



Transinformação

ISSN: 0103-3786

ISSN: 2318-0889

Pontifícia Universidade Católica de Campinas

SERRA, Liliana Giusti; SANTARÉM SEGUNDO, José Eduardo
Proposta de modelo para gestão de livros digitais por meio da Web Semântica
Transinformação, vol. 34, e210044, 2022
Pontifícia Universidade Católica de Campinas

DOI: <https://doi.org/10.1590/2318-0889202234e210044>

Disponível em: <https://www.redalyc.org/articulo.oa?id=384376929028>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em redalyc.org

UABM redalyc.org

Sistema de Informação Científica Redalyc
Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal
Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

Proposta de modelo para gestão de livros digitais por meio da Web Semântica

Model proposal for the management of e-books through the Semantic Web

Liliana Giusti SERRA¹  0000-0001-6788-2376

José Eduardo SANTARÉM SEGUNDO^{1,2}  0000-0003-3360-7872

Resumo

Este artigo é resultado de uma pesquisa qualitativa realizada por meio de análise conceitual com vistas a elaborar um modelo para gestão de dados bibliográficos e de licenciamento de livros digitais contratados junto a fornecedores utilizando a tecnologia da Web Semântica. Foram identificadas algumas dificuldades relatadas por bibliotecas na gestão dos livros licenciados, como a qualidade dos metadados, aspectos do desenvolvimento de uma coleção, gestão dos livros licenciados e a geração de indicadores. No aspecto da Web Semântica, foram estudados o *Linked Data*, o *Linked Open Data* e o *Linked Enterprise Data*, visando a troca de dados abertos e fechados. Como resultado da pesquisa foi possível estruturar um modelo onde dados bibliográficos podem ser trocados abertamente entre bibliotecas e fornecedores, e metadados de licenciamento são trocados somente entre as partes envolvidas. Com o modelo, os fornecedores podem disponibilizar *datasets* com dados bibliográficos e de licenciamento, favorecendo a atualização dos pacotes contratados e o processo de licenciamento como um todo. A biblioteca pode contribuir com a qualidade dos registros bibliográficos dos fornecedores, além de reduzir os ajustes nos dados do catálogo e de licenciamento, mantendo as estatísticas de utilização dos livros contratados. A aplicação desse modelo, além de confirmar sua possibilidade de uso, permite maior visibilidade e alcance aos catálogos das bibliotecas a partir do momento em que elas podem estruturar e publicar seus registros de acordo com as boas práticas para publicação de dados na web, posicionando as bibliotecas como provedoras de dados confiáveis e contribuindo com a web dos dados.

Palavras-chave: Livros digitais licenciados. Web Semântica. *Linked data*. *Linked Open Data*. *Linked Enterprise Data*.

Abstract

This paper is the result of qualitative research using the conceptual analysis method to develop a model for managing bibliographic and licensing metadata of digital books contracted through library providers using Semantic Web technology. Libraries identified some difficulties with licensed ebooks such as metadata quality, aspects of collection development, ebooks management, and indicators generation. Linked data, linked open data, and linked enterprise data were the Semantic Web aspects studied, aiming to exchange

¹ Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Programa de Pós-Graduação em Ciência da Informação. Marília, SP, Brasil. Correspondência para/Correspondence to: L. G. SERRA. E-mail: lgiustiserra@gmail.com.

² Universidade de São Paulo, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto. Ribeirão Preto, SP, Brasil.

Artigo elaborado a partir da tese de L. G. Serra, intitulada "A Web Semântica na gestão de livros digitais licenciados: uma proposta de modelo". Universidade Estadual Paulista, Marília, 2019.

Recebido em 8/6/21, versão final: 4/1/22 e aprovado: 20/3/22.

Apoio: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Código de Financiamento 001).

Como citar este artigo/How to cite this article

Serra, L. G.; Santarém Segundo, J. E. Proposta de modelo para gestão de livros digitais por meio da Web Semântica. *Transinformação*, v. 34, e210044, 2022. <https://doi.org/10.1590/2318-0889202234e210044>



open and closed data. As a result, it was possible to structure a model in which bibliographic data can be exchanged openly by libraries and providers, and the licensing metadata is exchanged only between involved parties. With this model, providers can offer datasets with bibliographic and licensing data, favoring the updating of contracted ebook packages. The library can contribute to the quality of bibliographic data of providers, in addition, to reducing adjustments in the catalog and the licensing data, keeping statistics on the use of ebooks. The application of this model, in addition to confirming its usability, allows visibility and reach to libraries' catalogs since they can structure and publish their records in accordance with the best practices for publishing data on the Web, positioning libraries as trusted data providers, contributing to the Web of data.

Keywords: *Licensed digital books. Semantic Web. Linked data. Linked Open Data. Linked Enterprise Data.*

Introdução

Livros digitais fazem cada vez mais parte dos acervos das bibliotecas. Essa demanda existia antes da pandemia de covid-19 e ampliou-se após ela, com atividades acadêmicas e profissionais transferidas em parte ao ambiente digital. As bibliotecas que ainda não trabalhavam com conteúdo digital enxergaram a necessidade de ofertar recursos a seus usuários e, com isso, situações foram observadas e ainda não foram totalmente equalizadas.

Neste artigo são destacadas questões relativas aos livros digitais licenciados nos catálogos das bibliotecas. Os livros licenciados, diferentemente dos livros de acesso aberto, são mediados somente pelas plataformas dos fornecedores. Entretanto, não é interessante aos usuários e aos bibliotecários que os livros digitais licenciados com terceiras partes fiquem segregados do catálogo, obrigando a consulta a cada plataforma contratada para localizar os títulos disponíveis. Conforme pontuam Armstrong e Lonsdale (2011), o *Online Public Access Catalog* (OPAC, catálogo de acesso público em linha) é a forma preferida dos usuários para acessar os livros digitais, e centralizar todos os recursos disponíveis em um único ponto favorece usuários e bibliotecários. Entretanto, ao analisar a seara de licenciamentos realizados por bibliotecas, observam-se os seguintes pontos:

- a) Qualidade dos metadados;
- b) Desenvolvimento da coleção;
- c) Gestão dos livros licenciados e geração de indicadores.

Para tentar solucionar os pontos identificados acima, este artigo analisa o uso da Web Semântica na gestão de livros digitais licenciados, permitindo a publicação, consumo, troca e reuso de metadados entre fornecedores e bibliotecas. Trata-se da proposta de um modelo de dados aberto e genérico, que possa ser utilizado por bibliotecas e fornecedores, independentemente dos sistemas utilizados por eles, e que permita a publicação de um conjunto de dados (*datasets*) na web.

Para realizar esta análise, o trabalho foi estruturado para aprofundar os problemas observados em relação à gestão dos livros digitais por bibliotecas e pela Web Semântica, resultando na elaboração de uma proposta de modelo que permita a troca de dados entre fornecedores e bibliotecas e proporcionando formas de gestão. Do modelo proposto são descritas as etapas de elaboração de política e seleção dos dados, formalização, formatos e licenças, conversão ferramental, processo de recuperação, marketing e *feedback*.

Procedimentos Metodológicos

Esta pesquisa foi desenvolvida com o intuito de analisar a possibilidade de utilização de tecnologias da Web semântica para a construção de um modelo de dados que permita a publicação e reutilização de *datasets* para troca de registros bibliográficos e de licenciamento entre bibliotecas e fornecedores de conteúdo digital.

Foi consultada literatura da área de Ciência da Informação e de Ciência da Computação disponível em livros, artigos, trabalhos apresentados em eventos e materiais disponíveis na web, com um recorte temporal entre 2001

e 2018. O método utilizado foi de análise conceitual, que consiste em levantar perguntas sobre um determinado tema e, a partir dessas perguntas, expandir a compreensão sobre ele, propondo o desenvolvimento de outros conhecimentos (Furner, 2004). A seleção do método se deu em decorrência da ausência de literatura sobre a aplicação da Web Semântica na gestão de livros digitais licenciados. Durante o período de seleção da literatura, foram identificados dois textos que discorriam sobre a possibilidade de uso do *Linked Data* na gestão de livros digitais (MacEwan, 2004; Byrne; Goddard, 2010); porém, são trabalhos publicados há mais de dez anos, com os conhecimentos sobre Web Semântica e *Linked Data* pouco discutidos em bibliotecas.

Para reconhecer as questões que envolvem a gestão de conteúdo digital, selecionou-se textos que discorrem sobre livros digitais licenciados em bibliotecas. Foram utilizados os termos “livro digital”, “ebook”, “e-book”, “livro eletrônico” e as formas desses termos no plural, sendo que os textos selecionados discorrem sobre os desafios da gestão de livros digitais licenciados por bibliotecas: qualidade dos metadados, desenvolvimento de coleção e gestão de licenciamentos e geração de indicadores.

Em relação à Web Semântica, foram selecionados textos que discorrem sobre *Linked Data* (LD), *Linked Open Data* (LOD) e *Linked Enterprise Data* (LED), buscando seus referenciais teóricos. Pesquisou-se, ainda, sobre *datasets* com dados abertos (LOD) e dados fechados (LED) em virtude da possibilidade de troca de dados comerciais entre fornecedores e bibliotecas.

A partir do levantamento dos referenciais teóricos identificados sobre as duas temáticas, partiu-se para a elaboração do modelo para troca de dados entre fornecedores e bibliotecas que também proporcionasse condições para realização de gestão de livros digitais por meio da Web Semântica.

O modelo para publicação dos dados selecionado foi o proposto por Santarém Segundo (2018) em virtude de tratar-se de um modelo neutro, sem detalhamento de área ou projeto específico, e alinhado com as recomendações da W3C sobre as melhores práticas para dados na web (Lóscio; Burle; Calegari, 2017).

Livros digitais licenciados por bibliotecas

As bibliotecas licenciam livros digitais por meio de fornecedores específicos. Invariavelmente, a mediação da leitura será realizada pela plataforma do fornecedor, que é desenvolvida de acordo com as necessidades desse, de forma a assegurar que os livros serão acessados somente por usuários cujas instituições fizeram a contratação.

Os fornecedores específicos para bibliotecas são, basicamente, três: os editores, os distribuidores e os agregadores de conteúdo. Editores e distribuidores são conhecidos pelas bibliotecas e a aquisição de recursos impressos normalmente é feita com eles. Os livros digitais trouxeram um novo *player* ao mercado livreiro, os agregadores de conteúdo, que podem ser compreendidos como um distribuidor, visto que também representam editores, mas possuem uma plataforma de leitura, por onde os livros serão disponibilizados.

Os editores também podem investir no desenvolvimento de uma plataforma própria, porém isso é oneroso. Se o editor possui a sua plataforma, o licenciamento com a biblioteca pode ser realizado de forma direta, sem intermediários. Caso o editor não tenha uma plataforma, a disponibilização de seus livros digitais para bibliotecas é feita por meio de um agregador. Mesmo que um editor tenha investido no desenvolvimento de uma plataforma, seus livros também podem ser representados por agregadores, aumentando as possibilidades de licenciamentos. Já os distribuidores não possuem plataformas, porém comercializam livros digitais de fornecedores que as possuem, sejam eles editores ou agregadores.

Difícilmente um único fornecedor atenderá plenamente uma biblioteca, resultando na contratação de diversos fornecedores. Assim, ao incluir os livros licenciados no catálogo, a biblioteca necessita identificar com quem foi feito o licenciamento e sob quais condições.

A seguir serão discutidas as questões relativas aos livros digitais licenciados nas bibliotecas.

Qualidade dos metadados

A qualidade dos metadados oferecidos por fornecedores às bibliotecas é, em geral, baixa (Rossmann; Foster; Babbitt, 2009; Martin; Mundle, 2010; Mugridge; Edmunds, 2009; Walters, 2013; Sapon-White, 2014; Van Kleeck *et al.*, 2017; Thompson; Traill, 2017). Existe, entretanto, demanda para que os fornecedores ofereçam dados de livros digitais no padrão utilizado pelas bibliotecas (Walters, 2013), afinal é inviável fazer catalogação manual ou mesmo importação dos metadados disponibilizados em outras bases de forma individual, visto que os pacotes assinados podem conter milhares de livros digitais. Os fornecedores não utilizam o formato *Machine-Readable Cataloging* (MARC 21), que é o aplicado ao ambiente das bibliotecas. Para fornecerem os registros no formato MARC 21, os fornecedores precisam recorrer a ferramentas terceiras, desenvolver o MARC 21 em suas próprias ferramentas, ou, ainda, contratar um bibliotecário para fazer a catalogação dos livros digitais.

Ao licenciar um pacote de livros digitais, normalmente o fornecedor encaminha um arquivo com todos os livros licenciados em MARC 21 para inclusão no catálogo. Como a quantidade de títulos costuma ser grande, nem sempre a biblioteca consegue conferir se todos os livros contratados estão no conjunto, podendo acontecer de a quantidade no pacote não coincidir com o número de registros bibliográficos no arquivo importado (Zhang; Jin, 2014).

Desenvolvimento da coleção

O desenvolvimento da coleção é bastante afetado com a inclusão de livros digitais licenciados em decorrência da constante entrada e saída de livros nos pacotes contratados (Frederick, 2015). Durante a vigência do licenciamento, os livros digitais presentes em um catálogo podem ser alterados, independentemente do modelo de negócio que foi aplicado. Essa situação exige um acompanhamento do que foi incluído ou removido dos pacotes e das plataformas e, conseqüentemente, a atualização do catálogo.

Nem sempre o fornecedor informa à biblioteca que ocorreram alterações nos pacotes (Rossmann; Foster; Babbitt, 2009; Wu; Mitchell, 2010; Mugridge; Edmunds, 2009; Hodge; Manoff; Watson, 2013; Walters, 2013; Georgas, 2015; Chen; Kim; Montgomery, 2016). Quando existente, a comunicação das alterações realizadas é feita por envio de mensagem relatando os itens que foram incluídos, alterados ou removidos (Chen; Kim; Montgomery, 2016), não ocorrendo atualização dinâmica entre contratado e contratante, o que demanda ajustes frequentes nos metadados de descrição bibliográfica e de licenciamento. Essa rotina é trabalhosa para fornecedores e bibliotecas.

A atualização dos dados bibliográficos e de licenciamento, além de serem questões relativas ao desenvolvimento da coleção, também são fundamentais para a gestão dos livros digitais contratados, tema que será discutido a seguir.

Gestão de livros digitais licenciados e geração de indicadores

A questão da gestão dos livros digitais licenciados suscita discussões sobre onde e como registrar os dados de licenciamento de forma que seja possível realizar buscas e produzir relatórios e estatísticas. Ao manter no catálogo recursos que não estão disponíveis, a biblioteca pode comprometer sua credibilidade perante a comunidade, além de ter dificuldades para mensurar o real tamanho de sua coleção. Se optar por inibir os livros digitais removidos, seu sistema de automação guardará a informação dos livros digitais que foram contratados e a quantidade de acessos realizados pelo OPAC, com os dados passíveis de utilização em estatísticas e relatórios. Se, por outro lado,

optar por excluir os registros, os dados de acesso são apagados do banco de dados, comprometendo a geração de indicadores e o histórico de utilização da coleção.

A centralização dos livros digitais em um único ponto possibilita uma gestão melhor dos recursos licenciados, identificando livros digitais de fornecedores, modelos de negócios adotados, datas de inclusão na coleção *etc.* Essa centralização contribui para a mensuração do uso que é feito do conteúdo a partir de um único ponto, independentemente do fornecedor.

A gestão de livros digitais é diferente da dos livros impressos. Enquanto os livros com volume físico ficam em estantes, os livros digitais estão em nuvem; em plataformas de terceiras partes. A relação entre fornecedores e biblioteca termina quando os livros impressos são entregues. Com os livros digitais, a dependência da plataforma é constante (Rossmann; Foster; Babbitt, 2009). Outro fator de destaque é que os livros são contratados junto a fornecedores diferentes. Isso exige distinguir quais livros foram licenciados com qual fornecedor, por qual modelo e com qual vigência, tendo, com isso, dados para avaliar se os livros estão sendo utilizados e se justificam a continuidade da contratação.

Os dados de licenciamento devem ser registrados em banco de dados, mas não no formato MARC 21. Dados como fornecedor, modelo de negócio aplicado, valor, período de validade, quantidade de acessos simultâneos contratados *etc.*, podem ser controlados e passíveis de atualizações, preservando dados de transações anteriores e permitindo o manejo de conjuntos de dados de forma global.

Web Semântica, *Linked Data*, *Linked Open Data* e *Linked Enterprise Data*

A web nasceu a partir da demanda de troca de informações entre cientistas em universidades e instituições, por iniciativas de Berners-Lee. Porém a web não poderia ser somente um local para publicar informações utilizando *hiperlinks* para criar um elo entre elas. Após quase dez anos da publicação do artigo que anunciou o desenvolvimento da web, Berners-Lee, Hendler e Lassila publicaram em 2001 o conceito da Web Semântica, que “[...] trará estrutura ao significado do conteúdo de sítios da Web, criando um ambiente onde agentes de softwares podem perambular de um sítio a outro, prontos para executarem tarefas para os usuários” (Berners-Lee; Hendler; Lassila, 2001, p. 1, tradução nossa).

O funcionamento da Web Semântica depende do uso de *datasets* estruturados e de um conjunto de regras de inferência capazes de compreender uma requisição. O desafio consiste em utilizar uma linguagem que expresse os dados e as regras para que a informação seja compreendida. Não basta, no entanto, formalizar os dados para que eles sejam legíveis por máquinas; é necessário que também sejam compreendidos por pessoas, apresentando as relações estabelecidas.

A ideia principal da Web semântica é suportar a distribuição da web em nível de dados ao invés do nível de apresentação. Ao invés de ter uma página da web apontando para outra, um dado pode apontar para outro, usando uma referência global chamada *Uniform Resource Identifiers* (URIs). A infraestrutura da web promove um modelo de dados onde a informação de uma única entidade pode ser distribuída na web (Allemang; Hendler, 2011, p. 6, tradução nossa).

Esta pesquisa selecionou o *Linked Data* (LD) como aplicação da Web Semântica para a elaboração do modelo de dados, onde o *Linked Open Data* (LOD) permite a troca de dados bibliográficos enquanto o *Linked Enterprise Data* (LED) é usado para dados dos licenciamentos.

Linked Data (LD)

De acordo com Bizer, Heath e Berners-Lee (2011), o *Linked Data* consiste em criar *links* entre dados provenientes de diferentes fontes na web, porém é diferente dos bancos de dados de organizações ou demais

sistemas. Para que ocorra a interligação entre os dados é necessário que eles sejam publicados de forma que sejam legíveis por máquinas, com os significados e as relações entre eles definidas explicitamente.

Ao conectar uma informação de um sítio a um conjunto de dados, uma atualização feita nesse conjunto acarretará em automática atualização do sítio, abrindo possibilidades de geração de novos dados a partir da integração com fontes diferentes. Assim, uma aplicação pode coletar dados de um único ou de diversos *datasets*, de acordo com os propósitos que se deseja alcançar. A comunicação entre máquinas ocorre com a adoção de padrões, modelos e de boas práticas que permitam a interconexão entre informações distribuídas na web, proporcionando que, além da integração entre fontes, ocorra o reuso dos dados. O LD representa o uso de boas práticas para publicação de dados na web, de forma organizada, utilizando padrões internacionais recomendados pelo *World Wide Web Consortium* (W3C) (Lóscio; Burle; Calejari, 2017). Essas boas práticas foram propostas para serem usadas como orientações para que sejam ampliadas as possibilidades de troca de dados na web, abordando aspectos como licenças, procedência, qualidade, versionamento, formatos, metadados, identificadores, vocabulários, formas de acesso, preservação, enriquecimento, atualização e republicação de dados. Foram definidas 35 boas práticas com o objetivo de alcançar os seguintes benefícios: reuso (*reuse*), compreensão (*comprehension*), vinculação (*linkability*), descoberta (*discoverability*), confiança (*trust*), acesso (*access*), interoperabilidade (*interoperability*) e processamento (*processability*).

Linked Open Data (LOD)

O *Linked Open Data* é resultado de um grupo de trabalho do W3C que teve como objetivo a criação de uma rede de dados interligados que podem ser utilizados livremente. Segundo Berners-Lee (2010), o LOD foi desenvolvido sob uma licença aberta exatamente para a reutilização livre dos dados. Nem todos os *datasets* precisam ser abertos, visto que existem informações que circulam somente entre um grupo, porém, para ser considerado LOD, os dados devem ser publicados para uso aberto, seguindo as boas práticas definidas pelo W3C.

São considerados dados abertos aqueles que correspondem à definição de dados bibliográficos abertos (*open bibliographic data*), que estabelece que os dados podem ser utilizados, reutilizados e redistribuídos livremente, com as mesmas atribuições de compartilhamento com as quais foram publicados. Para Wessels *et al.* (2017), os dados podem ser coletados de diversas fontes, como a web, mídias sociais, aplicativos para dispositivos móveis (APP), dados geoespaciais ou bancos de dados comerciais.

Os dados podem ser coletados separadamente e permitem a criação de inferências, ditando tendências, comportamentos e movimentos a partir do cruzamento de informações advindas de fontes distintas e que sozinhas poderiam não proporcionar o entendimento de um contexto. Para Berners-Lee (2010), o LOD é o *Linked Data* que permite o uso dos dados de forma livre, gratuita e sem restrições para reutilização.

Linked Enterprise Data (LED)

Para Allemang (2010) a adoção de aplicações da Web Semântica em ambientes empresariais é identificada como *Linked Enterprise Data*. As razões para manter dados fechados (*proprietary data*) podem ser diversas e normalmente são utilizadas por empresas privadas. As informações consumidas por esse público usualmente são relacionadas com segredos de negócios e estratégias da empresa. Portanto, é compreensível que os dados sejam de uso restrito aos participantes do ecossistema da empresa, com aplicação de níveis de acesso de acordo com a hierarquia e cultura institucional.

Segundo Hillerbrand (2016), as aplicações da Web Semântica podem ser utilizadas em empresas para integrar informações de fontes diversas diminuindo redundâncias, resolver ambiguidades em terminologias,

aumentar a recuperação de informações, identificar aspectos relevantes dentro de um domínio *etc.*, além de prover instrumentos para tomadas de decisão.

Segundo Villazon-Terrazas *et al.* (2017), as empresas podem construir ontologias para suportar as demandas de seus negócios, utilizando vocabulários existentes ou expandindo-os de acordo com necessidades e evoluções da instituição. Ao adotar um mesmo esquema para representar seus dados, é favorecida a integração entre outros sistemas, além da adoção de um único vocabulário, contribuindo com a comunicação entre os envolvidos. Essa característica faz com que os esquemas utilizados na web sejam mais genéricos, enquanto os utilizados pelas empresas são mais centrados no domínio específico.

Usualmente as empresas manterão seus *datasets* fechados ou com acesso restrito, enquanto a web luta para cada vez mais disponibilizar dados abertos. Normalmente os *datasets* fechados são construídos com foco em um domínio específico, contudo não existe impedimento para que empresas também utilizem *datasets* abertos.

Proposta de modelo de dados

A Web Semântica permite diversas aplicações no ambiente das bibliotecas. Seu uso pode proporcionar o enriquecimento do catálogo de forma dinâmica, vinculando dados a informações complementares dispostas em *datasets* na web. Também pode contribuir com ações para agilizar as atividades de catalogação e indexação a partir do momento em que fontes externas podem ser consumidas para essas atividades. Para Byrne e Goddard (2010), as dificuldades para implantação do LD pelas bibliotecas não são tecnológicas, mas de falta de consciência. Para as autoras, o perigo consiste em bibliotecários enxergarem o LD como mais um padrão de metadados, sem vislumbrarem as possibilidades de expansão e visibilidade dos catálogos. Para Shieh e Reese (2015), a inclusão de elementos da Web Semântica, como o uso de URLs, amplia o alcance das coleções das bibliotecas.

Segundo Byrne e Goddard (2010, não paginado, tradução nossa), “[...] com a experiência em pesquisa, geração de metadados e desenvolvimento de ontologias, os bibliotecários devem estar realmente bem-posicionados para entender e implementar o *linked data*”. Bibliotecas precisam se reconhecer como potenciais provedoras de dados confiáveis. Para tanto, precisam adequar seus metadados ao ambiente da web, permitindo interoperabilidade com outros sistemas e conjuntos de dados.

Como o formato MARC 21 não é utilizado na web dos dados, é necessária a conversão de registros para o modelo *Resource Description Framework* (RDF). Para Chen (2017), é necessária a conversão de dados do MARC 21 em banco de dados construídos com URLs.

A partir das informações levantadas, foi possível propor um modelo de dados para troca de registros entre bibliotecas e fornecedores de conteúdo digital. Cada fornecedor terá o seu *dataset*, mantendo, porém, o sigilo das transações comerciais realizadas com as bibliotecas. Por outro lado, as bibliotecas podem transformar seus catálogos bibliográfico e de autoridades em *datasets*, proporcionando uma fonte confiável para padronização descritiva para os fornecedores. Os *datasets* dos fornecedores podem ser atualizados e padronizados a partir dos dados existentes nos catálogos das bibliotecas, contribuindo com sua qualidade. A biblioteca ajuda o fornecedor a manter seus metadados padronizados, não comprometendo a qualidade de seu catálogo quando registros novos forem acrescentados. Assim, ao contratar um fornecedor, a biblioteca conecta-se ao *dataset* dele para receber os metadados que já foram limpos e padronizados pelo fornecedor a partir de *datasets* disponíveis na web, inclusive os próprios catálogos das bibliotecas. Com isso ocorre a troca e reuso de dados entre bibliotecas e fornecedores.

Para esta pesquisa foram adotadas as etapas de publicação de dados proposta por Santarém Segundo (2018), conforme Figura 1.

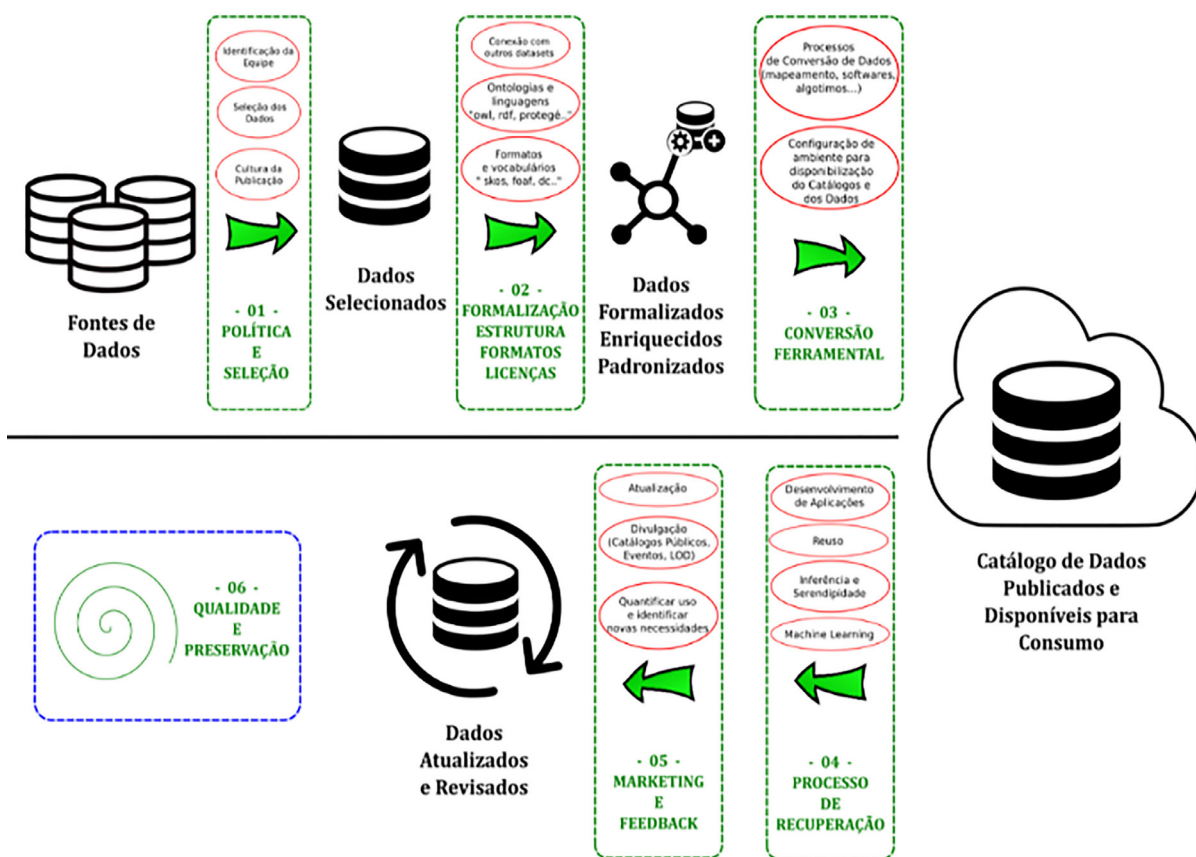


Figura 1 - Fluxo organizacional para publicação de dados.

Fonte: Santarém Segundo (2018, p. 125).

Para as etapas de publicação dos dados foi considerada aderência às 35 boas práticas para a publicação de dados abertos apresentadas por Lóscio, Burle e Calegari (2017).

A Figura 2 apresenta o modelo, destacando cada etapa do processo, iniciando pelos dados originais dos fornecedores, passando pela limpeza, conversão, enriquecimento, publicação dos *datasets* com dados em LOD e LED, unificação dos dados em um único ponto vinculado com o catálogo da biblioteca e para envio de notificações.

As etapas sinalizadas no modelo são:

1) Os dados de origem dos fornecedores podem ser obtidos de duas formas: formato *Online Information Exchange* (ONIX) ou planilhas de dados. Os fornecedores podem continuar a utilizar seus sistemas, exportando os dados para que sejam incluídos no *dataset*, podendo ser no formato de planilhas de dados ou pelo ONIX, utilizado comumente por fornecedores.

2) Os dados seleccionados dos fornecedores são tratados e normalizados;

3) Os dados tratados são convertidos ao modelo RDF;

4) Os dados são enriquecidos por *dataset(s)*;

5) Após o tratamento e enriquecimento dos dados, é possível publicá-los como *datasets*;

6) O LOD é empregado para os dados bibliográficos, que não são sigilosos;

7) Os dados de licenciamento entre fornecedores e bibliotecas são de acesso restrito aos envolvidos. O LED é utilizado para os dados de licenciamento;

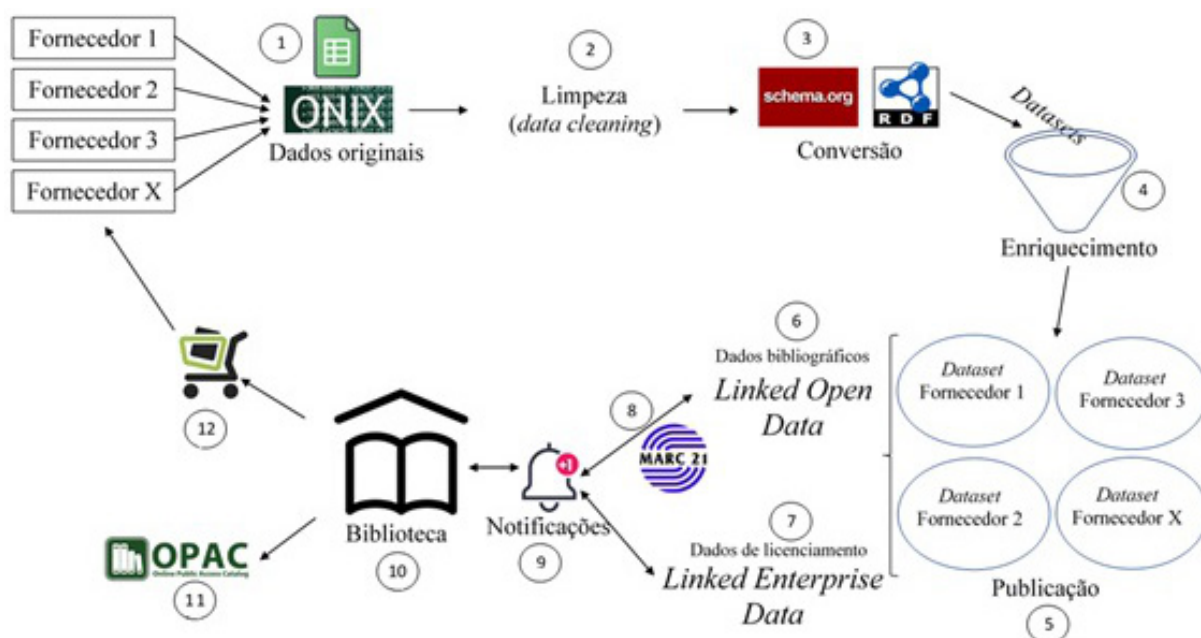


Figura 2 - Modelo para gestão de livros digitais licenciados.

Fonte: Elaborado pelos autores (2021).

8) Como bibliotecas utilizam o formato MARC 21 para descrição de registros bibliográficos e os fornecedores também precisam desse formato para atenderem as demandas de seus clientes, a disponibilização dos metadados é oferecida também no MARC 21;

9) Fornecedores e bibliotecas podem trocar notificações, sinalizando alterações na disponibilidade de livros digitais nas plataformas, mudanças nas formas de licenciamento, ou ainda alterações realizadas nas descrições;

10) Os dados que transitam entre fornecedores e bibliotecas são relacionados a três conjuntos: bibliográficos, de fornecedores e de licenciamentos. Fornecedores e bibliotecas trocam dados bibliográficos, de licenciamentos e notificações. A biblioteca centraliza os metadados de livros digitais dos fornecedores que foram selecionados;

11) A biblioteca disponibiliza em seu catálogo os livros digitais que foram licenciados, com as atualizações que ocorrerem durante a vigência do licenciamento;

12) A biblioteca pode enviar pedidos de novos licenciamentos ou troca de existentes diretamente ao(s) fornecedor(es), contribuindo para a agilidade das negociações.

Para proposta deste modelo, optou-se por não desenvolver uma ontologia específica, mas aproveitar vocabulários existentes, com vistas ao reaproveitamento e construção de um modelo de domínio genérico. Para este modelo são propostos os vocabulários *DC Terms* (mantido pelo *Dublin Core Metadata Initiative* (DCMI)), o *Friend of a Friend* (FOAF), códigos da lista de relações do MARC (*MARC code list for relators schema*), o *Simple Knowledge Organization System* (SKOS) e o *Schema*. O modelo proposto possui compatibilidade com os formatos ONIX e MARC 21, podendo receber e processar registros nesses dois formatos, além de dados constantes de planilhas.

Considerações Finais

Esta pesquisa propôs a elaboração de um modelo de dados para contribuir com a gestão de livros licenciados por bibliotecas, proporcionando outros benefícios além dos identificados. A proposta inicial era

usar a Web Semântica, especificamente o LOD e o LED, para propiciar a troca de metadados bibliográficos e de licenciamento entre bibliotecas e fornecedores de conteúdo digital. Essa abordagem foi feita visando sanar dificuldades observadas nas bibliotecas em relação à qualidade dos metadados oferecidos pelos fornecedores, contribuir com o desenvolvimento da coleção e fornecer indicadores para realização da gestão.

Entretanto, ao analisar as possibilidades de aplicação da Web Semântica em bibliotecas, constatou-se que, além de sanar as questões que eram o foco deste estudo, abriram-se outras oportunidades. A Web Semântica favorece a expansão e visibilidade dos catálogos, movimento que pode reconfigurar o posicionamento das bibliotecas como provedoras de dados confiáveis, contribuindo com a web dos dados.

O modelo proposto permite, além da gestão de livros digitais licenciados junto a fornecedores de conteúdo, a estruturação de um catálogo alinhado com as boas práticas para publicação de dados na web, descritas pelo W3C. Se a motivação para a elaboração de um modelo surgiu pelo desejo de troca de dados entre bibliotecas e fornecedores, vislumbram-se agora outras possibilidades - como a publicação do catálogo da biblioteca em formato semântico.

Num primeiro momento, cogitou-se o uso de tecnologias da Web Semântica para troca e reutilização de dados bibliográficos abertos. Isso por si só contribuiria com questões relativas a metadados oferecidos pelos fornecedores. Entretanto, dados de licenciamento não seriam abertos. O uso do LED mostra-se como uma opção para troca de dados fechados entre os envolvidos comerciais. Não foram identificados relatos de uso do LED em bibliotecas.

O modelo foi proposto para ser aberto e passível de utilização por fornecedores de conteúdo digital e bibliotecas, sem forçar a adoção de padrões ou formatos proprietários ou específicos para os segmentos analisados, como são hoje no caso do ONIX e do MARC 21 para editores e bibliotecas, respectivamente. Ao contrário, a ideia é exatamente garantir que formatos e padrões específicos possam ser utilizados e que sejam intercambiáveis. Na ausência de um formato específico, planilhas genéricas podem ser utilizadas para troca de dados.

Embora não tenha sido aprofundado nesta pesquisa, é possível o enriquecimento de dados abertos, destacando metadados bibliográficos como nomes de autoridades, editoras, locais, idiomas etc. O enriquecimento de dados bibliográficos pode ser feito por meio de vocabulários específicos, como o *Virtual International Authority File* (VIAF), *Library of Congress Subject Headings* (LCSH), ou mesmo a inclusão de identificadores de autoridades, como o *Open Research and Contributor Identifier* (ORCID) ou o Currículo Lattes, onde é possível enriquecer os dados de pessoas vinculadas aos registros a partir de fontes externas ao catálogo.

O ineditismo desta pesquisa centra-se no uso de tecnologias da Web Semântica, abrangendo dados abertos (LOD) e fechados (LED), visando a reutilização de dados e condições para solucionar dificuldades de gestão identificadas em bibliotecas. Não foram identificados, na literatura consultada, outros modelos para publicação de dados de bibliotecas; portanto, o modelo proposto nesta pesquisa é passível de aplicação em âmbito nacional e internacional.

Como foi demonstrado nesta pesquisa, a Web Semântica pode ser aplicada em bibliotecas em diversas esferas. A proposta deste texto debruçou-se sobre a gestão de livros digitais licenciados e o intercâmbio de dados entre bibliotecas e fornecedores, mas outras perspectivas são possíveis, fazendo da biblioteca um ator de destaque na web dos dados.

Referências

Allemang, D. Semantic Web and the linked data enterprise. In: Wood, D. (ed.). *Linking enterprise data*. New York: Springer, 2010. p. 3-23.

Allemang, D.; Hendler, J. *Semantic Web for the working ontologist: effective modeling in RDFS and OWL*. 2nd. ed. London: Elsevier, 2011. Disponível em: <https://nanopdf.com/>

- [download/1-what-is-the-semantic-web-chapter_.pdf](#). Acesso em: 29 dez. 2018.
- Armstrong, C.; Lonsdale, R. Introduction. In: Price, K.; Havergal, V. (ed.). *E-books in libraries: a practical guide*. London: Facet, 2011. p. xxi-xl.
- Berners-Lee, T. Is your Linked Open Data 5 star? In: Berners-Lee, T. *Linked Data*. Cambridge: W3C, 2010. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em 14 out. 2018.
- Berners-Lee, T.; Hendler, J.; Lassila, O. The semantic web. *Scientific American*, p. 24-30, May 2001. Disponível em: <https://www.scientificamerican.com/article/the-semantic-web/>. Acesso em: 27 jul. 2016.
- Bizer, C.; Heath, T.; Berners-Lee, T. Linked data: the story so far. *International Journal on Semantic Web and Information Systems*, v. 5, n. 3, p. 1-22, 2011. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 17 out. 2018.
- Byrne, G.; Goddard, L. The strongest link: libraries and linked data. *D-Lib Magazine*, v. 16, n. 11/12, 2010. p. 1-9, 2010. Disponível em: <http://www.dlib.org/dlib/november10/byrne/11byrne.html>. Acesso em: 19 nov. 2018.
- Chen, M.; Kim, M.; Montgomery, D. Ebook record management at The University of Texas at Dallas. *Technical Services Quarterly*, v. 33, n. 3, p. 251-267, 2016. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/07317131.2016.1169781?journalCode=wtsq20>. Acesso em: 30 jan. 2019.
- Chen, Y.-N. A review of practices for transforming library legacy records into linked open data. In: Garoufallou, E. et al. (ed.). *Metadata and semantic research*. Cham: Springer, 2017. p. 123 -133. (Communications in Computer and Information Science, 755).
- Frederick, D. E. *Managing ebook metadata in academic libraries: taming the tiger*. Waltham: Chandos Publishing, 2015. (Chandos Information Professional Series). Disponível em: <http://www.library.illinois.edu/proxy/go.php?url=http://www.sciencedirect.com/science/book/9780081001516>. Acesso em: 15 out. 2018.
- Furner, J. Conceptual analysis: a method for understanding information as evidence, and evidence as information. *Archival Science*, v. 4, p. 233-265, 2004. Disponível em: <https://link.springer.com/doi/10.1007/s10502-005-2594-8>. Acesso em: 27 set. 2018.
- Georgas, H. The case of the disappearing e-book: academic libraries and subscription packages. *College & Research Libraries*, v. 77, n. 11, p. 883-898, 2015. Disponível em: <https://crl.acrl.org/index.php/crl/article/view/16475>. Acesso em: 25 nov. 2018.
- Hillerbrand, E. Semantic web and business: reaching a tipping point? In: Workman, M. (ed.). *Semantic Web*. Wilmette, IL: Springer, 2016. p. 213-229.
- Hodge, V.; Manoff, M.; Watson, G. Providing access to e-books and e-book collections: struggles and solutions. *The Serials Librarian*, v. 64, n. 1-4, p. 200-205, 2013. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/0361526X.2013.760411>. Acesso em: 11 nov. 2018.
- Lóscio, B. F. F.; Burle, C.; Calegari, N. (ed.). Data on the Web best practices: W3C recommendation 31 jan. 2017. W3C, 2017. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 12 dez. 2018.
- MacEwan, A. Project InterParty: from library authority files to e-commerce. *Cataloging & Classification Quarterly*, v. 39, n. 1-2, p. 429-442, 2004.
- Martin, K. E.; Mundle, K. Notes on operations cataloging e-books and vendor records: a case study at the University of Illinois at Chicago. *Library Resources & Technical Services*, v. 54, n. 4, 227-237, 2010. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/4976/6015>. Acesso em: 12 dez. 2018.
- Mugridge, R. L.; Edmunds, J. Using batchloading to improve access to electronic and microform collections. *Library Resources & Technical Services*, v. 53, n. 1, p. 53-61, 2009. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5461/6700>. Acesso em: 13 dez. 2018.
- Rossmann, D.; Foster, A.; Babbitt, E. P. Ebook MARC records: do they make the mark? *Serials*, v. 22, n. 3, p. 546-550, 2009. Disponível em: <https://scholarworks.montana.edu/xmlui/bitstream/handle/1/323/ebookMARC.pdf>. Acesso em: 12 dez. 2018.
- Santarém Segundo, J. E. Web Semântica: fluxo para publicação de dados abertos e ligados. *Informação em Pauta*, Fortaleza, v. 3, número especial, 117-140, 2018. Disponível em: <https://doi.org/10.32810/2525-3468.ip.v3iEspecial.2018.39721.117-140>. Acesso em: 12 dez. 2018.
- Sapon-White, R. E-book cataloging workflow at Oregon State University. *Library Resources & Technical Services*, v. 58, n. 2, 127-136, 2014. Disponível em: <https://journals.ala.org/lrts/article/view/5335/6515>. Acesso em: 12 dez. 2018.
- Shieh, J.; Reese, T. The importance of identifiers in the new Web environment and using the Uniform Resource Identifier (URI) in subfield zero (\$0): a small step that is actually a big step. *Journal of Library Metadata*, v. 15, n. 3-4, p. 208-226, 2015. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/019386389.2015.1099981?journalCode=wjlm20>. Acesso em: 30 out. 2018.
- Thompson, K.; Traill, S. Leveraging Python to improve ebook metadata selection, ingest, and management. *Code4lib Journal*, n. 38, out. 2017. Disponível em: <https://journal.code4lib.org/articles/12828>. Acesso em: 12 dez. 2018.
- Van Kleeck, D. et al. Managing bibliographic data quality for electronic resources. *Cataloging & Classification Quarterly*, v. 55, n. 7-8, p. 560-577, 2017. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01639374.2017.1350777?journalCode=wccq20>. Acesso em: 28 out. 2018.
- Villazon-Terrazas, B. et al. Construction of enterprise knowledge graphs I. In: Pan, J. Z. et al. (ed.). *Exploiting linked data and knowledge graphs in large organizations*. Cham: Springer, 2017. p. 87-116. Disponível em: <http://dx.doi.org/10.1007/978-3-319-45654-6>. Acesso em: 29 out. 2018.
- Walters, W. H. E-books in academic libraries: challenges for discovery and access. *Serials Review*, v. 39, n. 2, p. 97-104, 2013.

Disponível em: <https://doi.org/10.1080/00987913.2013.10765501>. Acesso em: 12 dez. 2018.

Wessels, B. *et al. Open data and the knowledge society*. Amsterdam: Amsterdam University Press, 2017.

Wu, A.; Mitchell, A. M. Mass management of e-book catalog records. *Library Resources & Technical Services*, v. 54, n. 3, 164-174,

2010. Disponível em: <https://journals.ala.org/lrts/article/view/5558/6841>. Acesso em: 12 dez. 2018.

Zhang, L.; Jin, M. Cataloging e-books: dealing with vendors and various other problems. *The Serials Librarian*, v. 67, n. 1, p. 76-80, 2014. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/0361526X.2014.899295>. Acesso em: 12 dez. 2018.