



GT 2 – Organização e Representação do Conhecimento

ISSN 2177-3688

O ENRIQUECIMENTO SEMÂNTICO SOB UM PERSPECTIVA SOCIOCULTURAL DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

SEMANTIC ENRICHMENT UNDER A SOCIOCULTURAL PERSPECTIVE OF KNOWLEDGE ORGANIZATION AND REPRESENTATION

Maria Lígia Triques - Estadual de Londrina (UEL)

Ana Cristina de Albuquerque - Estadual de Londrina (UEL)

Mario Barité - Estadual de Londrina (UEL)

Modalidade: Trabalho Completo

Resumo: Comumente associado ao termo “semântico”, o enriquecimento denota um processo de atribuição de maior significação aos dados. Tal processo, no tocante às atividades de organização e representação da informação e do conhecimento, pode ser visto a partir de instrumentos, processos e produtos, que por sua vez, estão interligados às possibilidades de se tornar um conteúdo mais significativo e representativo de um ponto de vista das abordagens socioculturais. Logo, o objetivo é verificar os processos de enriquecimento de dados a partir de uma perspectiva sociocultural da Organização e Representação do Conhecimento (ORC). O estudo tem caráter qualitativo, exploratório-descritivo e bibliográfico, empregando a Análise de Conteúdo como procedimento para guiar a coleta e análise dos dados na construção dos resultados. Analisa-se um conjunto de publicações obtidas a partir de um levantamento sistematizado, fazendo um paralelo com o que pode ser abarcado como instrumentos, processos e produtos da ORC dentro de uma visão sociocultural. Os resultados demonstram as atividades vinculadas a esse processo em ambientes digitais, destacando seu potencial sociocultural e reforçando que o enriquecimento descentraliza o processo de representação dos dados, dando a possibilidade de atribuir garantias que justifiquem e validem os aspectos formalizados e, ao mesmo tempo, utilizar múltiplas perspectivas de atribuição de sentido em diversas camadas de informação.

Palavras-chave: enriquecimento semântico; organização e representação do conhecimento; abordagens socioculturais; Metadados.

Abstract: Commonly associated with the term "semantic", enrichment denotes a process of attributing greater meaning to data. This process, with regard to the activities of organizing and representing information and knowledge, can be seen in terms of instruments, processes, and products, which in turn are interconnected with the possibilities of making content more meaningful and representative from the point of view of sociocultural approaches. Therefore, the objective is to verify the processes of data enrichment from a sociocultural perspective of Knowledge Organization and Representation (ORC). The study is qualitative, exploratory-descriptive, and bibliographical, using Content Analysis as a procedure to guide data collection and analysis in the construction of the results. A set of publications obtained from a systematic survey is analyzed, drawing a parallel with what can be encompassed as ORC instruments, processes, and products within a sociocultural vision. The results demonstrate the activities linked to this process in digital environments, highlighting its sociocultural potential and reinforcing that enrichment decentralizes the data representation process, giving the possibility of

assigning guarantees that justify and validate the formalized aspects and, at the same time, using multiple perspectives to assign meaning to different layers of information.

Keywords: semantic enrichment; knowledge organization and representation; sociocultural approaches; metadata.

1 INTRODUÇÃO

O processo de enriquecimento que ocorre em ambientes digitais é um procedimento que abarca diferentes atividades que podem ocorrer sobre os dados. Na computação, onde suas proposições têm raízes, esse processo se refere a possibilidade de associar ou aprimorar dados com contexto relevante obtidos de fontes adicionais (KNAPP; LANGILL, 2015), por vezes complementando dados ausentes ou incompletos.

Para o campo da Ciência da Informação, que carrega em seus pressupostos teóricos e práticos muito do que se desenvolveu no escopo de reflexão das áreas de Biblioteconomia, Documentação, Arquivologia e mais recentemente, da Museologia, o processo de enriquecimento pode ser atrelado a possibilidade de melhorar os metadados que descrevem, contextualizam e caracterizam um objeto digital com elementos que lhe tragam maior significação. Esse processo pode ocorrer por meio de diferentes métodos e ferramentas, obtendo como resultado o enriquecimento dos metadados originais (ISAAC *et al.*, 2015, ZENG, 2019), isto é, aqueles elementos descritivos que já acompanhavam os dados.

Comumente associado ao termo “semântico”, o enriquecimento denota um processo de atribuição de maior significado aos dados, através do uso de uma semântica atribuída por meio de vocabulários pré-existentes, como sinônimos ou informações de proveniência, entre outros (LIRA, 2014). Com isso, o processo tem por objetivo facilitar a compreensão, a integração e o processamento dos dados tanto por pessoas, como por máquinas (LIRA, 2014, ISAAC *et al.*, 2015, ZENG, 2019).

Por essa razão, reconhece-se que esse não é um processo novo aos tradicionais campos de estudos que trabalham com metadados enquanto elementos que permitem a organização e a representação para um recurso de informação. O olhar voltado aos usuários e, mais precisamente, aos seus contextos sociais, ampliou a percepção de que a atribuição de metadados possui um caráter sociocultural. Como defende Capurro (2007) a CI desenvolve-se no atual século a partir de um paradigma pragmático e social dotado de cunho tecnológico-digital, que herda as discussões de seus predecessores, principalmente da biblioteconomia e da documentação.

Exemplo disso é a contínua preocupação com o desenvolvimento de estruturas capazes de representar a informação e o conhecimento de modo a ampliar a precisão e o significado de recursos informacionais a partir da estruturação e da associação de elementos representativos, isto é, a partir do emprego de padrões, que incluem, esquemas, vocabulários controlados, tesouros, ontologias, listas de controle, entre outros (GILLILAND, 2016). Portanto, a própria atribuição de metadados configura uma tarefa semântica.

Diante disso, percebe-se que, uma vez que as influências culturais podem se manifestar na ORC de várias maneiras (LEE, 2021), isto é, a partir de diferentes sistemas e estruturas e que as perspectivas socioculturais são indissociáveis da ORC, julga-se pertinente, portanto, elucidar como o processo de enriquecimento se dá no tocante das atividades de organização e representação da informação e do conhecimento, no que concerne seus instrumentos, processos e produtos em ambientes digitais.

Logo, a questão que norteia esse trabalho apresenta-se da seguinte forma: como se caracterizam os processos de enriquecimento de dados e como esse processo está interligado as possibilidades de se tornar um conteúdo mais significativo e representativo de um ponto de vista sociocultural que permeiam os estudos do campo da ORC?

Diante disso, o objetivo desse estudo é verificar os processos de enriquecimento de dados a partir de perspectiva sociocultural da ORC. Para isso, desenvolve-se um estudo pautado no método bibliográfico e na Análise de Conteúdo (BARDIN, 2011) para verificar à luz da fundamentação teórica, as possibilidades desse processo no que tange às abordagens socioculturais.

2 DESENVOLVIMENTO METODOLÓGICO

Em virtude de identificar os processos de enriquecimento de dados na perspectiva da ORC, recorreu-se a um levantamento da literatura de modo a formar um *corpus* de análise que viabilizasse o embasamento teórico necessário para a discussão. Diante disso, o estudo caracteriza-se como qualitativo, exploratório-descritivo e bibliográfico, se baseando na Análise de Conteúdo (BARDIN, 2011) como procedimento para guiar a coleta e análise dos dados na construção dos resultados.

Na primeira fase, denominada **Pré-análise**, ocorre o procedimento inicial de organização, na qual se dá a escolha, a formulação de critérios e a preparação do material para análise que fundamentará a interpretação dos dados. Sendo assim, para o levantamento

dos dados, foi selecionada, por intermédio do Portal de Periódicos da Capes, a base de dados internacional *Web of Science* pelo seu caráter de abrangência internacional, utilizando como estratégia de busca a *string*: "*semantic enrichment*" OR "*data enrichment*" OR "*metadata enrichment*", sendo a busca restringida ao campo Palavras-chave do autor, e tendo como delimitadores: Tipos de documento (Artigo de conferência or Artigo or Artigo de revisão) e Idiomas (*English*). A busca também foi delimitada pela área de pesquisa (*Information Science and Library Science*), de modo a obter um panorama de como a área trabalha com os termos. Por fim, foi considerado o intervalo temporal para as publicações a partir do ano de 2001, compreendendo, assim, o atual século. Para complementar o *corpus*, a Base de Dados em Ciência da Informação (BRAPCI) também foi utilizada, considerando seu caráter específico e de abrangência nacional. Apenas o termo “enriquecimento” foi utilizado como estratégia nesse caso.

O quadro 1 sintetiza o levantamento da literatura realizado e o número de itens recuperados.

Quadro 1 - Procedimentos para formação do corpus de análise

Base	Web of Science	BRAPCI
Data da Busca	04/06/2023	04/06/2023
String de Busca	AK=(" <i>semantic enrichment</i> " OR " <i>data enrichment</i> " OR " <i>metadata enrichment</i> ")	“enriquecimento”
Parâmetros e Campos de Busca	Palavras-chave de autor	Palavras-chave
Filtros e Delimitadores	Tipos de documento: Artigo de conferência or Artigo or Artigo de revisão; Idiomas: <i>English</i>	Não consta
Nº Itens	23	7

Fonte: Elaborado pelos autores (2023).

O processo de formação do *corpus* de análise restrito a apenas as palavras-chave das publicações se justifica tendo em vista enquadrar somente publicações que explicitamente tratam de processos de enriquecimento semântico, uma vez que se considera que um termo, enquanto descritor, tem relevância teórica para o estudo e espera-se que haja sua definição e descrição. Cabe ressaltar que se levou em conta a possibilidade de uma publicação tratar do processo sem utilizar explicitamente os termos selecionados, contudo, observou-se que essa estratégia resultou em itens mais representativos para a delimitação proposta. Também, julgou-se necessário o acesso integral aos textos das publicações, de modo que a análise fosse pertinente ao objetivo proposto.

Para construção das discussões, isto é, a caracterização dos processos de enriquecimento de dados, procedeu-se à próxima etapa da análise, a fase de **Exploração do material** que consiste em sistematizar o material, realizando a leitura das publicações selecionadas verificando as relações existentes entre os fundamentos encontrados a partir do emprego de categorias de análise definidas neste estudo *a posteriori*, isto é, com base nos padrões de reconhecimento explícito encontrados na leitura do material selecionado.

Das 30 publicações, 29 puderam ser analisadas considerando, em seu conteúdo integral, a presença ou ausência de definição e/ou descrição explícita, isto é, de um trecho identificável no texto que caracterize processos de enriquecimento. Considerando tal critério para a inclusão, como resultado foram obtidas um total de 22 publicações em que foi possível reconhecer claramente como os autores definem ou descrevem o processo.

A partir disso, seguiu-se para a fase de **Tratamento dos dados, inferência e interpretação**, na qual os dados são validados à luz da fundamentação teórica definida no escopo do trabalho, possibilitando fazer inferências e interpretações. Isso permitiu delinear os procedimentos indicados pelos autores como enriquecimento, bem como identificando as abordagens e os contextos considerados em cada trabalho, sendo estes analisados sob a perspectiva dos processos, instrumentos e produtos da ORC. Diante disso, procedeu-se às discussões fazendo um paralelo entre os resultados obtidos e as perspectivas teóricas nos estudos das abordagens socioculturais da ORC discutidas a seguir.

3 A PERSPECTIVA SOCIOCULTURAL NA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO (ORC)

A ORC, no âmbito da Ciência da Informação (CI), se dedica aos estudos e as atividades de organização e representação do conhecimento e de suas manifestações em registros de informação, considerando seus contextos socioculturais de produção, acesso, uso, apropriação e disseminação. Como pontuam Lara e Mendes (2017, p. 376), no núcleo da CI, a informação é a forma para compartilhar o conhecimento, sendo “[...] o resultado de uma articulação de um entendimento pragmático prévio sobre um mundo comum compartilhado apresentado sob determinada forma.”

Diante disso, a ORC tem direta relação com os aspectos socioculturais que moldam esse entendimento comum e compartilhado sobre um determinado contexto, resultando no que pode ser entendido como um conhecimento socializado (BARITÉ, 2001) que, em sua

materialização ou registro, pode ser acessado e interpretado, para que na interação com a consciência humana, possa gerar novos conhecimentos. Logo, parte-se do entendimento que o conhecimento é manifestado em informação, ou mesmo em conjuntos de dados, formando o que Frohmann (1995) considera como uma rede de elementos heterogêneos sujeitos a práticas sociais complexas e a tensões entre grupos de interesses.

Assim sendo, entende-se que falar em perspectivas socioculturais na ORC é compreender os processos, produtos e instrumentos sob a ótica de um diálogo heterogêneo, moldado pela dinâmica do momento e do lugar em que ocorrem. As posturas teóricas da ORC que seguem esse olhar, enfocam o estudo de domínio e buscam abandonar a perspectiva da existência de universo singular do conhecimento cuja pretensão é uma organização e representação ideal e universal, voltando-se a uma perspectiva pluralista na qual “[...] classificações, conceitos, categorias, significados e sujeitos são determinados pelo contexto histórico, social e cultural das comunidades usuárias.” (LEE, 2021, p. 1, tradução nossa).

Nessa perspectiva, entende-se os processos, produtos e instrumentos da ORC como sendo aqueles ligados, sobretudo, aos Sistemas de Organização do Conhecimento (SOC) em sua relação com as atividades de indexação, classificação e catalogação em função da recuperação da informação. E como abordagens socioculturais da ORC, a compreensão desses processos, produtos e instrumentos pensados a partir de múltiplas formas possíveis de organizar e representar o conhecimento e seus registros em função dos usuários e de determinados contextos tendo em vista uma postura ética e inclusiva do conhecimento.

A literatura aponta que as formas de manifestação dos aspectos socioculturais na ORC ocorrem principalmente por meio do estudo e observância das garantias (GUIMARÃES; EVANGELISTA; LUZ; OSAWA, 2019, BARITÉ, 2019, LEE, 2021). As garantias são o que Bullard (2017, p. 76) vê como um elemento comum a uma ampla variedade de sistemas, independentemente de assim ser nomeado, envolvendo “[...] desde a classificação tradicional de bibliotecas, até menus e categorias de aplicativos para coleções baseadas na Web [...].

Vistas como fontes de informação para a terminologia nos esquemas de organização (BULLARD, 2017), bem como um ferramental de validação e justificação para as terminologias que não apenas incide sobre recursos convencionais, mas também digitais (BARITÉ, 2019), as garantias “[...] carregam consigo um conjunto de fundamentos epistemológicos, pressupostos sobre o propósito do sistema, modos de construção de classificação e modos de avaliação.” (BULLARD, 2017, p. 76).

Os ambientes informacionais digitais permitem continuamente novas perspectivas e possibilidades de acesso e compartilhamento de informações, possibilitando um espaço reconfigurável e aberto de constante reescrita e adição de conhecimento registrado, que coloca em diálogo diferentes valores em uma cultura participativa (DALBELLO, 2015). Nesse contexto, como destaca Beghtol (2002, p. 526), “[...] a criação, busca, disseminação e uso da informação não são limitados por tempo, espaço ou comunidades discursivas estáticas”, indo ao encontro da concepção de transculturalismo em oposição ao multiculturalismo defendida por García-Gutiérrez (2002), na qual é necessário não simplificar a cultura frente a uma análise reducionista, uma vez que esta consiste em um sistema aberto com capacidades de interação.

Assim, ponderando as garantias no que se refere aos instrumentos, processos e produtos da ORC, estas são pensadas para abarcar, além da visão tradicional, enquanto elemento de autoridade nas decisões invocadas pelo responsável do sistema de organização do conhecimento, também as novas figuras de autoridade, tal como aponta Barité (2019, p. 649), como “[...] os líderes de certas culturas ou grupos minoritários, as terminologias alternativas propostas por pensadores ou movimentos sociais – especialmente os contraculturais – ou mesmo pelos fornecedores de novas tecnologias.”

Nesse sentido, olhar para qualquer sistema que intencione a organização e representação das manifestações informacionais do conhecimento constitui um desafio constante de articular diferentes noções, atores e garantias a partir de um conjunto de possibilidades abertas a múltiplas interpretações e aplicações. Assim, para a análise dos resultados, propõe-se evidenciar tais possibilidades socioculturais nos instrumentos, processos e produtos vinculados ao enriquecimento semântico.

4 APRESENTAÇÃO DOS RESULTADOS E DISCUSSÃO

Por vezes referido como enriquecimento de dados ou de metadados, o processo de enriquecimento semântico pode estar atrelado a diferentes procedimentos realizados sobre os dados, como anotação, reconciliação, mapeamento, alinhamento, correspondência, fusão, interligação, entre outros (LIRA, 2014, ISAAC *et al.*, 2015, ZENG, 2019).

De modo a reconhecer como esse processo se apresenta nas publicações selecionadas no recorte e permitir a formação de categorias para a análise, em cada publicação em que se identificou uma definição e/ou descrição, destacou-se o contexto de aplicação do processo na relação com o tema e, com base nisso, os aspectos que puderam ser associados aos

XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB
Aracaju-SE – 06 a 10 de novembro de 2023

instrumentos, processos ou produtos compreendidos dentro do escopo da ORC. O quadro 1 sintetiza a sistematização das publicações.

Quadro 1 – Sistematização das Publicações

AUTORES	TÍTULO	CATEGORIA TEMA/CONTEXTO	CATEGORIA ORC
Newman, D. <i>et al.</i> (2019)	Subject Metadata Enrichment using Statistical Topic Models	Enriquecimento de metadados heterogêneos com modelagem estatística de tópicos (técnica de aprendizado de máquina) em biblioteca digital que agrega conteúdo de várias fontes.	metadados de assuntos/ catálogos bibliográficos digitais
Kors, J. A. <i>et al.</i> (2015)	A multilingual gold-standard corpus for biomedical concept recognition: the Mantra GSC	Enriquecimento automático de terminologias multilíngue de conceitos biomédicos a partir de harmonização de anotações.	terminologias multilíngue/ vocabulários especializados
Yang, B.; Zhang, Y. (2015)	Pattern-mining approach for conflating crowdsourcing road networks with POIs	Abordagem de mineração de padrões para combinar redes rodoviárias de <i>crowdsourcing</i> (geradas pelo usuário) com pontos de interesse (POIs) geométrica e semanticamente.	padrões de dados georreferenciais e informações geográficas geradas pelo usuário/ vocabulários especializados e folksonomia
Manguinhas, H. <i>et al.</i> (2016)	Exploring Comparative Evaluation of Semantic Enrichment Tools for Cultural Heritage Metadata	Avaliações e comparações de ferramentas de enriquecimento atualmente desenvolvidas e utilizadas no contexto da Europeia.	metadados de patrimônio cultural digitalizado/ vocabulários especializados e multilíngues
Klein, E. <i>et al.</i> (2016)	Sustainable Linked Open Data Creation: An Experience Report	Publicação e reutilização de dados vinculados no setor público e nos negócios.	dados abertos de várias fontes/ vocabulários especializados
Medrek, J.; Otto, C.; Ewerth, R. (2018)	Recommending Scientific Videos Based on Metadata Enrichment Using Linked Open Data	Explorar e enriquecer as informações geradas automaticamente pela análise audiovisual automática, vinculando-as ao arquivo de autoridade integrada (GND) da Biblioteca Nacional Alemã para aprimorar a informação de metadados e a pesquisa semântica.	metadados de audiovisual/ vocabulários especializados
Al-Natsheh, H. T. <i>et al.</i> (2018)	Metadata Enrichment of Multi-disciplinary Digital Library: A Semantic-Based Approach	Marcação semântica de diferentes rótulos de artigos da biblioteca científica digital com uma rede de sinônimos.	palavras-chave de artigos científicos e sinônimos/vocabulários especializados
Luo, X.; Tan, X.; Wang, X. (2019)	Semantically Enriched Presentation for Cultural Heritage Image: A POI-Based Perspective	Visualização interativa de imagem do patrimônio cultural baseada na perspectiva do pois (pontos de interesse) a partir da agregação e apresentação de informações semânticas.	dados de visualização de imagens do patrimônio cultural
Candela, G. <i>et al.</i> (2019)	A linked open data framework to enhance the discoverability and impact of culture heritage	Estrutura para enriquecimento e desambiguação de localizações em texto (linguagem natural das informações do catálogo) usando bases de conhecimento abertas (Wikidata e Geonames) com informações da linguagem natural das informações do catálogo da Biblioteca Virtual Miguel de Cervantes.	dados de localização para catálogos de bibliotecas
Zeng, M. L. (2019)	Semantic enrichment for enhancing LAM data and supporting digital humanities	Métodos e esforços de enriquecimento semântico que podem ser aplicados aos dados fornecidos por bibliotecas, arquivos e museus (LAM) em vários níveis, com o objetivo de apoiar uma exploração mais profunda e ampla e o uso de dados LAM na pesquisa de humanidades digitais.	metadados fornecidos por bibliotecas, arquivos e museus (LAM)
Oliveira, L. F. R.; Martins, D. L. (2019)	Explorando a Reconciliação de Dados Culturais na Wikidata: experimento aplicado com o acervo museológico do Museu Histórico Nacional (MHN)	Aplicação de uma técnica de enriquecimento semântico em bases de dados culturais envolvendo <i>scripts</i> de reconciliação de dados para reconciliar parte do conjunto de dados do acervo museológico do MHN com objetos digitais da Wikidata.	dados do acervo museológico-cultural/ vocabulários especializados
Araújo, W. J.; Lima, G. Â. B. de O. (2019)	Em busca de uma metodologia para enriquecimento de ontologias de domínio	Metodologia com procedimentos detalhados para enriquecimento de todos os componentes das ontologias de domínio.	enriquecimento de ontologias
Ceriani, M.; Bernasconi, E.; Medella, M. (2020)	A Streamlined Pipeline to Enable the Semantic Exploration of a Bookstore	Pipeline (processo) que usa enriquecimento semântico e técnicas de visualização de grafo de conhecimento para permitir a exploração semântica de um <i>corpus</i> de texto.	dados de catálogos especializados
Lee, H.; Yoon, S.;	SEMANTIC in a Digital Curation Model	Modelo abstrato reunindo conceitos que podem focar na representação e descrição de recursos que	modelo de curadoria para objetos digitais

XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB
Aracaju-SE – 06 a 10 de novembro de 2023

Park, Z. (2020)		ênfatem o enriquecimento semântico em um modelo de curadoria digital.	
Torino, E. <i>et al.</i> (2020)	Enriquecimento semântico para a disponibilização de dados abertos: teoria e prática	Processo de formalização necessário à disponibilização de dados abertos no contexto do <i>linked open data</i> para estruturação e enriquecimento dos conjuntos de dados.	vocabulários/estruturação de dados
Vorndran, A.; Grund, S. (2021)	Metadata Sharing - How to Transfer Metadata Information among Work Cluster Members	Técnica de cluster para agregar obras do banco de dados Culturegraph, que coleta registros de metadados bibliográficos de todas as redes de bibliotecas regionais alemãs, de modo a melhorar a qualidade dos metadados enriquecendo os registros de metadados com informações adicionais.	elementos de metadados de registros bibliográficos
Nazário, C. E.; Dias, C. da C. (2021)	Avaliação do Europeana Data Model para o enriquecimento semântico de objetos	Avaliação das contribuições do EDM para o enriquecimento semântico de objetos publicados na WEB por meio do <i>Linked Data</i> , a partir da Matriz TRESO.	vocabulários/plataforma agregadora de dados
Georgiadis, H. <i>et al.</i> (2022)	Enriching the Greek National Cultural Aggregator with Key Figures in Greek History and Culture: Challenges, Methodology, Tools and Outputs	Método para enriquecer metadados agregados do Objetos de Patrimônio Cultural (CHO) com entidades pessoais de um vocabulário de dados vinculados.	dados de entidades pessoais/ plataforma agregadora de dados
Georgiadis, H. <i>et al.</i> (2022)	Publishing LOD Vocabularies in Any Schema with Semantics.gr	A plataforma Semantics.gr é examinada em sua funcionalidade e fluxos de trabalho para design personalizado de modelos de dados e esquemas de vocabulário e desenvolvimento e publicação de vocabulários.	vocabulários/plataforma agregadora de dados
Machado, D. O. F.; Arakaki, A. C. S. (2022)	Schema.org para catálogos digitais	Vocabulário Schema.org como uma alternativa para adequação dos catálogos digitais.	vocabulário schema.org/ catálogos digitais
Estevanovic, M. P.; Bax, M. P. (2022)	Dados abertos governamentais: uma metodologia para publicar dados semânticos no estado de Minas Gerais	Metodologia para enriquecer e publicar dados abertos governamentais aplicada a um estudo de caso envolvendo Emendas Parlamentares Impositivas no estado de Minas Gerais.	conjunto de dados abertos governamentais/ vocabulários especializados
Egan, P. (2023)	In search of the item: Irish traditional music, archived fieldwork and the digital	Abordagem experimental para enriquecer metadados de dados vinculados com base nas coleções de áudio da música tradicional irlandesa do American Folklife Center (AFC) da Biblioteca do Congresso em Washington DC.	metadados de áudio/vocabulários especializados
Bernasconi, E. <i>et al.</i> (2023)	Design, realization, and user evaluation of the ARCA system for exploring a digital library	Enriquecimento semântico do conteúdo textual em uma interface de usuário que possibilita a busca e exploração do corpus por meio da navegação visual de um grafo de conhecimento de tópicos.	representação visual/grafos de conhecimento de tópicos

Fonte: Dados de pesquisa (2023).

Em Categoria Tema/Contexto procurou-se sintetizar o contexto de aplicação do processo na relação com o tema de modo a identificar as possíveis correlações entre o conteúdo e a perspectiva sociocultural. Já as categorias elencadas em Categorias ORC demonstram os instrumentos, processos e produtos correlatos ao enriquecimento que ficam evidentes nas publicações, auxiliando na visualização das correspondências entre os conteúdos. A partir disso, inferências puderam ser destacadas com base no que a literatura aponta acerca das possibilidades socioculturais de organizar e representar informações e conhecimentos.

Ao sistematizar os aspectos destacados, foi possível identificar correspondências entre os temas abordados, sendo que duas delas se destacam. A primeira refere-se ao fato de que as temáticas envolvendo o enriquecimento semântico, sobretudo, estão atreladas ao reuso

de dados, englobando os aspectos concernentes ao contexto do *Linked Data*, quanto às melhores práticas para estruturar e ligar dados, bem como ao *Linked Open Data*, considerando tais práticas em formatos com licença de uso aberto. Na maioria dos trabalhos é possível verificar essa relação. No contexto internacional, trabalhos como o de Klein *et al.* (2016), Zeng (2019), Vorndran e Grund (2021) enfatizam claramente essa relação, seja para explicitar que o processo de enriquecimento se dá a partir de métodos de vinculação de dados, seja para avaliar os resultados desse tipo de enriquecimento em vocabulários de dados. Do mesmo modo, no contexto nacional, Torino *et al.* (2020), Nazário e Dias (2021) e Estevanovic e Bax (2022) destacam nessa relação o processo de formalização que é necessário para que os dados sejam reaproveitados no enriquecimento de demais conjuntos de dados.

Observa-se, assim, a ênfase na estruturação dos dados, criando vocabulários que possam ser semanticamente reaproveitados, isto é, que tenham suas conceituações explicitadas para que processos automatizados possam identificar as relações significativas entre os dados. Considerando uma visão sociocultural, explicitar as conceituações pretendidas por um domínio configura um processo complexo de mapeamento de noções acerca de termos e conceitos, que por sua vez devem ser pensados para serem reutilizados em diversos contextos, por diferentes atores com diferentes finalidades. Aqui, pode-se observar a possibilidade de uma postura mais ética e inclusiva do conhecimento, ao passo que são usados padrões flexíveis e expansíveis para acomodar diferentes significados, noções e visões sobre um tema, provenientes de diferentes fontes.

Outro ponto de destaque é a associação com o processo denominado anotação, que em si já é identificado como uma forma de enriquecimento, pois consiste em atribuir características identificadas ou extraídas de recursos informacionais sem a necessidade de modificá-lo, tal como notas, com a finalidade de garantir a interpretação. Um exemplo é o trabalho de Kors *et al.* (2015), em que o enriquecimento é visto pelo viés da harmonização de anotações referentes a terminologias multilíngues no domínio dos conceitos biomédicos, sendo essas anotações realizadas de modo automático e validadas por humanos. Mais uma vez é possível inferir que os aspectos socioculturais se materializam no enriquecimento ao ampliar a granularidade das informações atribuídas a partir de mecanismos como a harmonização, que não se baseia em um padrão estático para a organização e a representação, mas sim, busca acomodar os vários aspectos implicados no processo de combinação automática entre os valores dos metadados.

Ainda, é possível apontar que a validação humana em anotações automatizadas é mais uma das possibilidades de garantia, pois, em geral, é feita por atores especialistas ou diretamente implicadas no tema. Exemplo claro é visto na publicação de Yang e Zhang (2015), que trabalham uma abordagem de mineração de padrões para combinar informações de redes rodoviárias geradas por *Crowdsourcing*, isto é, em colaboração por usuários de um sistema.

As anotações também podem estar atreladas ao reuso de dados a partir do *Linked Data* ou *Linked Open Data*, utilizando os vínculos para compor ou ampliar as anotações criadas ou atribuídas. Novamente, essa ação se configura uma forma de atribuir maior significação, mas também se torna uma forma de proveniência em que se vincula a origem de uma perspectiva, dando garantias que justifiquem e validem os aspectos formalizados. Como destacam Estevanovi e Bax (2022) ao tratar da transparência de dados abertos governamentais, as anotações são úteis, pois, uma vez formalizadas em acordos e convenções, ou seja, em associações e relações entre os dados, ficam estabelecidas em um conjunto de regras para que façam sentido para o domínio. Logo, a proveniência e transparência dos dados é outro aspecto ético e inclusivo que pode ser destacado nas análises. Contudo, vale destacar que a aparente neutralidade de acordos e convenções não deve ser um aspecto tomado como inerente, visto o risco de reforçar ou reafirmar posturas dominantes.

Em se tratando de um levantamento delimitado ao universo de publicações da Ciência da Informação, era esperado que uma parcela dos trabalhos teria estreita relação com temáticas voltadas às bibliotecas digitais. Nesse aspecto, destaca-se o trabalho de Newman *et al* (2019), que trata do enriquecimento de metadados heterogêneos a partir de modelos de tópicos estatísticos que agregam conteúdo de várias fontes, demonstrando a utilidade para o enriquecimento automático de metadados de assuntos. Vê-se aqui a complementaridade de um método estatístico e de processamento de linguagem natural (PLN), que oferecem processos e instrumentos de ORC para o enriquecimento semântico.

Vorndran e Grund (2021) apresentam a técnica de *cluster* para agregar obras do banco de dados Culturegraph, que coleta registros de metadados bibliográficos de todas as redes de bibliotecas regionais alemãs, de modo a integrar e melhorar a qualidade dos metadados para toda a rede. Também no que diz respeito aos registros bibliográficos, Machado e Arakaki (2022) estudam o Schema.org como uma alternativa para adequação dos catálogos digitais, destacando sua capacidade de enriquecer a experiência de busca do usuário nesse

instrumento, ao enriquecer a interface de visualização dos dados com *links* para vídeos, notícias, compras, viagens, imagens e mapas. Nesse aspecto, destaca-se que a melhoria e a ampliação da experiência de busca do usuário como possibilidade sociocultural que se debruça sobre as condições e os aspectos em que os usuários procuram e acessam informações.

Além de dados bibliográficos, dados relacionados ao universo museológico e audiovisual também se destacam. Manguinhas *et al.* (2016) veem o enriquecimento no contexto da agregadora de dados Europeana, como um procedimento que coloca os objetos do patrimônio cultural em contexto, ligando-os a entidades relevantes, como pessoas, lugares, tipos de objetos etc. Por sua vez, Luo, Tan e Wang (2019) apresentam as possibilidades da visualização interativa de imagens do patrimônio cultural baseada na perspectiva do POIs (pontos de interesse) - uma localização especificada que é considerada útil - a partir da agregação e apresentação de informações semânticas antes isoladas que aprimoram a experiência de visualização. Ainda nesse contexto, Egan (2023) examina uma abordagem experimental para enriquecer metadados com base em coleções de recursos de áudio digital não comercial, levantando a discussão sobre as formas de representar a diversidade de tradições musicais. Nesse sentido, destaca-se a contextualização das informações para interpretação precisa de humanos e de agentes computacionais como um aspecto fundamental, sobretudo, no que concerne às questões técnicas de interoperabilidade semântica para adequação de sentido na articulação com o contexto sociocultural em que ocorrem.

Logo, observa-se que multidimensionar as formas representação a partir do enriquecimento implica em descrever informações com várias fontes de dados a partir de bases de conhecimento heterogêneas, buscando o melhor aproveitamento dos recursos disponíveis, sem, contudo, substituir os padrões específicos de um domínio. Ainda, é possível reconhecer que o enriquecimento, enquanto atividade que se desenvolve na interface com os processos, instrumentos e produtos da ORC, está diretamente vinculado às metodologias e as práticas de indexação e à construção, uso e reuso de ontologias e de vocabulários que estruturam e explicitam contextos de domínios dos mais diversos segmentos, que uma vez combinados a tecnologias e a ferramentas recomendadas pela *Web*, debruçam-se sobre o processamento semântico.

Em ambientes digitais, essas possibilidades se apresentam em bases de conhecimentos disponíveis via web constituídas de estruturas como vocabulários de dados, que uma vez presentes na nuvem de dados abertos ligados são denominados *Linked Open Data Knowledge Organization System* (LOD-KOS) admitindo sua reutilização semântica, a exemplo da Wikidata ou do Geonames (ZENG, 2019), podendo ser combinados para aumentar a expressividade de um significado.

Diante disso, como o intuito de responder à questão norteadora, observa-se que os processos de enriquecimento de dados se caracterizam em função do pressupostos da ORC tal como no quadro 2, que sintetiza essas percepções com base nas possibilidades identificadas nas publicações do recorte de se tornar um conteúdo mais significativo e representativo de um ponto de vista sociocultural.

Quadro 2 - Processos, instrumentos e produtos vinculados ao enriquecimento semântico

PROCESSOS	• Catalogação • Classificação • Indexação/Indexação social (<i>Crowdsourcing</i>) • Métodos estatísticos e Processamento de linguagem natural (PLN)
INSTRUMENTOS	• Modelos estatísticos com <i>machine learning</i> (ML) • Plataformas de <i>Crowdsourcing</i> • Protocolos de integração • Ontologias e Tesouros • Bases de dados/conhecimento (reuso) • Padrões, formatos e esquemas de metadados • Recomendações da <i>World Wide Web Consortium</i>
PRODUTOS	• Catálogos e índices digitais • Plataformas de busca integrada • Resumos e descritores • Bases de dados/conhecimento • Anotações de dados • Vocabulários especializados/multilíngues

Fonte: Elaborado pelos autores (2023).

Vale destacar que alguns dos elementos mencionados acabam por ter suas funções sobrepostas dada a natureza e o contexto de uso, a exemplo das ontologias, que podem tanto ser instrumentos como produtos dessas atividades, apresentando-se como bases de dados, metadados ou mesmo vocabulários.

Lidar com semântica, isto é, com significados atribuídos a itens, é lidar com aspectos simbólicos que não podem ser simplificados ou reduzidos a um único sistema ou estrutura. O enriquecimento cria condições para que as culturas representadas pela materialidade digital possam ser tratadas, tal como são, enquanto aparelhos sociais abertos com inúmeras

capacidades de interação em que diferentes atores podem ser vistos como autoridade para as decisões tomadas.

5 CONSIDERAÇÕES FINAIS

Com o intuito de contribuir com as discussões acerca das possibilidades e os desafios socioculturais e técnicos de organizar e representar informações e conhecimentos em ambientes digitais, esse trabalho buscou entender como os processos de enriquecimento de dados se caracterizam pela perspectiva da ORC, fazendo um paralelo entre o que pode ser abarcado como instrumentos, processos e produtos da ORC e um conjunto de publicações que abordam o processo como teoria e prática.

Apesar de não exaustivo, o levantamento e a análise das publicações permitiram delinear algumas percepções acerca da relação proposta, ficando evidente que o enriquecimento semântico, de um ponto de vista sociocultural nos estudos da ORC, pode ser considerado uma atividade basilar na medida que articula diferentes noções, atores e garantias para o usuário. Contudo, vale pontuar que ao olhar através de uma perspectiva sociocultural, é necessário reconhecer que os aspectos entorno do enriquecimento, apesar de possuírem o potencial para ampliar a diversidade de informações e conhecimentos, também incorrem na possibilidade de, tal como em qualquer processo que se utilize de sistemas que representam e organizam conhecimentos, reforçar ou reafirmar posturas dominantes que já estão consolidadas.

Portanto, o enriquecimento, para além dos aspectos técnicos como automatização, processamento de grandes volumes de dados e redução de custos, tem o potencial de descentralizar o processo de organização e representação dos dados, dando a possibilidade de utilizar múltiplas perspectivas de atribuição de sentido em diversas camadas de informação, sem, contudo, deixar de vincular a elas garantias que as justifique e valide na relação com o contexto.

REFERÊNCIAS

BARDIN, L. **Análise de conteúdo**. Lisboa, Portugal: Edições 70, 2011.

BARITÉ, M. Organización del conocimiento: un nuevo marco teorico-conceptual en bibliotecología y documentación. *In*: CARRARA, K. (Org.). **Educação, universidade e pesquisa**: textos completos do III simpósio em filosofia e ciência: paradigmas do conhecimento no final do milênio. Marília: Unesp; São Paulo: FAPESP, 2001. p. 35-60.

BARITÉ, M. Towards a general conception of warrants: first notes. **KO KNOWLEDGE ORGANIZATION**, [s.l.], v. 46, n. 8, p. 647-655, 2019. Disponível em: https://www.researchgate.net/publication/339174803_Towards_a_General_Conception_of_Warrants_First_Notes. Acesso em: 27 jun. 2023.

BEGHTOL, C. Universal concepts, cultural warrant, and cultural hospitality. In: LÓPEZ HUERTAS, M. J. (ed.). **Challenges in knowledge representation and organization for the 21st century**: integration of knowledge across boundaries. Würzburg: ERGON, 2002. p. 45-49.

BULLARD, J. Warrant as a means to study classification system design. **Journal of Documentation**, [s.l.], v. 73, n. 1, p. 75-90, 2017.

CAPURRO, R. Epistemología y ciencia de la información. **Enlace**, [s.l.], v.4, n.1, pp.11-29, 2007. Disponível em: http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-75152007000100002. Acesso em: 27 jun. 2023.

DALBELLO, M. Digital convergence: the past in the present. In: Richards, P. S. Wiegand, W. A. Dalbello, M. (Eds.). **A History of Modern Librarianship**: constructing the heritage of western cultures: constructing the heritage of western cultures. Santa Barbara, CA; Denver, CO: ABC-CLIO; LLC, 2015. p. 205.

EGAN, P. In search of the item: Irish traditional music, archived fieldwork and the digital. **Archival Science**, [s. l.], v. 23, n. 1, p. 45-63, 2023. Disponível em: <https://doi.org/10.1007/s10502-021-09382-z>. Acesso em: 27 jun. 2023.

ESTEVANOVIC, M. P.; BAX, M. P. Dados abertos governamentais: uma metodologia para publicar dados semânticos no estado de Minas Gerais. **Múltiplos Olhares em Ciência da Informação**, Minas Gerais, v. 12, 2022. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/40048>. Acesso em: 27 jun. 2023.

FROHMANN, B. Taking information policy beyond information science: applying the actor network theory. In: ANNUAL CONFERENCE: CANADIAN ASSOCIATION FORMIN FORMATION, 23., 1995. **Anais [...]**. Edmonton, Alberta: CAIS/ACSI, 1995. Disponível em: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=40176306291e2cf81cae4b4b6c9412853ae54031>. Acesso em; 27 jun. 2023.

GARCÍA-GUTIÉRREZ, A. Knowledge organization from a “culture of the border”: towards a transcultural ethics of mediation. In López-Huertas, M. J. (Ed.). **Challenges in knowledge representation and organization for the 21st century**: integration of knowledge across boundaries. Granada, Espanha: Ergon, 2002.

GILLILAND, A. J. Setting the stage. In: BACA, M. (Ed.). **Introduction to metadata**. Los Angeles, CA: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 30 jun. 2023.

GUIMARÃES, J. A. C.; EVANGELISTA, I. V.; LUZ, G. A. M.; OSAWA, H. F. A dimensão cultural da organização do conhecimento: uma análise de comunidades epistêmicas no contexto internacional da Ciência da Informação. **Scire: representación y organización del conocimiento**, v. 25, n. 1, p. 25-36, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/168185>. Acesso em: 30 jun. 2023.

ISAAC, A.; MANGUINHAS, H.; STILLER, J.; CHARLES, V. **Report on enrichment and evaluation**. the Hague, Netherlands: Europeana Task Force on Enrichment and Evaluation, 2015.

KLEIN, E.; HALLER, S.; GSCHWEND, A.; JOVANOVIK, M. Sustainable Linked Open Data Creation: an experience report. In: JANSSEN, M. F. W. H. A.; BANNISTER F.; GLASSEY, Olivier; SCHOLL, H. J.; TAMBOURIS, E.; WIMMER, M. A.; MACINTOSH, A. **Electronic Government and Electronic Participation**. Amsterdã: IOS Press, 2016. p. 99-110. Disponível em: <https://core.ac.uk/download/pdf/231206076.pdf>. Acesso em: 30 jun. 2023.

KNAPP, E. D.; LANGILL, J. T. Exception, anomaly, and threat detection. In: KNAPP, E. D.; LANGILL, J. T. (ed.). **Industrial Network Security** - securing critical infrastructure networks for smart grid, SCADA, and Other Industrial Control Systems. 2.ed. Boston: Syngress, 2015. p. 323-350. Disponível em: <https://www.sciencedirect.com/book/9780124201149/industrial-network-security>. Acesso em: 27 jun. 2023.

KORS, J. A.; CLEMATIDE, S.; AKHONDI, S. A.; VAN MULLIGEN, E. M.; REBHOLZ-SCHUHMANN, D. A multilingual gold-standard corpus for biomedical concept recognition: the Mantra GSC. **Journal of the American Medical Informatics Association**, Oxford, v. 22, n. 5, p. 948-956, 2015. Disponível em: <https://doi.org/10.1093/jamia/ocv037>. Acesso em: 27 jun. 2023.

LARA, M. L. G. de; MENDES, L. C. Referências socioculturais na Organização do Conhecimento. INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION - ISKO Brasil, 5, 2017. **Proceedings** [...]. Recife, PE: Ed. UFPE, 2017. p. 375-382. Disponível em: <https://isko.org.br/wp-content/uploads/2021/05/Proceedings-ISKO-Brasil-2017.pdf>. Acesso em: 30 jun. 2023.

LEE, W-C. **Operationalizing Cultural Warrant in Knowledge Organization**. 2021. Thesis (Doctor of Philosophy) – University of Washington: Seattle, Washington, 2021. Disponível em: <https://digital.lib.washington.edu/researchworks/handle/1773/47600>. Acesso em: 30 jun. 2023.

LIRA, M. A. B. de. **Uma abordagem para enriquecimento semântico de metadados para publicação de dados abertos**. 2014. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Pernambuco, Centro de Informática, Recife, 2014. Disponível em: <https://repositorio.ufpe.br/handle/123456789/11570>. Acesso em: 27 jun. 2023.

LUO, X; TAN, X.; WANG, X. Semantically enriched presentation for cultural heritage image: a POI-based perspective. In: ACM/IEEE JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL), 18, 2019. **Proceedings** [...]. IEEE, Champaign, Illinois, 2019. p. 410-411. Disponível em: <https://doi.org/10.1109/JCDL.2019.00094>. Acesso em: 30 jun. 2023.

MACHADO, D. O. F.; ARAKAKI, A. C. S. Schema.org para catálogos digitais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 27, n. 1, p. 1–21, 2022. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/87046>. Acesso em: 30 jun. 2023.

MANGUINHAS, H.; FREIRE, N.; ISAAC, A.; STILLER, J; CHARLES, V; SOROA, A; ALEXIEV, V. Exploring comparative evaluation of semantic enrichment tools for cultural heritage metadata. In: INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF DIGITAL LIBRARIES, 20, 2016, Hannover, September 5–9, 2016, **Proceedings** [...]. Springer International Publishing, Hannover, 2016. p. 266-278. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-43997-6_21. Acesso em: 30 jun. 2023

NAZÁRIO, C.; DIAS, C. C. Avaliação do Europeana Data Model para o enriquecimento semântico de objetos. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, n. Especial, 2021. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/37189>. Acesso em: 30 jun. 2023.

NEWMAN, D; HAGEDORN, K; CHEMUDUGUNTA, C; SMYTH, P. Subject metadata enrichment using statistical topic models. *In*: ACM/I IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL), 18. 2019. **Proceedings** [...]. IEEE, Champaign, Illinois, 2019. p. 366-375. Disponível em: <https://dl.acm.org/doi/abs/10.1145/1255175.1255248>. Acesso em: 27 jun. 2023.

TORINO, E.; TREVISAN, G. L.; CONEGLIAN, C. S.; BOTEAGA, L. C.; SEGUNDO, J. E. S.; VIDOTTI, S. A. B. G. Enriquecimento semântico para a disponibilização de dados abertos: teoria e prática. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 25, p. 1-19, 2020. Disponível em: <http://riut.utfpr.edu.br/jspui/handle/1/4844>. Acesso em: 30 jun. 2023.

VORNDRAN, A.; GRUND, S. Metadata Sharing—How to Transfer Metadata Information among Work Cluster Members. **Cataloging & Classification Quarterly**, [s. l.], v. 59, n. 8, p. 757-774, 2021.

YANG, Bisheng; ZHANG, Yunfei Zhang. Abordagem de mineração de padrões para fusãocrowdsourcing de redes rodoviárias com POIs, International Journal of Geographical Information Ciência, [s. l.], v. 29, n. 5, p. 786-80, 2015. Available in: https://www.researchgate.net/publication/273766395_Pattern-mining_approach_for_conflating_crowdsourcing_road_networks_with_POIs Access in: 2024.

ZENG, M. L. Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. **El profesional de la información**, [s. l.], v. 28, n. 1, 2019. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=6839005>. Acesso em: 27 jun. 2023.