDIA, DBPEDIA E WIKIDATA

São vocabulários inter-relacionados. A **Wikipedia**¹³ é uma enciclopédia geral nos moldes da *Encyclopedie*¹⁴ de Diderot e d’Alembert, disponível na Web, formada por entradas sobre os mais diferentes temas, em muitos idiomas. Seu conteúdo é em formato textual, legível por pessoas. É uma iniciativa coletiva e cooperativa, qualquer usuário pode editar, criar ou alterar qualquer verbete, os verbetes não são assinados.

A Wikipedia é estruturada tendo em vista as tecnologias LOD, suas entradas são todas identificadas por IRIs. Por isso ela vem se tornando o ponto central da Web de Dados, o recurso para o qual a maioria dos outros recursos disponíveis formatados como LOD fazem “links”. Devido ao fato de ser uma enciclopédia, um “link” para a Wikipedia (ou para suas irmãs, a DBpedia e a Wikidata) agrega informação a qualquer outro recurso. Esta posição especial da Wikipedia entre os outros recursos formatados segundo as tecnologias LOD pode ser ilustrada no diagrama da nuvem LOD¹⁵.

A **DBpedia**¹⁶ é a versão da Wikipedia em RDF, isto é, conteúdo estruturado, legível por máquinas. Como tal, é uma base de conhecimento que pode ser acessada diretamente por programas e consultada usando SPARQL. É obtida através do processamento automático das “knowledge box”, sumários que aparecem na parte superior direita de cada página da Wikipedia e que sumarizam o conteúdo da entrada. O conhecimento factual contido em cada “knowledge box” é extraído e representado em RDF segundo as tecnologias LOD. Por exemplo, temos a entrada sobre Miguel de Cervantes na Wikipedia¹⁷. Da “knowledge box” dessa entrada é extraída a informação de que o nascimento de Cervantes foi em “Alcalá de Henares, España” e que sua ocupação foi “Novelista, soldado, contable, poeta y dramaturgo”. Essas informações são representadas em RDF na entrada correspondente da DBpedia¹⁸ assim:

¹³ Ver <https://www.wikipedia.org/>.

¹⁴ Ver <https://en.wikipedia.org/wiki/Encyclopédie>.

¹⁵ Ver <http://lod-cloud.net/>.

¹⁶ Ver <http://wiki.dbpedia.org/>.

¹⁷ Ver https://es.wikipedia.org/wiki/Miguel_de_Cervantes.

¹⁸ Ver http://dbpedia.org/page/Miguel_de_Cervantes.

```
<?xml version="1.0">
<rdf:RDF
xmlns= http://www.w3.org/1999/02/22-rdf-syntax-ns
xmlns:dbo="http://dbpedia.org/ontology/">
<rdf:Description rdf:about "http://dbpedia.org/page/Miguel_
de_Cervantes">
<dbo:BirthPlace>http://dbpedia.org/page/Alcalá_de_Henares
    </dbo:BirthPlace>
...
<dbo:occupation>Novelist</dbo:occupation>
...
</rdf:Description>
</rdf:RDF>
```

A **Wikidata**¹⁹ é também uma versão da Wikipedia em formato RDF, como a DBpedia, só que invés de ser o resultado da extração automática do conteúdo das “knowledge boxes” da Wikipedia, é editada por pessoas. Como tal, é também uma base de conhecimento estruturado que pode ser acessada diretamente por programas e consultada usando SPARQL.

Como vocabulários Wikipedia, DBpedia e Wikidata podem funcionar de mesmas duas maneiras que outros vocabulários com o objetivo de controlarem e darem maior precisão semântica aos termos:

- agregando informação adicional sobre os mesmos, como no caso de substituir, no objeto do predicado dc:creator a cadeia de caracteres “Miguel de Cervantes” por um IRI para o item da DBpedia “http://dbpedia.org/page/Miguel_de_Cervantes” (neste caso seriam o objeto das triplas RDF), agregando toda a informação contida nesse item da DBpedia;

¹⁹ Ver <https://www.wikidata.org/>

- padronizando os nomes dos termos (que neste caso seriam os predicados das triplas RDF) pois seus diferentes tipos de termos (entradas, itens) são padronizados por uma ontologia própria. Esta ontologia contém, por exemplo, termos padronizados para diferentes tipos de agentes utilizados nas entradas da Wikipedia ou nos itens da DBpedia e Wikidata, como Artista, Cantor, Diretor de cinema, Escritor, Jogador de futebol, Político, etc²⁰.

Outra questão importante e recomendada no uso das tecnologias LOD é o reuso de vocabulários, sempre que possível (BIZER; HEATH; BERNERS-LEE, 2009; PATTUELLI; PROVO; THORSEN, 2015; SCHAIBLE; GOTTRON; SCHERP, 2014; VANDENBUSSCHE et al., 2017). Fontes importantes de informações sobre vocabulários que podem ser reusados são diretórios como os Linked Open Vocabularies, <http://lov.okfn.org/> e o vocab.org - A URI space for vocabularies, <http://vocab.org>. Ambos contém listas de vocabulários sobre diferentes temas, formatados para serem usados com as tecnologias LOD. Ambos são registros de vocabulários compatíveis com as tecnologias LOD; serão vistos com mais detalhes na seção 3.1.13.

3.1.8. GEONAMES²¹

É um vocabulários de lugares geográficos em que estes estão associados a suas coordenadas, fuso horário, código postal, nome em diferentes idiomas, etc. Cada lugar geográfico também tem “links” para a Wikipedia. Os lugares geográficos também estão organizados hierarquicamente (a cidade do Rio de Janeiro está sob “Brasil”, “Rio de Janeiro (estado)”) em classes como país, capital, estado, cidade, região administrativa, ou acidentes geográficos como montanha, rio, mar, etc. Cada lugar geográfico também tem um identificador persistente e pode ser exportado em RDF/XML. Estas características tornam o GEONAMES um vocabulário importante para acervos em memória e cultura, com grande potencial para fornecer um rico contexto a estes acervos.

²⁰ Ver <http://dbpedia.org/ontology/>

²¹ Ver <https://www.geonames.org/>.

3.1.9. SCHEMA.ORG

É um vocabulário criado por empresas como Google, Microsoft, Yahoo and Yandex, que operam grandes mecanismos de busca da Web. O vocabulário Schema é codificado como um micro formato, inserindo nas páginas html informações – os termos do vocabulário Schema - que permitem aos programas indexadores dos mecanismos de busca compreenderem o conteúdo das páginas e assim indexarem melhor cada página. O vocabulário cobre tipos de entidades, relações entre entidades e ações, e pode ser facilmente estendido para cobrir novas necessidades. Schema é o vocabulário utilizado no projeto de dados abertos da OCLC – Online Computer Library Center.

3.1.10. VRA – VISUAL RESOURCES ASSOCIATION

É um vocabulário de metadados descritivos mantido pela LC para a descrição de obras de arte e cultura e suas imagens. O conjunto de metadados descreve as diferentes propriedades desse tipo de objetos, organizadas em elementos, subelementos e atributos. O VRA é codificado em XML (atributos neste caso se referem aos atributos de um elemento XML). O vocabulário VRA se baseia no CDWA, já visto anteriormente. Os elementos principais descritos pelo VRA são work, image, collection, agent, location, culturalcontext, title e date.

3.1.11. AGROVOC²²

Tesouro mantido pelo AIMS - Agricultural Information Management Standards - da FAO - Food and Agriculture Organization -, é um vocabulário de valores para temáticas na área de agricultura e produção de alimentos. É multilingue, seus termos têm identificadores únicos - IRIs. O tesouro como um todo pode ser exportado em SKOS/RDF.

3.1.12. SKOS - SIMPLE KNOWLEDGE ORGANIZATION SYSTEM²³.

O SKOS é um vocabulário padronizado (uma recomendação do W3C) para representar sistemas de organização do conhecimento -

²² Ver <http://aims.fao.org/standards/agrovoc/functionalities/search>.

²³ Ver <http://www.w3.org/2004/02/skos/>.

taxonomias, sistemas de classificação, tesouros - no formato RDF, a serem usados com as tecnologias da Web Semântica. Como em um dicionário de sinônimos, o elemento principal do vocabulário SKOS é o conceito, `skos:Concept`. Cada conceito é expresso através de um rótulo `skos:Label` lingüístico. Os indivíduos na classes `skos:concept` são os conceitos de um tesouro. Também são fornecidos pelo vocabulário SKOS termos para os seguintes tipos de relacionamentos entre conceitos: hierárquico (termo genérico, termo específico), associativo (termo relacionado) ou terminológico (Uso, Usado por).

O exemplo a seguir mostra um conceito representado em SKOS no formato RDF.

```
<skos:ConceptScheme rdf: about="http://ClassificacaoFrutas/>
  <skos:Concept rdf: about = "http://ClassificacaoFrutas/
  Laranja">
    <skos:definition>Frutas cítricas ricas em vitamina C</
    skos:definition>
    <skos:prefLabel xml: lang = "pt">Laranja@pt</
    skos:prefLabel>
    <skos:broader rdf: resource = "http://ClassificacaoFrutas/
    Fruta"/>
    <skos:narrower rdf: resource = "http://
    ClassificacaoFrutas/Laranja_Lima"/>
    <skos:related rdf: resource = "http://ClassificacaoFrutas/
    Tangerina"/>
  </skos: Concept>.
```

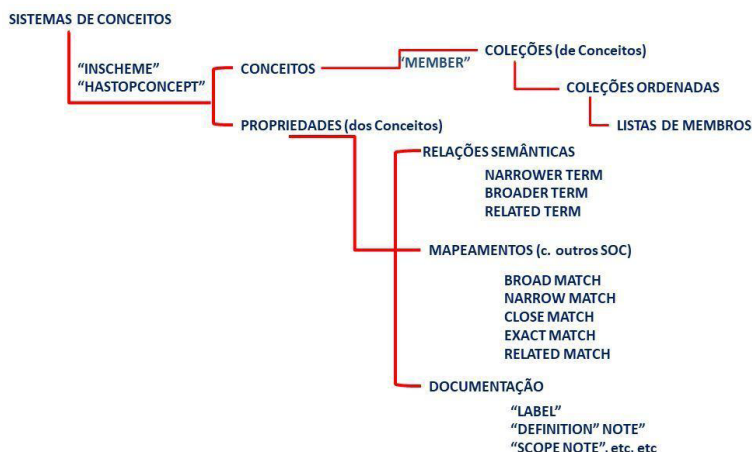
O SKOS, por ser codificado em RDF, a linguagem franca da Web Semântica (BERNERS-LEE; HENDLER; LASSILA, 2001), também é usado frequentemente intercambiar SOC entre sistemas que tem como um de seus componentes um SOC (por exemplo, bibliotecas digitais cujo sistema incorpora um tesouro). Além da função de intercâmbio entre sistemas, o SKOS permite também especificar o mapeamento e compatibilidade entre conceitos de um SOC em outro (SKOS PRIMER,

2009). Neste último caso, no exemplo anterior acrescentamos uma linha especificando que o conceito “Laranja” em nossa hipotética Classificação de Frutas é equivalente ao conceito “Orange jasmine” do Tesaurus AgroVoc²⁴.

<skos:exactMatch rdf:resource = “http://aims.fao.org/aos/agrovoc/c_25492”/ >

Um diagrama sintetizando todos os elementos do vocabulário SKOS e suas relações pode ser visto na figura seguinte, sintetizando o que é apresentado no SKOS Primer (2009).

Figura 7 - Diagrama mostrando os elementos do SKOS



Fonte: elaborado pelo autor

3.1.13. REGISTROS DE VOCABULÁRIOS COMPATÍVEIS COM AS TECNOLOGIAS DE DADOS ABERTOS INTERLIGADOS

O Getty Vocabulary Program dedicou quase três décadas para a criação de tesaurus que podem ser usados como bases de conhecimento, ferramentas de catalogação e documentação e assistentes de pesquisa on-line (BACA, 2016, p. 20).

²⁴ Ver em <http://aims.fao.org/vest-registry/vocabularies/agrovoc>.

Registros (“registries” em inglês) de vocabulários compatíveis com as tecnologias de dados abertos interligado são sítios que registram, documentam e alertam a existência de diferentes vocabulários que podem ser usados com as tecnologias de dados abertos interligados, além de fornecerem outros serviços aos desenvolvedores ou usuários. É conveniente lembrar, conforme alerta Baca na citação acima que, sempre que possível, deve-se optar pelo compartilhamento e reuso de vocabulários já existentes ou seu desenvolvimento cooperativo.

Registros de vocabulários compatíveis com as tecnologias de dados abertos interligado são os seguintes.

- BARTOC - Basel Register of Thesauri, Ontologies & Classifications - mantém cerca de 3000 vocabulários registrados, organiza os mesmos por disciplina e permite a busca por tipo de SOC (dicionários, ontologias, taxonomias, terminologias, etc.) , <https://bartoc.org/>.
- BioPortal, mantido pelo National Center for Biomedical Ontology, dos Estados Unidos, um portal especializado em ontologias biomédicas, <http://www.biportal.bioontology.org>.
- Heritage Data - Linked Data Vocabularies for cultural heritage - especialmente interessante para o tema deste livro, mantém um registro de vocabulários de interesse da área de memória e cultura, <http://www.heritagedata.org/blog/vocabularies-provided/>.
- LOV, Linked Open Vocabuaries, incluía 698 vocabulários que podem ser consultados por palavras-chave, <http://lov.okfn.org/>.
- RDA Registry, é um registro de vocabulários de elementos e relacionamentos que constam do padrão RDA²⁵ - Resource Description and Access - o padrão de descrição bibliográfica que sucede ao CCAA2 - Código de catalogação Anglo-americano 2 -, <https://www.rdaregistry.info/>.

²⁵ RDA Steering Committee, <http://www.rda-rsc.org/>.

- Linked Data Registry, um serviço de registro de vocabulário mantido pelo governo do reino Unido, <http://ukgovld.github.io/ukgovldwg/guides/registry.html>.
- VEST Registry (KOS), <http://aims.fao.org/vest-registry>, um serviço de registro mantido pela AIMS - Agricultural Information Management Standards -, mantém, além de vocabulários específicos para a área de agricultura e produção de alimentos, também ferramentas de software. conjuntos de metadados de elenco de campos e bases de dados específicas.
- Vocab.org - an URI space for vocabularies -, <http://vocab.org/>, é um registro que documenta e divulga vocabulários além de lhes emprestar um “namespace” sob a forma de um URI que pode ser utilizado pelos conceitos de qualquer vocabulário aí registrado para identificar seus conceitos, o que é muito útil para os desenvolvedores de vocabulários que muitas vezes não dispõem de um “namespace” próprio. Este URI é <http://purl.org/vocab/>. Se quiséssemos utilizar o vocab.org para a classificação hipotética de frutas do exemplo da subseção 3.19, teríamos o seguinte URI para o conceito “Laranja”: <http://purl.org/vocab/ClassificacaoFrutas/Laranja>.

3.2. MODELOS CONCEITUAIS NA ÁREA DE INFORMAÇÃO E DOCUMENTAÇÃO

Modelos Conceituais são os tipos de vocabulários mais estruturados e complexos. Estamos bastante acostumados em nossa conversação do dia a dia a mencionar o termo modelo: aeromodelos, as belas modelos que promovem os últimos lançamentos de moda, modelos de documentos como currículos vitae, contratos ou procurações, etc., modelos em escala, de automóveis, de prédios (maquetes), modelos bidimensionais de objetos tridimensionais, como plantas de prédios, mapas, etc. Esses sentidos de modelo são bem próximos do sentido que usamos o termo modelo conceitual, como uma representação simplificada de alguma outra coisa que, de alguma maneira, a substitui. As palavras-chaves desse sentido de modelo que devemos reter são “representação”, “simplificada” e “substitui”, que constituem a essência do significado de modelos.

Modelos são muito utilizados em ciência e nas engenharias. Uma equação de uma lei científica, como a Lei de Newton, é um modelo matemático de um fenômeno físico, que descreve este fenômeno, identifica as coisas envolvidas nele e como elas se inter-relacionam. O enunciado da Lei de Newton, $f = m_1 \times m_2 / d^2$ estabelece que a força de atração gravitacional f entre dois corpos de massas m_1 e m_2 é igual ao produto das massas de ambos os corpos dividido pelo quadrado da distância d entre eles. A Lei de Newton, ou modelo matemático do fenômeno da atração gravitacional, permite, entre outras coisas, compreender, explicar e *prever* o movimento dos planetas em torno do sol.

3.2.1. O QUE SÃO MODELOS CONCEITUAIS E SEU PAPEL NA REPRESENTAÇÃO CONTEÚDOS DE ARQUIVOS, BIBLIOTECAS E MUSEUS

A história das práticas de representação de registros do conhecimento – a história da catalogação (GARRIDO-ARILLA, 1999), a das descrições arquivísticas (COOK, 1998, p. 132), da documentação museológica (CAMARGO-MORO, 1986), é marcada por regras de criação de representações – descrições, registros - pouco formais, sem fundamentação ou princípios orientadores, baseadas em experiências pessoais. Com o surgimento de bases de dados, dos registros e catálogos em ambientes digitais, surge a necessidade de maior formalismo e rigor na construção desses registros. A proposta dos modelos conceituais é superar estas deficiências.

Os atuais modelos conceituais usados em diversas áreas e também nas áreas de arquivos, bibliotecas e museus, são baseados em uma metodologia da área de computação denominada modelo Entidades-Relacionamentos. Esta metodologia foi proposta inicialmente pelo cientista da computação Peter Chen em 1976 (CHEN, 1976), e tinha como finalidade o planejamento de bancos de dados. Pode-se considerar no entanto, que o primeiro uso de modelos conceituais atuais da área, foi o modelo UNISIST proposto pela Unesco (1971) para estabelecer um sistema de informação científica mundial.

Modelos conceituais vêm sendo utilizados cada vez mais na área de documentação, assim como em muitas outras áreas²⁶. Na área

²⁶ Ver Business Model, https://en.wikipedia.org/wiki/Business_model.

de documentação, exceto pela experiência do UNISIST, a experiência pioneira foi a criação do modelo conceitual FRBR pela IFLA em 1998 (IFLA, 1998). O modelo FRBR estabelece um modelo das necessidades de um usuário ao consultar um catálogo e de como estas necessidades deverão ser supridas pelos registros bibliográficos. Ele, a exemplo de outros modelos conceituais, não descreve *como* deve ser o catálogo, como ele deve funcionar, como deve ser programado ou resolvido em termos informáticos, mas tão somente *que* conteúdos deve os registros bibliográficos devem prover.

Ao FRBR se seguiram o modelo conceitual CIDOC CRM, cuja 1ª. versão é de 1999²⁷, tentativas de compatibilizar ambos os modelos, o modelo alternativo ao FRBR desenvolvido pela LC dos EUA, o BIBFRAME de 2012, o EDM – European Data Model (2010) e, recentemente, o modelo RiC-CM (INTERNATIONAL COUNCIL ON ARCHIVES, 2016) – Records in Context Conceptual Model.

Quando se fala em modelos conceituais o termo é utilizado em um sentido mais próximo do utilizado nas engenharias, como uma etapa de projeto, como uma representação simplificada e, na maioria das vezes, em escala menor, de algo que se deseja construir. Nesse sentido, um modelo, ao representar aquilo que futuramente será construído, tem a finalidade prática de antever, prever o funcionamento, comparar e avaliar aquilo que se deseja construir.

Na área de documentação, podemos dizer que, de uma certa maneira, um código de catalogação como o CCAA2 ou uma norma como a ISAD(G) são modelos dos registros a serem criados no catálogo; os registros, por sua vez, são modelos ou representações dos documentos físicos ou digitais do acervo.

Definiremos então modelos conceituais como uma REPRESENTAÇÃO, sob a forma de uma descrição abstrata e genérica, de determinado DOMÍNIO da realidade, para fins de compreender esta realidade, raciocinar sobre ela e fixar um entendimento comum dessa realidade dentro de uma comunidade. Como toda representação de algo, um modelo conceitual é algo *construído* por alguém, como alguma finalidade prática.

27 Conforme <http://network.icom.museum/cidoc/working-groups/crm-special-interest-group/>. Acesso em: 13 jul. 2017.

Um modelo conceitual estabelece que diferentes “COISAS” existem em um determinado domínio; como elas se distinguem umas das outras; como elas se relacionam entre si; e quais são suas características.

Uma realidade ou, mais especificamente, um domínio ou parte desta realidade, é percebida como um conjunto de coisas ou Indivíduos, ou seja, seres com identidade distinta, que agrupamos em classes de Entidades por algum critério de semelhança. Identificamos nessas entidades suas as Propriedades. As propriedades, isto é, tudo que é próprio, que caracteriza um entidade, podem ser ou as Características que constituem a essência da própria entidade (as características são de um indivíduo), ou os Relacionamentos entre uma entidade com outra (os relacionamentos são de pelo menos dois ou mais indivíduos, pertencentes a uma ou mais classes de entidades).

Ou seja, um modelo conceitual representa uma determinada realidade estabelecendo que tipos ou classes de coisas existem - Entidades – (também chamadas de classes) em um determinado domínio da realidade; quais são as Propriedades dessas classes de Entidades, isto é, suas Características e como essas Entidades se relacionam umas com as outras, seus Relacionamentos. Por exemplo, temos as entidades: pessoas, documentos, autores de documentos; as características: datas, nome, CPF, data de nascimento de uma pessoa; e os relacionamentos: matrícula em de alunos em determinado curso, casamento de uma pessoa com a outra, etc. Essas “coisas” distinguíveis, ou conceitos existentes em determinada domínio da realidade, para fins de um entendimento comum, são fixadas linguisticamente através de termos, muitas vezes acompanhados de suas definições.

Na literatura da área de vocabulários, modelos conceituais, ontologias existem variações terminológicas para identificar esses elementos constituintes de um modelo; ora se fala de **Indivíduos**, ora se fala de instâncias, ora se fala de **Entidades**, ora de classes de indivíduos, ora se fala de **Propriedades**, ora se fala de **Relacionamentos**, ora se fala de atributos, ora se fala de **Características**. Os conceitos que marcamos em negrito serão a terminologia que vamos adotar.

Outra questão importante. Se modelos estabelecem um entendimento comum de uma dada realidade, como eles são materializados, representados, para que essa representação (do entendimento comum) da realidade possa ser compartilhada por uma comunidade de interessados? A forma mais simples de representar um modelo é uma descrição em linguagem natural. Mas

existem linguagens mais formais, precisas, menos sujeitas a ambiguidades, para representar modelos, como a linguagem gráfica dos diagramas de classe da UML que será vista a seguir. Modelos também podem ser representados em formato legível por computadores através das linguagens RDF (RDF PRIMER, 2002), RDF Schema (RDF SCHEMA 1.1, 2014) e OWL (OWL ONTOLOGY WEB LANGUAGE OVERVIEW, 2004).

Veremos a seguir dois exemplos de representações de modelos em linguagem natural. Na Figura 1 seguinte são vistas duas tabelas do documento de referência do modelo conceitual RiC-CM onde são mostradas a definição da entidade *Record* e suas propriedades, e os relacionamentos em que está envolvida essa mesma entidade. Repare como a entidade é claramente definida, inclusive com seu escopo. O mesmo se dá com os relacionamentos; para cada relacionamento, o relacionamento inverso é especificado, e também o domínio (“domain”, a entidade de onde o relacionamento parte) e o escopo (“range”, a entidade onde o relacionamento chega) do relacionamento.

Figura 8. Entidade (ou classe) Record e seus relacionamentos.

3.2 Properties of Record		
3.2.1 Content Information		
#	Name	Definition
RIC-P5	Authenticity and Integrity Note	Description of features of the Record attesting to how it is genuine, reliable, trustworthy, whole and complete.
	Scope	--
	Data Type	Text
	Examples	the record missing its lower left corner of text and bears the seal of the King; the record is electronically signed and the signature is valid at the moment of description; the database (.dbf) file has the checksum SHA-1: bb165836546e13c8d80e97aa44e8f4369608e9f8
#	Name	Definition

4.1.1 Record Relations					
Relation Number	Relation Domain	Relation Name	Relation Range	Inverse Relation Name and Number	Comments
RIC-R1	Record	has copy	Record	is copy of (RIC-R6)	
RIC-R2	Record	has draft	Record	is draft of (RIC-R7)	
RIC-R3	Record	has original	Record	is original of (RIC-R8)	
RIC-R4	Record	has subject	Record	is subject of (RIC-R10)	
RIC-R5	Record	is associated with	Record	is associated with (RIC-R5)	See also RIC-R12
RIC-R6	Record	is copy of	Record	has copy (RIC-R1)	
RIC-R7	Record	is draft of	Record	has draft (RIC-R2)	

Fonte: INTERNATIONAL COUNCIL ON ARCHIVES (2016).

Outro exemplo, Mostrado na Figura 2, é a descrição da entidade **E24 Physical Man-Made Thing** e do relacionamento (“Property”) **P**

108 has produced (was produced by), ambos tirados do documento EUROPEANA DATA MODEL PRIMER (ISAAC, 2013).

Na descrição a entidade ou classe é definida através da especificação de suas subclasses e superclasses, da sua definição (“Scope Note”), de exemplos e dos relacionamentos ou propriedades nos quais essa entidade está envolvida. Na descrição dos relacionamentos são definidos seu domínio (“Domain”) e escopo (“Range”), suas subpropriedades (“Subproperty of”) e sua definição (“Scope Note”).

Figura 9. Entidade E24 Physical Man-Made Thing e relacionamento P108 has produced (was produced by).

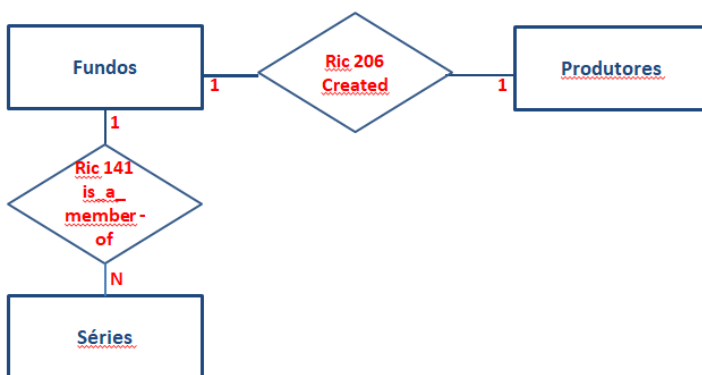
E24 Physical Man-Made Thing	
Subclass of:	E18 Physical Thing E71 Man-Made Thing
Superclass of:	E22 Man-Made Object E25 Man-Made Feature E78 Collection
Scope Note:	This class comprises all persistent physical items that are purposely created by human activity. This class comprises man-made objects, such as a swords, and man-made features, such as rock art. No assumptions are made as to the extent of modification required to justify regarding an object as man-made. For example, a “cup and ring” carving on bedrock is regarded as instance of E24 Physical Man-Made Thing.
Examples:	<ul style="list-style-type: none"> the Forth Railway Bridge (E22) the Channel Tunnel (E25) the Historical Collection of the Museum Benaki in Athens (E78)
Properties:	<p>P62 depicts (is depicted by): E1 CRM Entry (P62.1 mode of depiction: E55 Type)</p> <p>P65 shows visual item (is shown by): E36 Visual Item</p> <p>P128 carries (is carried by): E73 Information Object</p>
P108 has produced (was produced by)	
Domain:	E12 Production
Range:	E24 Physical Man-Made Thing
Subproperty of:	E11 Modification. F31 has modified (was modified by): E24 Physical Man-Made Thing E63 Beginning of Existence. P92 brought into existence (was brought into existence by): E77 Persistent Item
Quantification:	one to many, necessary, dependent (1..n;1..1)
Scope note:	This property identifies the E24 Physical Man-Made Thing that came into existence as a result of an E12 Production. The identity of an instance of E24 Physical Man-Made Thing is not defined by its matter, but by its existence as a subject of documentation. An E12 Production can result in the creation of multiple instances of E24 Physical Man-Made Thing.
Examples:	<ul style="list-style-type: none"> The building of Rome (E12) <i>has produced</i> The Colosseum (E22)

Fonte: EUROPEANA DATA MODEL PRIMER (ISAAC, 2013).

Outra forma de representar modelos é através de uma linguagem gráfica, de diagramas. Inicialmente, a partir da proposta pioneira de Chen (1976) modelos conceituais foram representados usando uma linguagem gráfica, o diagrama Entidades-Relacionamentos. O diagrama E-R evoluiu e se tornou uma das bases para o desenvolvimento da UML – Unified Modeling Language (RAMOS, 2006). A UML é uma linguagem, específica para representar modelos conceituais, com origem na computação. O Diagrama de Classes da UML, correspondente ao diagrama E-R, permite

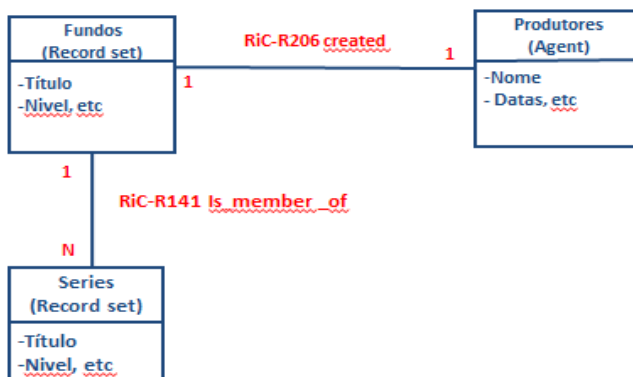
representar graficamente modelos conceituais através de dois elementos gráficos básicos: o retângulo para representar classes (entidades) e suas propriedades e a linha para representar relacionamentos. Nas figuras abaixo vemos a representação de três entidades (classes) do Modelo RiC-CM, Fundos, Séries e seus Produtores, com suas propriedades, e 2 relacionamentos (RiC-R206 created e RiC-R141 is_member_of) segundo o diagrama E-R na Figura 10 e segundo o diagrama de classes da UML na Figura 11; ambas as representações são equivalentes.

Figura 10. Exemplo de Diagrama E-R



Fonte: elaborado pelo autor

Figura 11. Exemplo de Diagrama de Classes UML



Fonte: elaborado pelo autor

Outra forma de representar modelos conceituais em formato legível por computadores é através de linguagens como RDF, RDF Schema e OWL, já mencionadas. Os modelos conceituais mais formalizados podem praticamente ser traduzidos para linguagens como estas. A seguir é mostrado um trecho da representação do modelo conceitual FRBR que mostra a especificação da classe Obra, em linguagem OWL. Pelo exemplo vemos que essa classe é uma subclasse da classe Endeavour; essa classe também diferente (“disjoint”) das classes Expression, Manifestation e Item, as outras classes que formam o Grupo 1 do modelo FRBR. Esse exemplo foi tirado de <http://vocab.org/frbr/core.rdf>.

```
<owl:Class rdf:about="&frbr;Work">
  <rdfs:label xml:lang="en">work</rdfs:label>
  <skos:definition xml:lang="en">An abstract notion of an artistic
  or intellectual creation.</skos:definition>
  <rdfs:comment xml:lang="en">This class corresponds to the
  FRBR group one entity 'Work'.</rdfs:comment>
  <rdfs:isDefinedBy rdf:resource="&frbr;" />
  <rdfs:subClassOf rdf:resource="&frbr;Endeavour" />
  <rdfs:subClassOf rdf:resource="http://xmlns.com/wordnet/1.6/
  Work-2" />
  <owl:disjointWith rdf:resource="&frbr;Expression" />
  <owl:disjointWith rdf:resource="&frbr;Manifestation" />
  <owl:disjointWith rdf:resource="&frbr;Item" />
  <dct:issued>2005-07-15</dct:issued>
  <skos:changeNote rdf:value="Added isDefinedBy property
  referring to FRBR core namespace" dc:date="2005-08-08"
  dc:creator="Ian Davis" />
  <skos:changeNote rdf:value="Added comment"
  dc:date="2005-08-08" dc:creator="Ian Davis" />
```

```
<skos:historyNote rdf:value="Made subclass of  
wordnet:Work-2" dc:date="2005-08-08" dc:creator="Ian  
Davis"/>  
<skos:changeNote rdf:value="Extracted definition from existing  
comment" dc:date="2005-08-09" dc:creator="Ian Davis"/>  
</owl:Class>
```

3.2.2. MODELO FRBR (BIBLIOTECAS)

O modelo FRBR, proposto pela IFLA em 1998, é a culminância de um longo processo desenvolvido pela Biblioteconomia no sentido de superar as práticas pouco formalizadas descritas anteriormente para a criação de registros bibliográficos que servissem de mecanismos de acesso ao acervo das bibliotecas. Pontos chave neste processo estão comentados a seguir.

O período anterior ao surgimento do modelo é o do surgimento da tecnologia dos computadores e sua aplicação à catalogação bibliográfica, com o formato MARC desenvolvido pela Library Of Congress, EUA, na década de 1960, como já foi mencionado.

A partir dos anos 1960 a IFLA iniciou uma discussão internacional sobre princípios e teorias que sustentassem as atividades de catalogação e representação bibliográficas. Em 1961 a IFLA promoveu uma conferência internacional sobre o tema em Paris. Dessa conferência surgiu um conjunto de princípios que estabelecem as bases da atividade de catalogação, conhecidos por Princípios de Paris. Em 1969 a IFLA promoveu em Copenhague um Encontro Internacional de Especialistas em Catalogação. Em 1990 em Stockholmo, o programa de Universal Bibliographic Control, o Programa International MARC (UBCIM) e a Division of Bibliographic Control da IFLA promoveram o Seminar on Bibliographic Records. Uma das resoluções do Seminário recomendava o desenvolvimento de um estudo para definir os requisitos funcionais que registros bibliográficos deveriam atender. O desenvolvimento dos FRBR é uma continuidade direta desse processo.

O modelo FRBR foi desenvolvido a partir da metodologia Entidades-Relacionamentos, proposta por Peter Chen (1976). No entanto,

antes de propor qualquer entidades ou relacionamentos o modelo FRBR estabelece como ponto de partida um conjunto de tarefas que usuários deveriam desenvolver na sua interação para busca por informações com catálogos bibliográficos, aos quais os registros bibliográficos deveriam atender; esses seriam os requisitos funcionais dos registros bibliográficos. Os registros bibliográficos deveriam prover dados que permitissem aos usuários desenvolver as seguintes tarefas:

- encontrar itens que correspondam aos critérios de busca estabelecidos pelos usuários.
- identificar uma determinada entidade, isto é, certificar que o registro recuperado corresponda ao documento procurado pelo usuário, por exemplo, permitindo distinguir dois documentos de mesmo título.
- selecionar a entidade mais adequada às necessidades do usuário, por exemplo, um documento em determinado idioma que o usuário domine, ou um programa de computador adequado ao “hardware” e “software” que o usuário dispõe.
- acessar o documento final.

Com base nesses requisitos a atender para permitir a realização das tarefas dos usuários, o modelo FRBR define entidades agrupadas em três grupos:

- Entidades do GRUPO I – os produtos da criação artística ou intelectual do acervo das bibliotecas e que são descritas nos registros bibliográficos. São 4, em organizadas em nível decrescente de abstração (do mais abstrato, a Obra, para o mais concreto, o Item):
 - Obra (Work) - Hamlet, a concepção ou ideia da tragédia de Shakespeare; a obra-prima Hamlet.
 - Expressão (Expression) – A realização da Obra em determinada forma ou linguagem: texto em inglês, áudio espanhol.
 - Manifestação (Manifestation) – a edição em inglês da Penguin Books, a edição em português da Editora LPM, traduzida por Millôr Fernandes, um audiobook em espanhol.

- Item (Item) – exemplar adquirido ao preço de E\$ 10,00 e registrado pela biblioteca sob o número 658678/2017.

Entidades do GRUPO II – aquelas entidades responsáveis pelo conteúdo artístico ou intelectual, pela produção, disseminação ou guarda das entidades do Grupo I

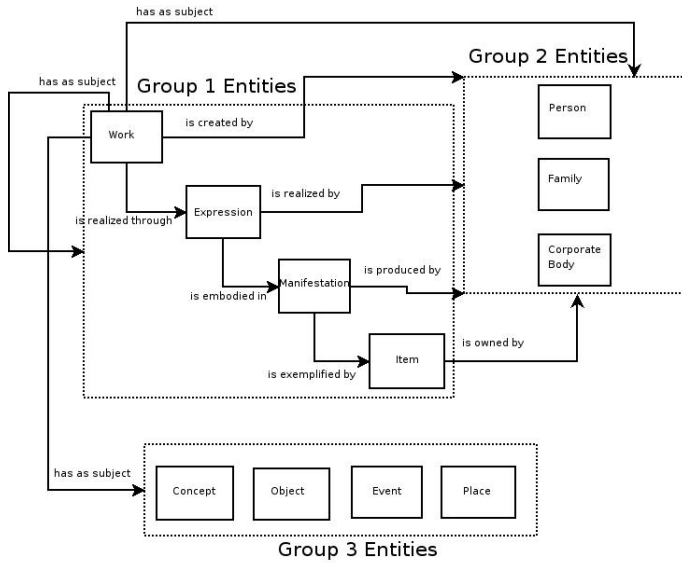
- Pessoas (Person) – o autor, como Willian Shakespeare, ou um tradutor como Millôr Fernandes.
- Instituições (Corporate body) - Penguin Books, Editora LPM, Biblioteca Nacional de España

Entidades do GRUPO III – aquelas entidades que servem de temas ou assuntos das entidades do Grupo I

- Conceito (Concept) – tragédia.
- Objeto (Object) – um punhal, um frasco de veneno, a vingança.
- Evento (Event) – a morte do Rei Hamlet.
- Lugar (Place) -> Dinamarca.

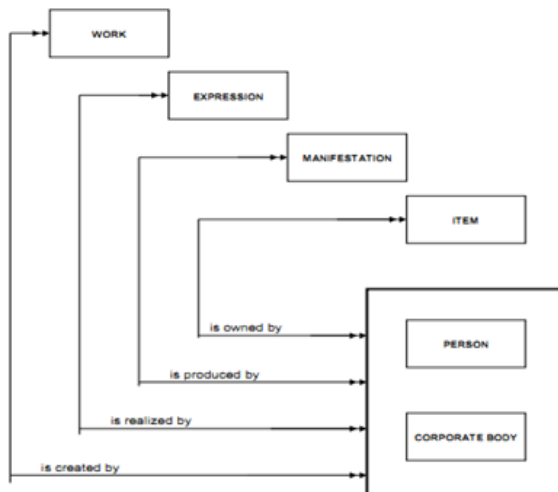
Além destas entidades o modelo FRBR prevê também diferentes tipos de relacionamentos entre elas. As figuras a seguir são tiradas do documento FRBR – FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS (IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS, 1998) e descrevem: Figura 3 – relacionamentos entre as entidades do Grupo 1 e entre as entidades do Grupo 1 e as entidades dos Grupos 2 e 3; Figura 4 - relacionamentos entre as entidades do Grupo 2; e Figura 5 – relacionamentos entre as entidades do Grupo 3.

Figura 12. Relacionamentos entre as entidades do Grupo 1 e entre as entidades do Grupo 1 com as entidades dos Grupos 2 e 3.



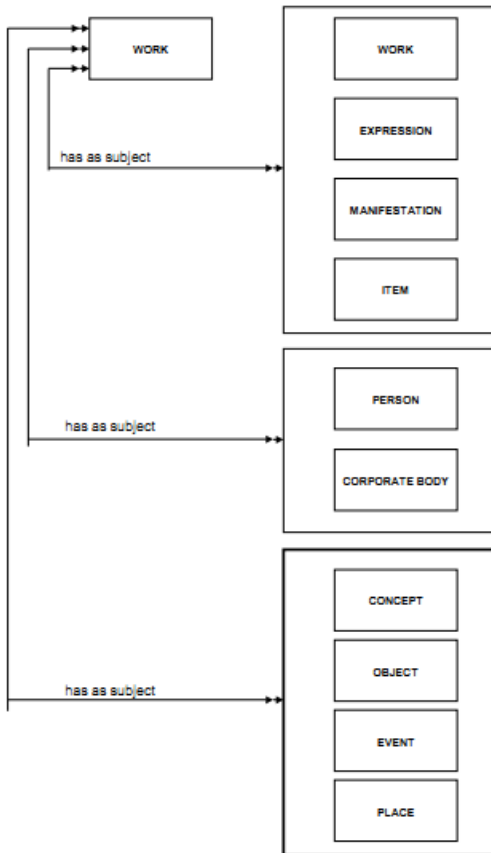
Fonte: retirado de Denton [2009].

Figura 13. Relacionamentos entre as entidades do Grupo 2



Fonte: retirado de IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS (1998)

Figura 14. Relacionamentos entre as entidades do Grupo 3.



Fonte: retirado de IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS (1998)

O modelo conceitual FRBR foi complementado e é usado hoje juntamente com dois outros modelos, o FRAD – Functional Requirements for Authority Data - (IFLA WORKING GROUP ON FUNCTIONAL REQUIREMENTS AND NUMBERING OF AUTHORITY RECORDS, 2008), e o FRSAD - Functional Requirements for Subject Authority Data (IFLA WORKING GROUP ON THE FUNCTIONAL REQUIREMENTS FOR SUBJECT AUTHORITY RECORDS (FRSAR), 2010).

O modelo FRBR teve uma grande influência, que ultrapassa os aspectos específicos da catalogação bibliográfica. Nele, de forma explícita, os usuários da biblioteca e suas tarefas aparecem como determinantes na construção do modelo. Isso significa um reposicionamento da biblioteca e de suas funções, não mais como um repositório de documentos, mas como um serviço voltado aos usuários.

O modelo FRBR vem também, desde 1998, influenciando esse reposicionamento da biblioteca. Motivado também pela evolução das tecnologias da informação, o surgimento de materiais de biblioteca em mídias diferentes do impresso – discos fonográficos, videocassetes, CD-ROM, DVD -, e pelo surgimento da Internet e da Web, o antigo Código de Catalogação Anglo-Americano segunda edição, vem sendo substituído por um novo código, o RDA - Resource Description and Access²⁸. O RDA é um padrão constituído por conjunto de diretrizes, itens de dados e regras para a criação de metadados bibliográficos descritivos. É totalmente baseado e organizado de acordo com as entidades, relacionamento e atributos do modelo FRBR. Segundo se pode ler no sítio do Comitê Diretor Internacional do RDA, este é concebido tendo em vista tecnologias LOD: “RDA is a package of data elements, guidelines, and instructions for creating library and cultural heritage resource metadata that are well-formed according to international models for user-focussed linked data applications”.

3.2.3. MODELO CIDOC CRM (MUSEUS)

O modelo CIDOC CRM é o resultado de esforços conjuntos da comunidade museológica, organizada no CIDOC – ICOM International Committee for Documentation – com o objetivo prover definições, estrutura e semântica para orientar a documentação das entidades e relacionamentos existentes na área de preservação do patrimônio cultural. Conforme é informado no seu sítio web²⁹, estes esforços tiveram um primeiro produto em 1994, um modelo de dados para museus, o CIDOC Relational Data Model. Esse modelo de dados evoluiu para um modelo conceitual, mais genérico e abstrato. Em 2000 o CIDOC CRM foi aceito

²⁸ Ver a página do Comitê Diretor Internacional do RDA, <http://www.rda-rsc.org/>.

²⁹ Ver <http://www.cidoc-crm.org/>.

como documento de trabalho do Comitê ISO TC46/SC4 – Interoperability -, o passo inicial para se tornar uma norma ISO, o que aconteceu em 2006, quando o CIDOC CRM se tornou a norma ISO 21127/2014. No ano de 2000 também foi criado pelo CIDOC o CRM Special Interest Group³⁰, um grupo de trabalho com o objetivo de manter e atualizar o CRM.

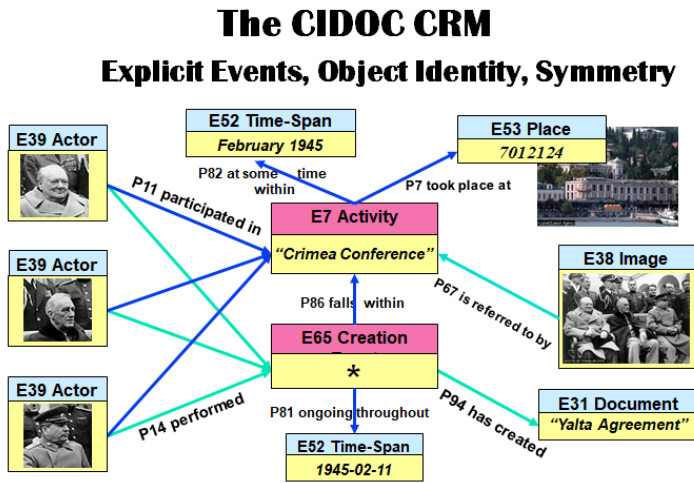
Ainda segundo seu sítio Web, o CIDOC CRM “is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to”. Como pode ser visto o objetivo declarado do CIDOC CRM tem um escopo mais amplo do que simplesmente a área de museus, abrangendo de instituições de memória e cultura em geral. O CRM é portanto mais amplo que o FRBR. Vários esforços têm sido desenvolvidos com o objetivo de compatibilizar ambos os modelos ³¹ (DOERR; LEBOEUF, 2007).

Ao contrário do modelo FRBR, centrado nas entidades que constituem o acervo das bibliotecas, o propósito do CRM é fornecer informação rica através da qual objetos de patrimônio cultural possam ser *contextualizados*. Informação contextual é essencial para a compreensão de objetos de acervos museológicos, uma vez que estes, ao contrário de livros ou documentos, não têm conteúdo, não são em si portadores de mensagens. O CRM é orientado por eventos - Temporal Entities E2 – acontecimentos do passado ocorridos em determinado lugar - Place E53 -, no bojo dos quais objetos de patrimônio cultural – Objects, Conceptual E28 and Physical E18 - são relacionados a atores - Actors/ People E39 - (OLDMAN, 2014, p. 19). Assim, por exemplo, um Evento – a Conferência de Yalta, ocorrida no fim da II Guerra Mundial na cidade de Yalta, Criméia, Ucrânia, em fevereiro de 1945, reunindo Atores como os líderes aliados Churchill, Roosevelt e Stalin, é registrada através de objetos como fotografias (E38 Image) dos três líderes e o documento (E31 Document) do Acordo de Yalta. A contextualização deste evento e dos objetos s de patrimônio cultural nele envolvidos fica mais clara na seguinte figura, tirada de Doerr, Stead (2009).

³⁰ Ver <http://www.cidoc-crm.org/node/1736>.

³¹ Ver <https://www.ifla.org/node/928>.

Figura 15. Relacionamento entre Atores e Objetos intermediados por Eventos situados em lugares e no tempo.



Fonte: elaborado pelo autor

A hierarquia de classes e subclasses do CRM é desenvolvida a partir dessas três categorias, Objetos, Eventos e Atores (ICOM/CIDOC, 2011). O modelo possui ao todo 86 classes e estas são relacionadas através de 137 propriedades. Classes e propriedades são organizadas em hierarquias de classes-subclasses e propriedades-subpropriedades. O CIDOC CRM tem também versões como uma ontologia computacional, nas linguagens RDF³² e OWL³³.

O CRM foi planejado segundo o paradigma da orientação a objetos – OO³⁴ -. Nesse paradigma os elementos do modelo - suas classes – são considerados objetos, o que significa constituem um todo, que herdam suas propriedades específicas dos objetos que são suas superclasses e qualquer nova subclasse que venha a ser criada herda também todas as propriedades específicas dessa classe. O CRM é também extensível, isto é, suas classes e propriedades podem ser estendidas através da criação de subclasses e subpropriedades mais específicas para atender necessidades

³² Ver <http://www.cidoc-crm.org/Resources/rdf-file-for-crm-core>.

³³ Ver <http://erlangen-crm.org/>.

³⁴ Ver https://en.wikipedia.org/wiki/Object-oriented_programming.

especiais. Por exemplo, a classe E39 Actor, subclasse E31 Person, poderia ser especializada, por exemplo, por uma classe como E 31.01 Painter. Existem várias experiências de extensões do CRM (DOERR; BEKIARI; LEBOEUF, 2008).

3.2.4. EDM - EUROPEANA DATA MODEL

O EDM é o modelo conceitual que estrutura os conteúdos da Biblioteca Europeana, iniciativa que visa criar um portal na Web para dar acesso às manifestações culturais dos países da Comunidade Européia. Essas manifestações incluem textos (livros, documentos), vídeos, objetos tridimensionais (imagens de obras de arte, de objetos), gravações de som, imagens (fotografias). Especificamente, a Europeana agrega registros – metadados - desses objetos mantidos nos acervos de mais 3500³⁵ instituições culturais europeias de diferentes tipos: bibliotecas, museus, arquivos, cinematecas, etc. Seu acervo totaliza cerca de 54.358.074 registros em 2017 segundo o Europeana Foundation Governing Board Meeting³⁶.

O mecanismo tecnológico básico para agregar ao acervo da Europeana registros de acervos de outras instituições é o protocolo OAI-PMH, visto na seção 1.3. No protocolo OAI-PMH o conteúdo original de um objeto – uma cópia digital do Don Quijote, de Cervantes, por exemplo – é mantido na instituição de origem, a BNE, e somente metadados são transferidos para a Europeana. Esses metadados são incluídos a partir de um formato baseado no conjunto de metadados Dublin Core, o formato padrão do protocolo OAI-PMH, chamado ESE – Europeana Semantic Elements. O ESE funciona como um máximo denominador comum entre as diferentes conjuntos de metadados de cada instituição que contribui para a Europeana. Uma vez incluídos registros segundo o ESE, a Europeana os processa, enriquecendo-os através de vocabulários de diferentes temáticas, tesouros e esquemas classificatórios, convertendo-os para o EDM (DOERR et al., 2010). O EDM tem a função de “The continuous development and

³⁵ Ver <https://pro.europeana.eu/page/new-provider>.

³⁶ Ver o Relatório em https://pro.europeana.eu/files/Europeana_Professional/Governing_Board/Public_Papers_of_Governing_Board_Meetings/Public%20Papers%20of%20Governing%20Board%20Meeting%202017-06-15/paper-05-ef-board-meeting-15jun17-dsi-2-progress-report-ii.pdf.

maintenance of the Europeana Data Model (EDM) enables data exchange of cultural heritage metadata and guarantees its interoperability”³⁷.

O EDM é fundamentalmente um modelo para representação e armazenamento interno de registros. Os problemas que motivaram o desenvolvimento do EDM são os colocados pelo armazenamento de registros provenientes de diversas instituições, com diversos níveis de detalhamento das suas descrições diferentes. A Europeana por exemplo, deve manter separados registros da Mona Lisa, de Da Vinci, provenientes tanto do catálogo do Museu do Louvre³⁸ quanto da base de dados Joconde³⁹, do Ministério da Cultura da França (base que contém registros dos acervos de todos os museus da França), ao mesmo tempo que informa aos usuários que a consultam, que ambos os registros se referem ao mesmo objeto físico. A Europeana deve também manter os registros originais enviados por diferentes instituições e ao mesmo tempo, agregar informações adicionais a estes registros, enriquecendo-os semanticamente e proporcionando a seus usuários maiores facilidades de recuperação e exibição de registros.

Contempla assim tanto objetos digitais que são representações centradas no objeto de patrimônio cultural em si, por exemplo, a Mona Lisa (chamadas de “object-centric approach”), quanto representações do seu contexto – sua autoria, quem está retratado, quem encomendou a obra, suas mudanças de propriedade, etc. (chamadas de “event-centric approach”). O objeto digital que centraliza estes diferentes conjuntos de informações é chamado de Provided CHO – Objeto de Patrimônio Cultura recebido.

A figura a seguir mostra esses detalhes, onde se podem ver dois registros do mesmo objeto, um proveniente do museu do Louvre e outro da base de dados Joconde. O exemplo foi tirado de **Isaac, Antoine** (2013).

³⁷ Ver Força Tarefa para Criação e Governança de Mapeamento, Perfis e Extensões do EDM, <https://pro.europeana.eu/project/creation-and-governance-of-edm-mappings-profiles-and-extensions-task-force>.

³⁸ Ver <http://www.louvre.fr/moteur-de-recherche-oeuvres>.

³⁹ Ver http://www.culture.gouv.fr/documentation/joconde/fr/recherche/rech_libre.htm.

Figura 16. Representação interna do EDM com dois registros (ore:Aggregation-ex1 e ore:Aggregation-ex2) do mesmo objeto (edm:ProvidedCHO)



Fonte: elaborado pelo autor

3.2.5. MODELO RiC-CM (ARQUIVOS)

O modelo conceitual RiC-CM (INTERNATIONAL COUNCIL ON ARCHIVES. EXPERTS GROUP ON ARCHIVAL DESCRIPTION, 2016) no momento em que este livro está sendo escrito encontra-se ainda em pleno processo de discussão. Uma versão inicial, um “Consultation Draft v0.1”, foi publicada para discussão, receber comentários e sugestões, em setembro de 2016.

Depois das bibliotecas e museus, a área de arquivos vem percebendo a necessidade de ter como base de suas práticas de representação e descrição um modelo conceitual mais geral e abstrato que as diferentes normas de descrição arquivísticas específicas criadas ao longo dos anos, dando consistência e compatibilidade às entidades, características e relacionamentos existentes nas diferentes normas. Semelhantemente ao processo ocorrido com as bibliotecas a partir da criação do modelo FRBR em 1998 pela IFLA e com os museus a partir da criação do CIDOC CRM em 2000 pelo ICOM, o ICA – International Council of Archives criou

em 2012 o EGAD - Experts Group on Archival Description – que se encarregou do desenvolvimento do RiC-CM.

Em fins da década de 1980 a UNESCO começou a apoiar o ICA para o desenvolvimento de uma norma de descrição arquivística. A primeira versão foi publicada em 1990, desenvolvida pela Comissão *Ad-Hoc* para as Normas de Descrição (ICA/DDS), que mais tarde se tornou um comité permanente. Esta primeira versão foi publicada em 1994. Nesse mesmo ano, o Comité de Normas de Descrição assumiu o processo de revisão e atualização da norma, concluído em 2000. Deste esforço surgiu a Norma ISAD(G) – General International Standard Archival Description, ou Norma Geral Internacional de Descrição Arquivística. À ISAD(G) seguiram-se a ISAAR (CPF) – International Standard Archival Authority Record for Corporate Bodies, Persons and Families, ou Norma Internacional de Registo de Autoridade Arquivística para Pessoas Coletivas, Pessoas Singulares e Famílias, em 1996; as ISDIAH: International Standard for Describing Institutions with Archival Holdings, ou Norma Internacional para a Descrição de Instituições com Acervo Arquivístico; e a ISDF - International Standard for Describing Functions, ou Norma Internacional para Descrição de Funções, ambas em 2008.

Outra motivação mencionada para o modelo é o desenvolvimento contínuo das tecnologias de informação como bancos de dados relacionais consultáveis pela linguagem SQL, documentos em XML, etc., que permitiram o surgimento de novos métodos de descrição e acesso aos registos arquivísticos. Em especial são mencionadas as tecnologias LOD, que tem o potencial de permitir a integração dos acervos arquivísticos com acervos de outras instituições de herança cultural. A esse respeito o documento do modelo RiC-CM (INTERNATIONAL COUNCIL ON ARCHIVES. EXPERTS GROUP ON ARCHIVAL DESCRIPTION, 2016, p. 2) diz: “RiC-O will provide the archival community with the ability to make archival description available using the techniques of Linked Open Data (LOD) and employ a conceptual vocabulary and structure that is specific to archival description”.

O modelo conceitual RiC-CM é um modelo centrado no documento de arquivo (entidade “RiC-E1 Record”). Possui 14 entidades que poderiam ser agrupadas como se segue.

- RiC-E1 Record e suas partes ou conjuntos: RiC-E2 Record Component e RiC-E3 Record Set, que representam os tradicionais agrupamentos hierárquicos de documentos como fundos, séries e dossiês, mas também novos agrupamentos dinâmicos, como os documentos resultados de uma consulta.
- RiC-E4 Agent, os diferentes tipos de agentes envolvidos com documentos.
- ações, funções, papéis e mandatos relacionados a agentes: RiC-E5 Occupation, RiC-E6 Position, RiC-E7 Function, RiC-E8 Function (Abstract), RiC-E9 Activity e RiC-E10 Mandate.
- RiC-E11 Documentary Form para as diferentes e variadas características formais de um documento.
- Entidades contextuais como RiC-E12 Date, RiC-E13 Place e RiC-E14 Concept/Thing.

Essas 14 entidades se relacionam através de 792 relacionamentos. Muitos desses relacionamentos incluem uma versão no tempo atual (“is”) e uma versão no passado (“was”), para registrar por exemplo, o histórico de mudanças na proveniência de documentos, como mostra o quadro a seguir.

Relation Number	Relation Domain	Relation Name	Relation Range	Inverse Relation Name and Number	Comments
RiC-R18	Record	is member of	Record Set	has member (RiC-R135)	See also RiC-R20
RiC-R20	Record	was member of	Record Set	had member (RiC-R134)	See also RiC-R18

Fonte: INTERNATIONAL COUNCIL ON ARCHIVES. EXPERTS GROUP ON ARCHIVAL DESCRIPTION (2016)

Até o momento, não existem nem uma versão gráfica do RiC-CM nem uma versão como ontologia computacional em RDFS ou OWL. Estas versões estão programadas para serem publicadas no início do ano de 2018.

3.2.6. OUTROS MODELOS CONCEITUAIS NA ÁREA DE INFORMAÇÃO E DOCUMENTAÇÃO

Por iniciativa de Library of Congress, EUA, vem sendo desenvolvido um modelo conceitual alternativo ao FRBR, o BIBFRAME anunciado pela LC em 2012⁴⁰. Esse modelo⁴¹, simplifica as relações entre Obra, Expressão, Manifestação e Item do modelo FRBR, prevendo somente três entidades, ou três níveis de abstração: Work (Obra), Instance (equivalente a Manifestação) e Item. Esta simplificação se deve ao caráter bastante abstrato que possuem os relacionamentos entre Obra, como a criação artística abstrata de um autor e sua Expressão, a codificação de uma Obra em um sistema simbólico (um idioma), que por sua vez se relaciona então com uma Manifestação concreta, uma determinada edição em um idioma. Além dessas entidades estão também previstas no BIBFRAME entidades como Agent, Subject e Event.

As motivações para o desenvolvimento do BIBFRAME são assim colocadas pela LC:

Initiated by the Library of Congress, BIBFRAME provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world that is grounded in Linked Data techniques. A major focus of the initiative is to determine a transition path for the MARC 21 formats while preserving a robust data exchange that has supported resource sharing and cataloging cost savings in recent decades. (LIBRARY OF CONGRESS, 2011).

É bastante ilustrativo e esclarecedor o relatório de 2011 (LIBRARY OF CONGRESS, 2011) que deu origem ao BIBFRAME, onde a LC analisa a situação das bibliotecas e da catalogação, diante da emergência de novas tecnologias. Aqui também as tecnologias LOD são mencionadas.

Uma iniciativa que propõe integrar o modelo conceitual BIBFRAME com outros vocabulários diversos, criando um ambiente que permita uma rede de interligações significativas que contextualizem a informação bibliográfica é o LD4L – Linked Data for Libraries⁴². A proposta

⁴⁰ Ver <http://www.loc.gov/bibframe/news/bibframe-052212.html>.

⁴¹ Ver <https://www.loc.gov/bibframe/>.

⁴² Ver <https://www.ld4l.org/>

é integrar vocabulários e ontologias bibliográficas (BIBFRAME, MODS, EAD), com autoridades (VIVO /CAP, ORCID, ISNI⁴³, VIAF) e com informação sobre utilização (citações, comentários, circulação). Trata-se de um projeto patrocinado pela Mellon Foundation para o período 2015-2015.

3.2.7. CONSIDERAÇÕES FINAIS

A maioria dos modelos conceituais na área de informação e documentação é representada também como ontologias computacionais, geralmente usando linguagens como RDFS (RDF SCHEMA 1.1, 2014) e OWL (OWL ONTOLOGY WEB LANGUAGE OVERVIEW, 2004). É importante discutir um pouco mais este ponto. As duas linguagens, embora derivadas da XML e RDF, se distinguem pela sua capacidade expressiva, isto é, que grau de restrições ou de precisão na especificação da semântica exata dos termos de uma ontologia elas suportam.

RDFS – RDF Schema language – oferece um vocabulário para descrever ontologias formadas por classes, propriedades entre classes, subclasses e subpropriedades, que classes compõem o domínio e escopo de uma propriedade, e relacionamentos classe-subclasse. OWL – Ontology Web Language – é mais expressiva, agrega ao vocabulário RDFS a possibilidade de especificar os elementos de uma classe, que classes não possuem nenhum elemento comum (são disjuntas), a cardinalidade, isto é, o número de indivíduos ou instâncias de cada classe que participa de uma propriedade, se uma propriedade é simétrica, além de possibilitarem um grande número de possíveis tipos (“type”) para cada classe. Geralmente as ontologias computacionais formatadas com RDFS são chamadas de ontologias leves (“lightweight ontologies”) em oposição às ontologias pesadas (“heavyweight ontologies”) formatadas em OWL.

É também bastante significativa a semelhança entre os processos nas três áreas, que levaram primeiramente ao desenvolvimento de normas e padrões específicos a partir dos métodos anteriores, com origem no século XIX e baseados na experiência pessoal e nas melhores práticas; entre esses pode-se mencionar o CCAM2, o MARC, as categorias de informação e o modelo de banco de dados relacional do ICOM, a ISAD(G) e suas normas complementares, o EAD. Esse esforço de padronização evoluiu, tornou-se mais formal e abstrato, resultando nos modelos conceituais das três áreas.

⁴³ International Standard Name Identifier (Norma ISO 27729), ver <http://www.isni.org/>.

4. PUBLICAÇÃO DE CATÁLOGOS NA WEB USANDO AS TECNOLOGIAS LOD

Este capítulo é voltado para a aplicação prática das questões, metodologias e tecnologias discutidas nos capítulos anteriores, isto é, para a publicação de acervos de arquivos, bibliotecas e museus como dados abertos interligados. Em especial, são destacadas as novas tarefas dos curadores de acervos digitais.

4.1. AS NOVAS TAREFAS DO CURADOR DIGITAL NA PUBLICAÇÃO DE ACERVOS COMO LOD

Como foi visto nos capítulos anteriores acervos ganham novas potencialidades ao serem convertidos para o formato digital. Para que estas potencialidades sejam desenvolvidas ao máximo, são necessárias toda um conjunto de novas tarefas de gestão que não existiam em relação aos acervos físicos. A estas novas tarefas a literatura vem denominando de curadoria digital.

Um curador, segundo o Oxford Dictionary, é “A keeper or custodian of a museum or other collection”¹. Uma definição mais abrangente e completa é a da entrada da Wikipedia:

¹ <https://en.oxforddictionaries.com/definition/curator>.

... (from Latin: *curare* meaning “take care”) is a manager or overseer. Traditionally, a curator or keeper of a cultural heritage institution (i.e., gallery, museum, library or archive) is a content specialist responsible for an institution’s collections and involved with the interpretation of heritage material.²

O Digital Curation Centre³, uma instituição voltada para a curadoria de dados digitais científicos e de pesquisa, define curadoria digital como “...involves maintaining, preserving and adding value to digital research data throughout its lifecycle”. Se substituíssemos “digital research data” por “acervos digitais em memória e cultura” teríamos uma boa aproximação do que seria a curadoria digital no contexto específico deste livro.

Por sua vez, o Digital Curation Institute⁴, da Universidade de Toronto, apresenta uma definição de curadoria digital mais ampla: “Digital curation is the active involvement in the management, including the preservation, of digital resources for future use”.

Com tudo que foi apresentado nos capítulos anteriores, pode-se relacionar como tarefas de um curador digital na publicação de acervos como dados abertos interligados, as seguintes:

- desenvolver um processo planejado de criar cópias ou imagens digitais dos itens de maior valor histórico, cultural e educacional de um acervo, segundo os padrões e melhores práticas. Estas cópias ou imagens vão se constituir no acervo digital, um *novo acervo*, tão valioso e importante quanto o acervo físico e que vai demandar cuidados e gestão tanto quanto o acervo físico.
- garantir, através de diferentes procedimentos técnicos e de gestão, a preservação dos objetos digitais resultantes do processo de criação do acervo digital.
- além dos metadados já assinalados aos itens físicos, assinalar aos objetos digitais do acervo novos e apropriados metadados (DAPERT; ENDERS, 2010). Entre esses novos e específicos

² <https://en.wikipedia.org/wiki/Curator>.

³ Ver <http://www.dcc.ac.uk/digital-curation/what-digital-curation>.

⁴ [http://dci.school.utoronto.ca/what-is-digital-curation/#\[2\]](http://dci.school.utoronto.ca/what-is-digital-curation/#[2]).

metadados deve ser assinado, conforme discutido na Seção 1.4, um *identificador persistente* ao objeto digital. A instituição e o curador do acervo devem planejar esta questão cuidadosamente de modo a se comprometerem com a permanência indefinidamente dos identificadores assinalados a cada objeto digital. Esta questão é fundamental para possibilitar a publicação do objeto digital como dados abertos interligados.

Além destas questões, que envolvem o que vem sendo chamado de curadoria digital, ligada principalmente à preservação do acervo digital, outras questões específicas devem ser observadas para a publicação do acervo como dados abertos interligados. Estas questões serão aprofundadas nas seções seguintes.

4.3. CONVERTENDO CAMPOS DE CATÁLOGOS PARA DADOS ABERTOS INTERLIGADOS

A conversão de registros de catálogos para RDF é uma atividade necessária para a publicação de acervos de arquivos, bibliotecas e museus segundo as tecnologias LOD. Devido ao volume de dados geralmente existentes em um catálogo essa operação deve ser automatizada. Programas leem os registros de um catálogo, convertendo-os e gerando registros equivalentes em RDF. Existem várias ferramentas que fazem essa tarefa como MARiMBA, MarcEdit, PHP MARC, etc., que serão vistos na seção 4.6.

O mapeamento de registros de um catálogo para triplas RDF é uma tarefa que consiste basicamente nos seguintes passos:

- 1º. O identificador do registro no catálogo será o IRI da declaração “about” do documento RDF;
- 2º. os nomes dos diferentes campos do registro no catálogo serão os diferentes predicados do documento RDF;
- 3º. os valores contidos nos diferentes campos do registro no catálogo serão os objetos dos respectivos predicados do documento RDF.

Veamos na figura seguinte como o registro de uma tabela em um catálogo pode ser convertido para RDF.

Figura 17. Registro de uma tabela convertido para RDF

PROFESSORES

MATRÍCULA	NOME	EMAIL	CVLATTES	DEPTO
311543	C. H. Marcondes	marcon@vm.uff.br	http://lattes.cnpq.br/6126712875843734	GCI


```

<?xml version="1.0">
<rdf:RDF
xmlns= http://www.w3.org/1999/02/22-rdf-syntax-ns
<rdf:Description rdf:about "http://www.uff.br/professores/311543">
<rdf:type>professor</rdf:type>
<nome>C. H. Marcondes</nome>
<email>marcon@vm.uff.br</email>
<cvlattes>http://lattes.cnpq.br/6126712875843734</cvlattes>
<depto>http://www.uff.br/departamentos/GCI</depto>
</rdf:Description>
</rdf:RDF>
    
```

Fonte: elaborado pelo autor

Neste segundo exemplo um registro MARC é convertido para RDF.

Figura 18. Registro MARC convertido para RDF

LC Online Catalog

LC Online Catalog Quick Search Search [Browse](#) [Advanced Search](#) [Keyword Search](#)

[Reverse Your Search](#) [Return to Titles List](#) [Search History](#) [Account Info](#) [Help](#) | [LC Authorities](#)

Dom Casmurro.

4 of 25

Full Record **MARC Tags**

```

000 00632cam a2200193u 4500
001 8952033
005 20000325131940.0
008 820510s1966 bi 000 0 por
035 _ |9 (DLC) 67076436
908 _ |a 0 |b cbc |c premium |d u |e no |f |g |h |i |j |k |l |m |n |o |p |q |r |s |t |u |v |w |x |y |z |AA |AB |AC |AD |AE |AF |AG |AH |AI |AJ |AK |AL |AM |AN |AO |AP |AQ |AR |AS |AT |AU |AV |AW |AX |AY |AZ |BA |BB |BC |BD |BE |BF |BG |BH |BI |BJ |BK |BL |BM |BN |BO |BP |BQ |BR |BS |BT |BU |BV |BW |BX |BY |BZ |CA |CB |CC |CD |CE |CF |CG |CH |CI |CJ |CK |CL |CM |CN |CO |CP |CQ |CR |CS |CT |CU |CV |CW |CX |CY |CZ |DA |DB |DC |DD |DE |DF |DG |DH |DI |DJ |DK |DL |DM |DN |DO |DP |DQ |DR |DS |DT |DU |DV |DW |DX |DY |DZ |EA |EB |EC |ED |EE |EF |EG |EH |EI |EJ |EK |EL |EM |EN |EO |EP |EQ |ER |ES |ET |EU |EV |EW |EX |EY |EZ |FA |FB |FC |FD |FE |FF |FG |FH |FI |FJ |FK |FL |FM |FN |FO |FP |FQ |FR |FS |FT |FU |FV |FW |FX |FY |FZ |GA |GB |GC |GD |GE |GF |GG |GH |GI |GJ |GK |GL |GM |GN |GO |GP |GQ |GR |GS |GT |GU |GV |GW |GX |GY |GZ |HA |HB |HC |HD |HE |HF |HG |HH |HI |HJ |HK |HL |HM |HN |HO |HP |HQ |HR |HS |HT |HU |HV |HW |HX |HY |HZ |IA |IB |IC |ID |IE |IF |IG |IH |II |IJ |IK |IL |IM |IN |IO |IP |IQ |IR |IS |IT |IU |IV |IW |IX |IY |IZ |JA |JB |JC |JD |JE |JF |JG |JH |JI |JJ |JK |JL |JM |JN |JO |JP |JQ |JR |JS |JT |JU |JV |JW |JX |JY |JZ |KA |KB |KC |KD |KE |KF |KG |KH |KI |KJ |KK |KL |KM |KN |KO |KP |KQ |KR |KS |KT |KU |KV |KW |KX |KY |KZ |LA |LB |LC |LD |LE |LF |LG |LH |LI |LJ |LK |LL |LM |LN |LO |LP |LQ |LR |LS |LT |LU |LV |LW |LX |LY |LZ |MA |MB |MC |MD |ME |MF |MG |MH |MI |MJ |MK |ML |MM |MN |MO |MP |MQ |MR |MS |MT |MU |MV |MW |MX |MY |MZ |NA |NB |NC |ND |NE |NF |NG |NH |NI |NJ |NK |NL |NM |NO |NP |NQ |NR |NS |NT |NU |NV |NW |NX |NY |NZ |OA |OB |OC |OD |OE |OF |OG |OH |OI |OJ |OK |OL |OM |ON |OO |OP |OQ |OR |OS |OT |OU |OV |OW |OX |OY |OZ |PA |PB |PC |PD |PE |PF |PG |PH |PI |PJ |PK |PL |PM |PN |PO |PP |PQ |PR |PS |PT |PU |PV |PW |PX |PY |PZ |QA |QB |QC |QD |QE |QF |QG |QH |QI |QJ |QK |QL |QM |QN |QO |QP |QQ |QR |QS |QT |QU |QV |QW |QX |QY |QZ |RA |RB |RC |RD |RE |RF |RG |RH |RI |RJ |RK |RL |RM |RN |RO |RP |RQ |RR |RS |RT |RU |RV |RW |RX |RY |RZ |SA |SB |SC |SD |SE |SF |SG |SH |SI |SJ |SK |SL |SM |SN |SO |SP |SQ |SR |SS |ST |SU |SV |SW |SX |SY |SZ |TA |TB |TC |TD |TE |TF |TG |TH |TI |TJ |TK |TL |TM |TN |TO |TP |TQ |TR |TS |TT |TU |TV |TW |TX |TY |TZ |UA |UB |UC |UD |UE |UF |UG |UH |UI |UJ |UK |UL |UM |UN |UO |UP |UQ |UR |US |UT |UU |UV |UW |UX |UY |UZ |VA |VB |VC |VD |VE |VF |VG |VH |VI |VJ |VK |VL |VM |VN |VO |VP |VQ |VR |VS |VT |VU |VV |VW |VX |VY |VZ |WA |WB |WC |WD |WE |WF |WG |WH |WI |WJ |WK |WL |WM |WN |WO |WP |WQ |WR |WS |WT |WU |WV |WW |WX |WY |WZ |XA |XB |XC |XD |XE |XF |XG |XH |XI |XJ |XK |XL |XM |XN |XO |XP |XQ |XR |XS |XT |XU |XV |XW |XX |XY |XZ |YA |YB |YC |YD |YE |YF |YG |YH |YI |YJ |YK |YL |YM |YN |YO |YP |YQ |YR |YS |YT |YU |YV |YW |YX |YY |YZ |ZA |ZB |ZC |ZD |ZE |ZF |ZG |ZH |ZI |ZJ |ZK |ZL |ZM |ZN |ZO |ZP |ZQ |ZR |ZS |ZT |ZU |ZV |ZW |ZX |ZY |ZZ
    
```

```

<?xml version="1.0">
<rdf:RDF
xmlns= http://www.w3.org/1999/02/22-rdf-syntax-ns
xmlns:dc="http://purl.org/dc/elements/1.1">
<rdf:Description rdf:about
"http://lccn.loc.gov/76076436">
<dc:creator>Machado de Assis</dc:creator>
<dc:title>Dom Casmurro</dc:title>
<dc:date>1966</dc:date>
</rdf:Description>
</rdf:RDF>
    
```

Request this Item

Print Record

Save Record

Email Record

Cite Record

Find It! **IRI**

Permalink: <https://lccn.loc.gov/76076436>

XML Formats:

Where to Request

CALL NUMBER [PQ9697.M18.D6.1966.LANDOV.R](#)
Copy 1

Request in Jefferson or Adams Building Reading Rooms - STORED OFFSITE

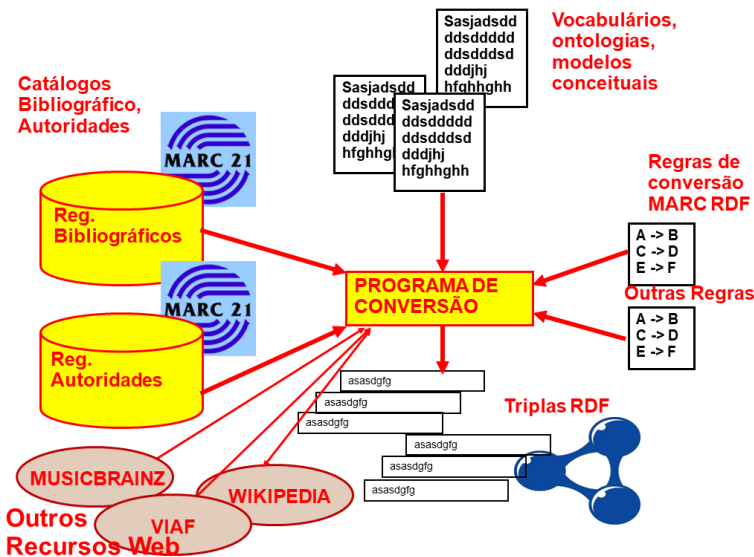
Status Not Charged

Fonte: elaborado pelo autor

Nesse exemplo, o Permalink⁵ do registro MARC da LC se transforma no IRI do registro RDF, o campo 100 – Entrada Principal, Nome pessoal, se transforma para o predicado dc:creator e seu conteúdo, “Machado de Assis”, se transforma no objeto deste predicado; o campo 245 – Título, se transforma no predicado dc:title e seu conteúdo, “Dom Casmurro”, se transforma no objeto deste predicado; o subcampo \$c do campo 260 – Publicação, Distribuição, Data de publicação, se transforma no predicado dc:date e seu conteúdo, “1966”, se transforma no objeto deste predicado.

Estas transformações, muitas vezes chamadas na literatura sobre o assunto, de *mapeamentos*, são simples e praticamente automáticas. Podem ser colocadas em tabelas como regras de conversão ou mapeamentos, e lidas pelos programas que convertem os catálogos para RDF para orientar o processamento dos registros do catálogo. Em linhas gerais a conversão de um catálogo para RDF se dá como ilustrado na figura seguinte, neste caso um catálogo bibliográfico.

Figura 19. Visão geral do processo de conversão de registros MARC de catálogos para RDF



Fonte: elaborado pelo autor

⁵ Ver em <https://lcn.loc.gov/lcnperm-faq.html>.

No processo ilustrado na figura registros MARC bibliográficos e de autoridades são lidos pelos programas de conversão, que consultam tabelas de regras específicas onde estão definidos os mapeamentos entre os campos MARC e RDF; esse mapeamentos podem incluir referências (IRIs) a vocabulários específicos ou a outros recursos Web em que sejam formatados como dados abertos interligados, como Wikipedia/DBpedia, VIAF, MusicBrainz; essas referências enriquecem os registros em RDF gerados com “links” semânticos. Maiores detalhes sobre a questão da promoção e enriquecimento dos serão vistos na próxima seção.

4.4. PROMOÇÃO E ENRIQUECIMENTO DE RELACIONAMENTOS ENTRE DE CATÁLOGOS PUBLICADOS COMO LOD E OUTROS RECURSOS

A valorização dos conteúdos de catálogos de arquivos, bibliotecas e museus pela sua publicação como dados abertos interligados está diretamente ligada aos “links” que podem ser feitos desses dados para outros e de terceiros para esses dados, fornecendo uma rica contextualização. Os curadores de acervos digitais destas instituições têm um papel fundamental em promover esses “links”, identificando potenciais relacionamentos entre registros de seu catálogo com registros de outros catálogos ou com outros recursos na Web.

A geração desses “links” acontece no momento da conversão do catálogo para RDF. Esta conversão geralmente é automática, feita por programas de conversão. Pode acontecer uma única vez ou periodicamente, na medida que novos “links” potenciais sejam identificados pelo curador. A conversão também pode acontecer com todos os registros do catálogo, ou somente com aqueles registros que o curador considere como os *destaques* da sua coleção; neste caso, esses destaques devem ser identificados e os registros respectivos marcados, para que sejam somente eles os registros que serão convertidos para RDF; essa marcação pode ser feita, por exemplo, no caso de um catálogo bibliográfico, usando o campo 590 – Notas locais, do formato MARC, para assinalar um registro como sendo “destaque”.

O curador do acervo, tendo conhecimento de sua coleção, dos seus destaques, e do potencial de fazer “links” para outros recursos, deve supervisionar a especificação dos programas que farão a conversão, interferindo no que foi chamado na figura 12, de Tabelas de Regras de

Conversão. Nestas tabelas o curador poderia especificar regras como as que estão exemplificadas a seguir.

TABELA DE REGRAS DE CONVERSÃO

CAMPO MARC para ->	No.	RECURSO WEB	REG RA	RDF resultante
100 – Entrada Principal, Nome pessoal, subcampo \$a ex:1001#\$aMiguel de Cervantes	1	DBpedia	Contruir um “link” para DBpedia com conteúdo do campo 100\$a, substituindo espaços por “_”: http://dbpedia.org/page/Miguel_de_Cervantes ; Se existe uma entrada, criar um “link” ⁶ .	<dc:creator> http://dbpedia.org/page/Miguel_de_Cervantes </dc:creator>
240 – Título Uniforme, subcampo \$a Ex: 24010\$aDon Quijote	2	The Movie Database	Fazer uma consulta com o conteúdo do campo 240\$a; Se for encontrado um registro, criar um “link”, Se foi encontrada uma entrada, criar um “link”, https://www.themoviedb.org/movie/31987-don-quijote-de-orson-welles	<frbr:has_a_transformation> https://www.themoviedb.org/movie/31987-don-quijote-de-orson-welles </frbr:has_a_transformation>
240 – Título Uniforme, subcampo \$a Ex: 24010\$aDon Quijote	3	MoMa – The Museum of Modern Art	Fazer uma consulta com o conteúdo do campo 240\$a; Se for encontrado um registro, criar um “link”, https://www.moma.org/collection/works/68157?locale=de	<wikim:inspiredby> https://www.moma.org/collection/works/68157?locale=de </wikim:inspired by>
651 – Nome geográfico, subcampo \$a Ex: 651#0\$aLa Mancha	4	Getty Thesaurus of Geographic Names	Fazer uma consulta com o conteúdo do campo 651\$a; Se for encontrado um registro, criar um “link” com o IRI http://vocab.getty.edu/tgn/7295350 .	<dc:coverage > http://vocab.getty.edu/tgn/7295350 </dc:coverage >

Fonte: elaborado pelo autor

⁶ A mesma regra pode ser usada para recursos como a DBpedia ou o arquivo de autoridades VIAF; podem também ser usadas as três regras, criando três “links” distintos no registro RDF.

A Tabela de Regras de Conversão pode trazer outros campos, como critérios adicionais de busca (no caso da Regra 2, “Orson Welles”), vocabulários utilizados na geração do documento RDF (Dublin Core, no caso da Regra 1, FRBR, no caso da Regra 2, wikim, no caso da Regra 3).

Naturalmente também os recursos para os quais serão feitos “links” deverão prover IRIs para seus recursos e permitir consultas em SPARQL; esse é o caso da DBpedia, usada na Regra 1, e do Getty Thesaurus of Geographic Names, usado na Regra 4, mas não o do The Movie Database, usado na Regra 2 e do The Museum of Modern Art, usado na Regra 3. A ideia aqui foi somente ilustrar o que seria uma Tabela de Regras de Conversão e como ela deveria funcionar no processo de conversão de registros de catálogos para RDF. O curador do acervo tem assim papel crucial na especificação das Tabelas de Regras de Conversão que vão orientar o programa que fará a conversão do seu acervo.

Entre os recursos importantes para que sejam feitos “links” a partir de registros de um catálogo estão a Wikipédia, e DBpedia, o VIAF, o Geonames, os vocabulários da Getty Foundation, entre outros. Esses vocabulários agregam semântica e informação contextual, enriquecendo os registros dos catálogos publicados como LOD, além de serem totalmente compatíveis com essa tecnologias. Além disso, a Wikipédia tem como característica poder ser editada por qualquer um, possibilitando que um curador do acervo faça “links” de entradas da Wikipedia para registros de sua coleção publicados como LOD.

Existe também a possibilidade do curador do acervo, usando ferramentas especiais como o RelFinder ou Silk, que serão vistas na próxima seção, possa descobrir “links” entre o seu acervo e outros. O próprio curador do acervo, ou outros curadores como historiadores da arte, jornalistas, professores, poderão também anotar⁷ e criar “links” específicos e até autorais, relacionando objetos digitais de acervos diversos. Pode-se, por exemplo, relacionar um objeto digital contendo imagens e metadados do MAC – Museu de Arte Contemporânea de Niterói⁸ com objetos digitais contendo imagens das plantas do museu ou fotografias da sua construção. Ou relacionar Guernica, de Picasso, que está no Museu Reina Sofia em Madrid, com seus esboços. Ou relacionar a Mona Lisa de

⁷ Anotações semânticas, ver <http://www.openannotation.org/spec/core/20130208/index.html>.

⁸ Ver https://pt.wikipedia.org/wiki/Museu_de_Arte_Contempor%C3%A2nea_de_Niter%C3%B3i.

Da Vinci, do Museu do Louvre em Paris, com suas várias versões “a partir de”, feitas por artistas como Fernando Botero, Andy Warhol ou Salvador Dalí, que estão em diferentes museus e galerias. As possibilidades de explorar relações culturalmente relevantes entre objetos de acervos digitais publicados como LOD para fins de criar exposições “virtuais” ou aulas “virtuais” são inúmeras (MARCONDES, 2020).

O curador do acervo deverá, entre outras tarefas, conhecer profundamente seu acervo, identificar e conhecer outros recursos com potencial para terem “links” de e para registros do seu catálogo, criar, juntamente com os especialistas em informática, as regras para conversão de campos específicos dos registros do seu catálogo, entrar em contato e cooperar com os gestores de outros acervos ou dos recursos para onde os registros do seu catálogo fazem “links”, para que estes recursos tenham também “links” recíprocos para os registros do seu catálogo. Estas são algumas das tarefas de um curador digital para gerir e promover seu acervo em um ambiente de dados abertos interligados.

4.5. FERRAMENTAS

Existem hoje disponíveis uma grande quantidade de ferramentas de software que auxiliam o curador digital na gestão, promoção e publicação de seu acervos segundo as tecnologias de dados abertos interligados. Apresentamos aqui as mais conhecidas, organizadas em categorias, como se segue.

- Navegadores de dados abertos

- GFacet, <http://www.visualdataweb.org/gfacet.php>, navegador gráfico que permite explorar e filtrar dados em RDF através de facetas.

- LodLive, <http://en.lodlive.it/>, navegador da Web de dados que permite que, fornecendo um IRI ou entrando palavras-chaves, navegar pelos dados, seguir os “links” existentes. Os datasets disponíveis são a DBpedia e a Freebase. Os dados são apresentados como grafos.

- Open Link Data Explorer, <http://demo.openlinksw.com/rdfbrowser2/>, permite navegar pela Web de dados

fornecendo um IRI ou a partir de palavras-chave. Os dados podem ser apresentados segundo diversas visões, como triplas, como grafos, como tabelas. Também permite que se faça várias perguntas acerca do dado que está sendo acessado: “What”, “Where”, “When”, “Who”.

- Quick and Dirty RDF browser, <http://graphite.ecs.soton.ac.uk/browser/>, navegador simples que permite explorar a Web de dados a partir de um IRI. OS dados são mostrados em formatos como RDF/XML, N-Triple e Turtle.

- Editores RDF.

- RDF Editor, <https://sourceforge.net/projects/rdfeditor/>, editor simples que permite criar documentos RDF. Pode ser usado por um curador digital para criar exemplos dos dados do seu catálogo quando convertidos para RDF, para orientar os especialistas em informática que forem desenvolver os programas de conversão.

- MarcEdit, <http://marcedit.reeset.net/>, é um editor de registros MARC que permite alterar campos e conteúdos de registros MARC. Pode ser usado como ferramenta para um projeto piloto de conversão/publicação de catálogos bibliográficos para RDF, graças a sua funcionalidade de converter registros MARC em RDF.

- Validadores.

- SPARQL Query validator, <http://sparql.org/query-validator.html>

- W3C RDF validator, <https://www.w3.org/RDF/Validator>

- Conversores de registros de catálogos

- MARiMBA, mayor2.dia.fi.upm.es/oeg-upm/index.php/en/technologies/228-marimba/, ferramenta construída pelo

Ontology Engineering Group, da Universidad Politecnica de Madrid, para o projeto de publicação de dados abertos da Biblioteca Nacional de España. O programa converte registros MARC em RDF.

- Catmandu, <http://librecat.org>, ferramenta que converte registros de catálogos bibliográficos em formato MARC, MOD, Dublin Core, para RDF.

- PHP MARC, <https://sourceforge.net/projects/phpmarc/>, rotina em linguagem PHP que lê registros MARC, podendo gerar saídas em diferentes formatos.

- Reformatadores, agregadores de dados

- KARMA, <http://usc-isi-i2.github.io/karma/>, do Information Science Institute da University of South California, para integração de dados de diferentes fontes, como bancos de dados, planilhas, arquivos CSV, XML, etc., tendo como base uma ontologia escolhida pelo usuário. A ferramenta permite editar, criar “links” para outros recursos e publicá-los os dados como LOD.

- OpenRefine, <https://openrefine.org/>, anteriormente Google Refine, agora um pode ser baixada livremente, permite “limpar” ou transformando-os de um formato em outro, enriquecendo-os com Web services ou dado externos de outros recursos Web.

- Bancos de dados.

- D2R Server, <http://d2rq.org/d2r-server>, ferramenta para acessar bancos de dados relacionais com SPARQL e receber resultados como dados abertos.

- Sparqlify, <http://aksw.org/Projects/Sparqlify.html>, ferramenta que funciona como uma interface a bancos de dados relacionais, permitindo definir visões dos dados como RDF e consultá-las usando SPARQL.

- OpenLink Virtuoso Universal Server, <https://virtuoso.openlinksw.com/universal-server/>, servidor de dados que combina dados relacionais, de documentos textuais, em XML, em triplas RDF, com Web services. Os dados podem ser consultados em SPARQL.

- GraphDB, <https://ontotext.com/products/graphdb/>, é um gerenciador de dados que estejam em formato de grafos RDF. Os dados podem ser carregados a partir de arquivos com triplas, exibidos como gráficos em rede e consultados usando SPARQL. É uma boa e simples alternativa para criar e testar uma versão piloto de um repositório de triplas, pois existe uma versão gratuita que opera em um computador PC, onde podem ser carregados e testados dados que resultam da conversão do catálogo.

- Ferramentas para descobrir “links”

- SILK, <http://silkframework.org/>, ferramenta gratuita para comparar dois datasets publicados como dados abertos interligados e identificar/sugerir possíveis “links” entre eles.

- RelFinder, <http://www.visualdataweb.org/relfinder/relfinder.php>, ferramenta que permite explorar possíveis relações – “links” – existentes entre dois conceitos. A ferramenta trabalha conceitos extraídos de entradas da DBpedia.

Com o RelFinder um curador digital pode explorar possíveis “links” entre conceitos existentes nos registros do seu catálogo e entre esses e recursos internos, e a partir daí criar “links” dos registros do seu catálogo. A ferramenta permite, por exemplo, identificar possíveis “links” existentes entre conceitos como “Miguel de Cervantes” e “Magic Realism” e, com os “links” descobertos, criar “links” adicionais para conceitos como “Günther Grass”, “Willian Faulkner”, “Gustave Flaubert”, “Janet Frame”, “Steve Erickson”, “C. K. Chesterton”, “Gabriel Garcia Marques”, “Patrick

Süsskind”, “Haruki Murakami”, entre outros, no conjunto de dados LOD do seu catálogo.

- Outras ferramentas

- Pundit Annotator, <http://thepund.it/>, ferramenta desenvolvida no âmbito do projeto “DM2E - Digitised Manuscripts to Europeana⁹”, é uma extensão que pode ser instalada em no Google Chrome, permitindo navegar por diferentes páginas Web, iluminar textos das páginas, anotar comentários e salvá-los como “favoritos”. Podem ser feitas relações semânticas na forma de triplas entre as anotações de diferentes páginas e de uma página com a Wikipedia/DBpedia e com outros recursos semânticos. As relações podem ser recuperadas e/ou filtradas por diferentes critérios.

Vários outras ferramentas, classificadas por categorias, podem ser encontradas em <http://semanticweb.org/wiki/Tools.html>.

⁹ <https://pro.europeana.eu/project/dm2e>.

5. CONSIDERAÇÕES FINAIS

As tecnologias LOD, muito mais que ampliarem as possibilidades de acesso quando comparadas às tecnologias associadas aos sistemas de catálogos convencionais e bibliotecas digitais, se constituem, por várias razões, em um novo paradigma tecnológico. Nesse paradigma os conteúdos são significativamente mais independentes de programas específicos, podem ser publicados diretamente na Web, acessíveis por identificadores persistentes, podem ser enriquecidos, integrados e contextualizados não só por seus curadores originais – instituições de memória e cultura como arquivos, bibliotecas e museus – mas por terceiros que considerem significativo fazer “links” para estes recursos.

Além disto, estes conteúdos passam a ser estruturados e representados em formatos que permitem sua utilização tanto por usuários humanos quando por programas (CONCORDIA; GRADMANN; SIEBINGA, 2010). Isso permite seu acesso em larga escala não só para usuários finais humanos, como nos sistemas de catálogos e bibliotecas digitais convencionais, mas através de APIs, abrindo possibilidades inusitadas de reuso, recriação, integração com outras aplicações e promoção das chamadas indústrias criativas (IKONOMOV; SIMEONOV; PARVANOVA; ALEXIEV, 2013).

Fica claro também que um novo perfil de habilidades dos profissionais de informação profissionais se delineia com todas as questões que foram vistas aqui. O que na nomenclatura tradicional de arquivos, bibliotecas e museus se chamava de gestor, ou catalogador, indexador, descrição arquivística, ou descrição museológica, o profissional

que trabalhava como promoção e “marketing”, o que trabalhava com preservação de acervos, o curador, todas essas atribuições ganham outra dimensão com acervos digitais e com a possibilidade de publicá-los com as tecnologias de dados abertos interligados e integrá-los com acervos de outras instituições. Novas questões se colocam também como as questões jurídicas relacionadas com as licenças de uso abertas, a nova forma de cooperação com outras instituições que têm acervos complementares ou inter-relacionados.

Acervos digitais publicados na Web e as tecnologias de dados abertos interligados implicam também em um domínio sólido das tecnologias da informação, não pelo viés do programador, desenvolvedor de “software”, suporte de redes, mas sim pelo viés do gestor, daquele que conhece profundamente estas tecnologias, suas limitações para além de qualquer modismo, seu emprego e, principalmente, suas potencialidades. Conhecer as tecnologias, mas conhecer também profundamente seu acervo, seu potencial cultural e educativo, compreendê-lo como um recurso social que deve ser ao máximo devolvido e fruído pela sociedade.

Estas novas atribuições não podem mais ser vistas separadamente, preservação digital, curadoria digital, curadoria informacional e curadoria cultural necessariamente devem se integrar em um único processo. O novo paradigma tecnológico não só amplia os papéis sociais, culturais das instituições de memória e cultura, juntamente com seus profissionais, como lhes atribui um novo papel econômico que fortalece e amplia seu papel social, cultural, educativo, como recurso de cidadania e sua auto-sustentação.

REFERÊNCIAS

ANDRESEN, Leif. Dublin Core as a tool for interoperability: Common presentation of data from archives, libraries and museums. *In: DC-2006 INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS*, Colima, Mexico, October, 2006. *Proceedings* [...]. Colima: DCMÍ, 2006. Disponível em: <http://dcpapers.dublincore.org/index.php/pubs/article/viewFile/844/840>. Acesso em: inserir data de acesso.

BACA, Murtha. Prefácio. *In: HARPRING, Patricia. Vocabulários Controlados: terminologia para arte, arquitetura e outras obras culturais*. São Paulo: Secretaria de Estado de Cultura, Pinacoteca de São Paulo, 2016. p. 20 – 21.

BERNERS-LEE, Tim. Cool URIs don't change. *W3C*, 1998. Disponível em: <https://www.w3.org/Provider/Style/URI>. Acesso em: 13 mai. 2012.

BERNERS-LEE, Tim. Linked data. *W3C*, 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 13 mai. 2012.

BERNERS-LEE, Tim; CAILLIAU, Robert. WorldWideWeb: Proposal for a HyperText Project. *W3C*, 1990. Disponível em: <https://www.w3.org/Proposal>. Acesso em: 23 mar. 2002.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. *The semantic web*. Scientific American, May, 2001. Disponível em: <http://www.scian.com/2001/0501issue/0501berners-lee.html>. Acesso em: 24 mai. 2001.

BESEK, Jone M. Copyright Issues Relevant to the Creation of a Digital Archive: a Preliminary Assessmentbody. *Council of Library and Information Resources*, 2003. Disponível em: <https://www.clir.org/pubs/reports/pub112/body.html>. Acesso em: 14 nov. 2013.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data: the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, Hershey, 2009. Special Issue on Linked Data. Disponível em: <https://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 25 fev. 2011.

CAMARGO-MORO, Fernanda de. *Museu, aquisição/documentação: tecnologias apropriadas para a preservação dos bens culturais*. Rio de Janeiro: Livraria Eça, 1986.

CANDELA, Leonardo et al. Moving digital library service systems to the grid. In: *Peer-to-Peer, Grid, and Service-Oriented in Digital Library Architectures*. Springer, Berlin, Heidelberg, 2005. p. 236-259.

Candela, L., Castelli, D., Ferro, N., Koutrika, G., Meghini, C., Pagano, P., ... & Dobрева, M. The DELOS Digital Library Reference model. Foundations for digital Libraries (Version 0.98), 2008. Disponível em: <<http://eprints.port.ac.uk/4104/>>. Acesso em: 12 jun. 2015.

CARACCILOLO, Caterina *et al.* Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case. *International Journal of Metadata, Semantics and Ontologies*, United Kingdom, v. 7, n. 1, p. 65-75, 2012. Disponível em: http://eprints.rclis.org/17735/1/IJMSO%207_1_Paper%206_PublishedVersion%5B1%5D.pdf. Acesso em: 19 mar. 2020.

CHEN, Peter Pin-Shan. The Entity-Relationship Model-Toward a Unified View of Data. *ACM Transactions on Database Systems* v. 1 n.1, p. 9-36, 1976.

CONCORDIA, Cesare; GRANDMANN, Stefan; SIEBINGA, Sjoerd. Not just another portal, not just another digital library: a portrait of Europeana as an application program interface. *IFLA Journal*, London, v. 36, n. 1, p.61-69, 2010. DOI: <http://dx.doi.org/10.1177/0340035209360764>.

COOK, Terry. Arquivos pessoais e arquivos institucionais: para um entendimento arquivístico comum da formação da memória de um mundo pós-moderno. *Estudos Históricos*, São Paulo, v. 11, n. 21, 1998.

CUENCA, Gonzalo Marco. Interoperabilidad en Europeana. In: SIMEÓN, Luis Fernando Ramos; AVILÉS, Rosario Arquero (coord.). *Europeana La Plataforma del Patrimonio Cultural Europeu*. Gijón: Ediciones Tres, 2014. p. 131-151.

DAPERT, Angela; ENDERS, Markus. Digital Preservation Metadata Standards. *ISQ – Information Standards Quarterly*, Baltimore, v. 22, n. 2, 2010. Special issue on digital preservation. Disponível em: https://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf. Acesso em: 20 nov. 2011.

DAVIDSON, Joy. *Persistent Identifiers. DCC Briefing Papers: Introduction to Curation*. Edinburgh: Digital Curation Centre, 2006. Handle: 1842/3368. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation>. Acesso em: 19 jan. 2010.

DENTON, William. FRBR and fundamental cataloguing rules. [Toronto, Ontario]: [W. Denton], [2009]. Disponível em: <http://www.miskatonic.org/library/frbr.html>>. Acesso em: 11 mai. 2011.

DIAZ, Gonzalo; ARENAS, Marcelo; BENEDIKT, Michael. Sparqlbye: Querying RDF data by example. *Proceedings of the VLDB Endowment*, New York, v. 9, n. 13, p.1533-1536, 2016. Disponível em: <http://www.vldb.org/pvldb/vol9/p1533-diaz.pdf>. Acesso em 20 de. 2018.

DIJKSHOORN, Chris; JONGMA, Lizzy; AROYO, Lora; VAN OSSENBRUGGEN, Jacco; SCHREIBER, Guus; TER WEELE, Wesley; WIELEMAKER, Jan. The Rijksmuseum Collection as Linked Data. *Semantic Web Journal*, Amsterdam, p. 1-10, 2017. Preprint. DOI: <http://dx.doi.org/10.3233/SW-170257>.

DOERR, Martin. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine*, Palo Alto, v. 24, n. 3, p. 75-92, 2003. Disponível em: <http://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1720/1618>. Acesso em: 7 abr. 2009.

DOERR, Martin; BEKIARI, Chryssoula; LEOEUF, Patrick. FRBRoo, a conceptual model for performing arts. *In: WORLD LIBRARY AND INFORMATION CONGRESS, 74th, IFLA GENERAL CONFERENCE AND COUNCIL, 10-14 August 2008, Québec, Canada. IFLA, 2008*. Disponível em: http://cidoc.mini.icom.museum/wp-content/uploads/sites/6/2018/12/29_papers.pdf. Acesso em: 7 abr. 2009.

DOERR, Martin; GRANDMANN, Stefan; HENNICKE, Steffen; ISAAC Antoine; MEGHINI, Carlo; VAN DE SOMPEL, Herbert. The Europeana Data Model (EDM). *In: WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY, Gothenburg, Sweden, August 2010. IFLA, 2010*. p. 10-15. Disponível em: <https://www.ifla.org/past-wlic/2010/149-doerr-en.pdf>. Acesso em: 18 abr. 2013.

DOERR, Martin; LEOEUF, Patrick. Modelling intellectual processes: the FRBR-CRM harmonization. *In: Digital libraries: Research and development*, First International DELOS Conference, Pisa, Italy, February 13-14, 2007. Springer, 2007. p. 114-123. Disponível em: http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/ConferencePapers/2006/Doerr_LeBoeuf.pdf. Acesso em 09 jun. 2015.

DOERR, Martin; STEAD, Stephen. *The CIDOC CRM, a Standard for the Integration of Cultural Information*. Imperial College, London, UK, May 22, 2009. Disponível em: http://old.cidoc-crm.org/docs/crm_for_imperial_2009.ppt. Acesso em: 3 fev. 2013.

DUBLIN CORE METADATA INITIATIVE. *Dublin Core Application Profile Guidelines*. 1995. Disponível em: <http://dublincore.org/usage/documents/profile-guidelines>. Acesso em: 25 nov. 2007.

DUBLIN CORE METADATA INITIATIVE. *Dublin Core Metadata Element Set. Version 1.1: Reference Description*, 02 June 2003. Disponível em: <http://dublincore.org/documents/dces>. Acesso: 25 nov. 2007.

EUROPEAN COMMISSION. *eContent: European digital content on the global network*. 2004. Disponível em: http://cordis.europa.eu/pub/econtent/docs/call1_brochure.pdf. Acesso em: 14 ago. 2014.

EUROPEAN COMMISSION. *eContent: European digital content on the global network. Customisation Projects*. 2009. Disponível em: http://cordis.europa.eu/econtent/customisation/cust_projects.htm. Acesso em: 14 ago. 2014.

EUROPEANA. *Definition of the europeana data model elements*. Version 5.2.7. Europeana, 2016. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.7_042016.pdf. Acesso em: 14 ago. 2014.

FOX, Edward A. Digital Libraries Initiative (DLI) Projects 1994-1999. *Bulletin of the Association for Information Science and Technology*, Silver Spring, v. 26, n. 1, 1999. Disponível em: <http://dx.doi.org/10.1002/bult.135>. Acesso em: 04 abr. 2002.

FOX, Edward A.; MARCHIONINI, Gary. Toward a worldwide digital library. *Communications of the ACM*, New York, v. 41, n. 4, p. 29-32, 1998. Disponível em: https://www.researchgate.net/profile/Edward_Fox/publication/27297582_Toward_a_Worldwide_Digital_Library/links/5574b0ee08aeb6d8c01945b3.pdf. Acesso em: 04 abr. 2002.

GARRIDO-ARILLA, María Rosa. *Teoría e historia de La catalogación de documentos*. Madrid: Síntesis, 1999.

HASHOFER, Bernhard; ISAAC, Antoine. data.europeana.eu - The Europeana Linked Open Data Pilot. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, The Hague. *Proceedings* [...]. DCMI, 2011. Disponível em: <http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/55>. Acesso em: 22 jul. 2015.

HUVILA, Isto. Participatory archive: towards decentralised curation, radical user orientation and broader contextualisation of records management. *Archival Science*, Netherlands, v. 8, n. 1, p. 15-36, 2008. Disponível em: <http://www.diva-portal.org/smash/get/diva2:287959/FULLTEXT01.pdf>. Acesso em: 25 out. 2015.

IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. *FRBR – Functional Requirements for Bibliographic Records: final report*. München: K. G. Sau, 1998. (UBCIM Publications New Series). Disponível em: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>. Acesso em: 02 fev. 2005.

IFLA WORKING GROUP ON THE FUNCTIONAL REQUIREMENTS AND NUMBERING OF AUTHORITY RECORDS _FANAR. *Functional Requirements for Authority Data A Conceptual Model. Final Report*. IFLA, 2008. Disponível em: https://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf. Acesso em: 02 fev. 2005.

IFLA WORKING GROUP ON THE FUNCTIONAL REQUIREMENTS FOR SUBJECT AUTHORITY RECORDS (FRSAR). *Functional Requirements for Subject Authority Data (FRSAD) A Conceptual Model*. IFLA, 2010. Disponível em: <https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>. Acesso em: 02 fev. 2005.

IKONOMOV, Nikola; SIMEONOV, Boyan; PARVANOV, Jana; ALEXIEV, Vladimir. Europeana Creative. EDM Endpoint. Custom Views. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, Frankfurt, v. 3, p. 1314-4006, 2013. Disponível em: <https://ontotext.com/documents/publications/2013/Ikonomov2013-EuropeanaCreative-EDM.pdf>. Acesso em: 20 set. 2018.

INTERNATIONAL COUNCIL OF MUSEUMS/INTERNATIONALS COMMITTEE FOR DOCUMENTATION (ICOM/CIDOC). *CIDOC-ICOM recommendation on Linked Open Data for museums –Draft*. [2013?] Disponível em: [http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/LoD_For_Museums_v1.7\(draft\)-en.pdf](http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/LoD_For_Museums_v1.7(draft)-en.pdf). Acesso em: 13 mar. 2014.

INTERNATIONAL COUNCIL OF MUSEUMS/INTERNATIONALS COMMITTEE FOR DOCUMENTATION (ICOM/CIDOC). *The Cidoc Conceptual Reference Model*. Version 5.0.4, November 2011. Disponível em: http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf. Acesso em: 24 set. 2015.

INTERNATIONAL COUNCIL OF MUSEUMS/INTERNATIONALS COMMITTEE FOR DOCUMENTATION (ICOM/CIDOC). *What is LIDO lightweight information describing objects*. 2010. Disponível em: <http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>. Acesso: 24 set. 2015.

INTERNATIONAL COUNCIL ON ARCHIVES. EXPERT'S GROUP ON ARCHIVAL DESCRIPTION. *Records in context a conceptual model for archival description*. (Consultation Draft v0.1). ICA, 2016. Disponível em: <http://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>. Acesso em: 02 nov. 2016

INTERNATIONAL ORGANIZATION FOR STANDARTIZATION. *ISO/DIS 25964-1*. Information and Documentation. Thesauri and interoperability with other vocabularies, Part 1: Draft for comment: Thesauri for information retrieval. International Standards Organization, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARTIZATION. *ISO2709:2008. Information and documentation -- Format for information exchange*. 2008. Disponível em: <https://www.iso.org/standard/41319.html>. Acesso em: 26 out. 2014.

ISAAC, Antoine. *Europeana Data Model Primer*. Europeana, 2013. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf. Acesso em: 11 mai. 2017.

JOINT STEERING COMMITTEE FOR THE DEVELOPMENT OF RDA. *RDA: Resource Description and Access*. 2010. Disponível em: <http://www.rda-jsc.org/rda.html>. Acesso em: 3 abr. 2015.

KAPSALIS, Effie. *The Impact of Open Access on Galleries, Libraries, Museums, & Archives*. Smithsonian Emerging Leaders Development Program. 2016. Disponível em: https://siarchives.si.edu/sites/default/files/pdfs/2016_03_10_OpenCollections_Public.pdf. Acesso em: 20 mar. 2018.

KLEIN, Martin; VAN DE SOMPEL, Herbert; SANDERSON, Robert; SHANKAR, Harihar; BALAKIREVA, Lyudmila; ZHOU, Ke; TOBIN, Richard. Scholarly context not found: one in five articles suffers from reference rot. *PLoS ONE*, San Francisco, v. 9, n. 12, 2014. DOI: <https://doi.org/10.1371/journal.pone.0115253>.

KOEHLER, Wallace. Web page change and persistence – a four-year longitudinal study. *Journal of the American Society for Information Science and Technology*, Hoboken, v. 53, n. 2, p. 162-171, 2002.

LARSON, C. e-Government Information: A Progress Report of Information Services. *Herald of Library Science*, v. 44, n. 1/2, p. 81, 2005. Disponível em: <https://www.loc.gov/loc/lcib/0411/egov.html>. Acesso em: 5 mar. 2011.

LIBRARY OF CONGRESS. *A Bibliographic Framework for the Digital Age*. Report, Library of Congress, 2011. Disponível em: <https://www.loc.gov/bibframe/news/framework-103111.html>. Acesso em: 8 ago. 2018.

LIBRARY OF CONGRESS. *Digital Libraries Initiative*. 1998. Disponível em: <https://memory.loc.gov/ammem/dli2/html/lcndlp.html>. Acesso em: 15 abr. 2018.

LIBRARY OF CONGRESS. *METS Metadata Encoding and Transmission Standard*. 2017. Disponível em: <http://www.loc.gov/standards/mets>. Acesso em: 15 abr. 2018.

LIBRARY OF CONGRESS. *MODS Metadata Object Description Schema*. 2017. Disponível em: <http://www.loc.gov/standards/mods>. Acesso em: 15 abr. 2018.

LIBRARY OF CONGRESS. *Technical Standards for Digital Conversion of Text and Graphic Materials*. 2006. Disponível em: <http://memory.loc.gov/ammem/about/techStandards.pdf>. Acesso em: 26 mai. 2010.

LUHN, Hans Peter. Keyword in Context Index for Technical Literature. *American Documentation*, Hoboken, v. 11, n. 4, p. 288-295, 1960.

MARCONDES, Carlos. H. Towards a Vocabulary to Implement Culturally Relevant Relationships between Digital Collections in Heritage Institutions. *Knowledge Organization*, Baden-Baden, v. 47, n. 2, p. 122-137, 2020. Disponível em: DOI:10.5771/0943-7444-15-2020-2-122. Acesso em: 07 mar. 2020.

MCKEMMISH, Sue; ACLAND, Glenda; WARD, Nigel; REED, Barbara. Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema. *Archivaria*, Ottawa, v. 48, Issue Fall, 1999. Disponível em: http://www.infotech.monash.edu.au/research/groups/rcrg/publications/archi_v01.html. Acesso em: 6 jun. 2018.

MCKENNA, Gordon; WYNS, Roxanne. Assessment of requirements for persistent identification of objects, collections and institutions. ECP-2007-DILI-517005 ATHENA, EContent Plus. Europeana, 2010. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ATHENA/Deliverables/D3.4%20Assessment%20of%20requirements%20for%20persistent%20identification%20of%20objects.pdf. Acesso em: 21 mai. 2013.

MENDO CARMONA, Concepción; TEJADA ARTIGAS, Carlos Miguel. Europeana: un recorrido desde su nacimiento hasta nuestros días. In: RAMOS SIMÓN, Luis Fernando; ARQUERO AVILÉS, Rosario (cord.). *Europeana la plataforma de patrimonio cultural europeo*. Gijón: Ediciones Trea, 2014. p. 25-44. (Biblioteconomia y Administración Cultural, 271).

MILLER, Paul. Interoperability. What is it and why should I want it?. *Ariadne*, Kassel, v. 24, 2000. Disponível em: <http://www.ariadne.ac.uk/issue24/interoperability/intro.html>. Acesso em: 17 abr. 2001.

MINTZER, Frederick C. *et al.* Toward on-line, worldwide access to Vatican Library materials. *IBM Journal of Research and Development*, Piscataway, v. 40, n. 2, p. 139-162, 1996. DOI: <http://dx.doi.org/10.1147/rd.402.0139>.

MOEN, William E. Mapping the interoperability landscape for networked information retrieval. In: *ACM/IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES, 1.*, ACM. *Proceedings* [...]. 2001, p. 50-51. Disponível em: https://digital.library.unt.edu/ark:/67531/metadc36320/m2/1/high_res_d/p50-moen.pdf. Acesso em: 5 fev. 2003.

MORGAN, E. L. An Introduction to the Search/Retrieve URL Service (SRU). 2004. Disponível em: <http://www.ariadne.ac.uk/issue40/morgan/intro.html>. Acesso em: 8 mai. 2015.

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION. *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*. U.S. National Archives and Records Administration, 2004. Disponível em: <https://www.archives.gov/files/preservation/technical/guidelines.pdf>. Acesso em: 15 ago. 2015.

NATIONAL INSTITUTE OF HEALTH. *Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research*. NIH, 2015. Disponível: <https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf>. Acesso em: 6 abr. 2018.

NATIONAL LIBRARY OF AUSTRALIA. *Persistent identifier scheme adopted by the National Library of Australia*, 2001. Disponível em: <http://pandora.nla.gov.au/pan/21336/20031011-0000/www.nla.gov.au/initiatives/nlapi.html>. Acesso em: 24 mar. 2017.

NATIONAL LIBRARY OF MEDICINE. *A brief history of NLM*. 2004. Disponível: <https://www.nlm.nih.gov/about/briefhistory.html>. Acesso em: 23 jan. 2017.

NATIONAL SCIENCE FOUNDATION. *Dissemination and Sharing of Research Results*. NSF, [2015?]. Disponível em: <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>. Acesso em: 20 mar. 2018.

NEVILLE, Liddy; LISSONNET, Sophie. Was CIMI too early? Dublin Core and Museum Information: metadata as cultural heritage data. In: DC 2005 INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, Madrid, Spain, 12-15 September 2005. *Proceedings* [...]. DCMI, 2005. p. 31-38. Disponível em: <http://dcpapers.dublincore.org/pubs/article/download/801/797>. Acesso em: 4 mar. 2006.

NILSONN, Mikael; POWEL, Andy; JOHNSTON, Pete; NAEVE, Ambjörn. *Expressing Dublin Core metadata using the Resource Description Framework (RDF)*. DCMI, 2010. Disponível em: <http://dublincore.org/documents/dc-rdf>. Acesso em: 23 ago. 2012.

OBRST, Leo. *The ontology spectrum & semantic models*. MITRE Corporation, 2006. (Presentación ppt). Disponível em: http://ontolog.cim3.net/file/resource/presentation/LeoObrst_20060112/OntologySpectrumSemanticModels--LeoObrst_20060112.ppt. Acesso em: 7 jun. 2011.

OLDMAN, Dominic. *The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER*. CRM Labs, 2014. Disponível em: http://www.cidoc-crm.org/sites/default/files/CRMPrimer_v1.1_1.pdf. Acesso em: 22 set. 2017.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OCDE, 2007. Disponível em: <http://www.oecd.org/science/scitech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>. Acesso em: 25 mar. 2018.

OWL WORKING GROUP. *OWL Ontology Web Language Overview*. 2004. Disponível em: <http://www.w3.org/TR/owl-features>. Acesso em: 13 abr. 2012.

PARK, Ok Nam. Development of Linked Data for Archives in Korea. *D-Lib Magazine*, Reston, v. 21, n. 3/4, 2015. DOI: <http://dx.doi.org/10.1045/march2015-park>. Acesso em: 6 set. 2018.

PATTUELLI, Cristina M; PROVO, Alexandra; THORSEN, Hilary. Ontology building for linked open data: a pragmatic perspective. *Journal of Library Metadata*, Philadelphia, v. 15, n 3/4, p. 265-294, 2015. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/19386389.2015.1099979>. Acesso em: 27 out. 2017.

PAYETTE, Sandra, *et al.* Interoperability for digital objects and repositories. *D-Lib Magazine*, Reston, v. 5, n. 5, p. 1082-9873, 1999. Disponível em: <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/may99/payette/05payette.html>. Acesso em: 9 mai. 2002.

- PEIRCE, Charles Sanders. *Collected papers of Charles Sanders Peirce*. Electronic Edition. 1994. Disponível em: <https://colorysemiotica.files.wordpress.com/2014/08/peirce-collectedpapers.pdf>. Acesso em: 23 abr. 2017.
- PERKINS, J.; SPINAZZE, A. T. Finding Museum Information in the Internet Commons: A Report on the CIMI Dublin Core Metadata Testbed Project. In: ICHIM, 1999. p. 175-177.
- RAMOS SIMÓN, Luis Fernando; ARQUERO AVILÉS, Rosario coords. *Europeana la plataforma de patrimonio cultural europeu*. Asturias: Ediciones Treas, 2014. (Biblioteconomia y Administración Cultural, 271).
- RAMOS, Ricardo A. *Treinamento prático em UML*. São Paulo: Digerati Books, 2006.
- RDF WORKING GROUP. *RDF PRIMER*. Manola, Frank; Miller, Eric (ed.). W3C, 2002. Disponível em: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210>. Acesso em: 17 ago. 2018.
- ROWLEY, Jennifer. *A biblioteca Eletrônica*. Brasília: Briquet de Lemos Livros, 2002.
- SCHAIBLE, Johann; GOTTRON, Thomas; SCHERP, Ansgar. Survey on common strategies of vocabulary reuse in linked open data modelling. In: *European SEMANTIC WEB CONFERENCE*, Springer, Cham, 2014. p. 457-472. Disponível em: https://www.researchgate.net/profile/Thomas_Gottron/publication/260870201_Survey_on_Common_Strategies_of_Vocabulary_Reuse_in_Linked_Open_Data_Modeling/links/00b495328a003c75f1000000.pdf. Acesso em: 5 jun. 2016.
- SKOS PRIMER. W3C, 2009. Disponível em: <https://www.w3.org/TR/skos-primer>. Acesso em: 5 nov. 2010.
- SOUZA, Renato Rocha; TUDHOPE, Douglas; ALMEIDA, Mauricio Barcellos. Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems. *Knowledge Organization*, Baden-Baden, v. 39, n. 3, p. 179-192, 2012. Disponível em: http://www.academia.edu/download/32446670/Souza_Tudhope_Almeida_-_KOS_Taxonomy.Submitted.pdf. Acesso em: 7 abr. 2014.
- TONKIN, Emma. Persistent Identifiers: Considering the Options. *Ariadne*, Kassel, v. 56. 2008. Disponível em: <http://www.ariadne.ac.uk/issue56/tonkin>. Acesso em: 6 ago. 2010.
- UNESCO. *UNISIST Study Report on the feasibility of a World Science Information System*. Paris: UNESCO, 1971. Disponível em: <http://unesdoc.unesco.org/images/0006/000648/064862eo.pdf>. Acesso em: 5 mai. 2002.
- VANDEBUSSCHE, Pierre-Yves; ATEMEZING, Ghislain A; POVEDA-VILLALÓN, María; VATANT, Bernard. Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Semantic Web*, Amsterdam, v. 8, n. 3, p. 437-452, 2017. Disponível em: <http://www.semantic-web-journal.net/system/files/swj1178.pdf>. Acesso em: 13 abr. 2018.

W3C LIBRARY LINKED DATA INCUBATOR GROUP. *Datasets, Value Vocabularies, and Metadata Element Sets*. W3C, 2011. Disponível em: http://www.w3.org/2005/Incubator/ld/wiki/Vocabulary_and_Dataset. Acesso em: 23 mai. 2014.

WEIBEL, Stuart. The state of the Dublin Core metadata initiative. *Bulletin of the Association for Information Science and Technology*, Silver Spring, v. 25, n. 5, p. 18-22, 1999. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/bult.127/ful>. Acesso em: 7 jan. 2001.

WORLD WIDE WEB CONSORTIUM (W3C). *Extensible Markup Language (XML)*. 2016. Disponível em: <https://www.w3.org/XML/>. Acesso em: 23 mai. 2018.

WORLD WIDE WEB CONSORTIUM (W3C). *RDF 1.1 Turtle/Terse RDF Triple Language*. 2014. Disponível em: <https://www.w3.org/TR/turtle>. Acesso: 22 mai. 2018.

WORLD WIDE WEB CONSORTIUM (W3C). *vCard Ontology - for describing People and Organizations*. 2014b. Disponível em: <https://www.w3.org/TR/vcard-rdf>. Acesso: Acesso: 22 mai. 2018.

WORLD WIDE WEB CONSORTIUM (W3C). *Web of Things at W3C*. 2017. Disponível em: <https://www.w3.org/WoT>. Acesso: Acesso: 22 mai. 2018.

XQUERY/SPARQL TUTORIAL. *Wikibooks*, 2017. Disponível em: https://en.wikibooks.org/wiki/XQuery/SPARQL_Tutorial. Acesso: 5 out. 2018.

XQUERY/SPARQL TUTORIAL. *Wikibooks*, 29 May 2018. Disponível em: https://en.wikibooks.org/wiki/XQuery/SPARQL_Tutorial. Acesso em: Acesso: 5 out. 2018.

ZENG, Marcia Lei. Interoperability. *Knowledge Organization*, Baden-Baden, v. 46, n. 2, p. 122-146, 2019.

SOBRE O LIVRO

CATALOGAÇÃO

Telma Jaqueline Dias Silveira
CRB 8/7867

NORMALIZAÇÃO

Maria Elisa Valentim Pickler Nicolino
CRB - 8/8292
Isabelle Ribeiro O. C. Lima

CAPA E DIAGRAMAÇÃO

Gláucio Rogério de Moraes

PRODUÇÃO GRÁFICA

Giancarlo Malheiro Silva
Gláucio Rogério de Moraes

ASSESSORIA TÉCNICA

Renato Geraldi

OFICINA UNIVERSITÁRIA

Laboratório Editorial
labeditorial.marilia@unesp.br

FORMATO

16 x 23cm

TIPOLOGIA

Adobe Garamond Pro

Papel

Polén soft 70g/m2 (miolo)
Cartão Supremo 250g/m2 (capa)

TIRAGEM

100

IMPRESSÃO E ACABAMENTO

Gráfica
unesp
Campus de Marília 

