26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
ISSN 2177-3688

GT 8 – Dados, Informação e Tecnologia

ARQUITETURA DE UM PORTAL SEMÂNTICO PARA PUBLICAÇÃO DE DADOS BIBLIOGRÁFICOS CONECTADOS: PROPOSTA PARA A REDE BIBLIODATA

A SEMANTIC PORTAL ARCHITECTURE FOR PUBLISHING LINKED BIBLIOGRAPHIC DATA: A PROPOSAL FOR THE BIBLIODATA NETWORK

Ana Carolina Simionato Arakaki – Universidade de Brasília (UnB)
Bruno Carlos da Cunha Costa – Instituto Federal do Rio de Janeiro (IFRJ)
Henrique Leal Tavares – Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp) Campus de Marília
Dayane Onaga Ferreira Machado – Universidade Federal de São Carlos (UFSCar)
Ana Carolina Novaes de Mendonça – Universidade Federal de São Carlos (UFSCar)

Modalidade: Trabalho completo

Resumo: A Rede Bibliodata é uma colaboração interinstitucional brasileira que, desde 1942, atua na geração e manutenção cooperativa de registros bibliográficos e de autoridade. Apesar da reestruturação tecnológica iniciada em 2019, a publicação e o acesso aos seus dados em formato aberto e conectado permanecem como desafio. Este artigo propõe uma arquitetura para um portal semântico da Rede Bibliodata, fundamentada em *Linked Open Data* (LOD) e tecnologias semânticas (RDF, SPARQL e OWL). A pesquisa é exploratória, baseada em revisão integrativa da literatura e análise de boas práticas internacionais. A arquitetura organiza um ecossistema de dados bibliográficos e de autoridade em oito camadas, da extração à publicação, visualização e integração. Um protótipo permitiu examinar preliminarmente a articulação entre transformação, armazenamento, publicação e consulta. Os resultados indicam a viabilidade da aplicação de LOD em redes cooperativas e evidenciam desafios de modelagem, governança, qualidade e manutenção dos dados.
Palavras-chave: dados bibliográficos conectados; linked open data; rede bibliodata; portal semântico; *web* semântica.

Abstract: The Bibliodata Network is a Brazilian inter-institutional collaboration that, since 1942, has supported the cooperative creation and maintenance of bibliographic and authority records. Despite the technological restructuring initiated in 2019, publishing and accessing its data in an open and linked format remains a challenge. This article proposes an architecture for a semantic portal for the Bibliodata Network, based on Linked Open Data (LOD) principles and semantic technologies (RDF, SPARQL, and OWL). The exploratory study is based on an integrative literature review and analysis of international best practices. The architecture organizes an ecosystem of bibliographic and authority data into eight layers, from extraction to publication, visualization, and integration. A prototype enabled a preliminary examination of the articulation among transformation, storage, publication, and query processes. The results indicate the feasibility of applying LOD to cooperative library networks while highlighting challenges involving modeling, governance, data quality, and maintenance.
Keywords: linked bibliographic data; linked open data; bibliodata network; semantic portal; semantic *web*.

1 INTRODUÇÃO

A Rede Bibliodata promove a catalogação cooperativa de registros bibliográficos e de autoridade no Brasil. Criada em 1942 como Serviço de Intercâmbio de Catalogação (SIC) do Departamento Administrativo do Serviço Público (DASP), sua trajetória acompanhou a evolução das tecnologias aplicadas ao domínio bibliográfico, incluindo CALCO, UNIMARC e a reestruturação conduzida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Desde 2019, a Rede passa por reestruturação conceitual e tecnológica, mas a publicação e o acesso aos dados em formato aberto e conectado ainda constituem um desafio.

A demanda por dados bibliográficos abertos e conectados visa ampliar o acesso, o uso e o reúso da informação (Isotani; Bittencourt, 2015; Jesus; Castro; Ramalho, 2021). A *Web* Semântica e o *Linked Open Data* (LOD) estabelecem fundamentos para essa transformação (Berners-Lee, 2006; Berners-Lee; Hendler; Lassila, 2001), já observada em iniciativas internacionais de publicação de dados bibliográficos (Silveira *et al.*, 2020).

Contudo, no Brasil, esta lacuna representa tanto um desafio quanto uma oportunidade para a Rede Bibliodata. Diante desse contexto, questiona-se: como pode ser estruturada uma arquitetura de portal semântico para a publicação dos dados bibliográficos e de autoridade da Rede Bibliodata como *Linked Open Data*?

Assim, o artigo objetiva propor uma arquitetura para um portal semântico da Rede Bibliodata, fundamentada em *Linked Open Data* e tecnologias semânticas. Para isso, sintetiza a literatura sobre dados bibliográficos conectados, estrutura uma arquitetura em camadas e examina sua implementação em um protótipo orientado à publicação conectada e interoperável dos dados do catálogo.

2 REFERENCIAL TEÓRICO

A *Web* Semântica foi proposta como uma extensão da *Web* voltada à descrição de recursos, conceitos e relações de modo processável por máquinas, permitindo integrar informações provenientes de diferentes sistemas e contextos (Berners-Lee; Hendler; Lassila, 2001). Na Ciência da Informação, essa perspectiva relaciona-se à organização, representação, recuperação e reutilização de dados em ambientes distribuídos e abrange ontologias, representação do conhecimento, mecanismos de consulta e integração. Também envolve processos de limpeza, conexão, publicação e interação com dados vinculados a instituições de memória (Arakaki; Simionato; Santos, 2017; Allemang; Hendler; Gandon, 2020; Cervone; Svensson, 2015; Hitzler; Krötzsch; Rudolph, 2009; Hogan *et al.,* 2021; Van Hooland; Verborgh, 2014).

O *Linked Data* constitui uma abordagem de publicação de dados na *Web* baseada em quatro diretrizes: uso de URIs para identificar recursos; URIs HTTP para recuperá-los; disponibilização por padrões abertos, como *Resource Description Framework* (RDF) e SPARQL; e estabelecimento de ligações entre conjuntos de dados (Berners-Lee, 2006; Bizer; Heath; Berners-Lee, 2009). Associado a condições técnicas e jurídicas de abertura, configura o *Linked Open Data* (LOD), com dados reutilizáveis e licenças compatíveis com acesso, redistribuição e reúso (Heath; Bizer, 2011; Isotani; Bittencourt, 2015).

No domínio bibliográfico, o *Linked Data* tem sido aplicado à organização, representação, integração, enriquecimento e publicação de dados, com discussões sobre requisitos, interoperabilidade e conexão dos catálogos à *Web* (Arakaki, 2016; Alemu, 2022; Gaitanou *et al*., 2024; Jesus; Castro; Ramalho, 2021; Jesus, 2018; Rautenberg, 2019; Serra; Santarém Segundo, 2021).

A infraestrutura de dados conectados combina tecnologias com funções distintas. O RDF representa dados em triplas formadas por sujeito, predicado e objeto; RDF *Schema* (RDFS) e *Web Ontology Language* (OWL) definem classes, propriedades, hierarquias e restrições; SPARQL consulta grafos RDF; e SKOS representa vocabulários controlados, tesauros e esquemas de classificação (Miles; Bechhofer, 2009; *World Wide Web Consortium*, 2014). Esses padrões sustentam a descrição, consulta e integração dos dados bibliográficos e permitem explicitar relações que, em formatos tradicionais, permanecem implícitas ou restritas ao contexto do registro.

A publicação de dados na *Web* ultrapassa a conversão para formatos abertos e envolve modelagem, identificadores persistentes, documentação, licenças, proveniência, atualização, mecanismos de acesso e ligação com outros conjuntos. As boas práticas indicam dimensões como compreensão, processabilidade, descoberta, reúso, confiabilidade, acesso e interoperabilidade (Lóscio; Burle; Calegari, 2017), especialmente relevantes em redes cooperativas, nas quais os dados circulam e são reutilizados por diferentes instituições.

No domínio bibliográfico, a adoção de *Linked Data* implica rever modelos centrados em registros. O MARC 21, concebido para intercâmbio bibliográfico, apresenta limitações em arquiteturas orientadas por entidades, relações e grafos. A literatura aponta a necessidade de migrar de estruturas baseadas em registros para modelos com entidades, relações e identificadores persistentes (Alemu; Stevens; Ross; Chandler, 2012; Gonzales, 2014; Ullah; Khusro; Ullah; Naeem, 2018).

Os modelos conceituais bibliográficos oferecem subsídios para essa transição. O *Functional Requirements for Bibliographic Records* (FRBR) introduziu uma estrutura de entidades e relações, posteriormente consolidada pelo IFLA *Library Reference Model* (IFLA LRM), que organiza o universo bibliográfico por entidades como Obra, Expressão, Manifestação e Item, além de agentes, lugares, tempos e conceitos (IFLA *Study Group on The Functional Requirements for Bibliographic Records*, 1998; IFLA, 2017). O IFLA LRM foi concebido em diálogo com novas demandas de reutilização dos dados e favorece a explicitação de relações em ambientes conectados (Simionato Arakaki, 2020). O *Library Reference Model object-oriented* (LRMoo), sua versão orientada a objetos, aproxima essa estrutura de ontologias e grafos RDF, ampliando possibilidades de formalização semântica (IFLA, 2024).

O BIBFRAME também integra a transição para modelos bibliográficos compatíveis com a *Web* de dados. Desenvolvido pela *Library of Congress* como alternativa ao MARC 21 em arquiteturas baseadas em *Linked Data*, utiliza entidades, propriedades e relações para representar recursos bibliográficos (*Library of Congress*, 2016). Estudos sobre modelos multi-entidade, como FRBR, RDA e BIBFRAME, destacam que a representação em ambientes semânticos depende de decisões sobre granularidade, relações entre entidades, correspondência entre elementos descritivos e compatibilidade com padrões de descrição (Baker; Coyle; Petiya, 2014).

Experiências internacionais mostram diferentes aplicações de dados bibliográficos conectados. O LIBRIS, da Suécia, emprega URIs persistentes, RDF, SPARQL e negociação de conteúdo para integrar o catálogo à *Web* Semântica (Malmsten, 2008). O data.bnf.fr, da *Bibliothèque Nationale* de *France*, articula registros bibliográficos e de autoridade, entidades, vocabulários e recursos externos em *Linked Open Data* (Silveira *et al*., 2020). Esses casos evidenciam que a publicação conectada depende da articulação entre modelagem conceitual, infraestrutura técnica, políticas institucionais, mecanismos de acesso e manutenção contínua dos dados.

A adoção de LOD em catálogos apresenta desafios conceituais, técnicos e institucionais, como conversão de registros legados, mapeamento entre formatos e modelos, definição da granularidade, qualidade dos dados, persistência dos identificadores e documentação das decisões de modelagem (Gonzales, 2014; Ullah; Khusro; Ullah; Naeem, 2018; Yee, 2009). Em redes cooperativas, somam-se a coordenação entre instituições produtoras, a padronização de procedimentos, a definição de responsabilidades e a governança dos dados ao longo de seu ciclo de atualização e publicação.

Assim, a publicação de dados bibliográficos conectados articula *Web* Semântica, *Linked Open Data*, tecnologias semânticas, modelos conceituais e governança. Esses fundamentos orientam a arquitetura proposta para transformar registros bibliográficos e de autoridade da Rede Bibliodata em dados conectados, interoperáveis e integráveis a outros ecossistemas de informação, preservando a continuidade dos processos cooperativos de catalogação.

3 PROCEDIMENTOS METODOLÓGICOS

Adotou-se a revisão integrativa da literatura para identificar métodos, técnicas e ferramentas de publicação de dados abertos conectados em bibliotecas. As buscas combinaram, em português, inglês e espanhol, termos relativos a publicação e abertura de dados, bibliotecas, registros bibliográficos e de autoridade, *Linked Data*, *Linked Open Data*, RDF, SPARQL e tecnologias semânticas. Foram consultadas *Scopus*, *Web of Science*, ACM Digital Library, IEEE *Digital Library*, BDTD, NDLTD, OATD, Oasisbr, BRAPCI e SciELO, considerando documentos publicados entre 2006 e 2026. Além dos estudos acadêmicos, a revisão permitiu identificar boas práticas e experiências internacionais de publicação de dados bibliográficos conectados, incluindo iniciativas europeias (Silveira *et al*., 2020). A busca inicial resultou em 315 registros; após a exclusão de 46 duplicatas, 269 documentos foram triados

por título e resumo. Os documentos remanescentes foram avaliados na íntegra segundo os critérios de inclusão e exclusão. Quatro estudos em italiano foram incorporados por apresentarem resumo em inglês e aderência temática. Ao final, 74 documentos compuseram o corpus, submetido à análise temática dos componentes recorrentes de extração, transformação, modelagem, representação, armazenamento, publicação, visualização e integração de dados.

Os estudos selecionados evidenciaram um ciclo de publicação envolvendo planejamento e coleta, curadoria e higienização, conversão para formatos abertos, enriquecimento semântico, publicação, divulgação e evolução contínua dos dados (Rautenberg, 2019; Santarém Segundo, 2018). Esses achados fundamentaram um modelo arquitetural baseado em *Linked Open Data* e tecnologias semânticas, organizado em oito camadas: Sistema Fonte; Extração e Transformação; Modelo Conceitual e Ontologias; Representação e Serialização Semântica; Armazenamento; Publicação e Acesso; Aplicação e Visualização; e Integração e Alinhamento. A organização em camadas busca separar responsabilidades e tornar explícitas as dependências entre transformação, modelagem, persistência, acesso e exploração dos dados.

A partir desse modelo, foi desenvolvido um protótipo que operacionaliza componentes selecionados das camadas para examinar, em caráter exploratório, a integração entre extração, transformação, representação RDF, armazenamento, publicação e consulta. O objetivo foi observar a articulação funcional desses componentes e identificar questões de implementação, sem pretender validar formalmente todas as camadas, uma ontologia completa ou o modelo conceitual em sua versão definitiva.

4 ARQUITETURA DO PORTAL SEMÂNTICO DA REDE BIBLIODATA

A arquitetura do Portal Semântico da Rede Bibliodata estrutura em oito camadas o fluxo desde a origem dos dados até sua publicação e exploração semântica. As funções identificadas na literatura foram agrupadas em níveis de origem, transformação, modelagem, representação, armazenamento, publicação, aplicação e integração.

O Quadro 1 sintetiza os componentes, entradas e saídas, protocolos, formatos e fundamentação de cada camada.

Quadro 1 - Camadas da arquitetura do Portal Semântico da Rede Bibliodata

Camada	Componentes	Entradas > Saídas	Protocolos e formatos	Fundamentação
1. Sistema Fonte	*Koha*; banco relacional (MySQL/MariaDB); catalogação cooperativa	Catalogação pelas instituições > registros bibliográficos e de autoridade	MARC 21 (ISO 2709), MARCXML, CSV; OAI-PMH, SRU, Z39.50; SQL (leitura)	Alemu, Stevens, Ross, Chandler (2012); Gonzales (2014); Rautenberg (2019) Santarém Segundo (2018); Ullah, Khusro, Ullah, Naeem, (2018); Yee (2009);
2. Extração e Transformação	Extração; limpeza e normalização; *crosswalk* MARC > modelo; conversão RDF (*pymarc*, XSLT, *OpenRefine*, RML/YARRRML)	Registros MARC 21/MARCXML > triplas RDF, relatórios de validação e proveniência	*Turtle*, N-*Triples*, JSON-LD; carga inicial e incremental	Alemu, Stevens, Ross, Chandler (2012); Castro e Jesus (2018); Lóscio, Burle e Calegari (2017); Santarém Segundo (2018); Rautenberg (2019) Van Hooland e Verborgh (2014); Gonzales (2014); Yee (2009);
3. Modelo Conceitual e Ontologias	IFLA LRM; LRMoo, BIBFRAME, RDA e Dublin Core como referências complementares; política de URIs (camada normativa)	Requisitos e análise dos campos MARC > perfil de aplicação e regras de mapeamento	Documentação do modelo; ontologias em OWL/RDFS	Alemu (2022) Arakaki, Simionato e IFLA (2017, 2024); IFLA Study Group (1998); Baker, Coyle e Petiya (2014); Library of Congress (2016); Santos (2017); Simionato Arakaki (2020);
4. Representação e Serialização Semântica	RDF 1.1, RDFS, OWL, SKOS, JSON-LD; grafos nomeados; verificação de conformidade	Triplas + vocabulários > grafos RDF nomeados e verificados	RDF/*Turtle*, JSON-LD; URIs HTTP	Allemang, Hendler e Gandon (2020); Berners-Lee, Hendler e Lassila (2001); Hitzler, Krötzsch e Rudolph (2009); Hogan *et al*. (2021) Miles e Bechhofer (2009);

				Santarém Segundo e Coneglian (2016); W3C (2014);
5. Armazenamento	Triplestore RDF (Virtuoso); processos de carga e publicação (Rodolf); backup e dumps RDF	Grafos RDF > grafo persistido e resultados de consulta	SPARQL 1.1 *Query*/*Update* sobre HTTP; *dumps* RDF	Heath e Bizer (2011); Hitzler, Krötzsch e Rudolph (2009); Hogan *et al*. (2021); Logilab (2024b); OpenLink Software (2024) Wang e Yang (2018);
6. Publicação e Acesso	URIs; negociação de conteúdo; *endpoint* SPARQL; API REST (consulta, exportação, inserção); *dumps*; metadados e licença	Consultas de clientes > recursos em HTML/RDF/JSON-LD, resultados SPARQL, conjuntos de dados	HTTPS; SPARQL *Protocol*; REST/JSON; autenticação na escrita	Berners-Lee (2006); Bizer, Heath e Berners-Lee (2009); Heath e Bizer (2011); Isotani e Bittencourt (2015); Lóscio, Burle e Calegari (2017); Malmsten (2008); Silveira *et al*. (2020)
7. Aplicação e Visualização	*CubicWeb*; alternativas: *React*/*Next.js*, *Tainacan*, *Omeka S*, *Wikibase*; busca, ficha por entidade, painel estatístico, visualização de grafo	Resultados SPARQL e da API > páginas, buscas e visualizações para o usuário	HTTPS; SPARQL sobre HTTP; REST/JSON	Cervone e Svensson (2015); Coneglian *et al*. (2019); Hyvönen (2023); IFLA (2017); Logilab (2024a)
8. Integração e Alinhamento	Reconciliação com VIAF, *Wikidata*, ORCID, ISNI; validação curatorial; identificadores ARK, *Handle*, DOI	Entidades do grafo > ligações validadas (owl:sameAs, skos:exactMatch) em grafo próprio	APIs HTTP das bases externas; RDF	Alemu e Garoufallou (2020); Arakaki (2016) Ullah, Khusro, Ullah, Naeem, (2018); Berners-Lee (2006); Bizer, Heath e Berners-Lee (2009); Gaitanou *et al*. (2024) Heath e Bizer (2011);

				Jesus, Castro e Ramalho (2021); Van Hooland e Verborgh (2014);

Fonte: elaboração própria (2026).

A Camada 1 (Sistema Fonte) corresponde ao ambiente em que as instituições cooperantes produzem e atualizam os registros, considerando Koha, MARC 21 e banco relacional como sistema de catalogação. As saídas podem ser obtidas em MARC 21, MARCXML ou CSV, por exportação, consulta ao banco ou protocolos como OAI-PMH, SRU e Z39.50. Como estruturas centradas em registros apresentam limitações para representar entidades, relações e identificadores persistentes (Alemu; Stevens; Ross; Chandler, 2012; Gonzales, 2014; Ullah; Khusro; Ullah; Naeem, 2018), a transformação semântica é realizada em camada própria, preservando os fluxos de catalogação (Yee, 2009). O planejamento da extração define escopo, campos e periodicidade (Rautenberg, 2019; Santarém Segundo, 2018), enquanto a qualidade dos registros de origem permanece sob responsabilidade das instituições catalogadoras e o acesso ao banco é restrito à leitura.

A Camada 2 (Extração e Transformação) converte MARC 21 ou MARCXML em triplas RDF, com serializações como Turtle, N-Triples ou JSON-LD e registro de validação e proveniência. O fluxo envolve extração, limpeza e normalização de nomes, datas e identificadores, *crosswalk* dos campos MARC para entidades, atributos e relações e geração das triplas, podendo utilizar Python/pymarc, XSLT, OpenRefine e RML/YARRRML (Castro; Jesus, 2018; Rautenberg, 2019; Santarém Segundo, 2018; Van Hooland; Verborgh, 2014). A mudança central é a passagem de uma estrutura orientada ao registro para outra baseada em entidades e relações (Alemu; Stevens; Ross; Chandler, 2012; Gonzales, 2014; Yee, 2009). A camada admite carga inicial e cargas incrementais, com scripts e mapeamentos versionados para documentar proveniência e garantir a reprodução das transformações (Lóscio; Burle; Calegari, 2017).

A Camada 3 (Modelo Conceitual e Ontologias) é normativa e define as entidades, atributos, relações, regras de mapeamento e política de identificadores que orientam as Camadas 2 e 4. O IFLA LRM (IFLA, 2017) é a referência conceitual principal; LRMoo (IFLA, 2024), BIBFRAME (*Library of Congress*, 2016), RDA e Dublin Core funcionam como referências

complementares, incorporadas conforme as decisões do perfil de aplicação. A adequação do LRM aos dados conectados é discutida por Simionato Arakaki (2020) e sua articulação com RDA, Linked Data e BIBFRAME por Alemu (2022). Decisões sobre granularidade, correspondência entre elementos e tratamento das relações são determinantes em modelos multi-entidade (Baker; Coyle; Petiya, 2014) e devem ser documentadas para assegurar consistência entre instituições e evolução do modelo (Gonzales, 2014; Ullah; Khusro; Ullah; Naeem, 2018).

A Camada 4 (Representação e Serialização Semântica) expressa os dados como grafos processáveis por máquinas. RDF 1.1 constitui o modelo de base; RDFS e OWL formalizam classes, propriedades e restrições e podem apoiar axiomas e inferências; SKOS representa vocabulários controlados; e JSON-LD facilita o consumo em aplicações *Web* (Allemang; Hendler; Gandon, 2020; Miles; Bechhofer, 2009; Santarém Segundo; Coneglian, 2016; *World Wide Web Consortium*, 2014). A organização em grafos nomeados para dados bibliográficos, autoridades, vocabulários e alinhamentos externos segue práticas de grafos de conhecimento e apoia a verificação de conformidade, a proveniência e o versionamento dos conjuntos (Hogan *et al*., 2021).

A Camada 5 (Armazenamento) persiste as triplas em um triplestore RDF responsável pela indexação e consulta SPARQL (Heath; Bizer, 2011; Hitzler; Krötzsch; Rudolph, 2009). Soluções como Virtuoso, *GraphDB* ou *Apache Jena Fuseki* podem desempenhar essa função; no protótipo, o Virtuoso atua como repositório RDF e o *Rodolf* apoia o fluxo de produção e publicação dos grafos (Logilab, 2024b; OpenLink *Software*, 2024). Os grafos podem ser carregados por importação em lote ou SPARQL Update. O *CubicWeb* permanece na camada de aplicação, separando o repositório semântico das funções de interface, enquanto operações de escrita, *backups* e *dumps* RDF são controlados e versionados para apoiar preservação e recuperação.

A Camada 6 (Publicação e Acesso) implementa princípios de *Linked Data* por URIs HTTP persistentes e desreferenciáveis, negociação de conteúdo e acesso em HTML, RDF/Turtle ou JSON-LD (Berners-Lee, 2006; Bizer; Heath; Berners-Lee, 2009). A proposta considera o padrão de identificação http://bibliodata.ibict.br/id/... e inclui *endpoint* SPARQL, *dumps* RDF, API REST para consulta, exportação e inserção, além de páginas por entidade, seguindo referências como LIBRIS e data.bnf.fr (Malmsten, 2008; Silveira *et al.,* 2020). A camada também concentra metadados dos conjuntos, licença, proveniência, versionamento e

documentação dos pontos de acesso, em consonância com boas práticas de dados na *Web* (Lóscio; Burle; Calegari, 2017). O tráfego ocorre sobre HTTPS; a leitura é pública, com controle de taxa, e a escrita exige autenticação institucional e auditoria.

A Camada 7 (Aplicação e Visualização) reúne interfaces de busca e exploração que consomem SPARQL e a API da Camada 6 para produzir páginas por entidade, buscas, painéis estatísticos e visualizações de grafo. Relaciona-se às tarefas de encontrar, identificar, selecionar, obter e explorar do IFLA LRM (IFLA, 2017) e à interação do usuário com dados conectados (Cervone; Svensson, 2015). O modelo Sampo (Hyvönen, 2023) orienta a separação entre infraestrutura de dados e aplicações de exploração com busca facetada. No protótipo, utiliza-se *CubicWeb* (Logilab, 2024a), embora a camada possa empregar front-ends próprios ou plataformas como *React/Next.js*, *Tainacan, Omeka S, DSpace e Wikibase,* sempre por meio dos serviços da Camada 6 e sem acesso direto ao repositório RDF.

A Camada 8 (Integração e Alinhamento) liga agentes, obras e conceitos do grafo a bases externas como VIAF, *Wikidata*, ORCID e ISNI, utilizando relações como *owl:sameAs* e *skos:exactMatch*. Serviços de reconciliação consultam APIs externas e as ligações passam por validação curatorial antes da publicação (Heath; Bizer, 2011; Van Hooland; Verborgh, 2014). A integração com identificadores e vocabulários compartilhados amplia a interoperabilidade e a inserção dos catálogos em um espaço global de conhecimento (Alemu; Garoufallou, 2020; Arakaki, 2016; Gaitanou *et al.,* 2024; Jesus; Castro; Ramalho, 2021). Identificadores como ARK, *Handle* e DOI podem ser adotados conforme políticas institucionais. O fluxo é bidirecional: as ligações validadas retornam ao armazenamento em grafo próprio e são publicadas pela Camada 6.

5 RESULTADOS E DISCUSSÃO DOS RESULTADOS

A implementação do protótipo foi orientada pela correspondência entre as camadas da arquitetura e os componentes necessários à publicação de dados bibliográficos conectados. A seleção das ferramentas considerou compatibilidade com tecnologias semânticas, possibilidade de integração com repositórios RDF e adequação às etapas de representação, armazenamento, consulta e visualização. O objetivo não foi testar exaustivamente cada tecnologia, mas observar sua articulação no fluxo proposto.

Na Camada 7, adotou-se o *CubicWeb*, *framework* semântico de código aberto para publicação e exploração de dados estruturados na *Web* (Logilab, 2024a). Sua escolha

relaciona-se à possibilidade de representar entidades, propriedades e relações em ambiente orientado a dados semânticos. No protótipo, o *CubicWeb* atua como componente de mediação entre os dados estruturados e as funcionalidades de consulta, navegação, páginas de detalhe e visualização voltadas ao usuário.

Na Camada 5, foi selecionado o Virtuoso como triplestore RDF com suporte a consultas SPARQL (*OpenLink Software*, 2024), permitindo persistir e recuperar entidades, atributos e relações do grafo. O Rodolf apoia o fluxo de produção e publicação RDF no ecossistema utilizado pelo protótipo (Logilab, 2024b), enquanto o *CubicWeb* permanece responsável pelas funções de aplicação e visualização. Essa distribuição de responsabilidades evita atribuir ao banco da aplicação o papel de repositório semântico autoritativo.

A implementação inicial contemplou a extração de registros MARC 21 do catálogo da Rede Bibliodata e sua transformação em RDF, visando uma representação orientada por entidades e relações. O modelo conceitual permanece em avaliação, tendo o IFLA LRM como referência principal e considerando futuro alinhamento ao LRMoo para formalização em ambiente RDF (IFLA, 2024). Essa etapa permitiu observar como elementos de registros legados podem ser reorganizados sem antecipar uma validação definitiva do perfil de aplicação.

O desenvolvimento evidenciou a complexidade do *crosswalk* entre MARC 21 e IFLA LRM, decorrente das diferenças entre campos e subcampos de um formato de intercâmbio e entidades, atributos e relações de um modelo conceitual. Isso exige decisões sobre granularidade, correspondência entre elementos descritivos, representação das relações bibliográficas e tratamento dos dados de autoridade. Essas escolhas afetam diretamente a consistência da transformação e precisam ser registradas no perfil de aplicação e nas regras de mapeamento.

Também foram observadas questões de instalação, configuração e integração dos componentes, especialmente do *CubicWeb* em contexto bibliográfico. A documentação disponível mostrou-se mais orientada a usos gerais do framework do que a catálogos e redes de catalogação cooperativa. Por isso, buscou-se contato com pesquisadores vinculados à *Bibliothèque Nationale de France* para conhecer experiências prévias de aplicação de tecnologias semânticas em ambientes bibliográficos e compreender alternativas de configuração e integração.

Os resultados convergem com experiências internacionais quanto ao uso de identificadores persistentes, RDF, mecanismos de consulta e integração com fontes externas.

A arquitetura, contudo, considera as especificidades de uma rede cooperativa, em que os dados são produzidos, atualizados e mantidos por múltiplas instituições, o que amplia a importância de regras comuns e responsabilidades compartilhadas.

Essa característica evidencia que a publicação de dados conectados depende não apenas das tecnologias escolhidas, mas também de decisões de modelagem, documentação, governança e manutenção. Na Rede Bibliodata, tornam-se necessários procedimentos comuns para conversão, alinhamento, publicação e atualização de registros bibliográficos e de autoridade produzidos por diferentes instituições, bem como mecanismos para registrar proveniência e acompanhar mudanças no fluxo.

A verificação da arquitetura foi conduzida, em caráter exploratório, com uma amostra de 20 registros bibliográficos. O teste percorreu a transformação de MARC 21 em RDF, a carga no Virtuoso, a consulta SPARQL e o consumo pelo portal. O objetivo foi verificar a integração funcional das camadas efetivamente implementadas, e não validar formalmente uma ontologia, medir desempenho em escala ou afirmar conformidade integral do protótipo com todas as camadas propostas.

Na transformação, observaram-se correspondências básicas entre campos MARC 21 e uma estrutura orientada por entidades e relações, preservando título, responsabilidades, dados de publicação, descrição física, ISBN e assuntos. O IFLA LRM serviu como referência conceitual, enquanto o alinhamento ao LRMoo permanece em avaliação. Também foram adotadas URIs HTTP persistentes para identificação dos recursos, conforme os princípios de *Linked Data* (Berners-Lee, 2006), preparando o fluxo para ligações e acesso externo.

As consultas SPARQL no *Virtuoso* recuperaram entidades e relações e alimentaram busca, páginas de detalhe e estatísticas do portal. A verificação de qualidade foi estrutural e manual, com restrições definidas no *schema* da aplicação e teste de ida e volta entre exportação e reinserção. Esses procedimentos fornecem evidências preliminares de consistência do fluxo, mas não substituem validação semântica formal nem testes de desempenho, escalabilidade, atualização incremental e comportamento em conjuntos de dados de maior volume.

À luz das boas práticas para publicação de dados na *Web* (Lóscio; Burle; Calegari, 2017), o protótipo apresenta evidências iniciais de processabilidade, descoberta, reúso, acesso e interoperabilidade técnica por meio de RDF, *endpoint* SPARQL, API e URIs HTTP. A confiabilidade e a interoperabilidade semântica ainda dependem do refinamento do perfil de

aplicação, da documentação da proveniência, da ampliação das cargas e do monitoramento das atualizações entre os componentes. Os resultados devem, portanto, ser interpretados como prova de integração funcional em processo de consolidação, e não como validação definitiva da arquitetura em escala de produção.

6 CONSIDERAÇÕES FINAIS

Este artigo apresentou uma arquitetura em oito camadas para o Portal Semântico da Rede Bibliodata, orientada à publicação conectada de dados bibliográficos e de autoridade. A proposta abrange sistema fonte, extração e transformação, modelagem conceitual, representação semântica, armazenamento, publicação e acesso, aplicação e visualização e integração externa. A organização em camadas procura separar responsabilidades técnicas e conceituais, mantendo a possibilidade de evolução independente dos componentes.

A arquitetura articula tecnologias como RDF, SPARQL, RDFS e OWL com modelos conceituais bibliográficos e princípios de *Linked Open Data*. O IFLA LRM constitui a principal referência conceitual do protótipo, enquanto LRMoo e BIBFRAME funcionam como referências complementares para formalização e interoperabilidade. Desse modo, o trabalho relaciona fundamentos da *Web* Semântica e da Ciência da Informação às etapas de conversão, estruturação, publicação e reutilização dos dados, sem pressupor a substituição imediata dos sistemas legados de catalogação.

O protótipo, com *CubicWeb*, Virtuoso e *Rodolf*, permitiu examinar a articulação entre aplicação, armazenamento e consulta em ambiente RDF, partindo de registros MARC 21. Os resultados preliminares indicam a viabilidade de organizar o fluxo de publicação da Rede Bibliodata em uma arquitetura em camadas, embora ainda sejam necessários refinamento do modelo, avaliação da qualidade, documentação dos mapeamentos, testes de atualização e ampliação das cargas com registros bibliográficos e de autoridade.

Entre os desafios estão o mapeamento MARC 21, modelos orientados a entidades, a definição da granularidade, a conversão de registros legados, a persistência de identificadores, a integração com bases externas de autoridade e a documentação das decisões de modelagem. Esses aspectos evidenciam dimensões técnicas, conceituais e institucionais da publicação de dados bibliográficos conectados e assumem importância adicional em um ambiente cooperativo com múltiplos produtores de dados.

Como continuidade, prevê-se aprofundar o *crosswalk* entre MARC 21 e IFLA LRM/LRMoo, ampliar o conjunto de dados, definir políticas de identificação persistente, alinhar entidades a VIAF, *Wikidata*, ORCID e ISNI e avaliar formas de consulta, navegação e visualização. Também será necessário acompanhar a qualidade e a atualização dos dados ao longo do fluxo. Essas etapas poderão consolidar procedimentos metodológicos aplicáveis à publicação de dados bibliográficos conectados em redes cooperativas de informação.

REFERÊNCIAS

ALEMU, Getaneh. The future of enriched, linked, open and filtered metadata: making sense of IFLA LRM, RDA, Linked Data and BIBFRAME. London: **Facet Publishing**, 2022. Disponível em: https://www.cambridge.org/core/books/future-of-enriched-linked-open-and-filtered-metadata/E6075126DE0B05E73BF7F67E7A5E701F. Acesso em: 10 set. 2026.

ALEMU, Getaneh; GAROUFALLOU, Emmanouel. The future of interlinked, interoperable and scalable metadata. **International Journal of Metadata, Semantics and Ontologies**, v. 14, n. 2, p. 81-87, 2020. em: https://www.inderscience.com/info/inarticle.php?artid=108340. Acesso em: 10 set. 2026.

ALEMU, Getaneh; STEVENS, Brett; ROSS, Penny; CHANDLER, Jane. Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models. **New Library World**, v. 113, n. 11/12, p. 549-570, 2012. Disponível em: https://doi.org/10.1108/03074801211282920. Acesso em: 10 set. 2026.

ALLEMANG, Dean; HENDLER, James; GANDON, Fabien. Semantic Web for the working ontologist: effective modeling for Linked Data, RDFS, and OWL. 3. ed. New York: **Association for Computing Machinery**, 2020. Disponível em: https://doi.org/10.1145/3382097. Acesso em: 10 set. 2026.

ARAKAKI, Felipe Augusto. Linked Data: ligação de dados bibliográficos. 2016. 144 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, **Universidade Estadual Paulista**, Marília, 2016. Disponível em: http://hdl.handle.net/11449/147979. Acesso em: 10 set. 2026.

ARAKAKI, Felipe Augusto; SIMIONATO, Ana Carolina; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Catalogação e tecnologia: interseções com a Web Semântica. **Informação@Profissões**, Londrina, v. 6, n. 2, p. 3-19, 2017. Disponível em: https://doi.org/10.5433/2317-4390.2017v6n2p03. Acesso em: 10 set. 2026.

BAKER, Thomas; COYLE, Karen; PETIYA, Sean. Multi-entity models of resource description in the Semantic Web: a comparison of FRBR, RDA and BIBFRAME. **Library Hi Tech**, v. 32, n. 4, p. 562-582, 2014. Disponível em: https://doi.org/10.1108/LHT-08-2014-0081. Acesso em: 10 set. 2026.

BERNERS-LEE, Tim. Linked Data. [S. l.]: **World Wide Web Consortium**, 2006. Disponível em: https://www.w3.org/DesignIssues/LinkedData.html. Acesso em: 10 set. 2026.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web. **Scientific American**, New York, v. 284, n. 5, p. 34-43, maio 2001. Disponível em: https://www.scientificamerican.com/article/the-semantic-web/. Acesso em: 10 set. 2026.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked Data: the story so far**. International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 1-22, 2009. Disponível em: https://doi.org/10.4018/jswis.2009081901. Acesso em: 10 set. 2026.

CASTRO, Fabiano Ferreira de; JESUS, Ananda Fernanda de. Adequando dados bibliográficos ao Linked Data: requisitos necessários, vantagens e desafios. **Informação & Sociedade: Estudos**, João Pessoa, v. 28, n. 3, p. 149-168, 2018. Disponível em: https://periodicos.ufpb.br/index.php/ies/article/view/38288. Acesso em: 10 set. 2026.

CERVONE, H. Frank; SVENSSON, Lars G. (ed.). Linked Data and user interaction. Berlin; München; Boston: **De Gruyter Saur**, 2015. Disponível em: https://doi.org/10.1515/9783110317008. Acesso em: 10 set. 2026.

CONEGLIAN, Caio Saraiva; ROA-MARTÍNEZ, Sandra Milena; FERREIRA, Ana Maria Jensen Ferreira da Costa; VIDOTTI, Silvana Aparecida Borsetti Gregório; SANTARÉM SEGUNDO, José Eduardo. Tecnologias da Web Semântica na arquitetura da informação. **Revista Interamericana de Bibliotecología**, Medellín, v. 42, n. 1, p. 23-35, 2019. em: https://revistas.udea.edu.co/public/journals/recursos_html/Rib-Html/rib-v42n1/v42n1a2.html. Acesso em: 10 set. 2026.

GAITANOU, Panorea; ANDREOU, Ioanna; SICILIA, Miguel-Angel; GAROUFALLOU, Emmanouel. Linked data for libraries: creating a global knowledge space, a systematic literature review. **Journal of Information Science**, v. 50, n. 1, p. 204-244, 2024. em: https://journals.sagepub.com/doi/10.1177/01655515221084645. Acesso em: 10 set. 2026.

GONZALES, Brighid M. Linking libraries to the web: Linked Data and the future of the bibliographic record. **Information Technology and Libraries,** Chicago, v. 33, n. 4, p. 10-22, 2014. Disponível em: https://ital.corejournals.org/index.php/ital/article/view/5631. Acesso em: 10 set. 2026.

HEATH, Tom; BIZER, Christian. **Linked Data**: evolving the web into a global data space. San Rafael: Morgan & Claypool, 2011. 122 p. Disponível em: https://books.google.com.br/books?id=OFv59Wcfkx8C. Acesso em: 10 set. 2026.

HITZLER, Pascal; KRÖTZSCH, Markus; RUDOLPH, Sebastian. Foundations of Semantic Web Technologies. Boca Raton: **Chapman & Hall/CRC**, 2009. 455 p. ISBN 978-1-4200-9050-5. Disponível em: https://www.semantic-web-book.org/. Acesso em: 10 set. 2026.

HOGAN, Aidan; BLOMQVIST, Eva; COCHEZ, Michael; D'AMATO, Claudia, *et al*. Knowledge

graphs. **ACM Computing Surveys**, New York, v. 54, n. 4, art. 71, p. 1-37, 2021. Disponível em: https://doi.org/10.1145/3447772. Acesso em: 10 set. 2026.

HYVÖNEN, Eero. Digital humanities on the Semantic Web: Sampo model and portal series. **Semantic Web,** v. 14, n. 4, p. 729-744, 2023. Disponível em: https://journals.sagepub.com/doi/10.3233/SW-223034. Acesso em: 10 set. 2026.

IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. Functional requirements for bibliographic records: final report. München: **K. G. Saur**, 1998. Disponível em: https://www.degruyterbrill.com/document/doi/10.1515/9783110962451/html. Acesso em: 10 set. 2026.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. IFLA Library Reference Model: a conceptual model for bibliographic information. Den Haag: **IFLA**, 2017. Disponível em: https://repository.ifla.org/handle/20.500.14598/40. Acesso em: 20 jun. 2026.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. LRMoo: object-oriented definition and mapping from IFLA Library Reference Model. Den Haag: **IFLA**, 2024. Disponível em: https://repository.ifla.org/items/94aedb49-2d6e-4a6d-9974-f33abb7e3c0e. Acesso em: 20 jun. 2026.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. Dados abertos conectados: em busca da Web do conhecimento. São Paulo: **Novatec**, 2015. 176 p. ISBN 978-85-7522-449-6. Disponível em: https://novatec.com.br/livros/dadosabertos/. Acesso em: 10 set. 2026.

JESUS, Ananda Fernanda de; CASTRO, Fabiano Ferreira de; RAMALHO, Rogério Aparecido Sá. O papel das bibliotecas no Linked Data. Encontros Bibli: **Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 26, p. 1-21, 2021. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/75909. Acesso em: 1 set. 2026.

LIBRARY OF CONGRESS. Overview of the BIBFRAME 2.0 Model. Washington, DC: **Library of Congress**, 2016. Disponível em: https://www.loc.gov/bibframe/docs/bibframe2-model.html. Acesso em: 20 jun. 2026.

LOGILAB. *CubicWeb*: semantic web application framework. Paris: **Logilab**, 2024. Disponível em: https://www.cubicweb.org/. Acesso em: 20 jun. 2026.

LOGILAB. Rodolf: RDF Data Production Monitoring. Paris: Logilab, 2024. Disponível em: https://forge.extranet.logilab.fr/cubicweb/cubes/rodolf. Acesso em: 20 jun. 2026.

LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton. Data on the Web Best Practices. **W3C Recommendation**, 31 jan. 2017. Disponível em: https://www.w3.org/TR/2017/REC-dwbp-20170131/. Acesso em: 20 jun. 2026.

MALMSTEN, Martin. *Making a library catalogue part of the Semantic Web*. *In*: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2008,

Berlin. Proceedings [...]. [*S. l.*]: **Dublin Core Metadata Initiative**, 2008. p. 146-152. Disponível em: https://dcpapers.dublincore.org/pubs/article/view/927. Acesso em: 20 jun. 2026.

MILES, Alistair; BECHHOFER, Sean (ed.). SKOS Simple Knowledge Organization System Reference. **W3C Recommendation**, 18 ago. 2009. Disponível em: https://www.w3.org/TR/skos-reference/. Acesso em: 20 jun. 2026.

OPENLINK SOFTWARE. Virtuoso Universal Server. Burlington: **OpenLink Software**, [s. d.]. Disponível em: https://virtuoso.openlinksw.com/. Acesso em: 20 jun. 2026.

RAUTENBERG, Sandro. Web Semântica e Bibliografia: um estudo para a publicação de registros bibliográficos como dados abertos conectados. **Em Questão**, Porto Alegre, v. 25, n. esp., p. 313-330, 2019. Disponível em: https://www.redalyc.org/journal/4656/465661027018/465661027018.pdf. Acesso em: 20 jun. 2026.

SANTARÉM SEGUNDO, José Eduardo. Web Semântica: fluxo para publicação de dados abertos e ligados. **Informação em Pauta**, Fortaleza, v. 3, n. esp., p. 117-140, nov. 2018. Disponível em: https://periodicos.ufc.br/informacaoempauta/article/view/39721. Acesso em: 10 set. 2026.

SANTARÉM SEGUNDO, José Eduardo; CONEGLIAN, Caio Saraiva. Web Semântica e Ontologias: um estudo sobre construção de axiomas e uso de inferências. **Informação & Informação**, Londrina, v. 21, n. 2, p. 217-244, 2016. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/26417. Acesso em: 10 set. 2026.

SERRA, Liliana Giusti; SANTARÉM SEGUNDO, José Eduardo. Dos silos de dados à web dos dados: bibliotecas e o Linked Data. **Informação & Informação**, Londrina, v. 26, n. 2, p. 625-645, 2021. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/41603. Acesso em: 13 set. 2026.

SILVEIRA, Lúcia da; SILVA, Fabiano Couto Corrêa da; SOUSA, Aline Trierweiler de; ZILIO, Sara Caselani *et al*. Convergência de práticas Linked Open Data na Bibliothèque Nationale de France (BNF DATA). Revista ACB: **Biblioteconomia em Santa Catarina**, Florianópolis, v. 25, n. 1, p. 21-40, 2020. Disponível em: https://revista.acbsc.org.br/racb/article/view/1643. Acesso em: 10 set. 2026.

SIMIONATO ARAKAKI, Ana Carolina. O Modelo IFLA Library Reference Model e o Linked Data. **Informação & Informação**, Londrina, v. 25, n. 3, p. 163-186, 2020. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/41868. Acesso em: 10 set. 2026.

ULLAH, Irfan; KHUSRO, Shah; ULLAH, Asim; NAEEM, Muhammad. An overview of the current state of Linked and Open Data in cataloging. **Information Technology and Libraries**, Chicago, v. 37, n. 4, p. 47-80, 2018. Disponível em:

https://ital.corejournals.org/index.php/ital/article/view/10432. Acesso em: 10 set. 2026.

VAN HOOLAND, Seth; VERBORGH, Ruben. Linked Data for libraries, archives and museums: how to clean, link and publish your metadata. London: **Facet Publishing**, 2014. Disponível em: https://www.routledge.com/Linked-Data-for-Libraries-Archives-and-Museums-How-to-clean-link-and-publish-your-metadata/vanHooland-Verborgh/p/book/9781856049641. Acesso em: 13 set. 2026.

WANG, Yongming; YANG, Sharon Q. Linked Data technologies and what libraries have accomplished so far. **International Journal of Librarianship**, v. 3, n. 1, p. 3-20, 2018. Disponível em: https://journal.calaijol.org/index.php/ijol/article/view/62. Acesso em: 10 set. 2026.

WORLD WIDE WEB CONSORTIUM. RDF 1.1 Concepts and Abstract Syntax. **W3C Recommendation,** 25 fev. 2014. Disponível em: https://www.w3.org/TR/rdf11-concepts/. Acesso em: 20 jun. 2026.

YEE, Martha M. Can bibliographic data be put directly onto the Semantic Web? **Information Technology and Libraries**, Chicago, v. 28, n. 2, p. 55-80, jun. 2009. Disponível em: https://doi.org/10.6017/ital.v28i2.3175. Acesso em: 10 set. 2026.