



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

Eixo 6 – 4º Encontro de Estudos e Pesquisas em Catalogação

PUBLICANDO E INTERLIGANDO ACERVOS DIGITAIS NA WEB ATRAVÉS DAS TECNOLOGIAS DE DADOS ABERTOS INTERLIGADOS

Carlos Henrique Marcondes

Docente do Programa de Pós-
graduação em Ciência da Informação
da Universidade Federal Fluminense.
E-mail: marcon@vm.uff.br

RESUMO

A partir das limitações dos atuais sistemas gerenciadores de catálogos o trabalho discute as potencialidades das tecnologias de dados abertos interligados para a publicação de acervos digitais de arquivos, bibliotecas e museus na Web, viabilizar seu reuso, integração e interligação entre si e a outros recursos disponíveis na Web.

Palavras-chave: acervos digitais, arquivos, bibliotecas, museus, interoperabilidade, reuso, dados abertos interligados

PUBLISHING AND INTERCONNECTING DIGITAL COLLECTIONS ON THE WEB USING LINKED OPEN DATA TECHNOLOGIES

ABSTRACT

Starting from the current catalog management systems faults this paper discusses the potential of linked open data technologies to publish digital collections of archives, libraries and museums on the Web, enable their reuse, integration and interconnection with each other and with other resources available on the Web.

Keywords: digital collections, archives, libraries, museums, interoperability, reuse, linked open data

1 INTRODUÇÃO



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

Arquivos, bibliotecas e museus têm sido guardiões da memória e cultura desde a Antiguidade. Catálogos são os instrumentos através dos quais essas instituições de memória e cultura divulgam seus acervos. Historicamente catálogos passaram de listas alfabéticas de títulos, a fichários desdobrados por títulos, autores e assuntos até, com o advento da tecnologia dos computadores, a partir da década de 1960, e da Web, a partir de 2000, aos catálogos em linha disponibilizados na Web, os OPACs – OnLinePublic Access Catalogs.

A Web têm sido considerada o mais abrangente mecanismo para a acesso e intercâmbio de conhecimento já desenvolvido pela humanidade (LYTRAS; SICILIA, 2006). No entanto a Web também trouxe sérios desafios para instituições de memória e cultura. Enquanto no mundo não digital elas eram praticamente as únicas fontes de informação, hoje sofrem a concorrência de outros recursos como Google Acadêmico, Google Books, YouTube, Amazon, Wikipedia, etc.

Esta situação tem colocado em questão o papel histórico destas instituições. Profissionais de informação/curadores de acervos de instituições de memória e cultura vêm, de forma crescente, percebendo esses desafios e discutindo alternativas para superá-los (DEMPSEY, 2006), (GORMAN, 2003), (HILLMANN, 2008) (MANN, 2006).

Nos anos recentes grandes projetos em diversos países também tornaram disponíveis na Web acervos importantes.

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, F. et al., 1996).

No artigo os autores chamam a atenção sobre as possibilidades abertas com a digitalização e disponibilização na Web do acervo da Biblioteca Vaticana. Antes do projeto, a Biblioteca só era capaz de atender cerca de 2000 acadêmicos anualmente.



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

Esta questão ilustra as potencialidades dos acervos digitais em memória e cultura publicados na Web. Quando comparados aos acervos tradicionais – manuscritos, documentos ou obras raras impressas, objetos físico - os objetos digitais publicados na Web têm novas 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.

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, reusados como materiais educativos, compõem publicações, exposições ou aulas virtuais. Estes objetos tem também um potencial econômico, vêm sendo considerados um insumo para as chamadas indústrias criativas.

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. (Projeto EuropeanaCreative¹)

Esse alcance e plasticidades dos acervos digitais publicados na Web se potencializa na medida que estes conteúdos possam ser realmente de livre acesso e ao máximo interligados, explorando suas sinergias e complementaridades.

A Web vem se desenvolvendo segundo a visão da Web Semântica (BERNERS-LEE, 2001). Para além de uma Web de documentos hipertextuais interligadas entre si e inteligíveis somente por humanos, teríamos uma Web de *recursos*, qualquer coisa que esteja disponível na Web e acessível através de um “link”, representada em formatos “inteligíveis” tanto por pessoas como por programa. A Web Semântica integra tanto os tradicionais

¹<https://euroclio.eu/projects/partner-europeana-creative/>.



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

recursos “nativos” digitais - documentos hipertextuais, imagens estáticas e em movimento, arquivos sonoros, etc. -, como também representações digitais de *coisas* reais como eu, você, um modelo de carro, a Mona Lisa de Da Vinci, o Pão de Açúcar, etc.; é a chamada *Web das coisas*².

As tecnologias LOD – “linked open data”, dados abertos interligados - são integrantes da Web Semântica e uma nova forma de representar conteúdos que permite a programas, mais que simplesmente publicarem estes conteúdos, também “compreenderem seus significados”, podendo processá-los de forma mais “inteligente”.

Se comparada à proposta da Web Semântica, a tecnologia atual dos sistemas de catálogos apresenta grandes limitações. Restringe o acesso e a interligação dos conteúdos ao escopo do sistema de catálogo, transformando este em um “silo” que aprisiona a informação. Além disso, os conteúdos dos acervos mantidos no catálogo só fazem sentido, só têm “semântica”, no escopo desse sistema, seu significado e inteligibilidade estão restritos 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 MARC, com seus identificadores de conteúdo, que só fazem sentido para catalogadores que os conheçam, e os exibe na tela em campos que fazem sentido aos usuários, como autor, título, editor, data, etc. Catálogos limitam a visibilidade desses conteúdos, as possibilidades de integração de acervos digitais entre si e aos demais conteúdos da Web, encerrando esses conteúdos em um mundo próprio, fechado, isolado dos outros fluxos de informação da Web.

A tecnologia dos catálogos atuais é da década de 1970. Baseada em arquivos invertidos e com algoritmos para comparar cadeias de caracteres e resolver operadores booleanos, ela 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:

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



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

- 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.

Arquivos, bibliotecas e museus pelo mundo afora veem desenvolvendo projetos de publicação de seus catálogos usando LOD. Entre os mais significativos estão os da Biblioteca Nacional da Espanha³, da British Library⁴, da Deutsche National Bibliothek⁵, da Biblioteca Nacional da França⁶, da Biblioteca Europeana⁷, do British Museum⁸, do Archives Hub⁹.

Os objetivos deste trabalho são 1- situar historicamente e discutir a importância cultural e econômica de disponibilizar na Web acervos digitais em memória e cultura; e 2- apresentar uma visão geral das tecnologias LOD e 3- discutir sua potencialidade e aplicação na publicação de acervos digitais de arquivos, bibliotecas e museus. Em especial o trabalho endereça a questão de como, a partir dos acervos digitais registrados em sistemas de catálogos, publicar e *interligar* objetos digitais representando itens de acervos de *diferentes* instituições como arquivos, bibliotecas e museus, a questão denominada *interoperabilidade*. O trabalho está organizado como se segue. Após a seção 1. Introdução, são apresentados na seção 2. os Métodos da Pesquisa. Segue-se a Seção 3, que endereça o primeiro objetivo, situando historicamente e discutindo a importância cultural e econômica

³ Ver <http://datos.bne.es>

⁴ Ver <http://bnb.data.bl.uk>

⁵ Ver http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html

⁶ Ver <http://data.bnf.fr>

⁷ Ver <http://labs.europeana.eu/api/linked-open-data-introduction>

⁸ Ver <https://datahub.io/dataset/british-museum-collection>

⁹ Ver <https://archiveshub.jisc.ac.uk/>



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

de disponibilizar na Web acervos digitais em memória e cultura; a seguir a Seção 4. Resultados, apresenta o segundo objetivo do trabalho, subdividida em – discussão de dois pressupostos iniciais, o que é interoperabilidade e conceituação do que são os objetos digitais que representam itens de acervos de arquivos, bibliotecas e museus; e - o que são e como funcionam as tecnologias LOD e como elas se diferenciam das tecnologias dos catálogos atuais. A Seção seguinte, 5. Discussão, apresenta o terceiro objetivo do trabalho, discutindo como esses objetos digitais poderiam ser interligados utilizando as tecnologias LOD. A última seção 6 traz as considerações finais e apresenta as conclusões do trabalho.

2 MÉTODO DA PESQUISA

Para elaborar, compilar e apresentar sistematicamente os tópicos propostos foi utilizado a pesquisa documental e bibliográfica. Documentos de políticas públicas e projetos de digitalização foram examinados; manuais como Heath e Bizer (2011) e RDF Primer (2004), documentos em linha, relatórios e padrões do W3C sobre o tema da tecnologia dos dados abertos interligados foram consultados, além das metodologias correntes de representação e descrição de acervos. A pesquisa usou como fontes o Google Acadêmico e anais de congressos da IFLA e ICOM, e como estratégia de busca *“linked open data” AND (archive OR library OR museum)*. Foi elaborada uma compilação das tecnologias LOD. Metodologias de descrição de acervos em arquivos, bibliotecas e museus como MARC (MARC STANDARDS, 2017), MODS, Dublin Core¹⁰, RDA (2010), ISAD(G), LIDO, e modelos conceituais como FRBR (IFLA, 1998), CIDOC CRM (2014), Ric-CM (INTERNATIONAL COUNCIL ON ARCHIVES, 2016), EDM (EUROPEANA DATA MODEL DOCUMENTATION), também foram examinados para identificar, a partir de registros de um catálogo, relações, implícitas ou explícitas, previstas nestas metodologias, que pudessem gerar “links” interligando objetos dos acervos ou destes com outros recursos existentes na Web.

¹⁰Ver <http://dublincore.org/>



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

3 ACERVOS DIGITAIS NA WEB E AS TECNOLOGIAS LOD

Internacionalmente, 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 Memory¹¹, 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 Initiative¹², 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 desde a criação de projetos-piloto até o 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.

Na Europa o programa eContent - European digital content on the global networks -, cuja primeira fase desenvolveu-se de 2001 a 2004¹³ e a segunda¹⁴, 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 (Carmona; Artigan,2014). Dessas políticas públicas resultaram também o desenvolvimento do Portal Europeu de Arquivos e a Biblioteca Europeia, esta, a antecessora da Biblioteca Europeana.

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

¹¹Ver <https://memory.loc.gov/ammem/dli2/html/lcndlp.html>

¹² Ver https://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=100660.

¹³ Ver http://cordis.europa.eu/pub/econtent/docs/call1_brochure.pdf.

¹⁴ Ver http://cordis.europa.eu/econtent/customisation/cust_projects.htm.



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

impactos para além da área cultural, em setores econômicos como educação, turismo, indústrias criativas.

Segundo Kapsalis (2016), comentando estudo feito em instituições que adotaram o livre acesso a seus acervos digitais na Web, essa decisão institucional jogou um papel fundamental para potencializar o reuso dos recursos, aumentar seus públicos, tanto no seu sítio Web quanto presenciais, aumentar sua exposição na mídia e facilitar o acesso a recursos de fomento.

As tecnologias LOD vêm justamente 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.

4 RESULTADOS

A partir de definições de interoperabilidade e dos objetos de acervos digitais, essa Seção expõe o que são as tecnologias de dados abertos interligados.

4.1 Pressupostos

Embora arquivos, bibliotecas e museus 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 através das tecnologias LOD. No entanto, os pressupostos da publicação e integração de acervos digitais na Web devem ser estabelecidos. O que, exatamente, estamos pretendemos interligar usando as tecnologias de dados abertos interligados?



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

O termo técnico para interligar objetos digitais de acervos de memória e cultura é *interoperabilidade*. Interoperabilidade como a capacidade técnica 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 convênios, serem capazes de operarem em conjunto e intercambiarem seus dados mantendo o significado original desses dados, a mesma semântica do sistema original.

O acesso a acervos digitalizados de patrimônio cultural através da Web – livros raros, manuscritos e outros documentos digitalizados, fotografias ou vídeos de objetos e artefatos 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. Essas 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 instituições como arquivos, bibliotecas e museus. Objetos de patrimônio cultural são valorizados não por seu valor em si, mas por seu valor cultural e histórico, atribuído principalmente por seus curadores, em última instância, pelas instituições de memória e cultura que os guardam e preservam. A citação seguinte destaca este aspecto com referência a objetos museológicos, mas poderia ser generalizadas para objetos aos quais é atribuído um valor histórico, cultural ou de memória. “Museum objects are objects separated from their original (primary) context and transferred to a new, museum reality in order to document the reality from which they were separated.” (VAN MENSCH, 1992, 104).

Desenvolver esta atividade de valorização de objetos com valor cultural, histórico e de memória é a missão e o papel que as sociedades atribuem às instituições como arquivos, bibliotecas e museus.

O objeto digital que será publicado, interligado e acessado através da Web utilizando as tecnologias LOD, é um *substituto* digital do objeto físico, é um objeto construído no



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

escopo das metodologias de documentação da arquivologia, biblioteconomia e museologia; especificamente, 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 trabalho de objeto digital de patrimônio – ODP. Um ODP tem, em geral, uma estrutura e componentes como são mostrados a seguir.

- OPD

- identificador persistente (sua importância será discutida a seguir, junto com as tecnologias LOD)

- 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(s) cópia(s) ou imagem(ns) do(s) objeto(s) digital(is)

- 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 METS¹⁵, ORE¹⁶ e Aggregation/ProvidedCHO, padrão de objeto digital complexo da Biblioteca Europeia (Isaac, 2013).

¹⁵ Ver METS – Metadata Encoding & Transmission Standard, <http://www.loc.gov/standards/mets/>.

¹⁶ Ver ORE - Object Reuse and Exchange, <https://www.openarchives.org/ore/>.



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

4.2 As tecnologias de dados abertos interligados

Em um texto de 2001, dirigido a um público amplo, Tim Berners-Lee outros, o cientista da computação criador da Internet e da Web, delineia a visão da Web Semântica. Esta seria uma evolução da Web atual, na qual os programas “compreenderiam” o conteúdo das páginas e assim poderiam nos auxiliar em tarefas muito mais complexas que simplesmente exibir esse conteúdo. Dados que compõe a Web e programas seriam estruturados de forma a poderem ser utilizados de forma generalizada 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 páginas hipertextuais, legíveis somente por humanos, interligados através de hiperlinks que servem para navegar entre uma página e outra. Esses “hiperlinks” são acessíveis através de URLs, o típico “link” da Web, 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 permitir a navegação de uma página para outra.

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, 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 objetos nativos 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, seu CV Lattes, ou uma universidade, representada através de sua página institucional, ou um sítio histórico, representado através de uma de imagem digital.

O modelo de dados da Web Semântica se baseia na ideia simples que recursos podem ser representados através de afirmações descrevendo suas propriedades. Cada



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

afirmação é composta de sujeito (o recurso que está sendo descrito), propriedades do recurso e os valores dessas propriedades. 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>. 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, públicas, isto é, também disponíveis na Web. 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 tem a forma de um grafo, ou rede semântica¹⁷, como na figura a seguir; nele, as coisas seriam ovais, os textos seriam retângulos e os “links” semânticos seriam setas.

¹⁷ Ver https://en.wikipedia.org/wiki/Semantic_network.

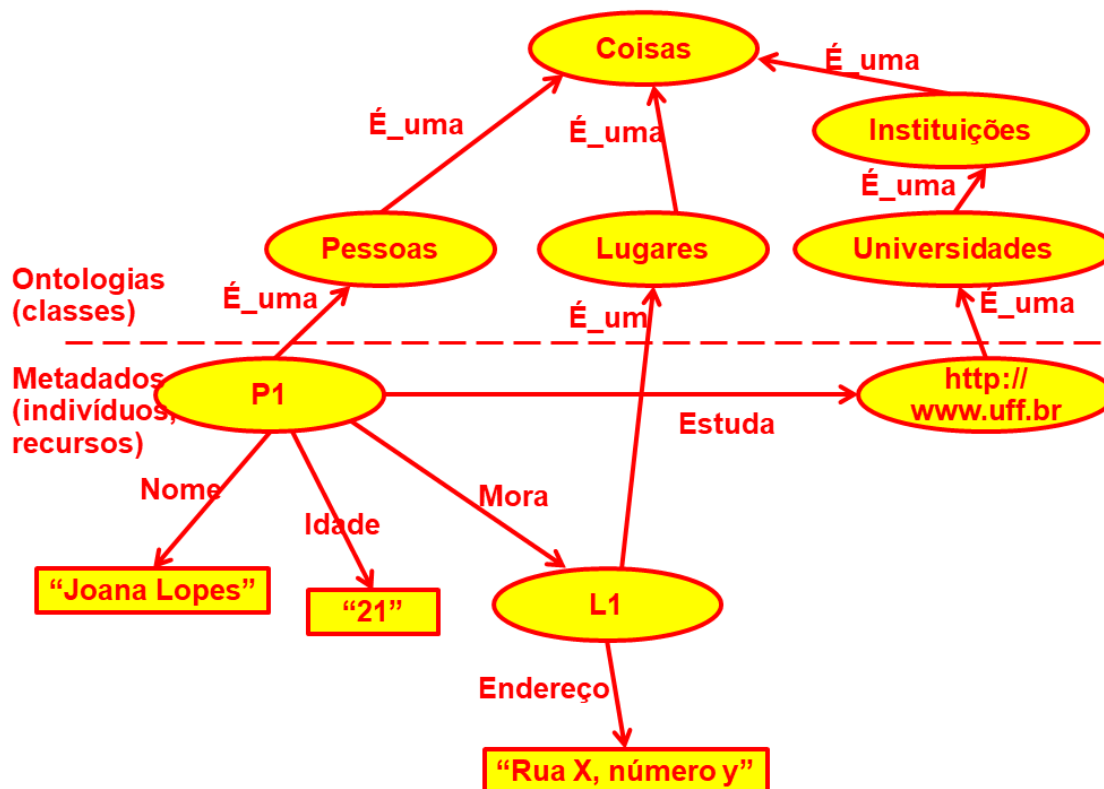


Figura 1. Modelo de dados da Web Semântica

Interligando um grafo a outros 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, usando a linguagem de consulta semântica SPARQL¹⁸ e não somente ser navegável como a Web atual.

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

- RDF - ResourceDescription Framework (RDF PRIMER, 2004)) – esquema baseada na linguagem XML para descrever e publicar diretamente na Web recursos de modo que estas descrições sejam “compreensíveis” por programas, além de por pessoas. A descrição de um

¹⁸Ver SPARQL Query language for RDF, <https://www.w3.org/TR/rdf-sparql-query/>



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

recurso em RDF se dá sempre através de afirmações de suas propriedades e dos valores dessas propriedades.

- “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)”.

RDF codifica uma afirmação como essa em um esquema específico, usando a linguagem XML. Ou seja, uma afirmação em RDF tem uma *sintaxe* específica. Recurso, propriedade e valor também são chamados de sujeito, predicado e objeto da afirmação. Veja a seguir.

```
<?xml version="1.0">
<rdf:RDF
<rdf:Descriptionrdf:about "http://www.uff.br"><-sujeito
    predicado -><autor> Carlos H. Marcondes </autor>
</rdf:Description>valor -^
</rdf:RDF>
```

- Uso de identificadores permanentes e válidos globalmente em toda a Web – IRIs¹⁹ 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”, “erro 404”, 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.

- Uso de vocabulários padronizados, reconhecidos e avalizados por comunidades de usuários em domínios específicos. Como o esquema RDF só possui como pressupostos semânticos a existência de três elementos distintos, o recurso, a propriedade e o valor da

¹⁹Anteriormente chamados de URIs¹⁹, dos quais são uma generalização, permitindo o uso 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. Ver https://en.wikipedia.org/wiki/Internationalized_Resource_Identifier.



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

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. Veja o exemplo na representação em triplas:

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

Nesse caso o vocabulário é o Dublin Core MetadataElements Set, um vocabulário de elementos que servem para descrever documentos digitais. Esse vocabulário tem um elemento, “dc:creator”, onde é especificado e definido o que é o dc:creator de um documento.

Usar IRIs para representar propriedades e valores de uma afirmação RDF ao invés de cadeias de caracteres permite usar (reusar) termos de vocabulários padronizados e largamente aceitos com semântica definida. Nas triplas RDF, uma referindo-se a um livro e outra, a um funcionário:

- libro0237	title	“Don Quijote”
- < http://catalogo.bne.es/libro0237 >	< http://purl.org/dc/elements/1.1/title >	<“Don Quijote”>
e		
- emp0027	title	“Presidente”
- < http://www.company.com/0027 >	< http://www.w3c.org/2006/vcard/ns/title >	<“Presidente”>

os predicados são expressos pela mesma cadeia de caracteres, “title”, mas como essa literal é parte de IRIs diferentes, um programa saberá que seus significados são distintos; na 1ª. o vocabulário referenciado é o Dublin Core e na 2ª. o vCard²⁰.

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

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



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

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.
- 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(ou seja, IRIs) 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)”²¹.

É essa combinação de atributos que tornam os dados realmente abertos, encontráveis e reutilizáveis. Em áreas como pesquisas científicas, a questão do reuso dos dados de pesquisa é bastante sensível 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, [s.d.]), (OECD, 2007). O potencial das tecnologias LOD para instituições de memória e cultura vem sendo crescentemente destacado (W3C Library Linked Data Incubator Group, 2011), (HANNEMANN, 2010), (ACENJO-BULLÓN, 2015).

4.3 Dados processáveis por programas, “semântica” e dados abertos interligados

O que significa exatamente dados “compreensíveis” por programas? Todo sistema informatizado consta de dois componentes interdependentes, um componente estático, os dados, os conteúdos, as informações significativas, e um componente dinâmico, os

²¹A explicação para as 5 estrelas foram tiradas de <http://5stardata.info/pt-BR/>.



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

programas que processam, transformam, recuperam os dados. Essa dicotomia dados/programas esta na raiz da própria arquitetura dos computadores, a chamada arquitetura de Von Neumann²² (Breton, 1991, p. 89), nome de um dos pioneiros da computação. A computação se desenvolveu por décadas baseada na interdependência desses dois componentes. No entanto, é esta interdependência que sustenta a questão dos formatos de arquivo proprietários, da qual tira partido a indústria de informática para manter seus usuário presos a seus programas e formatos. Esta interdependência começou a ser questionada por usuários no mundo todo.

Em oposição aos formatos proprietários, específicos de determinados programas, vêm se desenvolvendo então os formatos abertos. No caso dos catálogos, a própria semântica dos conteúdos recuperados depende dos sistemas de catálogos.

A proposta de publicar na Web conteúdos de catálogos usando as tecnologias de dados abertos interligados traz um avanço com relação à interdependência dados-programas. Essas tecnologias são uma evolução na direção de dados com semântica própria, autocontida, independentes de programas específicos. Formatos de dados/arquivos que têm esta propriedade são geralmente conhecidos como formatos estruturados. Quanto mais estruturados os dados, maior a capacidade de qualquer programa processá-los.

O W3C, em texto introdutório à linguagem OWL²³, discorrendo sobre dados estruturados, contrasta dados formatados em XML com dados formatados em RDF, uma das bases das tecnologias LOD: “XML provides a surfacesyntax for structureddocuments, butimposes no semanticconstraints on themeaning of thesedocuments.”; e com relação a RDF: “RDF is a datamodel for objects (“resources”) and relations between them, provides a simplesemantics for thisdatamodel, andthesedatamodelscanberepresented in an XML syntax”.

²² Ver https://en.wikipedia.org/wiki/Von_Neumann_architecture

²³ Ver <https://www.w3.org/TR/owl-features/>



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

Vejam esta questão a partir de uma perspectiva comparada e evolutiva em direção às tecnologias LOD. Imaginemos a combinação de conteúdo - um arquivo texto .txt contendo uma bibliografia - e programa - o Notepad, o simples editor de arquivos texto do Windows. Com esta combinação podemos recuperar um conteúdo específico, por exemplo, “Cervantes”, entre os registros da nossa bibliografia; mas o Notepad não consegue distinguir entre “Cervantes” como autor ou como título.

Vamos agora utilizar uma combinação diferentes, um programa gerenciador de bases de dados como o MS Access ou o LibreOffice base. Essa combinação, por estruturar dados como campos, nos permite recuperar um conteúdo em determinado contexto; podemos recuperar “Cervantes” como autor, ou como título. No entanto esse contexto, chamado de esquema do banco de dados, é *dependente* do específico programa gerenciador de bases de dados (as bases de dados do MS Access, por exemplo, têm um formato específico, o .mdb).

Um passo adiante são dados formatados em XML. XML é uma linguagem de marcação, conteúdos são “marcados” ou identificados por etiquetas especiais. Um arquivo XML conterà, juntamente com os dados, as etiquetas de identificação de cada campo. Os dados ficarão assim com um grau a mais de independência de programas com relação aos mesmos dados gerenciados pelo MS Access ou LibreOffice base. No entanto, embora independentes de programas específicos, dados em XML dependem ainda de programas que lhes deem semântica.

As tecnologias LOD são um passo adiante na direção de *dados que carreguem sua própria semântica*. A linguagem RDF traz uma semântica implícita na sua sintaxe: quem é o Sujeito, quem é o Predicado e quem é o Objeto de uma afirmação. Neste sentido, dados formatados segundo as tecnologias LOD são significativamente mais independentes de quaisquer programas que os processem – são mais “abertos”²⁴ - que dados disponibilizados através de sistemas de catálogos. A figura a seguir ilustra e sintetiza esta discussão.

²⁴ Ver em <http://5stardata.info/pt-BR/>.



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

Quijote.txt - Bloco de notas

Arquivo.txt

Arquivo Editar Formatar Exibir Ajuda

Cervantes Saavedra, Miguel de. El ingenioso hidalgo Don Quijote de la Mancha. Madrid : Grupo DILE, Universidad Autónoma de Madrid, 2016.
 Ungría, Alfonso. Cervantes y el Quijote : el autor y su obra. Madrid : Divisa Home Vídeo, 2016.
 Penco Valenzuela, Fernando. Un país llamado Cervantes : el origen judeoconverso del escritor. Madrid : Almu: **BD Relacional**

Id	AUTOR	TÍTULO	LOCAL	EDITOR	ANO
001	Cervantes, Miguel	El ingenioso hidalgo...	Madrid	Grupo DILE	2016
002	Ungría, Alfonso	Cervantes y el Quijote ...	Madrid	Divisa Home Vídeo	2016
003	Penco Valenzuela, Fernando	Un país llamado Cervantes ...	Madrid	Almuzara	2017

Arquivo XML

```
<registro>
  <id>001</id>
  <autor>Cervantes, Miguel</autor>
  <titulo>El ingenioso hidalgo ... </titulo>
  ...
</registro>
```

Arquivo RDF

```
<?xml version="1.0">
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1">
  <rdf:Description rdf:about
    "http://exemplos.org/001">
    <dc:creator>Cervantes, Miguel</dc:creator>
    <dc:title>El ingenioso hidalgo...</dc:title>
    ...
  </rdf:Description>
</rdf:RDF>
```

Sujeito →

Predicado →

Objeto →

Figura 2. Independência programas-dados e a questão da semântica

Como seriam publicados itens de acervos digitais como *recursos segundo as tecnologias LOD*? Essa questão será discutida na próxima Seção.

5 DISCUSSÃO

“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data” (BERNERS-LEE, 2006).

Embora existam várias experiências de publicação de acervos em arquivos, bibliotecas e museus utilizando as tecnologias LOD, estas experiências têm se limitado basicamente a publicar o conteúdo de catálogos; são poucas as que utilizam essas



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

tecnologias para tornar *interoperáveis* diferentes acervos. As tecnologias LOD mostram significativas vantagens em termos de interoperabilidade (MARCONDES, 2016).

Metodologias de descrição de acervos em arquivos, bibliotecas e museus – padrões de metadados, de catalogação e de descrição como MARC (MARC STANDARDS, 2017), MODS, Dublin Core, RDA (2010), ISAD(G), LIDO, e modelos conceituais como FRBR (IFLA, 1998), CIDOC CRM (2014), Ric-CM (INTERNATIONAL COUNCIL ON ARCHIVES, 2016), EDM (EUROPEANA DATA MODEL DOCUMENTATION), possuem diferentes exemplos de *relações* entre objetos de acervos. Muitas vezes essas relações são implícitas – aparecem como campos de um item – como a autoria ou o editor de uma obra, e são presas à descrição de cada item. O papel das relações para contextualizar e tornar mais rica a navegação pelos registros de catálogos é ressaltada por Tillet (2001). Muitas dessas relações já existentes nos catálogos e poderiam, uma vez que estes acervos fossem publicados segundo as tecnologias LOD, gerarem “links” entre objetos de acervos de outros arquivos, bibliotecas e museus e também “links” para e de outros recursos na Web, como universidades, enciclopédias, agências de turismo, eventos, etc., enriquecendo mutuamente esses acervos e integrando-os aos fluxos de informação da Web.

A valorização dos conteúdos de catálogos de arquivos, bibliotecas e museus pela sua publicação como dados abertos interligados esta 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 promoverem esses “links”, identificando potenciais relacionamentos entre registros de seu catálogo com registros de outros catálogos ou com outros recursos na Web.

Enquanto nos registros de um catálogo as relações são mais ou menos implícitas e servem principalmente para recuperar registros, na Web de dados as relações – “links” - têm um significado explícito e servem não só para recuperar um recurso, mas para navegar de uma recursos para outro.

Alguns exemplos da exploração das relações existentes nos padrões de descrição das três áreas que poderiam gerar “links” são: relações como um objeto e seu autor (campo 100

do MARC), entre uma publicação e seu ilustrador (campo 245 do MARC), entre uma obra de arte e suas versões segundo outros artistas (relação FRBR *is_an_adaptation_of*).

A transformação de campos de registros de catálogos em dados abertos interligados é muitas vezes chamada na literatura sobre o assunto, de *mapeamento*. Esses mapeamentos são simples e praticamente automáticos. Podem ser colocadas em tabelas como regras de conversão, e lidas pelos programas que fazem a conversão dos 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.

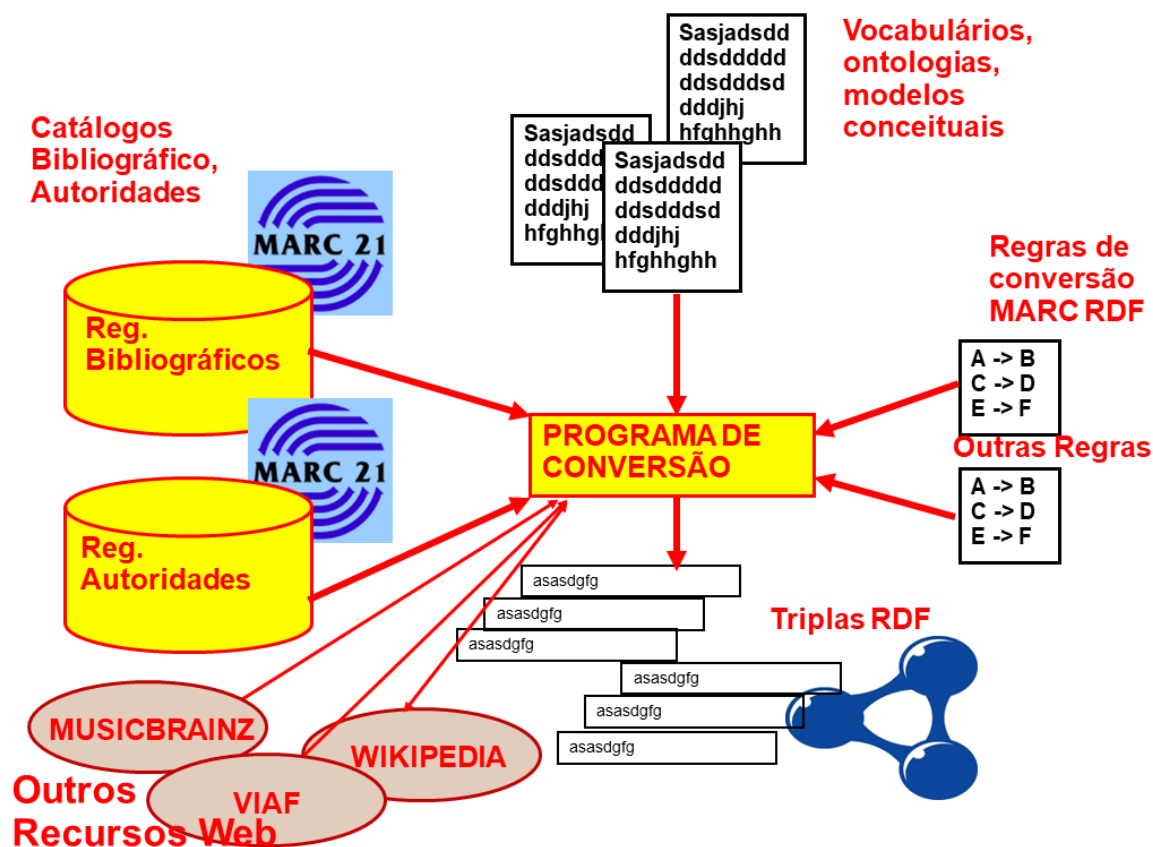


Figura 3. Visão geral do processo de conversão de registros MARC de catálogos para RDF

Na figura, registros MARC bibliográficos e de autoridades são lidos pelos programas de conversão, que consultam tabelas de regras específicas, onde estão definidos os



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

mapeamentos entre os campos MARC e RDF; esse mapeamentos podem incluir referências (IRIs) a vocabulários específicos ou a outros recursos Web publicados como dados abertos interligados, como Wikipedia/DBpedia²⁵, VIAF²⁶, MusicBrainz²⁷, etc; essas referências valorizam e enriquecem os registros em RDF gerados com “links” semânticos, fornecendo uma rica contextualização.

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 MARC 590 – Notas locais, , 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 3, de Tabelas de Regras de Conversão. Nestas o curador poderia especificar regras como as exemplificadas a seguir.

TABELA DE REGRAS DE CONVERSÃO

CAMPO MARC para ->	No.	RECURSO WEB	REGRA	RDF/XML resultante
100 – Entrada Principal, Nome pessoal, subcampo \$a ex: 1001#\$aMigu	1	DBpedia	Contruir um “link” para DBpedia com conteúdo do campo 100\$a, substituindo espaços por “ ”:http://dbpedia.org/page/Mig	<dc:creator> http://dbpedia.org/page/ Miguel_de_Cervantes </dc:creator>

²⁵ Ver <http://wiki.dbpedia.org/>

²⁶ Ver <http://viaf.org/>

²⁷ Ver <https://musicbrainz.org/>



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

el de Cervantes			uel_de_Cervantes; Se existe uma entrada, criar um “link” ²⁸ .	
240 – Título Uniforme, subcampo \$a Ex: 24010\$aDon Quijote	2	The MovieDatabase	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-quiote-de-orson-welles	<frbr:has_a_transformation> https://www.themoviedb.org/movie/31987-don-quiote-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>

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 GeographicNames, usado na Regra 4, mas não o do The MovieDatabase, usado na Regra 2 e do The Museum of ModernArt, 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 tem assim papel crucial na especificação das Tabelas de Regras de Conversão que vão orientar o programa que fará a conversão do seu acervo.

²⁸ 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.



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

Entre os recursos importantes para que sejam feitos “links” a partir de registros de um catálogo estão a Wikipédia²⁹, e DBpedia, o VIAF³⁰, o Geonames³¹, os vocabulários da Getty Foundation³², entre outros. Esses vocabulários agregam semântica e informação contextual, enriquecendo os registros dos catálogos publicados como LOD, além de serem totalmente compatíveis com essas 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 própria coleção publicados como LOD.

Existe também a possibilidade do curador do acervo, usando ferramentas especiais como o RelFinder³³ ou Silk³⁴, 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 anotar³⁵ 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 Niterói³⁶ 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 Da Vinci, no 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.

O curador deverá conhecer profundamente seu acervo, identificar e conhecer também outros recursos com potencial para terem “links” de e para registros do seu catálogo; criar, juntamente com os especialistas em informática, regras para conversão de campos específicos dos registros do seu catálogo, entrar em contato e cooperar com os

²⁹ Ver <https://www.wikipedia.org/>

³⁰ Ver <http://viaf.org/>

³¹ Ver <http://www.geonames.org/>

³² Ver <http://www.getty.edu/research/tools/vocabularies/>

³³ Ver <http://www.visualdataweb.org/relfinder.php>

³⁴ Ver <http://silkframework.org/>

³⁵ Anotações semânticas, ver <http://www.openannotation.org/spec/core/20130208/index.html>.

³⁶ Ver https://pt.wikipedia.org/wiki/Museu_de_Arte_Contempor%C3%A2nea_de_Niter%C3%B3i.



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

curadores de outros acervos ou dos recursos para onde os registros do seu catálogo fazem “links”, para que estes recursos possam também fazer “links” recíprocos para os registros do seu catálogo. Estas são algumas das tarefas de um curador digital para gerir e promover seu acervo publicado como dados abertos interligados.

6 CONSIDERAÇÕES FINAIS E CONCLUSÕES

“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 está associado a uma cópia digital da edição original do Quijote. O registro MARC exibe os campos **100 Entrada Principal Nome Pessoal** e **245 Título e campos relacionados ao Título**, como se segue.

1001#1\$aMiguel de Cervantes Saavedra\$d1547-1616

24510\$aDon Quijote\$cMiguel de Cervantes Saavedra

Certamente que este registro seria enriquecido se, ao invés de simplesmente informar que Cervantes é o autor de Don Quijote, informasse também dados biográficos sobre 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.



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

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; Campos, 2016) que podem ser estabelecidas entre objetos digitais pertencentes a estes acervos, integram 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, permitindo ainda criar exposições ou aulas “virtuais”. Acervos digitais ganham novos curadores, a medida que qualquer um pode criar um “link” para um ODP publicado segundo as tecnologias LOD.

As tecnologias LOD colocam a questão da publicação e interoperabilidade de acervos de instituições de memória e cultura em um novo patamar (ACENJO-BULLÓN, 2015). Muito mais que ampliam 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. Uma vez publicados na Web e acessíveis por identificadores persistentes, podem ser feitos “links” interligando diferentes conteúdos e assim, agregando valor aos mesmos. Os dados também ser enriquecidos, integrados e contextualizados não só por seus curadores originais – instituições de memória e cultura – mas por terceiros que considerem significativo fazer “links” para estes recursos.

Arquivos, bibliotecas e museus muito frequentemente possuem acervos complementares. Essas instituições podem usar as tecnologias LOD para tirarem partido dessas sinergias e complementariedades.



XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

REFERÊNCIAS

ACENJO-BULLÓN, X. Bibliotecas virtuales en 2014, año de consolidación del modelo *linked open data*. *Anuário Think EPI*, p. 9-21, 2015. Disponível em: <<http://dx.doi.org/10.3145/infor.2015.02>>. Acesso em 11 abr. 2017.

BERNERS-LEE, T. Linked data. 2006. Disponível em:
<<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 20 fev. 2017.

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 maio 2001.

BRETON, Philippe. *Historia da Informática*. São Paulo, Editora da UNESP, 1991.

CARMONA, Concepción Mendo; ARTIGAN, Carlos Miguel Tejada. Europeana: um recorrido desde su nacimiento hasta nuestros días". In: Simón, Luis Fernando Ramos; Avilés, RosarioArquero. (2014). *Europeana la plataforma de patrimonio cultural europeu*. EdicionesTreas, 2014. (Biblioteconomia y Administración Cultural, 271).

THE CIDOC CONCEPTUAL REFERENCE MODEL, version 5.1.12, January 2014. Disponível em:
<http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf>. Acesso em 15 nov. 2014.

DEMPSEY, L. The Library Catalogue in the New Discovery Environment: Some Thoughts. *Ariadne*, v. 48, July 2006. Disponível em <<http://www.ariadne.ac.uk/issue48/dempsey/>>. Acesso em: 1 julho 2009.

EUROPEANA DATA MODEL DOCUMENTATION. [s.d.] Disponível em:
<<http://pro.europeana.eu/page/edm-documentation>>. Acesso em 15 nov. 2014.

FOX, EDWARD A. "Digital Libraries Initiative (DLI) Projects 1994-1999". *Bulletin of the Association for Information Science and Technology*, v. 26, n. 1, 1999. Disponível em:
<<http://dx.doi.org/10.1002/bult.135>>. Acesso em 13 nov. 2001.

GORMAN, M. Cataloguing in an electronic age. *Cataloging & classification quarterly*, v. 36, n. 3-4, 2003. Disponível em: <<http://www.columbia.edu/~rjb57/electronic.pdf>>. Acesso em: 1 julho 2009.

HANNEMANN, J. Linked Data for Libraries. In: WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY, 10-15 August 2010, Gothenburg, Sweden, Proceedings... Disponível em: <<http://www.ifla.org/files/hq/papers/ifla76/149-hannemann-en.pdf>>. Acesso em: 4 fev. 2012.

HILLMANN, Diane I. The Emerging Cataloging Future. *Technicalities*, v. 28, no. 4, July/Aug. 2008. Disponível em <<http://hdl.handle.net/1813/10835>>. Acesso em 1 julho 2009.



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

HEATH, T.; BIZER, C. *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, v.1, n.1, p.1-136, 2011. Disponível em: <<http://linkeddatabook.com/editions/1.0/>>. Acesso em: 15 ago. 2016.

HILLMANN, D. I. The emerging cataloging future. *Technicalities*, July/Aug. 2008. Disponível em: <https://www.researchgate.net/profile/Diane_Hillmann/publication/277216971_The_Emerging_Cataloging_Future/links/562f7a7308ae4742240ae304.pdf>. Acesso em: 14 jun. 2013.

IFLA Study Group on Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records: final report*. München: K. G. Saur, 1998. INTERNATIONAL COUNCIL ON ARCHIVES. Experts 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: 23 nov. 2016.

ISAAC, Antoine. *EUROPEANA DATA MODEL PRIMER*, 2013. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf. Acesso em: 5 ago 2014.

ISAD(G): Norma geral internacional de descrição arquivística. Rio de Janeiro: CONARQ, 2000. Disponível em: <http://www.conarq.arquivonacional.gov.br/images/publicacoes_textos/isad_g_2001.pdf>. Acesso em 10 fev. 2001.

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

LIBRARY LINKED DATA INCUBATOR GROUP FINAL REPORT. W3C, 2011. Disponível em: <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>. Acesso em: 3 fev. 2012.

LIDO, Whatis LIDO, Lightweight Information Describing Objects. [s.d.]. Disponível em: <<http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>>. Acesso em: 8 jun. 2017.

LYTRAS, Miltiadis D.; SICILIA, Miguel Angel. The Knowledge Society: a manifesto for knowledge and learning. *International Journal of Knowledge and Learning*, v. 1, n. 1-2, p. 1-11, 2005.

MARC STANDARDS. Library of Congress, 2017. Disponível em: <<https://www.loc.gov/marc/>>. Acesso em: 23 jun. 2017.

MANN, T. *The Changing Nature of the Catalog and Its Integration with Other Discovery Tools*. Final Report: March 17, 2006. Prepared for the Library of Congress by Karen Calhoun.

MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. *Perspectivas em Ciência da*



**XXVII CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA,
DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO**

TEMA CENTRAL: Objetivos para o Desenvolvimento Sustentável das Nações Unidas:
como as bibliotecas podem contribuir com a implementação da Agenda 2030

Informação, v.21, n. 2, 2016. Disponível em:

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735>>. Acesso em: 2 dez. 2016.

MARCONDES, C. H; CAMPOS, M. L. A. Searching for a methodology to define culturally relevant relationships between digital collections in archives, libraries and museums. In: 14th International ISKO Conference 27-29 Sept.n2016 Rio de Janeiro, Brazil, Proceedings... Disponível em: http://www.ergonverlag.de/isko_ko/downloads/aiko_vol_15_2016_61_marcondes_campos.pdf. Acesso em: 23 nov. 2016.

VAN MENSCH, P. *Toward a methodology of museology*. Unpublished Ph.D. dissertation. Zagreb: University of Zagreb, 1992. Disponível em: <<http://xa.yimg.com/kq/groups/23466284/1995686355/name/Towards>>. Acesso em: 5 mar. 2016.

MINTZER, F. et al. (1996). Toward on-line, worldwide access to Vatican Library materials. *IBM Journal of Research and Development*, v. 40, n. 2, pp. 139-162, 1996. Disponível em: <<http://dx.doi.org/10.1147/rd.402.0139>>. Acesso em: 12 ago. 2016.

MODS Metadata Object Description Schema. Library of Congress, 2017. Disponível em: <<http://www.loc.gov/standards/mods/>>. Acesso em: 23 jun. 2017.

NATIONAL INSTITUTE OF HEALTH. *Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research*. NIH, 2015. Disponível em: <<https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf>>. Acesso em 21 mar. 2017.

NATIONAL SCIENCE FOUNDATION. *Dissemination and Sharing of Research Results*. NSF, ([s.d.]). Disponível em: <<https://www.nsf.gov/bfa/dias/policy/dmp.jsp>>. Acesso em: 21 mar. 2017. *RDA: Resource Description and Access*. 2010. Disponível em: <<http://www.rda-jsc.org/rda.html>>. Acesso em: 8 abr. 2012.

RDF PRIMER. MANOLA, F.; MILLER, E.(eds.). W3C, 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>>. Acesso em: 15 dez. 2011.

TILLET, B. B. Bibliographic relationships. In: Bean, C. A. and Green, R. (Eds.), *Relationships in the Organization of Knowledge*. Dordrecht: Kluwer Academic Publishers, 2001. p. 19-35. W3C Library Linked Data Incubator Group. *Datasets, Value Vocabularies, and Metadata Element Sets*. W3C, 2011. Disponível em: <http://www.w3.org/2005/Incubator/llid/wiki/Vocabulary_and_Dataset>. Acesso em: 31 de março de 2012.

Agências financiadoras: Projeto apoiado pelo CNPq através de bolsa de produtividade