

-XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-08 – Informação e Tecnologia

*LINKED DATA EM BIBLIOTECAS: INICIATIVAS E TENDÊNCIAS*¹

Felipe Augusto Arakaki (Universidade Estadual Paulista – UNESP)

Plácida Leopoldina Ventura Amorim da Costa Santos (Universidade Estadual Paulista – UNESP)

LIBRARY LINKED DATA: INITIATIVES AND TRENDS

Modalidade da Apresentação: Comunicação Oral

Resumo: A proposta da *Web Semântica* proporcionou novos desafios e o desencadeamento de tecnologias para o ambiente da *Web*. Nesse sentido, as informações necessitam de uma infraestrutura para que os computadores possam processá-las e analisá-las de forma eficiente. Diversas comunidades têm se preocupado em aprimorar seus sistemas e estão trabalhando na aplicação do *Linked Data*. O problema de pesquisa caracteriza-se em: quais são as iniciativas que estão trabalhando na aplicação do *Linked Data* em bibliotecas? O objetivo é apresentar um panorama dos principais esforços no uso do *Linked Data* no domínio bibliográfico, utilizando como método uma pesquisa qualitativa e teórica de caráter exploratória da literatura e dos principais estudos para disponibilização de dados ligados no contexto das bibliotecas. Apresenta os principais movimentos de bibliotecas na proposta do *Linked Data* como a *Dublin Core Metadata Initiative* (DCMI), a *Library of Congress* e o *Linked Data for Libraries* com o BIBFRAME, a *Online Computer Library Center* (OCLC) com o *WorldCat*. Aborda também, as iniciativas que buscam trabalhar no contexto de bibliotecas, arquivos e museus como o movimento *Linked Open Data Libraries, Archives and Museum* (LODLAM) e *Linked Open Data Galleries, Libraries, Archives and Museum* (LODGLAM) que inclui as galerias de artes. Como resultado apresenta as principais instituições do domínio bibliográfico no uso do *Linked Data* e as iniciativas que ligam dados de bibliotecas, arquivos, museus e galerias de arte, com propósito da preservação da herança cultural a partir da integração de dados.
Palavras-Chave: *Web Semântica*; *Linked Data*; Catálogo; Metadados; Catalogação.

Abstract: The proposal of the *Semantic Web* has provided new challenges and the producing of technologies for the *Web* environment. In this sense, the information requires an infrastructure so that computers can process them and analyze them efficiently. Several communities have been concerned to improve their systems and are working on application of *Linked Data*. The problem is characterized in research: what are the initiatives that are

¹ Trabalho resultado da dissertação: ARAKAKI, F. A. *Linked Data: ligação de dados bibliográficos*. 2016. 144 f. Dissertação (Mestrado em Ciência da Informação)–Universidade Estadual Paulista ‘Júlio de Mesquita Filho’, Marília/SP, 2016. Disponível em: <<https://repositorio.unesp.br/handle/11449/147979>>. Acesso em: 09 set. 2017.

working in the application of Linked Data in libraries? The goal is to present an overview of the major efforts in the use of Linked Data in Bibliographic Domain, using as a theoretical and exploratory research method of qualitative character of literature and of the main data-availability studies linked in the context of the libraries. Presents the main movements of libraries in the proposal Linked Data as the Dublin Core Metadata Initiative (DCMI), the Library of Congress and the Linked Data for Libraries with the BIBFRAME, the Online Computer Library Center (OCLC) with WorldCat.org. Addressed the initiatives seeking to work in the context of libraries, archives, museums such as the movement Linked Open Data Libraries, Archives and Museum (LODLAM) and Linked Open Data Galleries, Libraries, Archives and Museum (LODGLAM) that includes the art galleries. As a result, presents the main institutions of the domain in use of Linked Bibliographic Data and initiatives that link data from libraries, archives, museums and art galleries, with purpose of preserving the Cultural Heritage from the data integration.

Keywords: Semantic Web; Linked Data; Catalog; Metadata; Cataloging.

1 INTRODUÇÃO

A *Web Semântica* anunciada por Berners-Lee, Hendler e Lassila em 2001, proporcionou novos desafios e o desencadeamento de tecnologias para o ambiente da *Web*. Essa perspectiva está relacionada principalmente com a importância que os dados possuem para alcançar o enriquecimento semântico nos registros dos recursos informacionais. Nesse sentido, as informações necessitam de uma infraestrutura para que os computadores possam processar as informações e analisá-las de forma eficiente.

Para isso, foram criadas diversas soluções e tecnologias no intuito de estruturar e suportar essa infraestrutura. Segundo Santarém Segundo (2014, p. 3864) as tecnologias da *Web Semântica* “[...] estão diretamente relacionadas ao processo de construção da informação e armazenamento das mesmas, constituindo assim ambientes que possam ter conjunto de dados ligados semanticamente.” Entre essas tecnologias estão o *Resource Description Framework (RDF)*, *eXtensible Markup Language (XML)* e *Uniform Resource Identifier (URI)*, além de estarem associadas ao uso de metadados e padrões de metadados. O termo metadados é entendido como atributos referenciais que representam um recurso (ALVES, 2010), ou seja, é a junção de ‘propriedade + valor’ que resulta no metadado. O conjunto padronizado e metodologicamente construído é denominado de padrão de metadados. (ZENG; QIN, 2008; ALVES, 2010; ALVES; SANTOS, 2013; SIMIONATO, 2015).

Consequentemente, em 2006, Tim Berners-Lee apresentou o conceito *Linked Data* que consiste nas melhores práticas para ligação de dados estruturados. Nesse contexto, diversas comunidades têm se preocupado em aprimorar seus sistemas e aplicar os princípios do *Linked Data* em seus catálogos. Nesse cenário, o problema caracteriza-se em identificar quais são as iniciativas que estão trabalhando na aplicação do *Linked Data* em catálogos de bibliotecas?

O objetivo é apresentar um panorama dos principais esforços no uso do *Linked Data* em catálogos no domínio bibliográfico. A relevância desta pesquisa para comunidade científica está relacionada à disseminação e propagação de iniciativas que publicam dados estruturados na proposta do *Linked Data* como modelos para instituições e a fomentar uma discussão acerca da temática, no intuito que atingir gestores e especialistas em metadados e demonstrar as possibilidades de uso do *Linked Data* em catálogos.

Trata-se de uma pesquisa teórica e qualitativa, de caráter exploratória da literatura e dos principais estudos para disponibilização de dados ligados no contexto das bibliotecas. Nesse sentido, foram utilizadas como coleta dos dados a pesquisa bibliográfica, que “[...] é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos.” (GIL, 2002, p. 44).

Para localização desses documentos foram utilizadas as bases de dados: Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci), P@rthenon, Portal de Periódicos da Capes, *Scientific Electronic Library Online* (SciELO), *Scopus*, *Web of Science* e anais do Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ENANCIB). Nos idiomas Português, Inglês e Espanhol no período de 2006 a 2016.

O estudo do material levantado seguiu as seguintes etapas:

1. Seleção e análise de textos relevantes para a temática em apreço;
2. Mapeamento das características e os modelos das iniciativas do domínio bibliográfico na proposta do *Linked Data*;

Como resultado, na revisão de literatura é apresentado a definição dos principais termos relacionado como *Linked Data*, dados de bibliotecas e seus princípios para publicação de dados estruturados, por conseguinte são apresentadas as iniciativas para disponibilização de dados estruturados na proposta do *Linked Data*.

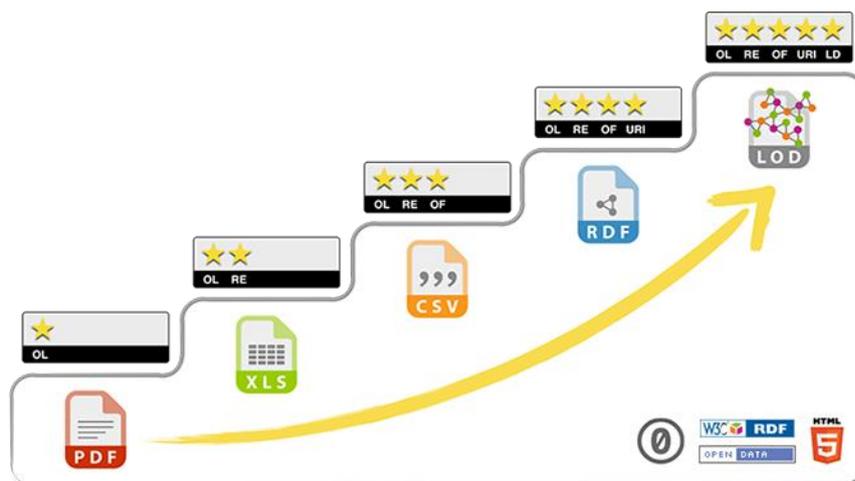
2 LINKED DATA E DADOS DE BIBLIOTECAS

O termo *Linked Data* teve origem em 2006 em um texto de Berners-Lee. O texto apresentou o conceito de *Linked Data* (Dados ligados) e *Linked Open Data* (Dados Abertos Ligados), os princípios e as melhores práticas para publicação e ligação de dados estruturados na *Web*. Segundo Berners-Lee (2006, não paginado, tradução nossa) para publicação de dados ligados e estruturados devem seguir os seguintes princípios:

1. Use URIs como nomes para as coisas;
2. Use HTTP URIs para que as pessoas possam procurar esses nomes;
3. Quando alguém procurar um URI, fornecer informações úteis, usando os padrões (RDF, SPARQL);
4. Incluir links para outros URIs, para que eles possam descobrir mais coisas.

Entretanto, para viabilizar a recuperação e a reutilização de dados no contexto dos dados abertos (*Open Data*), Berners-Lee (2006) utilizou um sistema que classifica em até cinco estrelas a disponibilização dos dados abertos ligados (*Linked Open Data*). A primeira estrela refere-se à disponibilização do recurso informacional em licença aberta (*Open Data*). A segunda estrela corresponde na estruturação de dados legíveis por máquina. Para conseguir as três estrelas é preciso disponibilizar os dados legíveis por máquina em formato não proprietário. Em seguida, para alcançar a quarta estrela é necessário utilizar padrões abertos propostos pelo W3C como RDF e SPARQL *Protocol and RDF Query Language* (SPARQL). Por fim, para conseguir as cinco estrelas, além de ter todas as quatro estrelas, precisa vincular os dados para outros ambientes e contextos, conforme apontado na figura 1.

Figura 1: Cinco estrelas do Linked Open Data.



Fonte: Hausenblas (2012, não paginado).

Bizer, Heath e Berners-Lee (2009, não paginado, tradução nossa) exemplificam que “Em resumo, *Linked Data* é simplesmente sobre como usar a *Web* para criar ligações entre os dados digitados a partir de diferentes fontes.” Baker et al. (2011, não paginado, tradução nossa), complementam que “Enquanto ‘*Linked Data*’ refere-se à interoperabilidade técnica dos dados, ‘*[Linked] Open Data*’ centra-se na sua interoperabilidade legal.”

Dessa forma, para construção de sistemas que suportem a aplicação do *Linked Data* / *Linked Open Data* são necessárias algumas mudanças nas estruturas e desenvolvimento dos catálogos. Assim, em maio de 2010 foi criado um grupo de trabalho em parceria com o

World Wide Web Consortium (W3C) para discutir a aplicação do *Linked Data* em bibliotecas denominado *Library Linked Data Incubator Group*. O objetivo do grupo foi

[...] ajudar a aumentar a interoperabilidade global dos dados da biblioteca na *Web*, reunindo pessoas envolvidas em atividades da *Web Semântica* - com foco em *Linked Data* - em que a comunidade de bibliotecas e além, com base em iniciativas já existentes, e identificar colaboração no futuro. (BAKER et al., 2011, não paginado, tradução nossa).

Segundo o relatório, recomenda-se que as bibliotecas discutam a disponibilização de quais dados podem ser abertos e seus direitos, com a perspectiva de: ampliar a participação das bibliotecas na padronização da *Web Semântica*, desenvolvendo padrões que sejam compatíveis com o *Linked Data*; utilizar as capacidades do *Linked Data* para criar URIs para identificar *datasets* de bibliotecas; desenvolver políticas para o gerenciamento de vocabulários RDF; e reutilizar ou mapear vocabulários *Linked Data* existentes. De modo que bibliotecários e arquivistas apliquem a experiência em curadoria digital para preservação a longo prazo dos *datasets* em *Linked Data*. (BAKER et al., 2011).

O termo *Library Linked Data* (Dados ligados em bibliotecas) é definido como “[...] qualquer tipo de dados de biblioteca [...] que é expresso em *Linked Data*.” (BAKER et al., 2011, não paginado, tradução nossa). No contexto das bibliotecas, Baker et al. (2011, não paginado, tradução nossa, grifo do autor) esclarecem que “‘**Dados de Biblioteca**’ referem-se a qualquer tipo de informação digital produzida ou com curadoria de bibliotecas que descreve recursos ou ajudam a sua descoberta.” Os autores distinguem que há três tipos de dados de bibliotecas, *datasets* (conjuntos de dados), *metadata element set* (conjuntos de elementos) e *value vocabularies* (vocabulários de valor). (BAKER et al., 2011).

Os *datasets* podem ser considerados como conjunto de registros de metadados de uma instituição e consistem em *value vocabularies* e no *metadata element set*. (ARAKAKI, 2016). Os *value vocabularies* definem os valores dos elementos para descrição de um recurso, ou seja, definem conceitos relacionados a um recurso como pessoas, assuntos, idiomas, países, etc., mas, não definem informações de um recurso. (ISAAC et al., 2011). São considerados *value vocabularies* o *Library of Congress Subject Headings (LCSH)*, o *Virtual International Authority File (VIAF)* e a Classificação Decimal de Dewey (CDD). *Metadata element set* (conjunto de elementos) é definido como conjunto de metadados que define as classes e atributos utilizados para descrever entidades. (ISAAC et al., 2011). Zeng e Qin

(2008) complementam que os *Metadata element set* referem-se ao conjunto de metadados e a codificação dos elementos e sua sintaxe e tem como exemplos o *Dublin Core Metadata Terms*; o *Simple Knowledge Organization System (SKOS)*, e o *Friend of a Friend vocabulary (FOAF)*. (BAKER et al., 2011).

3 RESULTADO E DISCUSSÕES: INICIATIVAS DO USO DO LINKED DATA NO DOMÍNIO BIBLIOGRÁFICO

Durante o levantamento, observou que algumas iniciativas destacam-se pelos esforços em disponibilizar e ligar dados no domínio bibliográfico. Essas iniciativas são referências mundiais por serem as primeiras a proporem princípios, diretrizes ou aperfeiçoamento dos metadados a partir do *Linked Data* em catálogos.

3.1 Dublin Core Metadata Initiative - DCMI

Com a meta de identificar metadados mínimos para localizar qualquer recurso no ambiente *Web*, o *Dublin Core* ganhou destaque internacional na década de 1990. Diversas parcerias foram criadas, em especial com o W3C e o *Friend of a Friend (FOAF) Project* para facilitar a ligação de dados de pessoas. (BAKER, 2012, ARAKAKI; SANTOS; ALVES, 2015). Frutos dessas parcerias destacam-se o desenvolvimento do RDF em 1998, a definição da gramática do *Dublin Core* proposto por Baker (2000) que conciliava o *Dublin Core* com o RDF. Posteriormente a proposta do *Dublin Core Abstract Model (DCAM - Modelo Abstrato Dublin Core)* em 2003 e o *Dublin Core Application Profiles (DCAP - Perfil de Aplicação Dublin Core)*.

O *Dublin Core* possui 15 elementos descritivos básicos e tem como características: simplicidade, interoperabilidade semântica, extensibilidade, flexibilidade, modularidade, independência de sintaxe, além de que todos os elementos são opcionais e repetíveis. (WEIBEL, 1995; BAPTISTA; MACHADO, 2001).

O uso do *Dublin Core* é diversificado e pode ser também qualificado com elementos de refinamento e esquemas de codificação que qualificam os metadados. Segundo Alves e Santos (2013, p. 76), os “Elementos de refinamento são qualificadores utilizados com o intuito de refinar ou restringir, especificando e ampliando o significado dos elementos descritivos.”, isto é, a especificidade do metadado. Os “Esquemas de codificação: qualificadores que identificam esquemas para o valor dos elementos [...]” (ALVES; SANTOS, 2013, p. 76). Os esquemas de codificação auxiliam entre outras funções, a padronização no

preenchimento do conteúdo dos elementos, proporcionando maior precisão das informações. O histórico e evolução do *Dublin Core* foi destacado por Baker (2012) e por Arakaki, Santos e Alves (2015).

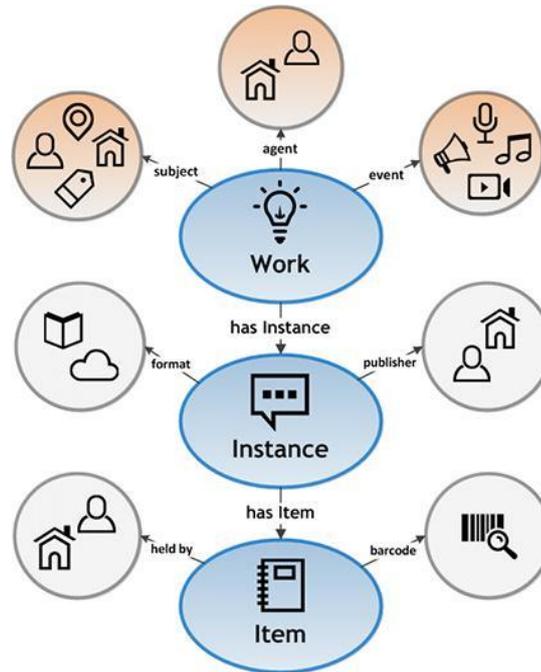
3.2 *Library of Congress* e o modelo BIBFRAME

Em 2011, a *Library of Congress*, dos Estados Unidos, iniciou a construção de um padrão de metadados para o domínio bibliográfico de acordo com a proposta do *Linked Data* e ainda, de acordo com a proposta do *Functional Requirements for Bibliographic Records* (FRBR) denominado *Bibliographic Framework* (BIBFRAME). (LIBRARY OF CONGRESS, 2012). Nesse novo cenário, autores como Silva (2013), Arakaki (2016) e Ramalho (2016) discutiram as transformações e influências que a *Web Semântica* tem realizado no contexto dos dados bibliográficos, em especial no impacto de uma nova estrutura para representação de registros bibliográficos em catálogos de bibliotecas com o BIBFRAME.

Segundo Ramalho (2016, p. 296) “A concepção do modelo BIBFRAME está baseada nas tendências e práticas contemporâneas que incentivam a disponibilização e interligação de dados na *Web* em formatos que sejam legíveis por máquinas.” Nesse contexto, o BIBFRAME tem a proposta de ser o substituto do MARC21, e segundo Silva (2013) suas principais características são: flexibilidade; possuir uma arquitetura para expressar e conectar informações; adoção para além da comunidade de bibliotecas. Diversas bibliotecas estão estudando a implementação do BIBFRAME, destacando a *Library of Congress* (EUA), Biblioteca Nacional de Cuba “José Martí” (BNJM), *German National Library* (Alemanha), *National Library of Medicine* (EUA), *Library of Alexandria* entre outras. (LIBRARY OF CONGRESS, 2015a).

Em abril de 2016, a *Library of Congress* publicou o modelo BIBFRAME 2.0, que está dividido em: *work* (obra), *instance* (instância) e *item* (item) conforme apresentado na figura 2.

Figura 2: Modelo BIBFRAME 2.0.



Fonte: Library of Congress (2016).

A obra é definida como o nível mais alto de abstração. No contexto do modelo 2.0 do BIBFRAME e reflete a essência conceitual do recurso catalogado: autores, idiomas e assuntos. A instância é caracterizada por possuir uma ou mais formas de realização de uma obra. Uma Instância reflete informações como seu editor, local de publicação e data de publicação e formato. Um item é definido como uma cópia real (analógica ou digital) de uma instância. Ele possui informações como a sua localização (presencial ou virtual), marca de prateleira, e código de barras. (LIBRARY OF CONGRESS, 2016).

O modelo BIBFRAME 2.0 aborda ainda algumas definições como *agents* (agentes), *subjects* (assuntos) e *events* (eventos). Os agentes são definidos como pessoas, organizações, jurisdições, etc., associados a uma obra ou instância através de funções como: autor, editor, artista, fotógrafo, compositor, ilustrador, etc. O assunto é caracterizado como o “sobre o quê” de uma obra e pode ter um ou mais conceitos. Esses conceitos incluem temas, lugares, expressões temporais, eventos, obras, instâncias, itens, agentes, etc. Eventos são definidos como ocorrências de gravações do conteúdo de uma obra. (LIBRARY OF CONGRESS, 2016).

Para completar o BIBFRAME, a *Library of Congress* disponibilizou também o vocabulário BIBFRAME que é composto por classes, propriedades e relacionamentos que são

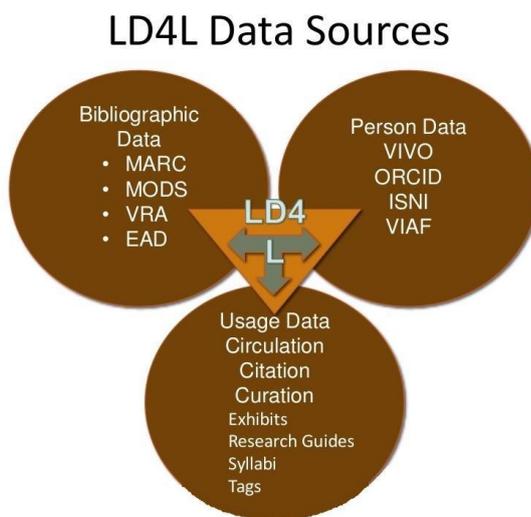
compatíveis com o RDF. O vocabulário BIBFRAME usa o modelo *Linked Data* e aproveita a estrutura do RDF para identificar recursos e todas as suas entidades, atributos e relacionamentos (ou seja, propriedades) entre entidades. (LIBRARY OF CONGRESS, 2015b).

3.3 *Linked Data for Libraries* - LD4L e *Linked Data for Production* - LD4P

O *Linked Data for Libraries* (LD4L) é um projeto idealizado pela parceria entre a *Cornell University Library*, a *Harvard Library Innovation Lab* e a *Stanford University Libraries*. O LD4L consiste em ligar informações de valor intelectual acrescentando por exemplo, notas e organizações, utilizados com outros recursos com valor social denominado como *Scholarly Resource Semantic Information Store* (SRSIS). (LINKED DATA FOR LIBRARIES, 2014).

A proposta do LD4L é fazer com que tanto os padrões bibliográficos, quanto os dados de pessoas e os dados de uso da biblioteca possam estar ligados conforme apresentado na figura 3.

Figura 3: *Linked Data for Libraries* - LD4L.



Fonte: Fallgren; Nelson; Lorimer (2015).

O projeto tem buscado ligar esses dados a partir da estrutura do BIBFRAME, que permitirá realizar as relações. Com a inclusão da *University of Iowa Library* os trabalhos ampliaram para desenvolver ferramentas, ontologias, serviços na proposta do *Linked Data* para melhorar a descoberta, o uso e entendimento de recursos de informações acadêmicas, entre outras atividades. Paralelamente, com a adesão de outras instituições como *Columbia*

University, Library of Congress e Princeton Universit, formou-se o *Linked Data for Production* (LD4P), tendo como objetivo:

- Reinventar a produção de metadados para qualquer recurso de uma biblioteca gerenciada, fazendo uso do *Linked Data* em um ambiente cooperativo, distribuído;
- Desenvolver uma ferramenta essencial e configurações de sistema para gerenciar essa produção de metadados;
- Desenvolver um quadro comunitário/colaborativa para coordenar a ontologia e os trabalhos técnicos. (LD4P..., 2016, não paginado).

Entre os domínios abordados pela iniciativa do LD4P estão as artes, cartografia e geolocalização, imagens em movimento, performance musical e obras raras. Informações sobre o projeto podem ser encontradas em www.ld4l.org e em <https://wiki.duraspace.org/display/ld4l>.

3.4 OCLC e o WorldCat

A *Online Computer Library Center* (OCLC) lançou o *WorldCat* com intuito de trabalhar com dados bibliográficos estruturados. Para isso, o *WorldCat* utiliza instrumentos internacionalmente reconhecidos como *Dewey Decimal Classification* (DDC) para classificação, o *Virtual International Authorities File* (VIAF) para o controle de autoridades de pessoas e entidades coletivas e o *Faceted Application of Subject Terminology* (FAST) utilizado para controle terminológico dos recursos. (ONLINE COMPUTER LIBRARY CENTER, 2015).

Além das ferramentas como o *Linked Open Data*, a OCLC utiliza também o conceito de *Linked Enterprise Data* (LED) que tem como objetivo estruturar dados ligados para fins comerciais. Dessa forma, a proposta da OCLC consiste em ligar dados de bibliotecas e do mercado editorial e livreiro. (ONLINE COMPUTER LIBRARY CENTER, 2015).

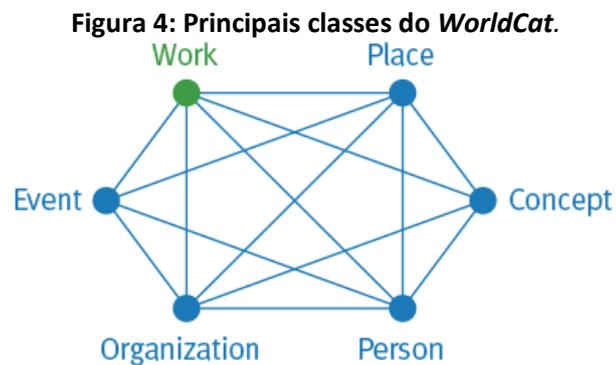
O *WorldCat* possui mais 330 milhões de registros bibliográficos e permitem localizar recursos na *Web*, além de identificar a biblioteca mais próxima do usuário onde o recurso está. (WALLIS, 2015). O “*WorldCat* é a maior rede do mundo de conteúdo e serviços de biblioteca. Bibliotecas do *WorldCat* são dedicados a fornecer o acesso aos seus recursos na *Web*, onde a maioria das pessoas começam sua busca por informações.” (WORLDICAT, 2015b, não paginado, tradução nossa).

Os dados do *WorldCat* estão estruturados com um padrão para *Web* denominado de *Schema.org*. A comunidade do *Schema.org* tem a missão de criar, manter e promover

esquemas estruturados de dados para *Internet*, páginas *Web*, mensagens de *e-mail* e entre outras aplicações. (SCHEMA.ORG, 2015). O *Schema.org* surgiu de uma iniciativa de buscadores como *Google*, *Yahoo*, *Bing* e *Yandex* (WALLIS, 2015), para desenvolver uma estrutura que seja capaz de melhorar a busca de informações.

Wills (2015) destacou as principais características do *Schema.org* como: é utilizado por diversos de domínios; tem como apoio grandes buscadores como *Google*, *Bing*, *Yahoo!*, *Yandex*; pertence uma comunidade W3C Grupos; é um vocabulário para compartilhamento de dados na *Web* e está estruturado em RDF; possibilita utilizar diversas sintaxes como: microdados, RDFa , JSON –LD; possibilita definir dados em um contexto *Web*; contribui para infraestrutura de pesquisas semânticas; é um mecanismo de extensão específica de domínio; e complementa - e não substitui vocabulários específicos de domínio.

A estrutura básica do *Schema.org* para o *WorldCat* é composta por classes e propriedades para representar recursos e ao todo são 640 Tipos (Classes) e 988 Propriedades. (WALLIS, 2015). As principais classes utilizadas pelo *WorldCat* estão representadas na figura 4.



Fonte: WorldCat (2015a).

Segundo o site do Schema.org (2015), as classes são definidas como: *CreativeWork*: O tipo mais genérico da obra criativa, ou seja, a concepção abstrata, incluindo livros, filmes, fotografias, programas de software, etc.; *Event*: Um evento acontecendo em um determinado momento e local, como um concerto, uma conferência, ou festival; *Organization*: Uma organização como uma escola, ONG, sociedade, clube, etc.; *Person*: Uma pessoa (vivo, morto, fictícia, entre outros); *Place*: Entidades que têm uma extensão física.

3.5 Movimentos entre galerias, bibliotecas, arquivos e museus

Outras iniciativas estão voltadas em ligar dados entre biblioteca e outros ambientes informacionais como de arquivos, museus e galerias de arte. O *Linked Open Data in Libraries, Archives and Museums* (LODLAM) é um movimento das bibliotecas, arquivos e museus em reunir e abordar práticas do *Linked Open Data*. O movimento é caracterizado principalmente por um evento que ocorre a cada dois anos, iniciado em 2011. A primeira edição ocorreu em São Francisco, Califórnia (EUA), a segunda edição ocorreu em Montreal em Quebec (Canadá) no ano de 2013. A terceira edição foi em Sidney na Austrália no ano de 2015 e a quarta edição foi em 2017 na Itália. O intuito das reuniões foram:

Identificar as ferramentas e técnicas para a publicação e trabalhar com *Linked Open Data*. Os projetos de precedentes e política de considerações de licenciamento e direitos autorais sobre a publicação de metadados biblioteca, arquivo e museu.

Publicar definições e promover casos de uso que lhe dão equipe LAM as ferramentas que precisam para utilizar *Linked Open Data* em suas instituições. (VOSS, 2012, não paginado, tradução nossa).

A área da Ciência da Informação possui alguns estudos com o propósito de aproximar ainda mais as estruturas que cada uma das instituições (bibliotecas, arquivos e museus) tem criado, a exemplo do que propõe Marcondes (2012; 2016) e Simionato (2015) com o modelo conceitual *Digital Images for Libraries, Archives and Museums* (DILAM). O destaque para o DILAM é o seu objetivo de aproximar e de reaproveitar a descrição realizada em arquivos, em bibliotecas e em museus, respeitando o contexto e a perspectiva que o recurso está sendo representado em cada uma das instituições.

O *Linked Open Data* aplicado em arquivos, bibliotecas e museus poderá ampliar as possibilidades de acesso a partir dos relacionamentos propostos e com isso criar outras possibilidades com sistemas de descoberta mais sofisticados. Assim, o catálogo poderá remeter a outros recursos que estão fora da biblioteca direcionando para um museu ou algum arquivo.

Além do LODLAM, outro movimento que engloba as galerias de arte vem surgindo no desenvolvimento de dados abertos conectados, como é o caso da iniciativa *Linked Open Data Galleries, Libraries, Archives and Museum* – LODGLAM. Sua principal característica está voltada a relação da herança e do patrimônio cultural.

A inserção das Galerias de Arte nesse contexto, abrange uma parceria fundamental para reuso de dados em bibliotecas, arquivos e museus. Com a diversidade de padrões e

esquemas para descrição de recursos, observa-se que há uma tendência de diversas instituições a colaborarem na construção de estruturas que possam trocar informações, mesmo utilizando padrões de metadados heterogêneos.

Outra vertente que vem se configurando nessa perspectiva é o caso da Europeana com o *Europeana Data Model* (EDM), no contexto do *Open Galleries, Libraries, Archives and Museum* (Open GLAM) e da *Digital Public Library of America* (DPLA).

Com intuito de disponibilizar e preservar o patrimônio cultural dos países europeus foi criado a Europeana que é uma biblioteca digital que reúne diversos acervos e sua disponibilização de vários países da Europa. A proposta é padronizar e criar uma estrutura mínima de descrição desses recursos a partir de um modelo denominado *Europeana Data Model* (EDM). O modelo EDM busca padronizar descrições de recursos de patrimônio cultural de instituições como bibliotecas, arquivos, museus e galerias no modelo *Linked Open Data*.

O modelo EDM tem como base, diretrizes e padrões internacionalmente reconhecidos como *Open Archives Initiative Object Resue & Exchange* (OAI-ORE) para organização de objetos e representações digitais, o padrão de metadados *Dublin Core* para descrição dos recursos, para representação dos vocabulários utiliza o *Simple Knowledge Organization System* (SKOS) e ainda o *Conceptual Reference Model* CIDOC-CRM que possibilita a relação de objetos e eventos.

O uso e integração de vocabulários internacionalmente reconhecidos e nascidos de padrões de metadados já sedimentados, torna o EDM um rico modelo de descrição dos objetos da Europeana, fortalecendo a estrutura de descrição semântica dos recursos. (CONEGLIAN; SANTAREM SEGUNDO, 2017, p. 91)

A Europeana também está relacionada com o OpenGLAM (*Galleries, Libraries, Archives and Museum*) que promove o uso e acesso aberto do patrimônio cultural. *OpenGLAM* é uma iniciativa Fundada pela *European Commission* e coordenada pela *Open Knowledge*.

Inserida no âmbito Norte-Americano, a DPLA é uma biblioteca digital que agrega coleções digitais e que incorpora conteúdo instituições como *National Archives and Records Administration* e da *HathiTrust Digital Library*, que é uma biblioteca digital para fins de preservação digital e longo prazo. (GREGORY; WILLIAMS, 2014). A DPLA utiliza um perfil de

aplicação *Dublin Core* denominado *DPLA Metadata Application Profile* (DPLA MAP) e foi baseado no EDM.

Observa-se ao longo do estudo que cada vez mais há iniciativas para evitar o retrabalho da descrição dos recursos. Com o surgimento do *Linked Data*, reforçou as propostas de cooperação entre bibliotecas, arquivos, museus e ainda foram incluídas as galerias de arte que desempenham um papel fundamental na disseminação da informação e que também necessita de uma representação adequada para que seus recursos possam ser localizados.

4 CONSIDERAÇÕES FINAIS

Como visto, o artigo apresenta os principais movimentos de bibliotecas na proposta do *Linked Data*. Diversas iniciativas já estão mobilizando esforços para publicar dados ligados em bibliotecas entre elas pôde-se destacar as iniciativas como a DCMI; a *Library of Congress* e o BIBFRAME que aborda uma perspectiva focada em bibliotecas e utiliza como base o FRBR. A iniciativa do LD4L que busca ligar não apenas recursos e registros informacionais, mas ligar informações e comentários referentes ao próprio recurso. A OCLC com *WorldCat.org*, que visa a construção de um catálogo universal e que possam ser ligados dados de bibliotecas, livrarias e editoras, apresentando assim a biblioteca mais perto do usuário, ou ainda, onde o usuário pode adquirir o recurso de interesse.

Ao verificar as iniciativas encontradas, observa-se grande preocupação para estruturação dos dados em RDF e na abertura dos registros de metadados para recuperação na *Web*, conforme os princípios do *Linked Open Data*. Isso impacta principalmente no reuso dos dados para alimentação de catálogos, possibilitando assim, os princípios propostos por Jewett.

O *Dublin Core* foi a primeira iniciativa para localização de recursos no ambiente da *Web*, dessa forma, influenciou outras iniciativas como a Europeia e o DPLA que utilizam o *Dublin Core* como base para descrição de seus recursos.

Observou-se também uma tendência nas estruturas dos instrumentos de descrição que estão utilizando classes e propriedades para descrever recursos e seus relacionamentos com outros recursos. Ou seja, das estruturas analisadas, há uma tendência do BIBFRAME, do

EDM, do DPLA MAP e do *Schema.org* dividir em classes e propriedades para representar os recursos informacionais.

A integração de domínios também esteve presente nas iniciativas, apresentando como tendência a integração entre bibliotecas, arquivos, museus, galerias de arte, arte, cartografia e geolocalização, imagens em movimento, performance musical e obras raras, além da integração com livrarias e editoras que em alguns casos, há um viés de aproximação entre o catálogo do *e-commerce*.

Outra perspectiva apresentada, foram as iniciativas que envolvem as bibliotecas e outras unidades de informação como museus, arquivos e galerias de arte. Nesse ponto, destaca-se a iniciativa Europeia que busca reunir os mais diversos tipos de recursos do patrimônio cultural. Foi abordado também o movimento LODLAM que busca a integração entre biblioteca, arquivo e museus e ainda o movimento LODGLAM que inclui as galerias de artes com os casos da Europeia com o EDM e da DPLA com o DPLA MAP.

Conclui-se o estudo proporcionando um panorama das principais instituições do domínio bibliográfico com foco em disponibilizar conteúdos no ambiente *Web*, com a possibilidade de ligar dados de bibliotecas, arquivos, museus e galerias de arte, e algumas iniciativas preocupadas com a preservação da herança cultural a partir da integração de dados.

REFERÊNCIAS

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: <http://www.athena.biblioteca.unesp.br/exlibris/bd/bma/33004110043P4/2010/alves_rcv_dr_mar.pdf>. Acesso em: 25 jun. 2017.

ALVES, R. C. V.; SANTOS, P. L. V. A. da C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.

ARAKAKI, F. A. **Linked Data: ligação de dados bibliográficos**. 2016. 144 f. Dissertação (Mestrado em Ciência da Informação)–Universidade Estadual Paulista ‘Júlio de Mesquita Filho’, Marília/SP, 2016. Disponível em: <<https://repositorio.unesp.br/handle/11449/147979>>. Acesso em: 16 jul. 2017.

ARAKAKI, F. A.; SANTOS, P. L. V. A. da C.; ALVES, R. C. V. Evolution of Dublin Core Metadata Standard: an analysis of the literature from 1995-2013. In: INTERNATIONAL CONFERENCE ON

DUBLIN CORE AND METADATA APPLICATIONS, 15., 2015, São Paulo. **Anais...** São Paulo: DCMI; Unesp, 2015. p. 220 - 222. Disponível em: <<http://dcpapers.dublincore.org/pubs/article/view/3787>>. Acesso em: 09 ago. 2017.

BAKER, T. A grammar of Dublin Core. **D-Lib Magazine**, v. 6, n. 10, out., 2000. Disponível em: <<http://dlib.org/dlib/october00/baker/10baker.html>>. Acesso em: 24 jun. 2017.

BAKER, T. Libraries, languages of description, and Linked Data: A Dublin Core perspective. **Library Hi Tech**, v. 30, n. 1, Mar, 2012. p. 116-133. DOI: 10.1108/07378831211213256

BAKER, T. et al. **Library Linked Data Incubator Group Final Report**. W3C Incubator Group Report, 2011. Disponível em: <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>. Acesso em: 20 jun. 2017.

BAPTISTA, A. A.; MACHADO, A. B. Um gato preto num quarto escuro: falando sobre metadados. **Revista de Biblioteconomia de Brasília**, v. 25, n. 1, p. 77-90, 2001.

BERNERS-LEE, T. **Linked Data: Design Issues**. [S.l.]: W3C, 2006. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 25 jun. 2017.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, May 2001, p. 29-37.

CONEGLIAN, C. S.; SANTAREM SEGUNDO, J. E. Europeana no Linked Open Data: conceitos de Web Semântica na dimensão aplicada das Humanidades Digitais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 22, n. 48, p. 88-99, dez. 2016. ISSN 1518-2924. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n48p88/33031>>. Acesso em: 04 ago. 2017.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas: 2002.

GREGORY, Lisa; WILLIAMS, Stephanie. On being a Hub: some details behind providing metadata for the Digital Public Library of America. **D-Lib Magazine**, EUA, v. 20, n. 7/8, Jul./Aug. 2014. Disponível em: <<http://www.dlib.org/dlib/july14/gregory/07gregory.html>>. Acesso em 13 jul. 2017

HAUSENBLAS, M. **5 estrelas dos dados abertos**. Disponível em: <<http://5stardata.info/pt-BR/>>. Acesso em 09 ago. 2017.

ISAAC, A. et al. **Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets: W3C Incubator Group Report 25 October 2011**. W3C, 2011. Disponível em: <http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/#Published_Datasets>. Acesso em: 21 jun. 2017.

LINKED DATA FOR LIBRARIES. 2014. Disponível em: <<https://www.ld4l.org/>>. Acesso em: 18 jun. 2017.

LIBRARY OF CONGRESS. **Bibliographic Framework as a Web of Data**: Linked Data Model and Supporting Services. Washington, DC. Nov./2012. Disponível em:
<<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>> Acesso em: 5 ago. 2017.

LIBRARY OF CONGRESS. **BIBFRAME Implementation Register**. Washington, DC. 2015a. Disponível em: <<http://www.loc.gov/bibframe/implementation/register.html>>. Acesso em: 15 jun. 2017.

LIBRARY OF CONGRESS. **BIBFRAME Vocabulary**: Terminology and Conventions. EUA: LC, 2015b. Disponível em: <<http://www.loc.gov/bibframe/docs/vocab-conventions.html>>. Acesso em: 14 jul. 2017.

LIBRARY OF CONGRESS. **Overview of the BIBFRAME 2.0 Model**. Washington, DC. 2016. Disponível em: <<https://www.loc.gov/bibframe/docs/bibframe2-model.html>>. Acesso em: 15 jul. 2017.

ONLINE COMPUTER LIBRARY CENTER. **OCLC Linked Data**. 2015. Disponível em:
<<https://www.oclc.org/developer/develop/linked-data.en.html>>. Acesso em: 15 jul. 2017.

RAMALHO, R. A. S. Bibframe: modelo de dados interligados para bibliotecas. **Informação & Informação**, [S.l.], v. 21, n. 2, p. 292–306, dez. 2016. ISSN 1981-8920. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/26425>>. Acesso em: 09 ago. 2017. doi: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/26425>>. 10.5433/1981-8920.2016v21n2p292.

SANTAREM SEGUNDO, J. E. Web Semântica: introdução a recuperação de dados usando SPARQL. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: além das nuvens, expandindo as fronteiras da Ciência da Informação, 15., 2014. Belo Horizonte, MG. **Anais...** Belo Horizonte, MG: ECI, UFMG, 2014. p. 3863-3882. Disponível em: <<http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>>. Acesso em: 16 jul. 2017.

SCHEMA.ORG. **Welcome to Schema.org**. 2015. Disponível em: <<http://schema.org/>>. Acesso em 30 jun. 2017.

SILVA, R. E. da. **As tecnologias da Web semântica no domínio bibliográfico**. 134 f. 2013. Dissertação (Mestrado em Ciência da Informação)–Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Filosofia e Ciências, Marília, 2013. Disponível em: <<http://hdl.handle.net/11449/93653>>. Acesso: 24 jun. 2017.

SIMIONATO, A. C. **Modelagem conceitual DILAM**: princípios descritivos de arquivos, bibliotecas e museus para o recurso imagético digital. 2015. 200 f. Tese (doutorado) - Universidade Estadual Paulista Julio de Mesquita Filho, Faculdade de Filosofia e Ciências, 2015. Disponível em: <<http://hdl.handle.net/11449/123318>>. Acesso em: 03 jul. 2017.

VOSS, J. Radically Open Cultural Heritage Data on the Web. In: MUSEUMS AND THE WEB 2012. **Anais...** San Diego, CA, USA, 2012. Disponível em: <http://www.museumsandtheWeb.com/mw2012/papers/radically_open_cultural_heritage_data_on_the_w>. Acesso em 14 jul. 2017.

WALLIS, R. Schema.org Structured Data on the Web: an extending influence. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 15., 2015, São Paulo. **Anais...** . São Paulo: DCMI; Unesp, 2015. Disponível em: <<http://dcevents.dublincore.org/IntConf/dc-2015/paper/view/321>>. Acesso em: 13 ago. 2017.

WEIBEL, S. Metadata: the foundations of resource description. D-Lib Magazine. EUA, jul., 1995. Disponível em: <<http://www.dlib.org/dlib/July95/07weibel.html>>. Acesso em: 30 jun. 2017.

WORLDCAT. **What is WorldCat?** Estados Unidos: OCLC, 2015a. Disponível em: <<https://www.worldcat.org/whatis/default.jsp>>. Acesso em: 20 jun. 2017.

ZENG, M. L.; QIN, J. **Metadata**. New York: Neal-Schuman Publishers, 2008.