


# CATALOGAÇÃO EM DADOS CONECTADOS ABERTOS: UMA EXPERIÊNCIA DE BIBLIOTECA UNIVERSITÁRIA COM A WIKIDATA

Cataloguing in linked open data: a university library's experience with Wikidata

**Lilian Viana**

Universidade de São Paulo, Escola de Comunicações e Artes  
São Paulo, SP, Brasil  
lilianviana@usp.br

<https://orcid.org/0000-0002-8574-5268> 

A lista completa com informações dos autores está no final do artigo 

## RESUMO

**Objetivo:** A partir de experiência realizada na biblioteca da Escola de Comunicações e Artes (ECA) da Universidade de São Paulo, objetiva abordar os aspectos operacionais da importação de registros de artigos científicos, catalogados em formato MARC21 e disponibilizados em repositório institucional, para a Wikidata, base de dados conectados abertos. Espera-se contribuir para discussões sobre o uso da Wikidata para replicar registros de catálogos de bibliotecas universitárias e promover a interligação entre distintos conjuntos de dados, aumentando o alcance da produção científica no contexto digital.

**Método:** Estudo de caso descritivo, utilizando-se também pesquisa bibliográfica.

**Resultado:** A importação de registros em formato MARC21 para dados conectados abertos demandou a coleta e a estruturação de novos dados, para viabilizar a representação da trama institucional da produção científica na Wikidata. Ao criar metadados para vincular o repositório institucional à Wikidata, a iniciativa expôs a interligação entre conjuntos de dados como estratégia para ampliar o alcance da produção científica no contexto digital. No domínio da visualização de dados, identificou-se a necessidade de desenvolver interfaces que dialoguem com necessidades específicas da instituição.

**Conclusões:** O emprego de softwares abertos despontou como estratégia relevante ao desenvolvimento de ações em bibliotecas, pois caracterizam-se pela existência de uma comunidade de práticas visando tanto o seu desenvolvimento como sua apropriação pelos interessados. Além disso, sobressaiu a relevância de que a escolha por recursos tecnológicos reconheça a relevância de sistemas e dados abertos para o desenvolvimento da ciência aberta.

**PALAVRAS-CHAVE:** Acesso Livre. Artigos de periódicos. Metadados. Registros MARC.

## ABSTRACT

**Objective:** Based on an experience conducted at the library of the School of Communications and Arts (ECA) of the University of São Paulo, the objective is to address the operational aspects of importing records of scientific articles cataloged in MARC21 format into Wikidata, an open knowledge structured database. It aims to contribute to discussions on the use of Wikidata to replicate records from university library catalogs and promote connection between data sets, increasing the reach of scientific production.

**Methodology:** Descriptive case study, bibliographic research.

**Results:** Transforming MARC21 records into linked open data involved the collection and structuring of new data to depict the scientific production network of the institution within Wikidata. By creating metadata that links the institutional repository to Wikidata, this initiative unlocks the interconnectedness of datasets to amplify the reach of scientific production in the digital sphere. In the field of data visualization, the study identifies the importance of developing interfaces that dialogues with specific demands from the institution.

**Conclusions:** Libraries benefit by adopting open-source software. These tools are supported by development communities, ensuring ongoing improvement and user support. Furthermore, by prioritizing open systems and open data, libraries actively contribute to the open science movement.

**KEYWORDS:** Open Access. Journal Articles. Metadata. MARC records.

## 1 INTRODUÇÃO

Instituições privilegiadas de produção de conhecimentos, compete às universidades ocupar-se de dinâmicas referentes à sua circulação para que o conhecimento impacte e beneficie o meio social de forma ampla. No fluxo da comunicação científica fundamentada pelo conhecimento aberto (Gomes; Rosa, 2010), as instituições de ensino superior (IES), frequentemente, recorrem aos repositórios institucionais abertos para viabilizar o acesso à sua produção científica (Machado, 2015).

Cabe às bibliotecas universitárias garantir a organização desta produção por meio de atividades como a catalogação, que compreende o desenvolvimento e emprego de formatos e padrões garantidores da consistência e recuperação das informações (Jesus; Castro; Ramalho, 2021). Apesar de assegurar a consistência dos dados, sua publicação e acesso na *World Wide Web* são limitados quando estes formatos e padrões prescindem de tecnologias semânticas, que dizem respeito ao significado dos dados disponíveis no contexto digital (Allemang; Hendler, 2011). Diante disso, os dados conectados abertos (em inglês, *Linked Open Data - LOD*), uma prática da Web semântica que possibilita maior recuperabilidade da informação e garante condições de livre utilização dos dados (Bizer; Heath; Berners-Lee, 2009), configuram território potencial a ser explorado por bibliotecários.

No ecossistema do conhecimento aberto – passível de acesso, uso, modificação e compartilhamento por todos (Open Knowledge Foundation, [202-?]) –, a Wikidata é representativa desta prática. Nesse sentido, é abordada neste estudo de caso descritivo que apresenta a atuação de biblioteca universitária em práticas de catalogação de produção científica na Wikidata, tendo como objeto a experiência de bibliotecária da Escola de Comunicações e Artes (ECA) da Universidade de São Paulo (USP) em projeto de importação de registros de artigos científicos do Repositório da Produção USP, catalogados em formato MARC21, para a Wikidata.

Além de ser aberta e de amplo alcance, a Wikidata foi escolhida por ser um recurso disponível para uso que elimina a necessidade de espaço para armazenamento de dados e suporte técnico para sua instalação. Ademais, é sustentada por uma comunidade de práticas e dispõe de documentação para o aprendizado, elementos que criaram um caminho factível para a iniciativa em foco.

De caráter expositivo, este texto objetiva explorar os aspectos operacionais dessa importação de registros para a Wikidata, abordando os metadados utilizados. Com isso, espera-se contribuir para discussões sobre o uso da Wikidata para replicar registros de catálogos de bibliotecas universitárias, promovendo a conexão entre distintos conjuntos de dados e aumentando o alcance da produção científica no contexto digital.

## **2 REPOSITÓRIO DA PRODUÇÃO USP: PRESERVAR E COMUNICAR A MEMÓRIA CIENTÍFICA DA UNIVERSIDADE**

No contexto do acesso aberto uma estratégia adotada pelas universidades para gerir e disponibilizar a produção intelectual de sua comunidade é a implementação e manutenção de repositórios institucionais (Robinson-Garcia; Costas; Leeuwen, 2020). Embora a responsabilidade operacional pelos repositórios varie conforme o contexto institucional, sua efetividade se dá a partir da colaboração entre distintos profissionais, como bibliotecários, analistas de sistemas, programadores, docentes e gestores responsáveis pela formulação de políticas institucionais (Lynch, 2003).

Na USP – universidade pública brasileira fundada em 1934 que dispõe de 333 cursos de graduação, 264 de pós-graduação (mestrado e doutorado), 42 unidades de ensino e pesquisa, cerca de 97 mil estudantes e mais de 5 mil docentes distribuídos entre as diversas áreas do conhecimento (DEPAR, [2024]) –, o repositório institucional da produção intelectual foi inaugurado em 2012. A partir da coleta, organização e registro da produção científica, artística, acadêmica e técnica, objetiva ampliar o alcance e acesso aos resultados de atividades acadêmicas e ao conhecimento produzido pela comunidade da USP, configurando eixo fundamental à constituição de sua memória científica. Assim, concentra o registro da produção intelectual e armazena as publicações, quando disponíveis on-line (ABCD, [2024]).

O Repositório da Produção USP possui mais de 1.106 milhões de registros (Repositório..., [2024]) e é responsabilidade das equipes das bibliotecas da universidade catalogar a produção dos docentes vinculados a sua unidade de ensino ou pesquisa. Realizada no software Aleph, a catalogação é no formato MARC21 e os dados não são conectados; consistem em conjuntos de caracteres sem significado atribuído, o que limita sua compreensão por máquinas, impactando na eficácia da recuperação da informação.

Sem desconsiderar sua relevância, o Repositório da Produção USP não tem seus dados abertos ao público em geral. Embora passíveis de visualização, apenas a

comunidade universitária pode fazer o download de conjuntos de dados do repositório em arquivos em formato CSV (*Comma-Separated Values*), RIS (*Research Information Systems*) ou Bibtex. Além disso, seus filtros não são customizáveis, limitando possibilidades de visualização e busca.

Tais aspectos evidenciam desafios impostos por padrões que não operam com a conexão de dados e limitam a sua livre utilização (Montgomery *et al.* 2020), apontando para a relevância de que bibliotecas explorem o potencial dos dados conectados abertos para organizar e comunicar informações.

### 3 WIKIDATA: A PRÁTICA DOS DADOS CONECTADOS ABERTOS

Os dados conectados são a principal aplicação da Web semântica; possibilitam a publicação de dados na Web com significados definidos, de modo que possam ser lidos por máquinas e conectados a conjuntos de dados externos (Bizer; Heath; Berners-Lee, 2009). Representando este contexto, surge em 2012 a Wikidata, uma base de dados conectados abertos disponibilizados sob a licença livre CC0 1.0 Universal e desenvolvida com o software Wikibase (Wikidata, 2023a). Ou seja, além de seguir os padrões para dados conectados, a Wikidata também estabelece sua abertura.

Inscrita no movimento Wikimedia, pautado no princípio do conhecimento aberto, a Wikidata é um projeto sociotécnico desenvolvido a partir de dinâmicas colaborativas entre uma comunidade global que vem se expandindo (Shenoy *et al.*, 2022) e inclui profissionais da informação, interessados em seu potencial para organizar e comunicar conteúdos (ARL, 2019).

Conquanto seu caráter colaborativo e aberto, com livre edição dos registros, possa configurar entrave às práticas de bibliotecas, que prezam pelo controle da qualidade e confiabilidade dos seus registros, o histórico de edições do item e o sistema de alertas por email possibilitam o monitoramento das alterações realizadas. Além disso, a Wikidata dispõe de regras e de uma comunidade voltada ao seu cumprimento, aspectos que contribuem para que, embora todos possam editá-la, não seja possível sustentar ações que descumpram suas normativas.

De caráter multilíngue, a Wikidata possibilita que uma mesma entidade, concreta ou abstrata, seja representada em diferentes idiomas, pois utiliza o Identificador Uniforme de Recursos (em inglês, *Uniform Resource Identifier* - URI). Além das contribuições humanas, dispõe de robôs – ferramentas de edição que eliminam a necessidade de tomada de

decisão humana – e é composta por um conjunto de declarações factuais, ampliado pela comunidade de editores (Shenoy *et al.*, 2022; Wikidata, 2024a, 2023b).

A estruturação dos dados se dá a partir da relação entre itens, propriedades e valores. Suas declarações podem ser enriquecidas com qualificadores para prover referências ou informações adicionais que contextualizam o dado (Mora-Cantallops; Sánchez-Alonso; García-Barriocanal, 2019). Seguindo os princípios dos dados conectados para publicação de dados na Web, a Wikidata tem como práticas:

- a) utilizar URIs para nomear as entidades e utilizar HTTP URIs para que tais entidades possam ser pesquisadas e localizadas: URIs configuram meio para identificar qualquer entidade – abstrata ou concreta. Os URIs via Protocolo de Transferência de Hipertexto (em inglês, *Hypertext Transfer Protocol* - HTTP) são mecanismos que possibilitam atribuir a uma entidade um identificador único na Web (Bizer; Heath; Berners-Lee, 2009);
- b) fornecer informações por meio dos padrões RDF (em português, Estrutura de Descrição de Recursos; em inglês, *Resource Description Framework*) e SPARQL (em português, Protocolo SPARQL e Linguagem de Consulta RDF; em inglês, *SPARQL Protocol and RDF Query Language*) quando alguém procura um URI: O RDF é um formato padrão para representação de dados na Web que utiliza triplas, uma estrutura composta por assunto, predicado e objeto. Nela, o predicado é responsável por estabelecer a relação entre assunto e objeto. Na Wikidata a estrutura assunto-predicado-objeto se dá pela relação item-propriedade-valor, expressa em uma declaração. As propriedades são equivalentes aos predicados e tem como função estabelecer a relação entre o item (assunto) e determinado(s) valor(es) (objetos), sendo possível criar várias declarações sobre um mesmo item. A partir de uma série de relacionamentos é possível estabelecer a interligação exponencial dos dados, viabilizando a recuperação de resultados face a buscas bastante específicas. É possível que o item (assunto) e o valor (objeto) de uma tripla sejam URIs, cada qual identificando um recurso específico ou, então, que correspondam a um URI e a uma *string* – sequência de caracteres não estruturados (ARL, 2019; Bizer; Heath; Berners-Lee, 2009) (ver Figura 1). A SPARQL é uma linguagem de consulta RDF para conjuntos de triplas. Por meio dela é possível realizar buscas combinando os dados de formas variadas (Lemus-Rojas; Pintscher, 2018). A Wikidata Query Service emprega

a linguagem SPARQL para realizar consultas na Wikidata a partir da combinação de triplas. O mecanismo dispõe do recurso Wikidata query builder, uma interface facilitadora da consulta para pessoas com pouca ou nenhuma experiência em SPARQL. Assim, ao catalogar itens na Wikidata podemos obter resultados para buscas como “Quantos docentes do sexo ou gênero feminino a USP possui?”;

- c) incluir links para outros URIs para viabilizar a descoberta de mais conteúdos: as triplas em formato RDF podem vincular itens de diferentes conjuntos de dados. Os links RDF “assumem a forma de triplas RDF em que o assunto da tripla é uma referência de URI no domínio de um conjunto de dados, enquanto o objeto da tripla é uma referência de URI em outro conjunto de dados” (Bizer; Heath; Berners-Lee, 2009, tradução nossa) (ver Figura 1).

Figura 1 - Assunto (item), propriedades e valores na Wikidata

The image shows a Wikidata item page for 'Ana Mae Barbosa' (Q11679823). The page is divided into several sections:

- Assunto (Item):** 'Ana Mae Barbosa' with the identifier '(Q11679823) ... URI'. Below this, it says 'professora universitária e pesquisadora ativa no Brasil' and 'Ana Mae Tavares Bastos Barbosa | Anna Mae Barbosa'.
- Propriedades (Properties):** A list of properties including 'sexo ou gênero', 'data de nascimento', 'ORCID', and 'empregador'.
- Valores (Values):** A list of values corresponding to the properties: 'feminino ... URI' for 'sexo ou gênero', '17 julho 1936 ... STRING' for 'data de nascimento', '0000-0002-4966-2043 ... STRING' for 'ORCID', and 'Universidade de São Paulo ... URI' for 'empregador'.
- Qualificador (Qualifier):** A section showing 'data inicial' with the value '1982'.

Fonte: Elaborado pela autora (2024).

Na Figura 1, elaborada a partir da Wikidata, temos uma declaração em que ORCID estabelece a relação entre um URI na Wikidata, representando uma pessoa, e uma sequência numérica que corresponde ao registro desta pessoa na base de dados ORCID. Além de viabilizar a descoberta de mais conteúdos, a conexão entre conjuntos de dados contribui para a melhoria das informações sobre um item, aumentando a qualidade do



registro e sua confiabilidade, ao possibilitar a checagem de informações em conjuntos de dados externos à Wikidata (ARL, 2019).

Diante disso, a Wikidata foi escolhida para a catalogação de artigos científicos de docentes da USP em dados conectados abertos, opção reforçada por dispensar instalação e espaço para armazenamento de dados, além de dispor de documentação para o aprendizado. A possibilidade de criar metadados (propriedades) configurou aspecto de interesse, ao contribuir para uma catalogação em diálogo com as necessidades da biblioteca. Embora a Wikidata seja construída no software Wikibase (Wikidata, 2023a), disponível para download e instalação por interessados em implementar suas próprias bases de dados, essa alternativa não foi adotada por demandar suporte tecnológico e espaço de armazenamento de dados. Ademais, a decisão de utilizar a Wikidata incluiu a compreensão de que este seria um passo essencial para o diálogo com sua comunidade de práticas, favorecendo o aprendizado da catalogação em dados conectados abertos.

#### **4 IMPORTAÇÃO DE REGISTROS DO REPOSITÓRIO DA PRODUÇÃO USP PARA A WIKIDATA**

O projeto de importação de registros de artigos científicos do Repositório da Produção USP para a Wikidata integra a iniciativa GLAM Wiki Bibliotecas da USP, iniciada em 2020 com a participação das bibliotecas da ECA, Instituto de Matemática e Estatística (IME) e Faculdade de Arquitetura e Urbanismo (FAU) da USP, com o propósito de desenvolver ações com projetos Wikimedia, dentre eles a Wikidata. Com atuação colaborativa e autônoma, as bibliotecas participantes do GLAM Bibliotecas da USP têm como um de seus objetivos "comunicar informações e conhecimentos produzidos pela comunidade acadêmica da USP nos diferentes projetos Wikimedia, contribuindo com o desenvolvimento do conhecimento aberto" (Wikipédia, 2024). A iniciativa conta com o suporte técnico-especializado do Wiki Movimento Brasil (WMB), grupo de usuários brasileiro afiliado à Fundação Wikimedia, formalizado em parceria institucional com as unidades de ensino envolvidas na iniciativa.

O contato com referenciais indicativos de possibilidades de atuação de bibliotecas com a Wikidata foi essencial ao desenvolvimento da experiência em causa neste estudo. Dentre eles, destacam-se o trabalho com perfis acadêmicos realizado na Indiana University-Purdue University Indianapolis (Lemus-Rojas; Odell, 2018) e o protótipo com Wikibase, desenvolvido pela Online Computer Library Center (OCLC) para que bibliotecários de certas

instituições estadunidenses experimentassem uso e potencialidades dos dados conectados (Godby *et al.*, 2019). Destaca-se, também, a apresentação de Jens Ohlig (2020) na conferência LD4 de 2020 abordando a temática dos dados conectados para bibliotecas.

Tais referenciais catalisaram a iniciativa de importação de registros de artigos científicos para a Wikidata, que será abordada a partir da experiência da biblioteca da ECA com registros relativos aos docentes desta unidade de ensino. Os aspectos operacionais desta importação serão sistematizados com base nos eixos: recorte da ação; coleta de dados e correspondência de metadados; estratégias para importação de dados para a Wikidata; visualização de dados.<sup>1</sup>

#### 4.1 Recorte da Ação

O Repositório da Produção USP concentra o registro da diversificada produção intelectual da universidade. Neste projeto, os artigos científicos foram objeto da ação pois seria possível visualizar os dados por meio da ferramenta Scholia, que cria e exibe perfis acadêmicos referentes a “autor, obra, instituição, local, evento, projeto e prêmio”, a partir de consultas na Wikidata Query Service (Scholia, [202-?]a).

Iniciada em 2020, a princípio, a ação compreendeu artigos publicados em 2017 e registrados no repositório. O recorte visou garantir que expressiva quantidade de artigos do período selecionado estivesse cadastrada no repositório, pois o fluxo de catalogação nem sempre acompanha o fluxo de publicação dos artigos. Ademais, seria viável trabalhar com dados de artigos publicados ao longo de um ano, recorte que totalizou, à época da ação, 183 artigos de docentes da ECA. Os dados dos artigos foram extraídos do repositório em planilha em formato CSV, ordenada conforme os campos MARC21 utilizados na catalogação.

#### 4.2 Coleta de Dados e Correspondência de Metadados

Criar um item na Wikidata implica interligar a trama de elementos que o representa, atribuindo significado aos dados. Portanto, antes de cadastrar os artigos científicos na Wikidata, algumas entidades deveriam ser registradas e foi necessário definir critérios

---

<sup>1</sup> Cabe ressaltar que diferentes etapas desta atividade foram planejadas em colaboração com bibliotecárias das demais unidades envolvidas e cada equipe desenvolveu ações em seu contexto, conforme condições de possibilidade.



quanto ao nível de detalhamento dos registros. Na medida em que o projeto buscou dar visibilidade à produção intelectual de docentes da ECA, seria essencial explicitar o vínculo entre autores e unidade de ensino, assim, optou-se por cadastrar unidade de ensino, seus departamentos e programas de pós-graduação. Além disso, a revista em que o artigo foi publicado deveria ser cadastrada na Wikidata, caso seu registro não constasse na base. Face a essa extensa trama, foram adotados os seguintes critérios para estabelecer o nível de detalhamento dos registros:

- Unidade de ensino: cadastro detalhado. Dados coletados no site da instituição;
- Departamento de ensino: cadastro detalhado. Os dados foram coletados no site da instituição;
- Docente: cadastro detalhado de docentes ativos e aposentados ou mortos, contratados em regime de dedicação exclusiva. Coleta de dados realizada em recursos administrativos internos e em plataformas de identificação de autor (Lattes, ORCID, FAPESP ID e Enciclopédia Itaú Cultural para docentes das Artes);
- Revistas de publicação dos artigos científicos de 2017 de docentes da ECA: cadastro sintético, apenas para viabilizar o registro dos artigos. Coleta de dados no Portal do ISSN.

Os dados foram organizados em planilhas Excel, com colunas nomeadas com as propriedades da Wikidata que iriam estruturá-los, conforme discriminado a seguir:

Quadro 1 - Entidades relativas à produção científica e propriedades na Wikidata

Entidade	Propriedades Wikidata (metadados)	Exemplo na Wikidata
Unidade de ensino	Instância de (P31); Parte de (P361); Imagem (P18); Data de criação ou fundação (P571); País (P17); Localizado na unidade administrativa (P131); Coordenadas geográficas (P625); Organização-mãe (P749); Subsidiárias (P355); Idioma usado (P2936); Página web oficial (P856); Categoria no commons (P373); Identificador VIAF (P214); Identificador da FAPESP para instituição (P4597); Identificador de canal do Youtube (P2397); Na lista de interesse do Wikiprojeto (5008)	<a href="https://www.wikidata.org/wiki/Q10274592">https://www.wikidata.org/wiki/Q10274592</a>
Departamento de ensino	Instância de (P31); Parte de (P361); Nome curto (P1813); Campo de trabalho (P101); País (P17); Localizado na unidade administrativa (P131); Coordenadas geográficas (P625); Organização-mãe	<a href="https://www.wikidata.org/wiki/Q101114337">https://www.wikidata.org/wiki/Q101114337</a>

	(P749); Página web oficial (P856); Na lista de interesse do Wikiprojeto (5008)	
Programa de pós-graduação	Instância de (P31); Parte de (P361); Nome curto (P1813); Campo de trabalho (P101); País (P17); Localizado na unidade administrativa (P131); Coordenadas geográficas (P625); Organização-mãe (P749); Página web oficial (P856); Na lista de interesse do Wikiprojeto (5008)	<a href="https://www.wikidata.org/wiki/Q106660948">https://www.wikidata.org/wiki/Q106660948</a>
Docente	Instância de (P31); Imagem (P18); Sexo ou gênero (P21); País de cidadania (P27); Nome no idioma nativo (P1559); Línguas faladas ou escritas (P1412); Ocupação (P106); Campo de trabalho (P101); Empregador (P108); Alma mater (P69); Orientador(a) de doutorado (P184); Residência (P551); Afiliação (P1416); Número na Plataforma Lattes (P1007); Identificador ORCID (P946); Identificador ResearcherID (1053); Identificador de autor Scopus (P1153); Identificador de pessoa no Repositório da Produção da USP (P9612); Identificador no Google acadêmico (P1960); Identificador de pesquisador da FAPESP (P4598); Itaú Cultural ID (P4399); Na lista de interesse do Wikiprojeto (5008)	<a href="https://www.wikidata.org/wiki/Q101235609">https://www.wikidata.org/wiki/Q101235609</a>
Revista	Instância de (P31); ISSN (P236)	<a href="https://www.wikidata.org/wiki/Q50814376">https://www.wikidata.org/wiki/Q50814376</a>

Fonte: Elaborado pela autora (2024).

A profundidade da coleta de dados para efetuar a catalogação considerou a disponibilidade de funcionários da biblioteca. Além disso, os itens criados podem ter sido complementados por outros editores e, quando uma entidade já existia na Wikidata, seu registro foi editado preservando-se os dados existentes.

No registro dos docentes, foi inserida a propriedade Identificador de Pessoa no Repositório da Produção da USP (P9612), criada para vincular o item correspondente ao docente na Wikidata à página do repositório em que constam os registros de sua produção intelectual. Com isso, objetivou-se contribuir para ampliar o alcance do Repositório da Produção da USP a partir do ecossistema Wikimedia. Além disso, a propriedade Na Lista de Interesse do Wikiprojeto (P5008), acrescida do valor GLAM Bibliotecas da USP (Q97013571), foi utilizada para identificar os registros editados.

Em seguida, teve início o trabalho com os dados dos artigos. A partir de planilhas em formato CSV, extraídas do repositório com os dados dos artigos organizados conforme os campos MARC21, foi feita a correspondência entre metadados, conforme quadro a seguir:

Quadro 2 - Correspondência entre metadados MARC21 e Wikidata para artigos científicos

Metadados MARC21	Metadados Wikidata	Decisões
Tipo de Material (945)	Instância de (P31)	--
DOI (024)	DOI (P356)	--
Título principal (245 - subcampos a, b)	Rótulo; Título (P1476); Subtítulo (P1680)	Nessas propriedades foi utilizado o qualificador de idioma. Para identificar o idioma das publicações, as listas de artigos extraídas do Repositório da Produção USP foram geradas a partir de filtros de idioma e trabalhadas isoladamente.
Autoria (100; 700 - subcampo a)	Autor(a) (P50); Nome de autor (P2093)	Na propriedade P2093 é possível inserir uma cadeia de caracteres; ela é utilizada quando o item da Wikidata para a autoria não existir ou for desconhecido. Autores que não eram docentes da ECA foram inseridos nesta propriedade, opção adotada para viabilizar o trabalho, pois o campo Autor (P50) demanda que a pessoa tenha registro na Wikidata e cadastrar todas as autorias não era viável, por limitações de disponibilidade de funcionários para a tarefa. Em ambas propriedades, utilizou-se o qualificador Ordem na série (P1545) para indicar a ordem de autoria do artigo.
Local de publicação (260 - subcampo a)	--	O dado referente ao local da revista de publicação do artigo foi usado para compor uma planilha de dados criada para o cadastro das revistas na Wikidata.
Data de publicação (260 - subcampo c)	Data de publicação (P577)	Optou-se por indicar somente o ano da publicação, sem a especificação do mês, dado que poderia ser extraído a partir do campo 773, subcampo h.
Assunto (6XX)	Tema principal (P921)	Optou-se por cadastrar no mínimo 1 e no máximo 3 assuntos, priorizando aqueles que já existiam na Wikidata. A opção está relacionada ao tempo despendido com a tarefa. A inclusão de assunto foi tida como essencial, pois possibilita filtros de busca a partir deste dado.
Entrada analítica - fonte da publicação (773 - subcampo t)	Publicado em (P1433)	--

Entrada analítica - descrição física (773 - subcampo h)	Página (P304); Volume (P478); Número (P433).	–
Entrada analítica - ISSN (773 - subcampo x)	--	O dado referente ao ISSN da revista foi usado para compor a planilha de dados utilizada para efetuar o cadastro das revistas na Wikidata.
Acesso e localização eletrônica (856 - subcampo u)	Obra completa disponível em (P953)	–
Código de idioma (041 - subcampo a)	Idioma da obra ou do nome (P407)	Para identificar o idioma das publicações, as listas de artigos extraídas do Repositório da Produção USP foram geradas a partir de filtros de idioma e trabalhadas isoladamente.
Número de sistema	Repositório da Produção USP ID (P10492)	O número de sistema é um identificador único do recurso no catálogo das bibliotecas da USP e foi utilizado para interligar o item da produção intelectual na Wikidata com seu registro no Repositório da Produção USP
--	Na lista de interesse do wikiprojeto (P5008)	Propriedade usada para identificar os registros editados no escopo do projeto GLAM Bibliotecas da USP e viabilizar filtros a partir deste valor.

Fonte: Elaborado pela autora (2024).

Para definir as propriedades utilizadas foram consultados registros de entidades similares na Wikidata e alguns Wikiprojetos<sup>2</sup>, dentre eles: Wikiprojeto Periodicals (Wikidata, 2024b) e Wikiprojeto IUPUI University Library (Wikidata, 2023c). Além disso, o recurso *Recoin* foi habilitado – via opção Preferências – para auxiliar a identificar propriedades relevantes que faltam em um registro e a Lista de propriedades foi consultada para identificar metadados pertinentes (Wikidata, 2024c). A consulta à página de discussão das propriedades para conhecer a justificativa para sua criação, a discussão da comunidade, os tipos de dados que estruturam e exemplos de uso foi relevante para definir quais propriedades seriam utilizadas.<sup>3</sup>

<sup>2</sup> Um WikiProjeto é formado por um grupo de pessoas que trabalham colaborativamente para aprimorar a Wikidata. Frequentemente, os grupos concentram-se em tópicos de uma área específica. Na página a seguir é possível consultar os Wikiprojetos: <https://www.wikidata.org/wiki/Wikidata:WikiProjects>. Acesso em: 29 mar. 2024.

<sup>3</sup> As páginas de discussão podem ser acessadas a partir da página da propriedade, clicando na opção Discussão, no topo da página. Como exemplo: [https://www.wikidata.org/wiki/Property\\_talk:P10492](https://www.wikidata.org/wiki/Property_talk:P10492). Acesso em: 29 mar. 2024.

A correspondência de metadados ocorreu num percurso marcado por testes de catalogação manual na Wikidata, a partir dos quais foi possível verificar sua adequação e testar filtros de busca na Wikidata Query Service, o que permitiu efetuar ajustes antes do início da importação automatizada de grande quantidade de dados.

### 4.3 Estratégias para Importação de Dados para a Wikidata

Após o cadastro manual de unidade de ensino, departamentos de ensino e programas de pós-graduação, recorreu-se ao QuickStatement (QS) e ao OpenRefine, ferramentas que possibilitam a importação de grande quantidade de dados para a Wikidata. Os dados de docentes, revistas e artigos foram importados em lote, após testes de edição manual para verificar a consistência da estruturação dos dados.

Para cadastrar docentes, foi utilizado o QuickStatement, ferramenta de importação de dados baseada num conjunto de comandos de texto para realizar edição de itens da Wikidata. O QS possibilita inserir e excluir declarações, rótulos, descrições, dados de "também conhecido como" e qualificadores nas declarações (Ajuda, 2024). A opção metodológica ocorreu em meio à parceria com o WMB, que prestou suporte ao desenvolvimento da ação e executou a importação do lote de dados de docentes via QS, além de ministrar oficina (Stocco, 2021) para ensinar o uso da ferramenta às equipes das bibliotecas que integram o GLAM Wiki Bibliotecas da USP. Ademais, a página do QS dispõe de um menu de ajuda que contribuiu à realização da importação dos demais lotes de dados pela equipe da biblioteca.

Importar dados para a Wikidata implica reconciliá-los, processo de transposição do dado para a linguagem da Wikidata também conhecido como wikidatificação. Inicialmente, os dados dos docentes foram reconciliados em planilhas do Google Sheets e depois importados via QS. No Google Sheets, foi testado o uso da função WikidataQid, integrante do complemento Wikipedia Tools, para realizar a reconciliação dos dados. Contudo, o recurso foi descartado devido à inviabilidade de tratar homônimos a partir da planilha. Como exemplo, para criar um item sobre um docente da área de Arte especificando seu campo de trabalho, é preciso fazer tal declaração na linguagem Wikidata. Ao utilizar o complemento para reconciliar a palavra Arte com a Wikidata, este poderá retornar Q735, código relativo à Arte descrita como um processo criativo, ou, Q36901508, relativo ao nome de família Arte. Logo, optou-se por pesquisar os códigos diretamente na Wikidata e adicioná-los à planilha.

O trabalho com o QS expôs a necessidade de checar previamente na Wikidata se as entidades possuíam cadastro, pois a ferramenta não identifica duplicidades ou bloqueia sua criação. Na primeira importação de dados de docentes, alguns itens foram duplicados; em decorrência, foi preciso usar o recurso "Fundir com" – disponível na página de cada item na Wikidata –, que possibilita a fusão dos itens duplicados. A duplicidade pode ser verificada a partir de consultas via Wikidata Query Service ou diretamente em consultas na Wikidata.

Abaixo, a Figura 2 ilustra o uso do QuickStatement para inserir propriedades e respectivos valores em um item existente na Wikidata, referente ao docente José Marques de Melo.

Figura 2 – Exemplo de uso do QuickStatement

QuickStatements língua portuguesa Novo lote Últimos lotes Chat Git ajuda

Criar novo lote de comandos para Wikidata como nome do lote

Q101242598	P1559	pt-br: "José Marques de Melo"
Q101242598	P9612	"33143"
Q101242598	P103	Q750553
Q101242598	P1412	Q750553

Importar comandos da Versão 1 Importe comandos CSV

Fonte: Elaborado pela autora (2024).

A wikidatificação e importação via QS inscreveu-se num quadro em que a bibliotecária responsável pela ação na biblioteca da ECA não tinha domínio avançado de Excel e os desafios para lidar com a grande quantidade de declarações sobre uma mesma entidade foram consideráveis, levando a investigar outras estratégias para importação de dados.

Assim, após concluir o trabalho com os dados de docentes, foi feita a importação dos dados dos artigos com o OpenRefine, um recurso aberto para trabalhar com grandes quantidades de dados que possibilita reconciliá-los e importá-los para a Wikidata. Além disso, permite gerar dados em formato QS (OpenRefine, [202-?]).



A consulta a recursos de aprendizagem sobre o OpenRefine (Rossenova; Sohmen, 2022; Santarem Segundo, 2020; Wikidata, 2023d; Wikidata, 2024d) evidenciou aspectos centrais à decisão de utilizá-lo. De um lado, as funcionalidades facilitadoras da limpeza e estruturação de dados, de outro, as possibilidades de reconciliar os dados com a Wikidata e criar itens a partir do OpenRefine.

Antes de iniciar o trabalho com o OpenRefine, optou-se por criar o registro de cada artigo via QuickStatement, apenas com o rótulo em português do Brasil e uma declaração com a propriedade Instância de (P31) acrescida do valor Artigo científico (Q13442814). Criar um registro via QS, apenas com o rótulo e uma declaração, era tarefa de execução simples para a bibliotecária, pois implicou estruturar uma tabela com uma única propriedade e valor combinados ao título de cada artigo. Além disso, não seria preciso reconciliar nenhum dado. A decisão também recaiu na agilidade conferida ao processo de importação quando o item já consta na Wikidata, pois, na experiência em foco, criar vários registros a partir do OpenRefine, mostrou-se um processo moroso, em termos de resposta da ferramenta aos comandos.

Após essa etapa, teve início o trabalho no OpenRefine. As planilhas em formato CSV extraídas do Repositório da Produção USP foram inseridas no OpenRefine para iniciar os processos de limpeza, reconciliação e estruturação dos dados. Por meio do recurso de reconciliação, foi checado se as revistas em que os artigos foram publicados existiam na Wikidata e, em caso negativo, foi feito um cadastro sintético via QS com o rótulo, a descrição “publicação periódica”, a propriedade Instância de (P31) com o valor Publicação periódica (Q1002697) e a propriedade ISSN-L (P7363) preenchida com o ISSN-L da obra. De modo geral, as funcionalidades do OpenRefine para separar e ordenar conteúdos, unir colunas, fazer facetas e substituir conteúdos em lote facilitaram a limpeza dos dados.

Na sequência, procedeu-se à reconciliação dos dados, exceto para números – paginação, volume, número e ano da obra – e URLs, pois são sequências de caracteres. Somente os autores que eram docentes da ECA foram reconciliados, os demais foram cadastrados na propriedade Nome de autor (P2093), com as autorias tratadas como cadeia de caracteres. A escolha se deu pela inviabilidade de cadastrar todas as autorias, ação que demandaria rigorosa coleta de dados, sobretudo considerando a ocorrência de homônimos. Ademais, o risco de homônimos levou a desconsiderar a possibilidade de reconciliar tais nomes com dados já existentes na Wikidata.

No processo com o OpenRefine a reconciliação foi a etapa que demandou maior esforço, pois a existência de termos iguais ou similares demandou checagem e validação

do dado reconciliado. A ferramenta mostrou-se eficaz nesta fase, pois sua interface apresenta as diferentes ocorrências do termo para que o editor selecione aquela que julgar adequada, além de redirecionar para a Wikidata. Além disso, é possível reconciliar cada coluna da planilha com uma entidade de tipologia que podemos especificar. Assim, a reconciliação será restrita à tipologia indicada, aumentando as chances de sucesso. O maior trabalho na reconciliação foi com os assuntos dos artigos, dada a inviabilidade de restringi-los a uma tipologia para fazer o processo. Em muitos casos, foi preciso verificar as sugestões fornecidas pela ferramenta de reconciliação e selecionar a entidade adequada. A Figura 3, a seguir, traz exemplo de entidades reconciliadas (em cor azul), termos não reconciliados (em cor preta), sugestões para reconciliação (entidades listadas em cor azul, com a opção de selecionar a adequada).

Figura 3 - Processo de wikidatificação no OpenRefine

Assuntos 1	Assuntos 2	Assuntos 3
TRANSMISSÃO AO VIVO <input checked="" type="checkbox"/> em direto (72) <input checked="" type="checkbox"/> Corpo a corpo (38) <input checked="" type="checkbox"/> DLive (22) <input checked="" type="checkbox"/> Azubu (8) <input checked="" type="checkbox"/> Create new item Search for match	Plataforma digital (infraestrutura) Choose new match	consumo Choose new match
teatro Choose new match	happening Choose new match	RELAÇÕES ÉTNICAS E RACIAIS <input checked="" type="checkbox"/> sociologia de raça e relações étnicas (90) <input checked="" type="checkbox"/> Create new item Search for match
BIG DATA	PRODUÇÃO CIENTÍFICA	REPRESENTAÇÃO DA INFORMAÇÃO
INDEPENDÊNCIA DO BRASIL	SOBERANIA NACIONAL	MODERNIZAÇÃO

Fonte: Elaborado pela autora no OpenRefine (2023)

Após a reconciliação, se deu a estruturação dos dados por meio da criação de um esquema Wikibase, no qual foram elencadas as propriedades para estruturá-los. O esquema possibilita inserir referências e qualificadores nas declarações, além disso, sinaliza erros para que sejam corrigidos antes da importação dos dados, última etapa do processo realizada com o login de usuário da Wikidata a partir do OpenRefine.

De modo geral, o processo com o OpenRefine revelou a pertinência de realizar testes de importação com uma pequena quantidade de artigos para verificar a adequação da estrutura dos dados e de executar consultas via Wikidata Query Service, para checar a qualidade da recuperação da informação a partir da estrutura proposta.

Além disso, neste projeto estava em causa o uso da ferramenta Scholia para visualizar os dados, portanto, foi indispensável verificar como eram apresentados neste

recurso para avaliar a adequação dos metadados propostos, previamente à importação de grande quantidade de dados.

#### 4.4 Visualização de Dados: a Ferramenta Scholia

As possibilidades oferecidas pelo Scholia para visualização de dados contribuíram para a decisão de catalogar a produção científica na Wikidata, pois consistia em um recurso pronto para experimentar as possibilidades que os dados conectados abertos oferecem para comunicar a produção intelectual da universidade.

O foco recaiu sobre a possibilidade de exibição do perfil institucional da ECA a partir do qual é possível visualizar os seguintes dados: empregados e afiliados; gráfico de co-autoria; gráfico de vinculação entre orientador de doutorado e orientando; gráfico de temas sobre os quais pessoas vinculadas à instituição publicaram; publicações mais recentes, categoria que engloba o limite de 200 publicações; componentes utilizados em projetos descritos nas publicações; produção científica, quantitativamente expressa em número de páginas publicadas por autoria; citações; prêmios; distribuição por gênero (Escola ..., 2024).

Esses eixos compõem o perfil institucional, independentemente de possuírem dados vinculados, pois não é possível alterar a interface do Scholia. Embora o recurso seja aberto e esteja disponível no Github (Scholia, [202-?]b) para download do código fonte e dos arquivos para implementação e adaptação, este caminho não foi adotado por falta de conhecimento técnico-especializado da bibliotecária para executar a tarefa.

Considerando a disponibilidade de tempo e equipe, optou-se por não cadastrar os dados de total de páginas dos artigos, pois seria necessário realizar o cálculo, tampouco os componentes utilizados nos projetos descritos nos artigos – aspecto que demandaria a leitura dos textos. A título de experimentação, certas premiações foram vinculadas ao nome de alguns docentes e a relação orientador de doutorado e orientando foi estabelecida entre pessoas que são docentes na ECA. Além disso, apesar de haver interesse, não foi feito um trabalho em torno dos dados de citações.

O Scholia realçou o potencial da catalogação em formato aberto e semântico. Além do desenvolvimento de bases que possibilitam buscas bastante específicas, num contexto de significação dos dados, o recurso demonstrou que os dados conectados abertos configuram caminho à promoção de diferentes formas de visualização, contribuindo para comunicar coleções de bibliotecas e instituições afins na Web de maneiras variadas.

Por outro lado, alguns limites foram identificados no uso do Scholia. Dentre eles, sua interface exclusivamente em inglês e a impossibilidade de alterar as consultas pré-definidas do Wikidata Query Service, que podem desconsiderar aspectos relevantes aos diversos contextos institucionais. Como exemplo, o Scholia não apresenta dados da produção artística de docentes da ECA cadastrada na Wikidata. Diante disso, ficou evidente a importância de que a continuidade do projeto comporte avanços em direção à implementação de interfaces delimitadas pela equipe da biblioteca, em diálogo com questões inerentes ao seu contexto institucional.

## 5 RESULTADOS E DISCUSSÃO

A importação dos registros da produção científica de docentes da ECA para a Wikidata exigiu a transição do formato MARC21 para dados conectados abertos, ampliando o escopo da catalogação para incluir o cadastro de diversas entidades. Além dos artigos, seus autores e as revistas em que foram publicados deveriam estar cadastrados em dados conectados abertos. Para explicitar as tramas de produção do conhecimento, a ECA, seus respectivos departamentos e programas de pós-graduação foram cadastrados e estabelecida a relação com os docentes. Desse modo, foi possível explicitar os vínculos institucionais e sobressaiu a necessidade de levantamento e organização de dados que extrapolam aqueles existentes no catálogo da biblioteca, já que muitos não existiam na Wikidata ou não estavam conectados.

Para viabilizar a ação, face ao grande volume de dados, foi essencial estabelecer o nível de detalhamento dos registros para as diferentes entidades. Assim, a clareza quanto ao objetivo do projeto – comunicar a produção científica de docentes da ECA, circunscrevendo-a na esfera institucional – constituiu aspecto fundamental à catalogação na Wikidata. Como uma entidade pode ser descrita de diferentes maneiras, foi preciso estabelecer modelos com metadados para cada tipo de entidade, pertinentes face ao objetivo e viáveis para execução. Assim, compreender a estruturação dos dados na Wikidata foi indispensável.

Iniciada com artigos publicados em 2017 e cadastrados no Repositório da Produção USP, a ação expandiu-se para comportar, até o momento da escrita deste texto, artigos de docentes da ECA do período de 2017-2022, existentes no repositório no momento da importação. Como resultado, foram editados registros de 6 programas de pós-graduação; 8 departamentos de ensino; 345 docentes; 936 artigos e 22 revistas publicadas pela ECA.

Além de refletir as condições de efetuar a ação, a expansão decorreu da viabilidade de trabalhar com grandes quantidades de dados na Wikidata, por meio de recursos que conferem celeridade ao processo, como o QuickStatement e o OpenRefine. Ademais, sobressaiu o potencial da linguagem SPARQL para buscas e visualização de dados na Web em interfaces pré-definidas, como o Scholia.

Por sua vez, ressaiu a necessidade de avançar para outras formas de visualização dos dados, desenvolvendo interfaces em diálogo com questões específicas ao contexto da ECA. Nesse sentido, o domínio Wikimedia *Toolforge* poderia ser investigado, pois, é “ambiente de hospedagem para desenvolvedores que trabalham em serviços que agregam valor ao movimento Wikimedia” (Wikimedia, [202-?], tradução nossa), que viabiliza criar e hospedar ferramentas. Embora seja um facilitador para a ação, seu uso requer domínio de programação de sistemas, aspecto indicativo da importância de interlocuções entre as equipes de bibliotecas e de setores de tecnologia para estabelecer e planejar projetos em torno da produção científica, dimensão inexistente na iniciativa GLAM-Wiki em foco. Diante disso, a parceria com o Wiki Movimento Brasil é estratégica, pois o grupo contribuiu para o desenvolvimento da ação e tem potencial para colaborar com o crescimento do projeto, auxiliando no planejamento e execução de ações para a visualização de dados.

Destaca-se que alguns elementos aprendidos ao longo do percurso não se manifestam na totalidade dos registros editados, por desconhecimento no momento de realizar a tarefa. É o caso da inserção de referências como qualificadores para validar os dados. Embora a Wikidata não exija inserção de referências em todas as declarações, sempre que possível é pertinente inseri-las para qualificar o dado, atribuindo confiabilidade à informação, sobretudo num contexto em que dados divergentes sobre um mesmo item podem coexistir no mesmo registro.

O caráter colaborativo da Wikidata facilitou o desenvolvimento da ação, pois a comunidade de práticas gera documentação essencial ao aprendizado e mantém instâncias de interlocução fundamentais para esclarecer dúvidas. Todavia, a comunicação nessa comunidade global ocorre, majoritariamente, no idioma inglês, o que configurou empecilho em alguns momentos, sobretudo considerando que estava em causa dialogar em vocabulário técnico.

A abertura dos registros para edição por outras pessoas não configurou aspecto problemático; foi utilizado o recurso de monitoramento que envia email em caso de novas edições nos itens monitorados e, por meio do histórico é possível identificá-las e revertê-

las. Nos registros editados pela comunidade, houve seu aprimoramento, portanto, as alterações não foram desfeitas.

O projeto em pauta consistiu na replicação de um conjunto de dados na Wikidata e sua integração ao Repositório da Produção USP, por meio da criação de metadados. Estava em causa explorar o potencial da Wikidata para aumentar o alcance da comunicação da produção científica, logo, não foi objetivo substituir seu uso pela Wikidata, pois isso significaria a perda de controle sobre a preservação dos dados aspecto que, de antemão, anulou o interesse em explorar tal possibilidade.

A opção metodológica de utilizar exclusivamente uma base de dados conectados abertos levaria a explorar a Wikibase, software utilizado pela Wikidata. É possível fazer a instalação da Wikibase num servidor local e desenvolvê-la, mantendo o controle sobre edição e preservação dos dados. Contudo, tal opção demandaria uma equipe de tecnologia para implementar a ação em parceria com a equipe da biblioteca, algo não vislumbrado no horizonte de médio prazo. Ademais, em busca de apoio no contexto institucional, é relevante inicialmente avaliar o impacto da ação com a Wikidata em busca de justificar investimentos e equipe necessários para implementações futuras.

Em termos gerais, importar registros para a Wikidata mostrou-se uma tarefa viável, todavia, até o momento, não foi mensurado o alcance dos registros criados na Wikidata, atividade que demanda novos aprendizados. Ou seja, é preciso avaliar e gerar dados sobre o potencial da Wikidata para aumentar a descoberta da produção científica da ECA na Web, inclusive como forma de justificar a manutenção deste projeto na esfera institucional.

A iniciativa explicitou o potencial dos dados conectados abertos, reforçando o valor de conectar conjuntos de dados por meio da criação de metadados para vincular o Repositório da Produção USP à Wikidata. Além de direcionar para os registros do repositório, essa integração possibilita que o caminho inverso seja feito, ou seja, seria possível que o repositório direcionasse à Wikidata, especificamente aos itens editados pelas bibliotecas e que poderiam ter sido aprimorados pela comunidade de editores e por robôs. Além disso, é possível que outras bases de dados conectados estabeleçam a conexão com o Repositório da Produção USP a partir dos metadados de conexão criados, pois a integração da Wikidata a outros conjuntos de dados já é uma realidade no contexto bibliotecário, destacando-se, como exemplo, a conexão que o *Virtual International Authority File* (VIAF) e o Portal do ISSN fazem com a Wikidata.

## 6 CONSIDERAÇÕES





A experiência demonstrou o potencial da publicação de dados com significados definidos na Web e de sua interligação a conjuntos externos de dados, na medida em que possibilita realizar buscas bastante específicas e promover visualizações em formatos variados, além de ampliar o alcance dos conteúdos no contexto digital.

Por meio de vinculações entre pesquisadores, instituições e suas subdivisões, publicações e campos do conhecimento, a dimensão semântica inerente à Wikidata permitiu efetuar catalogação que comportou a trama institucional da produção do conhecimento em que se inserem os docentes da ECA. Com isso, sobressaiu que a importação de registros em formato MARC21 para dados conectados abertos na Wikidata poderá demandar a coleta e a estruturação de novos dados, que extrapolam aqueles presentes no registro em formato MARC21.

Nessa perspectiva, é fundamental compreender as possibilidades semânticas que a Wikidata oferece à estruturação de dados para planejar uma catalogação que descreva entidades e o relacionamento entre elas, promovendo novas possibilidades de buscas e a interoperabilidade com outros conjuntos de dados.

A dimensão social da Wikidata – desenvolvida por uma comunidade de voluntários – revelou que o território a ser explorado por bibliotecários vai além da inserção de dados, incluindo sua participação afirmativa na comunidade ao colaborar com suas políticas e normas, propostas de estruturas de metadados e sua documentação, assim como criação de metadados. Ademais, a adoção de software aberto sobressaiu como estratégia relevante para garantir a manutenção de ações em bibliotecas, historicamente marcadas pela falta de recursos financeiros e tecnológicos no país. Isto porque, os recursos abertos caracterizam-se pela existência de uma comunidade de práticas, que prevê instâncias colaborativas visando tanto o seu desenvolvimento como sua apropriação pelos interessados.

De outro lado, a escolha de aparatos tecnológicos em bibliotecas implica considerações sobre seus princípios norteadores. Nesta experiência, de biblioteca universitária ocupada com a preservação e a comunicação da memória científica do país em perspectiva democrática, a ciência aberta foi conceito orientador da práxis, direcionando ao uso de uma base de dados representativa do ecossistema do conhecimento aberto. Portanto, cabe realçar a importância de postura crítica na adoção de recursos tecnológicos. É fundamental que as decisões considerem os impactos que podem gerar no campo

científico e no meio social mais amplo, reconhecendo a relevância de sistemas e dados abertos para o desenvolvimento da ciência aberta.

## REFERÊNCIAS

ABCD - AGÊNCIA DE BIBLIOTECAS E COLEÇÕES DIGITAIS DA UNIVERSIDADE DE SÃO PAULO. **Repositório da Produção Intelectual da USP**. In: USP. ABCD. São Paulo, [2024]. Disponível em: <https://www.abcd.usp.br/bibliotecas/digitais-sistemicas/producao-usp>. Acesso em: 29 mar. 2024.

AJUDA: QuickStatements. In: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2024]. Disponível em: <https://www.wikidata.org/wiki/Help:QuickStatements/pt-br>. Acesso em: 31 jan. 2024.

ALLEMANG, Dean; HENDLER, James. **Semantic web for the working ontologist: effective modeling in RDFS and OWL**. 2. ed. Waltham, USA: Elsevier, 2011. Disponível em: <https://www.sciencedirect.com/book/9780123859655/semantic-web-for-the-working-ontologist>. Acesso em: 30 jan. 2024.

ARL - ASSOCIATION OF RESEARCH LIBRARIES. **ARL white paper on wikidata: opportunities and recommendations**. Washington: ARL, 2019. Disponível em: <https://www.arl.org/resources/arl-whitepaper-on-wikidata/>. Acesso em: 31 jan. 2024.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data: the story so far. **International Journal on Semantic Web and Information Systems**, [S. l.], v. 5, n. 3, p. 1–22, 2009. Disponível em: [link.gale.com/apps/doc/A209477051/AONE?u=anon~96826906&sid=bookmark-AONE&xid=1fc31238](http://link.gale.com/apps/doc/A209477051/AONE?u=anon~96826906&sid=bookmark-AONE&xid=1fc31238). Acesso em: 30 jan. 2024.

DEPAR - ESCRITÓRIO DE DESENVOLVIMENTO DE PARCERIAS DA USP. **USP em números**. In: USP. DEPAR. São Paulo, [2024]. Disponível em: <https://depar.usp.br/num/>. Acesso em: 29 mar. 2024.

ESCOLA DE COMUNICAÇÕES E ARTES. In: SCHOLIA. [San Francisco, CA: Wikimedia Foundation, 2024]. Disponível em: <https://scholia.toolforge.org/organization/Q10274592>. Acesso em: 31 jan. 2024.

GODBY, Jean ... [et al.]. **Creating library linked data with wikibase: lessons learned from project passage**. Dublin: OCLC, 2019. Disponível em: <https://www.oclc.org/research/publications/2019/oclcresearch-creating-library-linked-data-with-wikibase-project-passage.html>. Acesso em: 18 jun. 2024.

GOMES, Maria João; ROSA, Flávia. **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: EDUFBA, 2010.

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**, [S. l.], v. 26, p. 01–21, 2021.



Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/75909>. Acesso em: 29 jan. 2024.

LEMUS-ROJAS, Mairelys; ODELL, Jere D. Creating structured linked data to generate scholarly profiles: a pilot project using wikidata and scholia. **Journal of Librarianship and Scholarly Communication**, v. 6, p. 1-23, 2018. Disponível em: <https://doi.org/10.7710/2162-3309.2272>. Acesso em: 18 jun. 2024.

LEMUS-ROJAS, Mairelys, PINTSCHER, Lydia. Wikidata and Libraries: facilitating open knowledge. In: PROFFITT, Merrilee (Ed.) **Leveraging Wikipedia: connecting communities of knowledge**. Chicago: ALA, 2018. p. 143–158. Disponível em: <https://www.oclc.org/research/publications/2018/leveraging-wikipedia.html>. Acesso em: 31 jan. 2024.

LYNCH, Clifford A. Institutional repositories: essential infrastructure for scholarship in the digital age **ARL: a bimonthly report on research library issyes and actions from ARL, CNI, and SPARC**, n. 226, fev. 2003. Disponível em: <https://doi.org/10.1353/pla.2003.0039>. Acesso em: 30 jan. 2024.

MACHADO, Jorge. Dados abertos e ciência aberta. In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 201-227. Disponível em: <https://livroaberto.ibict.br/handle/1/1060>. Acesso em: 30 jan. 2024.

MONTGOMERY, Lucy ... [et al.]. Who puts the ‘open’ in open knowledge? **Cultural Science Journal**, v. 12, n. 1, p. 13–22, 2020. Disponível em: <https://sciendo.com/article/10.5334/csci.136>. Acesso em: 30 jan. 2024.

MORA-CANTALLOPS, Marçal; SÁNCHEZ-ALONSO, Salvador; GARCÍA-BARRIOCANAL, Elena. A systematic literature review on Wikidata. **Data Technologies and Applications**, v. 53, n. 3, p. 250–268, jan. 2019. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/DTA-12-2018-0110/full/html?skipTracking=true>. Acesso em: 30 jan. 2024.

OHLIG, Jens. Linked data for libraries with Wikibase. [S.l.: s.n.], 2020. 1 vídeo (30 min). Publicado pelo canal LD4. Disponível em: <https://www.youtube.com/watch?v=FVmxpaB-Th4&list=PLx2ZluWEZtIAETxLY-TaqJWsRMNY59r9v&index=8>. Acesso em: 18 jun. 2024.

OPEN KNOWLEDGE FOUNDATION. **Open definition**: defining open in open data, open content and open knowledge. In: OPEN Knowledge Foundation. [S.l., 202-?]. Disponível em: <https://opendefinition.org/od/2.1/en/>. Acesso em: 30 jan. 2024.

OPENREFINE. [S.l., 202-?]. Disponível em: <https://openrefine.org/>. Acesso em: 31 jan. 2024.

REPOSITÓRIO da Produção USP, [2024]. Disponível em: <https://repositorio.usp.br/index.php>. Acesso em: 08 abr. 2024.

ROBINSON-GARCIA, Nicolas; COSTAS, Rodrigo; LEEUWEN, Thed N. Van. Open access uptake by universities worldwide. **PeerJ**, v. 8, p. e9410, jul. 2020. Disponível em: <https://doi.org/10.7717/peerj.9410>. Acesso em: 30 jan. 2024.

ROSSENOVA, Lozana; SOHMEN, Lucia. OpenRefine to Wikibase: data upload pipeline. In: WIKIVERSITY. [San Francisco, CA: Wikimedia Foundation, 2022]. Disponível em: [https://en.wikiversity.org/wiki/OpenRefine\\_to\\_Wikibase:\\_Data\\_Upload\\_Pipeline](https://en.wikiversity.org/wiki/OpenRefine_to_Wikibase:_Data_Upload_Pipeline). Acesso em: 31 jan. 2024.

SANTAREM SEGUNDO, José Eduardo. **Mini tutorial**: OpenRefine.. [S.l.: s.n.], 2020. 1 vídeo (39 min). Publicado pelo canal José Eduardo Santarem Segundo. Disponível em: <https://www.youtube.com/watch?v=lbBmXC0ADRQ>. Acesso em: 31 jan. 2024.

SCHOLIA. In: SCHOLIA [San Francisco, CA: Wikimedia Foundation, 202-?a]. Disponível em: <https://scholia.toolforge.org/>. Acesso em: 31 jan. 2024.

SCHOLIA. In: GITHUB [S.l.: s.n., 202-?b]. Disponível em: <https://github.com/WDscholia>. Acesso em: 31 jan. 2024.

SHENOY, Kartik *et al.* A study of the quality of Wikidata. **Journal of Web Semantics**, v. 72, p. 100679, abr. 2022. Disponível em: <https://doi.org/10.1016/j.websem.2021.100679>. Acesso em: 30 jan. 2024.

STOCCO, Verônica. **Compreendendo a importação de dados no Wikidata via Quickstatements**. [S.l.: s.n.], 2021. 1 vídeo (88 min). Publicado pelo canal Biblioteca IME-USP. Disponível em: <https://www.youtube.com/watch?v=PbvtAqZqXeY>. Acesso em: 31 jan. 2024.

WIKIDATA: Introduction. In: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2024a]. Disponível em: <https://www.wikidata.org/wiki/Wikidata:Introduction>. Acesso em: 31 jan. 2024.

WIKIDATA: List of properties. In: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2024c]. Disponível em: [https://www.wikidata.org/wiki/Wikidata:List\\_of\\_properties](https://www.wikidata.org/wiki/Wikidata:List_of_properties). Acesso em: 29 mar. 2024.

WIKIDATA: Robôs. In: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2023b]. Disponível em: <https://www.wikidata.org/wiki/Wikidata:Bots/pt-br>. Acesso em: 29 mar. 2024.

WIKIDATA: Tools/OpenRefine. In: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2023d]. Disponível em: <https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine>. Acesso em: 31 jan. 2024.

WIKIDATA: Tools/OpenRefine/Editing. In: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2024d]. Disponível em: <https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine/Editing>. Acesso em: 31 jan. 2024.

WIKIDATA: Wikibase. *In*: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2023a]. Disponível em: <https://www.wikidata.org/wiki/Wikidata:Wikibase>. Acesso em: 31 jan. 2024.

WIKIDATA: WikiProject IUPUI University Library. *In*: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2023c]. Disponível em: [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_IUPUI\\_University\\_Library](https://www.wikidata.org/wiki/Wikidata:WikiProject_IUPUI_University_Library). Acesso em: 31 jan. 2024.

WIKIDATA: WikiProject Periodicals. *In*: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 2024b]. Disponível em: [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Periodicals](https://www.wikidata.org/wiki/Wikidata:WikiProject_Periodicals). Acesso em: 29 mar. 2024.

WIKIMEDIA Toolforge [S.l., 202-?]. Disponível em: <https://admin.toolforge.org/>. Acesso em: 31 jan. 2024.

WIKIPÉDIA: GLAM/Bibliotecas da USP. *In*: WIKIPÉDIA. [San Francisco, CA: Wikimedia Foundation, 2024]. Disponível em: [https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:GLAM/Bibliotecas\\_da\\_USP](https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:GLAM/Bibliotecas_da_USP). Acesso em: 31 jan. 2024.

## NOTAS

### CONTRIBUIÇÃO DE AUTORIA

**Concepção e elaboração do manuscrito:** L. Viana

**Coleta de dados:** L. Viana

**Análise de dados:** L. Viana

**Discussão dos resultados:** L. Viana

**Revisão e aprovação:** L. Viana

### CONJUNTO DE DADOS DE PESQUISA

Não se aplica.

### FINANCIAMENTO

Não se aplica.

### CONSENTIMENTO DE USO DE IMAGEM

Não se aplica

### APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

### CONFLITO DE INTERESSES

Não se aplica.

### LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico

(ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

#### **PUBLISHER**

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

#### **EDITORES**

Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert, Genilson Geraldo, Jônatas Edison da Silva, Mayara Madeira Trevisol, Edna Karina da Silva Lira e Luan Soares Silva.

#### **HISTÓRICO**

Recebido em: 12-04-2024 – Aprovado em: 22-07-2024 - Publicado em: 14-10-2024.

