

WEB SEMÂNTICA E BIBLIOTECAS: ASPECTOS RELACIONADOS COM A NORMALIZAÇÃO, GESTÃO E RECUPERAÇÃO DA INFORMAÇÃO NA WEB



<https://doi.org/10.22533/at.ed.491112501047>

Data de aceite: 11/09/2025

Thiago Giordano de Souza Siqueira

Universidade Estadual Paulista (Unesp),
Faculdade de Filosofia e Ciências, Marília
Programa de Pós-graduação em Ciência
da Informação

Resumo: Trata-se de um levantamento teórico-bibliográfico que ilustra conceitos da web semântica no campo da Ciência da Informação que emerge como uma potencialidade nos novos padrões de sistemas de recuperação da informação. Apresenta, ademais, uma ilustração da possibilidade de uso da proposta dos dados abertos e da tecnologia como mecanismos para tornar interoperáveis as informações em acervos de bibliotecas.

Palavras-chave: Web semântica. Linked Data. Biblioteconomia e Ciência da Informação. Bilbliotecas.

INTRODUÇÃO

A Web Semântica é uma web mais organizada onde os recursos são descritos de forma mais inteligente, para que seja possível conectar dados e conteúdo. Desde a sua implantação tecnológica, o seu

desenvolvimento tem sido complicado, pois representa uma oportunidade crescente de experiências do usuário na recuperação da informação de forma ágil.

A ideia é que os documentos sejam enriquecidos com informações sobre os conteúdos / formatos e estes sejam interligados por nós. Devendo ser levado em consideração que a informação, hoje, nasce comumente no formato digital.

A biblioteca, dentro de sua função, sempre busca direcionar esforços para atender aos sujeitos informacionais e comunidades de interesse; o catálogo, por exemplo, deve ser recuperado por todas as ferramentas de busca, como o Google ou similares, por exemplo. Todavia, para que isso seja possível os dados devem ser rastreáveis e, para isso, devem ser visíveis.

Percebe-se a diversidade de conteúdos na internet, muitos deles disponíveis e agrupando diferentes formatos. Diante desse cenário, não é mais suficiente ter um catálogo automatizado de acesso público online (OPAC - do inglês *On-line Public Access Catalog*) dos materiais informativos de uma biblioteca, é necessário enriquecê-lo.

Sabe-se que, como profissional da informação, especialmente o bibliotecário tem a responsabilidade de potencializar o uso de acervos, para justificar sobretudo, os investimentos que a instituição mantenedora aporta para esse espaço, mas não existem atividades eficientes desenvolvidas no ambiente digital, considerando as limitações dos *softwares* de gestão que, no máximo, possuem o padrão MARC¹ e o protocolo Z39.50², que se restringem ao nosso campo técnico de atuação. Em suma, o registro dos materiais é fixo e restrito ao catálogo on-line, mas na maioria das vezes sem a integração com a *web*.

O ensaio tem o objetivo de expor em que medida as contribuições de dados abertos podem ser úteis, viáveis e aplicáveis no serviço de bibliotecas. Como específico apresentam-se: a) Conhecer os novos padrões e tecnologias de dados abertos que estão a ser desenvolvidos na web semântica; b) Identificar os trabalhos práticos desenvolvidos e c) Descrever os sucessos e possíveis dificuldades existentes na aplicação da web semântica para recuperação de informação.

Esse caso serve para ilustrar a possibilidade proveniente da interligação de dados provenientes de diferentes fontes, partindo do catálogo da Biblioteca Nacional Argentina. Tal escolha se deu porque este estudo foi recuperado e aprofundado a partir de discussões propostas durante a disciplina *Gestão de conteúdos e tecnologias de redes sociais* no Mestrado em Biblioteconomia e Ciência da Informação na Universidade de Buenos Aires (UBA) e aproximando com discussões retomadas no âmbito da Disciplina *Web: trajetória e perspectivas para a Ciência da Informação* no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (Unesp).

A escolha do autor e da obra ocorreu pela relevância na literatura Argentina, e por conseguinte, partir da hipótese de que teríamos dados suficientes para a interligação de dados de arquivos, bibliotecas e museus. Assim, por meio de vocabulários já existentes para ampliar a semântica dos conteúdos e descrevê-los em RDF, evidenciando a utilidade desta aplicabilidade nas bibliotecas.

O percurso metodológico segue um método de análise das possibilidades de aplicação da web semântica para o gerenciamento da informação em bibliotecas.

Desta forma, configurando-se como um estudo documental, não exaustivo, a fim de conhecer e ilustrar o estado da questão.

1. Os formatos MARC 21 são padrões para a representação e comunicação de informações bibliográficas e relacionadas em formato legível por máquina. O registro MARC envolve três elementos: a estrutura do registro, a designação do conteúdo e o conteúdo dos dados do registro. (LIBRARY OF CONGRESS, 2006)

2. O Z39.50 é um protocolo de comunicação entre computadores projetado para permitir a busca e recuperação de informação - documentos com textos completos, dados bibliográficos, imagens, multimídia - em redes distribuídas de computadores. Baseado na arquitetura cliente / servidor e operando na Internet, o protocolo viabiliza um número crescente de aplicações. E como esse ambiente é muito dinâmico, no qual o protocolo é aplicado, é necessário que a norma seja constantemente revisada e atualizada para proporcionar as mudanças que os criadores, provedores e usuários das informações precisam. (AMERICAN NATIONAL STANDARDS ORGANIZATION, 2003)

CONCEITOS-CHAVE

Para entender melhor o que este ensaio propõe, devemos ser muito claros sobre alguns conceitos como interoperabilidade, web semântica, dados vinculados (*linked data*), RDF e ontologias – pois existe uma ligação entre cada um deles.

Interoperabilidade

Interoperabilidade é a capacidade de vários sistemas heterogêneos de trocar informações e interpretá-las da mesma maneira. No mundo das aplicações tecnológicas e digitais, pode-se dizer que é aquela que suporta a comunicação entre dois sistemas através da utilização de sinais e protocolos de comunicação que implica um legado comum. Ou seja, além da capacidade dos dados serem trocados entre diferentes sistemas (portabilidade).

Embora haja abundante literatura sobre modelos de interoperabilidade e em seus respectivos níveis, os autores Manso, Wachowicz, Bernabé, Sanchez e Rodriguez (2008) fizeram um extenso estudo sobre metadados e descreve os padrões de interoperabilidade em diferentes níveis e sendo analisados sob três dimensões fundamentais, de acordo com o modelo proposto pela EIF - *European Interoperability Framework* (2004) com a colaboração organizacional, técnica e semântica.

Assim, a dimensão organizacional - refere-se à definição de objetivos orientados aos processos do negócio e para a concretização da colaboração entre gestores que desejam trocar informações e podem ter diferentes estruturas e processos internos. Por outro lado, visa atender às necessidades dos usuários, tornando os serviços disponíveis facilmente identificáveis, acessíveis e orientados para o usuário. A dimensão técnica corresponde aos aspectos técnicos de conexão de sistemas e serviços de computador. Inclui aspectos como interfaces abertas, serviços de interconexão, integração de dados e *middleware*, apresentação e troca de dados, serviços de acessibilidade e de segurança.

E a dimensão semântica preocupa-se em garantir que os dados transferidos compartilhem o mesmo significado para os outros sistemas vinculados. O objetivo é garantir que a consistência de representação e recuperação do conteúdo trocado seja compreensível por qualquer outro, uma vez que outro aplicativo que não foi inicialmente desenvolvido para esse fim.

Web semântica

A proposta da web semântica surgiu em um momento em que os sistemas de recuperação da informação apresentavam dificuldades para lidar com a real necessidade do usuário no momento da busca. Com ênfase nas tecnologias e linguagens, passou a focar nos aspectos de exibição e apresentação dos dados para que as informações não fiquem mal descritas e consequentemente pouco passíveis de serem inteligíveis por máquinas e consumidas por humanos (Souza; Alvarenga, 2004).

Por semântica entendemos dar significados a uma coisa. Em termos de representatividade, podemos atribuir significados que darão sentido e compreensão a um recurso, informação e objeto. E transportamos isso para os recursos no ambiente web.

Berners-Lee; Hendler e Lassila (2001) apresentaram uma ideia de como esse modelo conceitual funcionaria e seria implementado. Conceituando a Web Semântica como uma extensão da Web existente até aquele momento, com o diferencial que cada informação é dada por um significado bem definido e estruturado, fazendo com que a interação humano-computador funcione melhor a medida em que a informação legível por máquina seja capaz de compreender a lógica do pensamento do humano e o sistema computacional capaz de analisar os padrões de interação existente.

A Web Semântica não se trata apenas da publicação de dados na Web, mas estes podem estar ligados a outros, para que pessoas e máquinas possam explorar a web de dados, podendo chegar a informações relacionadas que são referenciadas a partir de outros dados iniciais. Possui uma arquitetura de informação cujos recursos representam conhecimentos que concordam que podem ser compreendidos por máquinas, não apenas por pessoas.

O modelo de publicação incorpora uma série de *tags* que descrevem o formato do documento e o conteúdo do documento, com um conjunto de *tags* que descrevem as entidades do documento, por exemplo: quando há uma pessoa, quando há uma organização etc.

Desta forma, é criado um gráfico que as máquinas podem interpretar e então criar operações, ligadas entre si de acordo com os interesses das pessoas. É uma rede de dados que pode tornar os conjuntos de dados estruturados que ali estão disponíveis uma possível solução para as dúvidas dos usuários por meio da recomendação.

Linked data ou dados vinculados e Linked Open Data

Linked Data ou Dados vinculados é a forma que a Web Semântica tem de vincular os diferentes dados que são distribuídos na Web, de modo que sejam referenciados da mesma forma que os *links* das páginas na web (WORLD WIDE WEB CONSORTIUM, 2016).

De forma simplificada, referem-se a um conjunto de princípios e boas práticas de publicação e vinculação de dados estruturados na web. Isso não significa que os dados devam ser de acesso, de uso livres ou abertos como postulados pelo *Open Data*. A ideia de dados vinculados oferece grande vantagem porque, por meio de *links* semânticos, eles podem ser processados e explorados.

Numa escala maior de desenvolvimento e abertura, à medida que os dados ligados forem publicados na web, por meio de linguagens explícitas, sob uma licença aberta e descritos de modo padronizado, são chamados de *Linked Open Data* (LOD) ou “Dados Abertos e Ligados”, recomendando que seu reuso e compartilhamento sejam livres e gratuitos (Berners-Lee, 2006).

A adoção do LOD, apresenta benefícios tanto para as bibliotecas e pode ser extensivo para enriquecer informações bibliográficas na *Web*, e parte desse movimento nesse sentido já pode ser observado devido a adoção do Dublin Core em catálogos de

bibliotecas e contribuindo para a redução das barreiras para que os serviços incorporem mais dados de forma massiva utilizando algumas APIs³ e consequentemente tornará as suas fontes de dados bem conhecidas e promoverá o seu uso (Saorin, 2012).

RDF (Resource Description Framework)

Segundo Souza e Alvarenga (2004, p.136) “A idéia do RDF é a descrição dos dados e dos metadados por meio de um esquema de ‘triplas’ de recurso propriedade-valor, e uma forma coerente de acesso aos padrões de metadados (namespaces) publicados na Web”. Ou seja, é a infraestrutura necessária para a descrição de recursos, que pode ser compreendida como uma base de dados para processar metadados e ao mesmo tempo que pode ser utilizada para a recuperação dos recursos *web*. Fornece melhores utilidades aos buscadores e catálogos nos quais podem descrever os conteúdos e ao mesmo tempo, as relações com os conteúdos livres na *web*, em bibliotecas digitais; graças ao *software* inteligente que facilita esta troca e desta forma promove uma troca de conhecimento.

Em resumo, alguns dos benefícios do padrão RDF são:

- Fornecer um ambiente consistente para publicação e uso de metadados na *web* usando a infraestrutura XML;
- Oportunizar a descrição de significados sobre o recurso;
- Permitir a interação de informações à medida em que tem como princípio a ligação de um recurso a outro dentro na rede;
- Fornecer uma sintaxe padrão para a descrição de recursos e propriedades de documentos na *Web*;
- Permitir que os aplicativos atuem de forma inteligente e automatizada sobre as informações veiculadas na *Web*, já que seu significado é mais facilmente compreendido;

XML permite a criação de marcações definidas pelo próprio usuário, podendo ser considerada como um sistema gramatical para construção de linguagens de marcação personalizadas. (Ramalho, 2006). Tendo isso em conta, podemos dizer que o objetivo geral do RDF é encontrar uma forma de definir um mecanismo de descrição de recursos sem gerar uma ausência em algum comando de aplicação estabelecido, muito menos definir a semântica de qualquer comando; portanto, o referido mecanismo tende a ser neutro para poder descrever as informações sobre qualquer comando. Para facilitar a definição de metadados. É por isso que o RDF será composto por um sistema de classificação semelhante aos sistemas de programação.

Graças ao fato de que RDF pode representar as sentenças sobre os recursos da *web* e compartilhar diferentes esquemas, assim os agentes provenientes desses dados podem rastrear a origem dos metadados. Isso revela a extensibilidade e o compartilhamento que o RDF permite aos criadores de metadados misturar e apresentar dados usando o trabalho de outros.

3.. API é uma interface de programação de aplicações que permite que sua solução ou serviço se comunique com outros produtos e serviços sem precisar saber como eles foram implementados

Ontologias e Inteligência Artificial

No contexto da Inteligência Artificial (IA), as ontologias cumprem um papel essencial como mecanismos que permitem a representação formal do conhecimento, possibilitando que sistemas computacionais infiram, recomendem ou interajam de forma mais eficiente com os usuários. Em unidades de informação, elas sustentam desde o funcionamento de assistentes virtuais até sistemas de recomendação personalizados, ao oferecer uma base semântica que orienta a interpretação automática de conteúdos e relações conceituais. Isso contribui para a organização inteligente de dados e para a mediação informacional em ambientes digitais.

A ontologia, nesse sentido, configura-se como uma técnica avançada de organização da informação que tem recebido especial atenção nos últimos anos, principalmente no que diz respeito à representação formal do conhecimento e à interoperabilidade semântica entre sistemas. Guarino (1998) um dos pioneiros nessa área de estudos, e relata que a pesquisa em ontologias começa na comunidade científica da informática, posicionando-se a respeito da IA, da linguística computacional e da teoria dos bancos de dados.

No contexto da *Web*, Ramalho (2006, p.97), define a ontologia no contexto da Web Semântica como um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica estruturada e formal. Essa descrição dos aspectos semânticos de conteúdos informacionais, possibilita a realização de inferências automatizadas por sistemas computacionais, tornando possível a atuação de agentes inteligentes no tratamento da informação.

Sua construção responde à necessidade de um vocabulário comum para a troca de informações entre humanos e máquinas, conforme reforçam Souza e Alvarenga (2004). Essa característica torna as ontologias elementos essenciais para a interoperabilidade e para a padronização do significado dos dados trocados em redes complexas de informação.

Smith (2004) salienta o caráter transcategorístico das ontologias, dada sua capacidade de lidar com relações entre entidades oriundas de diferentes áreas do conhecimento, incluindo aquelas reconhecidas pelo senso comum. E é precisamente essa riqueza de descrição (de todas as várias partes e dimensões da realidade em todos os níveis de granularidade) que permitirão o acesso à realidade do senso comum (coisas).

Nas bibliotecas digitais, por exemplo, ontologias permitem que conteúdos sejam vinculados de forma mais contextualizada, associando obras a temáticas correlatas mesmo quando expressas em terminologias distintas. Isso amplia a eficácia da recuperação da informação e proporciona uma navegação mais significativa para os usuários.

A principal contribuição das ontologias para a IA está na representação do conhecimento e na modelagem conceitual dos Sistemas de Informação (SI) das comunidades de ontologias. Isso porque aqui, o objetivo é integrar os sistemas automatizados existentes e se apresenta como o principal problema denominado por Smith (2004) como “Torre de Babel”, pois permite a união de múltiplas bases de dados além de poder utilizar as mesmas identificações, mas com significados diferentes; em alternativa, o mesmo significado pode

ser expresso por nomes diferentes. Isto é, diferentes bases de dados utilizam os mesmos termos com significados distintos ou significados semelhantes expressos por nomes diferentes – o que dificulta a integração entre sistemas e a interoperabilidade plena dos dados.

Evidentemente, isso deve ser considerado um imenso problema, principalmente quando a fusão de ontologias ou sistema que permite sua utilização, pois os conceitos trazidos por especialistas ou grupos disciplinares, também significa que se trata de conceitos e significados particulares e correspondentes para cada um dos indivíduos, bem como para cada um dos ramos em que os conceitos evoluíram independentemente uns dos outros.

Com relação a essas estruturas formais existentes e os obstáculos enfrentados por tais sistemas de informação ontológica em geral, eles são produzidos em grande parte por motivos de eficiência computacional. Com o propósito de estabelecer regras de inferência e outros formalismos linguagem mais utilizada é a OWL, que pode ser entendida como um idioma complementar ao RDF para definir as relações e “[...] apresentar propriedades que inserem lógica e axiomas nas relações existentes[...]” (Santarém Segundo; Coneglian; Lucas, 2017, p. 299).

Apesar de seu potencial expressivo, a adoção de ontologias em ambientes informacionais ainda enfrenta obstáculos, sobretudo em instituições com poucos recursos técnicos ou com dados pouco estruturados. Ainda assim, elas constituem um dos pilares conceituais e operacionais mais promissores para o desenvolvimento de soluções baseadas em IA no âmbito das bibliotecas.

A PROPOSTA DE APLICAÇÃO DE DADOS VINCULADOS PARA COLEÇÕES DE BIBLIOTECAS

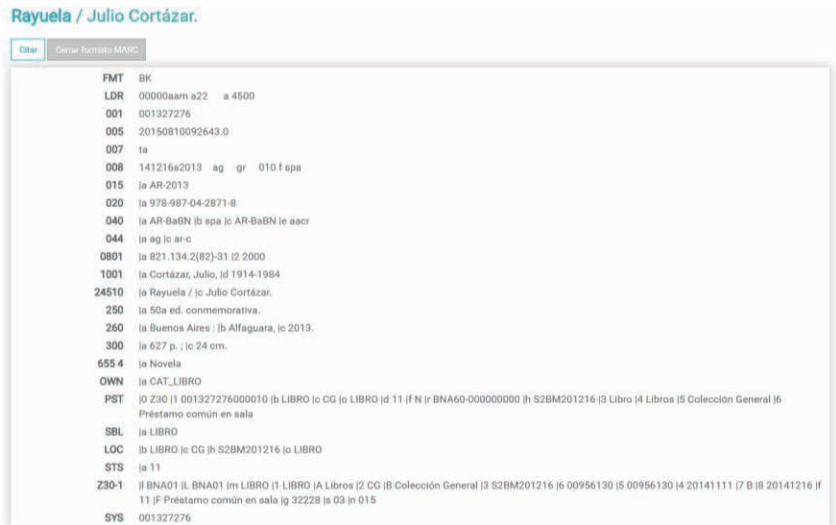
As bibliotecas têm grande potencial para compartilhar seu conteúdo na web se as entendermos como instituições culturais e de memória, ou seja, há um interesse comum à humanidade.

A primeira invenção que se aproxima da tentativa foi criada na década de 70, os Sistemas Integrados de Bibliotecas (ILS, do inglês *Integrad Library Systems*) que a partir dos anos 80 passaram a oferecer seus catálogos online ou OPACs. O critério de eficiência não é totalmente atendido para os dias atuais, visto que os catálogos online existentes são amplamente limitados a ambientes sistêmicos e organizacionais. Com a tecnologia aplicada aos catálogos das bibliotecas, é possível criar vínculos (*links*) de um registro de catálogo para um recurso externo da *web*, mas dificilmente se aceita como prática comum a incorporação de links para registros de outras bibliotecas. O que causa uma perda de agrupamento das informações para o usuário ao idealizar que se tudo fosse interoperável, seria possível conectar diferentes conteúdo do mesmo assunto, como: vídeos, músicas, mapas, fotos, entre outras coisas relacionadas.

Assim, o primeiro passo para uma biblioteca se adaptar à proposta de dados abertos interconectados seria converter seu catálogo de formato MARC para RDF. Onde cada obra será considerada entidades de interesse com suas respectivas propriedades e desdobradas em classes que na estrutura do RDF são identificadas com os espaços de nomes (*namespaces*).

Obviamente, um livro não possui um URI devido à sua aparência como um objeto físico, portanto o que vamos atribuir como um atributo que lhe dará uma característica única será o *International Standard Book Number* (ISBN - Número Internacional Padrão do Livro, em português). Mas vamos saber que se a biblioteca quiser aderir à proposta de dados abertos, um URI deve ser criado para o arquivo do livro. **Rayuela**, então terá uma representação no ciberespaço.

A título de ilustração, segue abaixo um registro do MARC referente a uma obra de Julio Cortázar recuperada no catálogo da Biblioteca Nacional da República Nacional da Argentina “Mariano Moreno” (Figura 1), e mapeou-se manualmente os campos com seus respectivos equivalentes em elementos em *Dublin Core*⁴ (Figura 2) que possui vínculos semânticos e finalmente em sua conversão final para um gráfico RDF validado no serviço do Grupo de Trabalho RDFCore da W3C⁵ (Figura 3) onde é possível verificar e visualizar documentos RDF, que neste caso foi gerado a partir do documento original estruturado.



| | |
|-------|---|
| FMT | BK |
| LDR | 00000aam a22 a 4500 |
| 001 | 001327276 |
| 005 | 20150810092643.0 |
| 007 | ta |
| 008 | 141216a2013 ag gr 010 f spa |
| 015 | la AR-2013 |
| 020 | la 978-967-04-2871-8 |
| 040 | la AR-BaBN lb spa lc AR-BaBN le aacr |
| 044 | la aq lc ar-c |
| 0801 | la 821.134.2(82)-31 (2 2000 |
| 1001 | la Cortázar, Julio, ld 1914-1984 |
| 24510 | la Rayuela / lc Julio Cortázar. |
| 250 | la 5da ed. conmemorativa. |
| 260 | la Buenos Aires : lb Alfaguara, lc 2013. |
| 300 | la 627 p. : lc 24 cm. |
| 655 4 | la Novela |
| OWN | la CAT_LIBRO |
| PST | ld Z30 l1 001327276000010 lb LIBRO lc CG lc LIBRO ld 11 lf N lr BNA60-000000000 lh S2BM201216 lj Libro l4 Libros l5 Colección General l6 Préstamo común en sala |
| SBL | la LIBRO |
| LOC | lb LIBRO lc CG lh S2BM201216 lo LIBRO |
| STS | la 11 |
| Z30-1 | lf BNA01 ll BNA01 lm LIBRO l1 LIBRO lA Libros l2 CG lB Colección General l3 S2BM201216 l6 00956130 l5 00956130 l4 20141111 l7 B l8 20141216 lf 11 lf Préstamo común en sala lg 32228 lx 03 ln 015 |
| SYS | 001327276 |

Figura1. Registro do MARC referente a obra Rayuela

Fonte: Captura de tela do registro MARC da obra no Catálogo da Biblioteca Nacional Mariano Moreno (2021).

4. DCMI Metadata Terms (também conhecido como “o Dublin Core”) mais várias dezenas de propriedades, classes, tipos de dados e esquemas de codificação de vocabulário. Esses termos devem ser usados em combinação com termos de metadados de outros vocabulários compatíveis no contexto de perfis de aplicativos.

5. World Wide Web Consortium (W3C) é uma comunidade internacional onde as organizações membros, uma equipe em tempo integral e o público trabalham juntos para desenvolver padrões da web.

Observa-se na Figura 2 que os *namespaces*, são definidos a partir de declarações representadas em linguagem XML, permitindo identificar sem ambiguidades a localização dos vocabulários correspondentes ao conjunto de conceitos utilizados na ontologia e declarados entre etiquetas do tipo *rdf:RDF*.

```
1: <?xml version="1.0" encoding="utf-8" ?>
2: <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3: xmlns:dc11="http://purl.org/dc/elements/1.1/">
4: <dc11:simpledc rdf:nodeID="genid1">
5: <dc11:title>Rayuela</dc11:title>
6: <dc11:creator>Julio Cortázar</dc11:creator>
7: <dc11:publisher>Alfaguara</dc11:publisher>
8: <dc11:date>2013</dc11:date>
9: <dc11:identifier>isbn:978-987-04-2871-8</dc11:identifier>
10: </dc11:simpledc>
11: </rdf:RDF>
```

Figura 2. Mapeamento dos campos com seus respectivos equivalente em elementos no Dublin Core

Fonte: Elaborada pelo autor (2021).

Na Figura 3 temos marcações semânticas dos documentos e das relações trípticas, onde de recurso (sujeito), propriedade (predicado) e valor (objeto). Aproximando a organização para os sistemas de informação de forma mais intuitiva e coerente a partir da lógica do funcionamento cognitivo dos seres humanos.

| Number | Subject | Predicate | Object |
|--------|---------------|---|--|
| 1 | genid:Ugenid1 | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://purl.org/dc/elements/1.1/simpledc |
| 2 | genid:Ugenid1 | http://purl.org/dc/elements/1.1/title | "Rayuela" |
| 3 | genid:Ugenid1 | http://purl.org/dc/elements/1.1/creator | "Julio Cortázar" |
| 4 | genid:Ugenid1 | http://purl.org/dc/elements/1.1/publisher | "Alfaguara" |
| 5 | genid:Ugenid1 | http://purl.org/dc/elements/1.1/date | "2013" |
| 6 | genid:Ugenid1 | http://purl.org/dc/elements/1.1/identifier | "isbn:978-987-04-2871-8" |

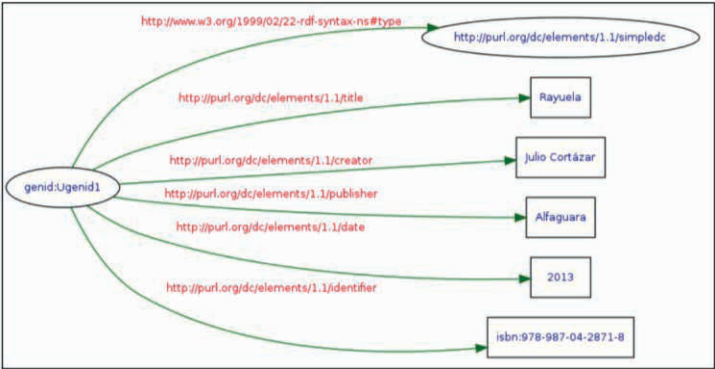


Figura 3. Tripla e Gráfico do modelo de dados RDF da obra Rayuela

Fonte: Validation Service W3C (2021).

Obviamente, é preciso agregar valor aos registros bibliográficos, associando cada es, quando possível, a um conjunto de *links* semânticos. Por exemplo: para o campo istro de autoridades (dc: criador), existe um arquivo como o Arquivo de Autoridadesacionais Virtuais (VIAF) com informações detalhadas sobre vários autores, como pode to a seguir no caso de Cortázar e auxilia na padronização das fontes de autoridades s devidas filiações (Figura 4).



O VIAF recolhe os registros de autoridade de agências nacionais e regionais, disponibilizando e ligando muitos dos registros para o DBPedia⁶. Dessa forma é possível consultar quais são as bibliotecas que contêm cada obra e cada versão do autor.

No catálogo baseado na proposta de dados abertos, ao mesmo tempo é possível mudar o campo **<dc11:creator>Julio Cortázar</dc11:creator>** pelo URI do registro VIAF sobre o autor: **<dc11:creator><http://viaf.org/viaf/12306822></dc11:creator>**.

A obra *Rayuela* (“Jogo da amarelinha, em português) foi escrita em Paris e publicada no ano de 1963 pela primeira vez na Espanha. A primeira parte da história se passa em

Comunicação e conexões digitais: Transformações e estratégias na Era da Informação

Paris, e um dos primeiros locais mencionados no romance são os passos dos passos dos personagens Maga e Horácio Oliveira na capital francesa, onde destaca-se a *Le Pont Neuf* (Ponte Nova, em português) patrimônio da humanidade reconhecida pela Unesco por ser a primeira ponte em pedra construída no país, localizada na região próxima do *Quai de Conti* - uma doca localizada ao longo do Sena.

Dessa forma, poderia ser agregado ao conjunto de informações o mapa da ponte adicionado a partir do banco de dados geográfico GeoNames.

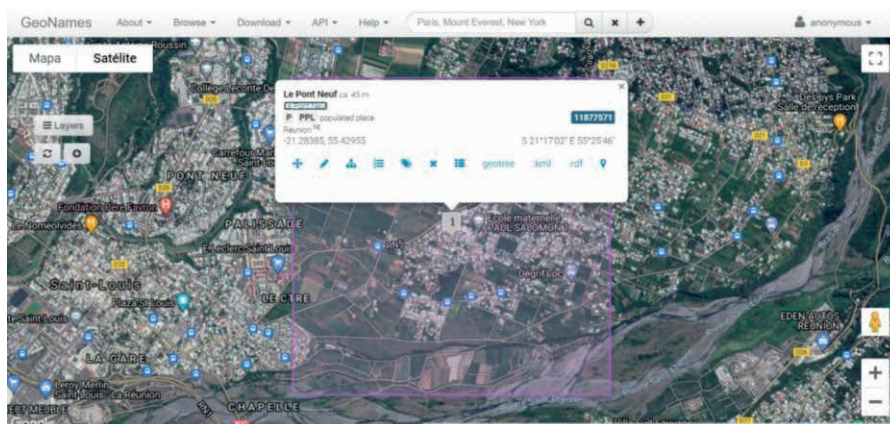


Figura 5. Registro da *Le Pont Neuf*, Paris no GeoNames

Fonte: Captura de tela no GeoNames (2021).

Além disso, pode-se pensar em vídeos do YouTube e outras coisas relacionadas ao trabalho como versões de publicação, exposições, ou seja, há um conjunto maior, é mesmo pensar na informação inserida em um ecossistema que pode ser identificado e navegável na web.

Pensar nas funcionalidades semânticas num catálogo de biblioteca é essencial, desta forma, apresenta-se a seguir, casos em que foram ultrapassados o nível de modelos conceituais e sendo serviços em funcionamento.

CASOS REAIS APLICADOS

Neste ponto relata-se brevemente a experiência que a Europeana - biblioteca digital europeia de acesso livre. Esta biblioteca tem no seu acervo um compêndio de documentos que já são contribuições digitalizadas de outras instituições culturais dos

28 estados que estão das nações europeias (Europeana, 2016), nesta coleção também podemos encontrar livros, pinturas, jornais, mapas, manuscritos, áudios, entre outros grande número de documentos.

Nota-se que a Europeana se preocupa com o patrimônio cultural da Europa e é um portal que iniciou o seu trabalho com um pequeno conjunto de documentos e em pouco tempo conseguiu multiplicar o seu acervo em milhões de documentos, aos quais foram cedidos para mais de duas mil instituições relacionadas com a Ciência da Informação (biblioteconomia, arquivos, galerias e museus, entre outras), isso resulta em um acervo muito variado de documentos em diferentes idiomas.

Tudo isto é o resultado de esforços conjuntos, da aplicação e da experimentação evolutiva constante de novas tecnologias, aplicando-as ao acervo, à preservação - mas principalmente à divulgação de todo o seu material.

O trabalho da Europeana tem sido dificultado pelo grande número de formatos de metadados utilizados para digitalizar os documentos, motivo pelo qual foi necessário promover a adoção de um padrão de normalização, ou seja, um único formato de metadados, embora neste processo ocorra a perda de informação útil. Nota-se também algumas dificuldades quanto a conexão de objetos com a informação contextual. Para obter melhores resultados, a Europeana tenta oferecer uma contextualização semântica mais rica no que diz respeito à representação de objetos de forma a brindar suporte às operações semânticas complexas que não seriam possíveis de realizar na interface tradicional de uma biblioteca digital (Ríos-Hilario; Martín-Campo; Ferreras-Fernández, 2012).

Marcondes (2015) apresentou os resultados de seus estudos sobre a questão da interoperabilidade entre coleções digitais frente as possibilidades apresentadas pelas tecnologias da Web Semântica e dados abertos vinculados. Uma vez que as instituições entendam que unidades de informação podem oferecer seus objetos digitais de suas coleções na *Web* por meio da tecnologia de dados *links* abertos, talvez favoreça a exploração de outros conteúdos relacionados em diferentes ambientes, como turismo, música, notícias, pesquisas, exposições, vídeos, entre outros.

Os metadados dos objetos contidos na Europeana são processados e enriquecidos com novos metadados e outros recursos disponíveis, como tesouros e arquivos de autoridades. Geram um formato interno capaz de representar objetos digitais complexos (isto é, formados por diversos arquivos digitais interligados) baseado no padrão *Object Exchange and Reuse* - (ORE, em português, Troca e Reutilização de Objetos), que, por sua vez, é baseado na tecnologia de dados abertos interligados, RDF. (Marcondes, 2015).

Outro caso estudado foi apresentado por Silveira; Silva; Sousa; Zilio; Cordeiro (2020) e apresenta o resultado da convergência de práticas *Linked Open Data* (LOD) na *Bibliothèque Nationale de France* (Biblioteca Nacional da França), denominado BnF Data.

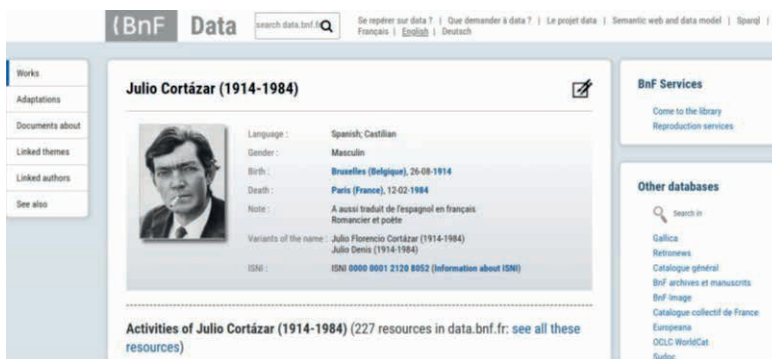


Figura 6. Captura de tela da BnF Data com dados sobre Julio Cortázar

Fonte: BnF Data (2021).

Observa-se na Figura 6 no registro do autor informações relacionadas a biografia do autor argentino Julio Cortázar – considerado um dos mais importantes autores latino-americanos - tais como: idioma, gênero, local e data de nascimento e de falecimento, nota informando que há obras deste romancista e poeta traduzidos do espanhol para o francês.

Ademais, como código identificador único adota-se o ISNI (código internacional padronizado para nomes, do inglês *International Standard Name Identifier*) usado para identificar exclusivamente pessoas e organizações a longo prazo e internacionalmente, definido pela norma ISO 27729: 2012.

Ao clicar no registro que se refere à categoria “author”, o sistema lista os 227 recursos vinculados ao autor, distribuídos em: trabalhos textuais (163), 40 trabalhos mistos (40), obras musicais (12), manuscritos e arquivos (2), programas (4). Expandido o menu disponível encontra-se ainda adaptações das obras (22), além de Documentos sobre o autor (150), Temas relacionados a Julio Cortázar (97) e autores relacionados com Julio Cortázar (174).

No canto inferior direito, há uma lista de outras bases de dados, é dizer que conectam os diversos conteúdos internos e externos, *links* e serviços que a instituição disponibiliza na web dispersos nas diversas aplicações do BnF, como a) Catálogo geral BnF⁷; b) BnF arquivos e manuscritos⁸; e parte dos dados é combinada com vocabulários de valores externos: Gallica⁹; CCFr¹⁰; VIAF e ISNI são utilizados como fontes para padronizar, compartilhar e vincular os dados de autoridades às suas devidas filiações; IdRef¹¹; SUDOC (*Système universitaire de documentation*)¹²; id.loc.gov¹³ para idiomas e nacionalidades,

7. Catálogo Geral da Biblioteca da França, disponível em: <https://catalogue.bnf.fr/index.do>

8. Catálogo de Arquivos e Manuscritos da Biblioteca Nacional da França, disponível em: <https://archivesetmanuscrits.bnf.fr/>

9. Gallica - Biblioteca digital da Biblioteca Nacional da França, disponível em: <https://gallica.bnf.fr/>

10. Centro Nacional de Literatura Infantil, disponível em: <https://cnli.bnf.fr/>

11. IdRef - Registros de autoridade IdRef e referências bibliográficas relacionadas de Sudoc, disponível em: <https://data.idref.fr/>

12. Catálogo coletivo francês do Sistema Universitário de Documentação, disponível em: <http://www.sudoc.abes.fr/cbs/xslt/>

13. Serviço de linked data da Library of Congress - acesso interativo e de máquina a ontologias comumente usadas, vocabulários controlados e outras listas para descrição bibliográfica, disponível em: <https://id.loc.gov/>

dewey.info¹⁴ para assuntos, tipo DCMI para tipos de documentos. Combinam-se também conjuntos de dados identificados por CKAN¹⁵: DBpedia e VIAF. As páginas dos cabeçalhos dos temas RAMEAU são comparadas com outros tesouros, de bibliotecas (*Library of Congress Subject Headings*¹⁶, *Deutsche Nationalbibliothek*¹⁷, BNE) ou mais especializados (Agrovoc¹⁸, Geonames¹⁹, Thesaurus W²⁰).

Dessa maneira, em 2021 a BnF Data apresenta uma cobertura quase completa de registros de catálogos de boa qualidade, alcançando mais de 2 milhões de autores (BIBLIOTECA NACIONAL DA FRANÇA, 2021).

Diante o exposto, verifica-se que começa a ganhar materialidade o conceito do projeto Web Semântica no contexto da Ciência da Informação à medida que uso intensivo das novas tecnologias amplia as possibilidades de classificação dos recursos informacionais disponíveis no ambiente Web. Aproveitando o uso dos instrumentos de representação utilizados tradicionalmente no âmbito da Biblioteconomia e Ciência da Informação (tesouros, lista de cabeçalhos de autores, lista de cabeçalho de assunto, taxonomias), possibilitando o desenvolvimento de ontologias, e desta forma, permitindo a representação formal dos relacionamentos existentes entre os termos e conceitos e rotulando-os a partir de categorias que interpretadas automaticamente pelos computadores (Ramalho, 2006).

CONSIDERAÇÕES FINAIS

O advento das Tecnologias de Informação e Comunicação têm provocado avanços consideráveis mudanças consideráveis em diversos setores da sociedade e práticas de trabalho. No âmbito das tarefas e serviços desempenhados por profissionais da informação a Web Semântica incorpora inovações a partir de modelos conceituais que possibilitam melhorias significativas nos processos de organização e recuperação de informações em ambiente digital.

A geração de dados em ambientes digitais é um fato tanto no campo científico, acadêmico, educacional, artístico e cultural. Portanto a riqueza desses dados e de suas ferramentas permitiria que pesquisadores, estudantes e artistas pudessem apropriar-se, interpretar e compartilhar um maior número de dados, à medida que consigam estabelecer relações significativas e que contribuam para a colaboração nos distintos espaços de organização ou disseminação de conhecimento. Para isso, dados e recursos precisam ser interligados em um contexto digital.

14. Dewey.info é uma iniciativa de disponibilizar o conjunto de dados aberto das Classificações Decimais de Dewey, disponível em: https://old.datahub.io/pt_BR/dataset/dewey_decimal_classification

15. CKAN é um DMS (sistema de gerenciamento de dados) de código aberto para alimentar hubs e portais de dados, disponível em: <https://ckan.org/>

16. Cabeçalhos de assuntos da Biblioteca do Congresso, disponível em: <https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html>

17. Biblioteca Nacional da Alemanha, disponível em: https://www.dnb.de/DE/Home/home_node.html

18. Agrovoc é o tesouro da Organização das Nações Unidas para Alimentação e Agricultura (*Food and Agriculture Organization*) especializado em agricultura, silvicultura, pesca e outros domínios do conhecimento relacionados com alimentação, disponível em: <https://www.fao.org/agrovoc/>

19. O banco de dados geográfico GeoNames cobre todos os países e contém mais de onze milhões de nomes de lugares, disponível em: <http://www.geonames.org/>

20. <https://www.thesaurus.com/>

O presente caracterizou-se como uma discussão de literatura de caráter interdisciplinar e delineamento transversal, buscando identificar os principais conceitos inerentes as contribuições da Web Semântica para a Biblioteconomia e Ciência da Informação afim de clarificar como ocorrem os relacionamentos conceituais a partir da sustentação teórica dos conceitos identificados.

É reconhecido que publicar documentos na Web não é suficiente para torná- los efetivamente descobríveis, especialmente para mecanismos de busca, recuperação e análise. No entanto, com a convergência das práticas, sobretudo com a possibilidade de estruturar os dados abertos de forma vinculada, temos uma realidade possível no ambiente da web a medida em que permitem acesso permanente aos recursos e seus respectivos dados no ambiente digital.

É importante não perder de vista as consequências que a migração de dados acarreta, a necessidade da interoperabilidade dos formatos – sendo necessário, portanto, a adoção de padrões/formatos abertos livre de limitações legais no uso, garantindo o acesso aos dados em longo prazo sem incertezas atuais ou futuras no que diz respeito às direitas legais ou à especificação técnica para que estes processos não se percam informações que podem ser valiosas para catalogação.

REFERÊNCIAS

AMERICAN NATIONAL STANDARDS ORGANIZATION. **Information Retrieval (Z39.50):** Application Service Definition and Protocol Specification. Washington, DC, Estados Unidos: NISO Press, 2003. Disponível em: <https://www.loc.gov/z3950/agency/Z39-50-2003.pdf>

BIBLIOTECA NACIONAL DA FRANÇA. **BnF Data**. 2021. Disponível em: <https://data.bnf.fr/>. Acesso em: 26 out. 2024.

BIBLIOTECA NACIONAL MARIANO MORENO (Argentina). **Catálogo bibliográfico**. 2021. Busca pela obra Rayuela, de Julio Cortázar. Disponível em: https://catalogo.bn.gov.ar/F/?func=direct&doc_number=001327276&local_base=GENER. Acesso em: 09 nov. 2024.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 284, n. 5, p. 34-43, 2001.

BERNERS-LEE, Tim. Linked Data. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 12 jul. 2019.

DUBLIN CORE METADATA INITIATIVE. Association For Information Science and Technology. **DCMI Metadata Terms**. 2020. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. Acesso em: 02 nov. 2024.

EUROPEANA (União Europeia). Connecting Europe Facility da União Europeia. **Europeana**. 2016. Disponível em: <https://www.europeana.eu/pt/about-us>. Acesso em: 06 nov. 2024.

EUROPEAN INTEROPERABILITY FRAMEWORK (União Europeia). **European Interoperability Framework for Pan-european eGovernment Services**. Luxemburgo: Office for Official Publications of The European Communities, 2004. 26 p. Disponível em: https://elfarchive1718.foi.hr/pluginfile.php/55018/mod_resource/content/1/EIF.pdf. Acesso em: 08 nov. 2024.

GEONAMES. 2021. Disponível em: <http://geonames.org>. Acesso em: 08 nov. 2021.

LIBRARY OF CONGRESS (Washington). Library Of Congress. **MARC 21 Format for Bibliographic Data**. 2006. Desenvolvido pelo setor Network Development and MARC Standards Office. Disponível em: <https://www.loc.gov/marc/bibliographic/bdintro.html>. Acesso em: 06 nov. 2024.

MANSO, Miguel A.; WACHOWICZ, Monica; BERNABÉ, Miguel A.; SANCHEZ, Alejandra; RODRIGUEZ, Antonio F. Modelo de Interoperabilidad Basado en Metadatos (MIBM). In: JORNADAS TÉCNICAS DE LA IDE DE ESPAÑA (JIDEE), 5., 2008, **Proceedings [...]**. Terenife: IDE, 2008. Disponível em: https://www.idee.es/resources/presentaciones/JIDEE08/ARTICULOS_JIDEE2008/articulo36.pdf

MARCONDES, Carlos H. O papel dos modelos conceituais para interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus. In: ENCONTRO IBÉRICO EDICIC, 7., 2015, Madri. **Anais [...]**. Madri: Universidad Complutense de Madrid, 2015. p. 1-14. Disponível em: <https://eprints.ucm.es/id/eprint/34550/>. Acesso em: 26 out. 2021.

GUARINO, Nicola. Formal ontology in information systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS (FOIS'98), 1., 1998, **Proceedings [...]**. Trento, Itália: IOS press, 1998, p. 3-15.

RAMALHO, Rogério A. S. **Web Semântica**: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação. 2006. 120 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2006. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/ramalho_ras_me_mar.pdf. Acesso em: 20 out. 2023.

SANTARÉM SEGUNDO, José E.; CONEGLIAN, Caio. S.; LUCAS, Elaine. R. de O. Conceitos e tecnologias da Web semântica no contexto da colaboração acadêmico- científica: um estudo da plataforma Vivo. **Transinformação**, Campinas, v. 29, n. 3, set./dez. 2017, p. 297-309.

SAORÍN, Tomás. Cómo linked open data impactará en las bibliotecas a través de la innovación abierta”. **Anuario ThinkEPI**, 2012, v. 6, p. 288-292.

SOUZA, Renato Rocha; ALVARENGA, Lúcia. A Web Semântica e suas contribuições para a Ciência da Informação. **Ciência da Informação**, v. 33, n. 1, p. 132-141, abr. 2004. <http://dx.doi.org/10.1590/s0100-19652004000100016>.

SILVEIRA, L.; SILVA, F. C. C.; SOUSA, A. T.; ZILIO, S. C.; CORDEIRO, L. S. Convergência de práticas linked open data na Bibliothèque Nationale de France (BNF Data). **Revista ACB: Biblioteconomia em Santa Catarina**, v. 25, n. 1, p. 21-40, 2020. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/138106>. Acesso em: 25 out. 2024.

SMITH, Barry. Beyond concepts: ontology as reality representation. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY AND INFORMATION SYSTEMS, 3., 2004, Turin. **Proceedings [...]**. Turin: IOS Press, 2004. v. 1, p. 73-84. Disponível em: <http://ontology.buffalo.edu/bfo/BeyondConcepts.pdf>. Acesso em: 25 out. 2024.