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

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

GT 8 – Dados, Informação e Tecnologia

DADOS LIGADOS E NANOPUBLICAÇÕES EM CONTEXTOS GLAM: UM *WORKFLOW* METODOLÓGICO

LINKED DATA AND NANOPUBLICATIONS IN GLAM CONTEXTS*: A METHODOLOGICAL WORKFLOW*

Henrique Monteiro Cristovão – Universidade Federal do Espírito Santo (UFES)

Modalidade: Trabalho Completo

Resumo: Galerias, bibliotecas, arquivos e museus (GLAM) lidam com acervos heterogêneos, descritos por diferentes sistemas, planilhas, catálogos, vocabulários e práticas institucionais. Essa diversidade dificulta a integração, a interoperabilidade, a recuperação, a rastreabilidade e o reuso dos dados culturais. Nesse contexto, dados ligados, ontologias operacionais e nanopublicações podem contribuir para estruturar, conectar, consultar e publicar informações culturais de forma mais interoperável e verificável. O objetivo deste artigo é apresentar um *workflow* metodológico para orientar a representação, conexão, recuperação e publicação semântica de dados em contextos GLAM, articulando preparação de dados, modelagem conceitual, vocabulários controlados, ontologias operacionais, alinhamento semântico, grafos RDF, consultas SPARQL e nanopublicações. A proposta foi construída a partir de cinco abordagens metodológicas de referência: Ontology Development 101, OntoForInfoScience, Linked Open Terms, Linked Data for Libraries, Archives and Museums e SABiOx. Metodologicamente, trata-se de uma pesquisa aplicada, exploratória e propositiva, cujo resultado é um *workflow* organizado em cinco blocos analíticos, 11 camadas e 32 etapas. O *workflow* vem sendo experimentado em uma disciplina de mestrado em Ciência da Informação, mas o artigo não se configura como relato de experiência concluído. Conclui-se que a principal contribuição da proposta está na integração entre desenvolvimento de ontologias, práticas de dados ligados em GLAM e publicação granular de afirmações com proveniência, oferecendo um percurso metodológico sintético, documentado e aplicável tanto a pequenos recortes experimentais quanto a iniciativas incrementais de organização semântica de dados culturais.

Palavras-chave: dados ligados; ontologia operacional; nanopublicação; GLAM; interoperabilidade semântica.

Abstract: Galleries, libraries, archives, and museums (GLAM) deal with heterogeneous collections described through different systems, spreadsheets, catalogues, vocabularies, and institutional practices. This diversity hinders the integration, interoperability, retrieval, traceability, and reuse of cultural data. In this context, linked data, operational ontologies, and nanopublications can contribute to structuring, connecting, querying, and publishing cultural information in a more interoperable and verifiable way. The objective of this article is to present a methodological workflow to guide the semantic representation, connection, retrieval, and publication of data in GLAM contexts, articulating data preparation, conceptual modeling, controlled vocabularies, operational ontologies, semantic alignment, RDF graphs, SPARQL queries, and nanopublications. The proposal was developed based on five methodological reference approaches: Ontology Development 101, OntoForInfoScience, Linked Open Terms, Linked Data for Libraries, Archives and Museums, and SABiOx. Methodologically, this is an applied, exploratory, and propositional study, resulting in a workflow organized into five analytical blocks, 11 layers, and 32 stages. The workflow is being tested in a master's-level course in Information Science, but the article is not configured as a completed experience report. It concludes that the main contribution of the proposal lies in integrating ontology development, linked data practices in GLAM, and the granular publication of assertions with provenance, offering a concise and documented methodological path applicable both to small experimental subsets and to incremental initiatives for the semantic organization of cultural data.
Keywords: linked data; operational ontology; nanopublication; GLAM; semantic interoperability.

1 INTRODUÇÃO

Instituições GLAM (sigla derivada de *galleries*, *libraries*, *archives* and *museums*, em português, galerias, bibliotecas, arquivos e museus) desempenham papel central na organização, preservação, representação e circulação de objetos culturais, documentos, obras, coleções, agentes, lugares, eventos e memórias. Entretanto, os dados produzidos ou mantidos por essas instituições frequentemente se encontram distribuídos em planilhas, catálogos, sistemas legados, bases locais, inventários e registros institucionais com diferentes níveis de padronização, documentação e interoperabilidade.

A heterogeneidade desses dados dificulta sua integração e reutilização. Registros de objetos culturais podem apresentar variações terminológicas, campos multivalorados, ausência de identificadores, inconsistências, lacunas documentais, pouca explicitação de proveniência e dificuldades de alinhamento externo. Em contextos GLAM, esses problemas também envolvem decisões curatoriais, documentais, classificatórias, históricas e institucionais. Nesse cenário, a Web Semântica propõe a representação de dados com

significado explícito e processável por máquinas, ampliando possibilidades de integração e descoberta de informação (Berners-Lee; Hendler; Lassila, 2001). Os dados ligados, por sua vez, fundamentam-se no uso de URIs, padrões como RDF e links entre recursos distribuídos na Web (Berners-Lee, 2006), constituindo práticas para publicar e interligar dados estruturados e favorecer descoberta, integração e reuso (Bizer; Heath; Berners-Lee, 2009).

As ontologias são relevantes nesse processo porque permitem explicitar conceitos, classes, propriedades e relações de um domínio. A noção de ontologia como especificação explícita de uma conceitualização tornou-se referência para o desenvolvimento de artefatos formais de representação (Gruber, 1993). Em perspectiva complementar, uma ontologia pode ser entendida como uma especificação formal e explícita de uma conceitualização compartilhada (Studer; Benjamins; Fensel, 1998). No contexto deste artigo, adota-se a distinção entre modelagem conceitual, ontologias de referência e ontologias operacionais. A modelagem conceitual antecede a implementação e organiza, em nível abstrato, os principais conceitos, relações, restrições e decisões de representação do domínio. O workflow não prescreve uma linguagem gráfica específica nem torna obrigatória a elaboração prévia de uma ontologia de referência, admitindo diferentes níveis de formalização da modelagem conceitual. A ontologia operacional, por sua vez, corresponde à formalização computacional dessa modelagem, voltada à interpretação por máquina e à aplicação em determinado recorte de domínio (Falbo, 2014). Neste artigo, ela é compreendida como uma representação computacional pequena, documentada e suficiente para orientar a construção de grafos RDF, consultas SPARQL, alinhamentos semânticos e nanopublicações.

Os princípios FAIR também sustentam a proposta, ao indicarem que dados e metadados devem ser localizáveis, acessíveis, interoperáveis e reutilizáveis, com ênfase na possibilidade de processamento por máquinas (Wilkinson *et al*., 2016). A interoperabilidade semântica, contudo, exige mais do que formatos compatíveis, pois depende da explicitação de categorias, relações e compromissos conceituais por meio de ontologias (Guizzardi, 2020).

Em contextos GLAM, a construção de dados ligados depende de processos de limpeza, reconciliação, enriquecimento e publicação de metadados culturais (Hooland; Verborgh, 2014). Além disso, modelos de domínio, como o CIDOC CRM, contribuem para a interoperabilidade semântica de informações de patrimônio cultural ao representar entidades, agentes, eventos, lugares, períodos, produção, aquisição, custódia e proveniência (CIDOC CRM Special Interest Group, 2024; Doerr, 2003).

As nanopublicações completam esse cenário ao possibilitar a publicação de afirmações específicas acompanhadas de proveniência e informações de publicação. Sua estrutura distingue, em termos gerais, a asserção, a proveniência e as informações da própria publicação (Groth; Gibson; Velterop, 2010). No campo da Ciência da Informação, essa abordagem amplia as possibilidades de representação do conhecimento ao permitir que afirmações sejam tratadas como unidades granulares, rastreáveis, verificáveis e reutilizáveis (Sayão; Sales, 2024), aspecto especialmente relevante para contextos GLAM, nos quais dados culturais frequentemente dependem de fontes, interpretações, atribuições e decisões curatoriais.

Apesar da existência de metodologias para construção de ontologias, desenvolvimento de vocabulários e publicação de dados ligados, observa-se que essas abordagens tendem a privilegiar partes específicas do processo, como desenvolvimento ontológico, publicação de vocabulários, limpeza e reconciliação de metadados GLAM ou distinção entre ontologias de referência e ontologias operacionais. A ausência de um percurso que articule essas contribuições, estendendo-as à construção e consulta de grafos RDF e à publicação de nanopublicações, justifica a proposição de um workflow metodológico integrador aplicável à organização semântica de dados culturais em contextos GLAM.

Este artigo parte da seguinte questão: como estruturar um *workflow* metodológico, fundamentado em abordagens já existentes, para orientar a representação, conexão, recuperação e publicação semântica de dados em contextos GLAM, articulando dados ligados, ontologias operacionais e nanopublicações?

O objetivo geral é apresentar um *workflow* metodológico para orientar a representação, conexão, recuperação e publicação semântica de dados em contextos GLAM. Como objetivos específicos, busca-se: sintetizar abordagens metodológicas de referência relacionadas ao desenvolvimento de ontologias e dados ligados; identificar contribuições e limites dessas abordagens frente ao percurso metodológico pretendido; descrever o *workflow* proposto em blocos analíticos organizados em camadas e etapas; explicitar os artefatos produzidos em cada camada; e discutir a contribuição da proposta em relação às abordagens que a fundamentam. A construção do *workflow* dialoga com cinco abordagens metodológicas de referência, sintetizadas na seção de procedimentos metodológicos.

O *workflow* vem sendo experimentado em uma disciplina de mestrado em Ciência da Informação dedicada a dados ligados e nanopublicações para Ciência Aberta e Patrimônio

Cultural. A disciplina utiliza uma situação-problema inspirada na metodologia de ensino-aprendizagem denominada de *problem-based learning* (PBL), na qual uma instituição GLAM busca melhorar a organização, integração e reuso de informações sobre parte de seu acervo. No entanto, o artigo não se configura como relato de experiência concluído, nem apresenta avaliação dos resultados dos estudantes. A disciplina é mencionada apenas como contexto inicial de experimentação, enquanto o foco do artigo recai sobre a proposição e discussão do *workflow* metodológico.

2 PROCEDIMENTOS METODOLÓGICOS

Este artigo apresenta uma pesquisa aplicada, exploratória e propositiva. O principal resultado é um *workflow* metodológico para orientar a representação, conexão, recuperação e publicação semântica de dados em contextos GLAM, articulando dados ligados, ontologias operacionais e nanopublicações.

A proposta foi construída por seleção intencional de cinco abordagens metodológicas, identificadas em levantamento bibliográfico dirigido e não exaustivo sobre ontologias, vocabulários, dados ligados e aplicações GLAM. Adotaram-se quatro critérios: relação com desenvolvimento de ontologias ou vocabulários; aderência a dados ligados; pertinência para GLAM; e articulação com preparação de dados, interoperabilidade, recuperação e proveniência. Compararam-se objetivos, atividades, artefatos e cobertura dessas dimensões, sintetizando contribuições complementares e lacunas por correspondência funcional.

Com base nesses critérios, selecionaram-se cinco abordagens. A Ontology Development 101 contribui com definição de domínio e escopo, reuso, termos, classes, propriedades e instâncias (Noy; McGuinness, 2001). A OntoForInfoScience aproxima desenvolvimento ontológico, Ciência da Informação, Organização do Conhecimento e análise de domínio (Mendonça; Almeida, 2016). A Linked Open Terms (LOT) enfatiza requisitos, implementação, publicação, documentação e manutenção de ontologias e vocabulários em dados ligados (Poveda-Villalón, 2018). Linked Data for Libraries, Archives and Museums articula limpeza, reconciliação, enriquecimento e publicação de metadados culturais (Hooland; Verborgh, 2014). A SABiOx sistematiza a passagem entre conceitualização e implementação, distinguindo ontologias de referência e operacionais (Aguiar; Souza, 2024; Falbo, 2014;). Nenhuma cobre isoladamente o percurso da preparação dos dados culturais à

publicação de afirmações com proveniência; por isso, suas contribuições foram integradas no workflow.

A partir dessa síntese metodológica, as contribuições foram reorganizadas em um workflow composto por cinco blocos analíticos, 11 camadas e 32 etapas. A ordenação adota uma progressão incremental e didático-operacional, considerando que equipes envolvidas em projetos GLAM podem reunir profissionais com diferentes níveis de familiaridade com tecnologias da Web Semântica. Por isso, prevê-se uma experimentação preliminar do modelo com dados e consultas SPARQL antes dos alinhamentos externos, seguida da consolidação posterior do grafo, conforme detalhado na Seção 3.1.

3 RESULTADOS E DISCUSSÃO

Esta seção apresenta o workflow metodológico proposto e discute sua contribuição em relação às abordagens metodológicas que o fundamentam.

3.1 APRESENTAÇÃO DO *WORKFLOW* METODOLÓGICO PROPOSTO

O workflow metodológico proposto, apresentado no Quadro 1, orienta a construção incremental de aplicações semânticas em contextos GLAM, articulando preparação de dados, modelagem conceitual, ontologias operacionais, alinhamento semântico, grafos RDF, consultas SPARQL e nanopublicações.

Quadro 1 - Síntese do *workflow* metodológico organizado por blocos, camadas, etapas e artefatos

Bloco A: Preparação e compreensão do domínio			
Camada	**Etapas**	**Função metodológica**	**Artefatos resultantes**
Camada I: Análise do domínio, caracterização da base e preparação dos dados	**Etapa 1**: Análise e delimitação do domínio GLAM e caracterização da base de dados; **Etapa 2**: Formulação de perguntas de competência; **Etapa 3**: Preparação e avaliação da qualidade dos dados; **Etapa 4**: Reconciliação preliminar de entidades e valores com fontes externas confiáveis.	Definir o escopo empírico e informacional, qualificando a base de dados que sustentará a modelagem semântica.	Recorte do domínio, perguntas de competência, base preparada, análise de qualidade e registros de reconciliação.

Camada II: Modelagem conceitual do domínio	**Etapa 5**: Levantamento e definição preliminar dos conceitos do domínio; **Etapa 6**: Modelagem conceitual de entidades, atributos, relações e eventos; **Etapa 7**: Classificação preliminar dos elementos segundo sua representação prevista: classes ou propriedades na TBox, conceitos controlados na CBox, indivíduos ou fatos na ABox, ou valores literais.	Construir uma representação conceitual do domínio antes de sua formalização, identificando entidades, atributos, relações, eventos e o tipo de representação semântica previsto para os elementos.	Dicionário preliminar de conceitos, modelo conceitual e quadro dos tipos de representação semântica previstos.
Bloco B: Organização e formalização semântica			
Camada	**Etapas**	**Função metodológica**	**Artefatos resultantes**
Camada III: Organização terminológica (CBox - *Category Box*)	**Etapa 8**: Delimitação dos conceitos destinados aos vocabulários controlados; **Etapa 9**: Definição da política geral de URIs para os artefatos semânticos; **Etapa 10**: Estruturação e implementação da CBox em SKOS.	Organizar a dimensão categorial e terminológica do domínio, distinguindo-a da estrutura ontológica e dos indivíduos, e implementar a CBox em SKOS.	Política de URIs e **vocabularios.ttl** (CBox, com esquemas de conceitos, conceitos, rótulos e relações conceituais internas em SKOS).
Camada IV: Ontologia operacional (TBox - *Terminological Box*)	**Etapa 11**: Criação de classes e propriedades locais; **Etapa 12**: Definição dos axiomas necessários ao recorte do domínio; **Etapa 13**: Implementação da TBox em RDF/OWL e inclusão dos metadados da ontologia.	Formalizar computacionalmente o modelo conceitual como TBox local, definindo classes, propriedades e axiomas suficientes para o recorte do domínio.	**ontologia.ttl** (TBox), metadados da ontologia e documentação das decisões de formalização.
Bloco C: Construção preliminar e interoperabilidade semântica			
Camada	**Etapas**	**Função metodológica**	**Artefatos resultantes**
Camada V: Mapeamento dos dados para RDF, povoamento preliminar e	**Etapa 14**: Mapeamento dos dados de origem para o modelo semântico em RDF; **Etapa 15**: Povoamento preliminar da ABox em RDF;	Estabelecer correspondências entre campos e valores da base de origem e elementos da TBox e da CBox,	Tabela de mapeamento inicial, **recursos.ttl** (ABox preliminar), consultas SPARQL

consulta exploratória (ABox - *Assertional Box*)	**Etapa 16**: Consultas SPARQL exploratórias; **Etapa 17**: Revisão preliminar dos dados, do modelo e/ou do mapeamento a partir dos resultados das consultas.	definir regras de criação de recursos e literais da ABox, produzir uma ABox preliminar e submetê-la a consultas SPARQL exploratórias.	exploratórias e registro de ajustes.
Camada VI: Interoperabilidade e alinhamento semântico	**Etapa 18**: Seleção de referenciais semânticos externos pertinentes, como ontologias, modelos de referência, vocabulários controlados, bases de autoridade e outros recursos semânticos externos; **Etapa 19**: Definição dos alinhamentos ontológicos (TBox), terminológicos (CBox) e de entidades (ABox); **Etapa 20**: Implementação dos alinhamentos em **interoperabilidade.ttl**.	Estabelecer alinhamentos ontológicos, terminológicos e de entidades entre elementos locais e referenciais externos pertinentes, ampliando a interoperabilidade semântica.	Quadro de alinhamentos e **interoperabilidade.ttl**, reunindo alinhamentos ontológicos (TBox), terminológicos (CBox) e de entidades (ABox).
Bloco D: Consolidação, povoamento final e recuperação da informação			
Camada	**Etapas**	**Função metodológica**	**Artefatos resultantes**
Camada VII: Revisão do mapeamento e povoamento final da ABox	**Etapa 21**: Revisão do mapeamento e sua complementação, quando necessária, após os alinhamentos semânticos; **Etapa 22**: Povoamento final da ABox conforme o mapeamento consolidado.	Revisar e complementar o mapeamento dos dados após os alinhamentos semânticos e executar o povoamento final da ABox, produzindo a versão consolidada dos recursos RDF que comporão o grafo semântico integrado.	Tabela de mapeamento revisada e **recursos.ttl** (ABox final).
Camada VIII: Consulta e recuperação	**Etapa 23**: Consultas SPARQL orientadas às perguntas de competência;	Avaliar a capacidade informacional do grafo semântico	Consultas SPARQL, resultados, registro das respostas às

da informação no grafo RDF	**Etapa 24**: Consultas federadas, quando viáveis.	consolidado por meio de consultas vinculadas às perguntas de competência e, quando pertinente, explorar a recuperação integrada com fontes externas.	perguntas de competência e consultas federadas, quando realizadas.
Bloco E: Nanopublicação, proveniência e rastreabilidade			
Camada	**Etapas**	**Função metodológica**	**Artefatos resultantes**
Camada IX: Seleção, contextualização e proveniência das afirmações	**Etapa 25**: Seleção das afirmações candidatas à nanopublicação; **Etapa 26**: Modelagem da proveniência, do contexto, das evidências e das responsabilidades associadas; **Etapa 27**: Definição dos elementos e vocabulários de proveniência e informação de publicação.	Selecionar afirmações relevantes e explicitar sua proveniência, contexto, evidências e responsabilidades, preparando-as para publicação granular.	Afirmações candidatas, modelo de proveniência e contexto e especificação dos elementos e vocabulários de proveniência e informação de publicação.
Camada X: Estruturação, implementação e publicação das nanopublicações	**Etapa 28**: Estruturação em Head, Assertion, Provenance e Publication Info; **Etapa 29**: Implementação das nanopublicações como RDF Dataset serializado em TriG; **Etapa 30**: Publicação ou disponibilização das nanopublicações em infraestrutura apropriada, quando viável.	Estruturar e implementar as nanopublicações, utilizando o Head para referenciar os grafos nomeados de Assertion, Provenance e Publication Info, e, quando viável, publicá-las ou disponibilizá-las em infraestrutura apropriada.	**nanopublicacoes.trig**, padrão documentado de estruturação e registro de publicação/disponibilização, quando realizada.
Camada XI: Recuperação, rastreabilidade e reuso	**Etapa 31**: Recuperação dos componentes e informações das nanopublicações, incluindo asserções, proveniência e informações de publicação;	Recuperar os componentes das nanopublicações e avaliar sua capacidade de sustentar	Consultas e resultados de recuperação, avaliação da rastreabilidade,

	Etapa 32: Avaliação da rastreabilidade, citabilidade e potencial de reuso das nanopublicações.	rastreabilidade, citabilidade, verificabilidade e potencial de reuso das afirmações.	citabilidade e potencial de reuso.

Fonte: elaboração própria (2026).

A organização do workflow combina três níveis complementares. O primeiro nível é composto por cinco blocos analíticos: o Bloco A, dedicado à preparação e compreensão do domínio; o Bloco B, à organização e formalização semântica; o Bloco C, à construção preliminar e interoperabilidade semântica; o Bloco D, à consolidação, povoamento final e recuperação da informação; e o Bloco E, à nanopublicação, proveniência e rastreabilidade. O segundo nível é composto por 11 camadas distribuídas entre esses blocos. O terceiro compreende 32 etapas, que detalham o percurso de cada camada.

Na tradição das Lógicas de Descrição, bases de conhecimento distinguem TBox (*Terminological Box*) e ABox (*Assertional Box*) (Baader; McGuinness; Nardi; Patel-Schneider., 2003): a primeira reúne a estrutura terminológica e ontológica; a segunda, indivíduos e afirmações factuais. Para explicitar a dimensão categorial dos vocabulários controlados, o workflow incorpora a CBox (Category Box), denominação usada por McComb (2023) para categorias e taxonomias. Aqui, a CBox é operacionalizada em SKOS, separando conceitos controlados da TBox e da ABox.

As camadas e etapas são sucessivas, mas flexíveis, permitindo revisões de escopo, dados, conceitos, alinhamentos e implementação. O *workflow* documenta progressivamente decisões metodológicas e curatoriais. Seus dois ciclos de mapeamento, povoamento e consulta reforçam o caráter incremental: o primeiro é exploratório e diagnóstico, retroalimentando dados, modelo e mapeamento; após os alinhamentos, o segundo consolida o povoamento e avalia o grafo por perguntas de competência. A verificação sintática pode ocorrer na geração ou importação dos dados; validações estruturais, quando necessárias, podem ser incorporadas transversalmente com SHACL, sem constituir etapa autônoma (Knublauch; Kontokostas, 2017).

Nesse percurso, a ontologia operacional, correspondente à TBox local, resulta da modelagem conceitual anterior. Essa modelagem pode assumir diferentes níveis de formalização e notações gráficas, desde que explicite entidades, atributos, relações, eventos

e decisões de representação. Em aplicações que demandem maior rigor ontológico, pode ser apoiada por ontologias de fundamentação ou de topo, como a Unified Foundational Ontology (UFO) e a Basic Formal Ontology (BFO), e por linguagens de modelagem ontologicamente fundamentadas, como OntoUML, baseada em UFO (Guizzardi; Wagner; Guizzardi, 2015; ISO/IEC, 2021); também pode recorrer a ontologias de referência ou modelos conceituais consolidados pertinentes ao domínio, como o CIDOC CRM no patrimônio cultural. O workflow, entretanto, não condiciona a construção da ontologia operacional à elaboração prévia de uma ontologia de referência formalmente fundamentada. CBox e ABox permanecem separadas, respectivamente, para conceitos controlados e dados de instância.

Para materializar essa arquitetura de forma modular, o workflow organiza os principais artefatos semânticos em cinco arquivos. O arquivo **ontologia.ttl** contém a TBox local, com classes, propriedades, axiomas e metadados da ontologia; **vocabularios.ttl** contém a CBox, representada em SKOS por esquemas de conceitos, conceitos, rótulos e relações conceituais internas; **recursos.ttl** contém a ABox, formada pelos indivíduos, valores literais e fatos derivados da base de dados; **interoperabilidade.ttl** registra separadamente os alinhamentos ontológicos, terminológicos e de entidades entre elementos locais e referenciais externos; e **nanopublicacoes.trig** reúne as nanopublicações como RDF Dataset serializado em TriG, preservando a separação entre Head, Assertion, Provenance e Publication Info. Essa organização física explicita a distinção entre estrutura ontológica, organização terminológica, dados de instância, alinhamentos externos e publicação granular de afirmações. Por exemplo, um registro **OBJ001** pode ser instanciado em **recursos.ttl** segundo classes e propriedades definidas em **ontologia.ttl**, utilizar um conceito de **vocabularios.ttl**, receber alinhamentos externos registrados em **interoperabilidade.ttl** e ter uma afirmação selecionada publicada em **nanopublicacoes.trig** acompanhada de sua proveniência.

O Bloco A reúne as Camadas I e II, voltadas à preparação, compreensão e modelagem conceitual do domínio. Nesse bloco, o workflow delimita o recorte GLAM, caracteriza a base de dados, avalia sua qualidade, formula perguntas de competência, levanta os conceitos relevantes e constrói uma representação conceitual de entidades, atributos, relações e eventos. A análise de domínio fundamenta a delimitação conceitual e contextual do recorte trabalhado (Hjørland, 2002), enquanto a avaliação da qualidade dos dados orienta a identificação de problemas que podem comprometer uso, integração e reuso (Wang; Strong,

1996). A documentação e os metadados sustentam a interpretação, a gestão e a reutilização dos dados (Riley, 2017; Sayão; Sales, 2015).

O Bloco B reúne as Camadas III e IV, dedicadas à organização e formalização semântica. A Camada III delimita os conceitos destinados aos vocabulários controlados, estabelece uma política geral de URIs que privilegie identificadores estáveis e persistentes e implementa a CBox em SKOS, aproximando o workflow das práticas de Organização do Conhecimento e da representação de sistemas terminológicos (Isaac; Summers, 2023; Zeng, 2008). A Camada IV formaliza a modelagem conceitual como ontologia operacional, isto é, como TBox local composta por classes, propriedades e axiomas suficientes para o recorte do domínio. A distinção entre modelagem conceitual e ontologia operacional dialoga com a separação entre ontologias de referência e operacionais (Aguiar; Souza, 2024; Falbo, 2014).

O Bloco C reúne as Camadas V e VI, dedicadas à construção preliminar do grafo e à interoperabilidade semântica. Na Camada V, os dados de origem são mapeados para o modelo semântico em RDF, produzindo uma ABox preliminar que é submetida a consultas SPARQL exploratórias; seus resultados podem motivar ajustes nos dados, no modelo ou no próprio mapeamento. Na Camada VI, são selecionados referenciais externos e implementados, em **interoperabilidade.ttl**, alinhamentos ontológicos da TBox, terminológicos da CBox e de entidades da ABox. Essa separação explicita que mapeamento de dados e alinhamento semântico são operações distintas, embora articuladas, e que a interoperabilidade depende de correspondências justificadas entre estruturas, vocabulários, significados e entidades (Zeng, 2019).

O Bloco D reúne as Camadas VII e VIII, voltadas à consolidação, ao povoamento final e à recuperação da informação. Após os alinhamentos, o mapeamento é revisto e complementado quando necessário, e a ABox é novamente povoada conforme o mapeamento consolidado. Em seguida, consultas SPARQL orientadas às perguntas de competência avaliam a capacidade informacional do grafo semântico consolidado; consultas federadas podem complementar essa recuperação quando tecnicamente viáveis.

O Bloco E reúne as Camadas IX, X e XI, dedicadas à nanopublicação, proveniência e rastreabilidade. Na seleção, priorizam-se afirmações relacionadas às perguntas de competência ou cuja fonte, evidência, atribuição, incerteza, responsabilidade ou potencial de reuso justifiquem sua publicação granular. A proveniência pode ser representada computacionalmente com PROV-O, explicitando entidades, atividades, agentes e relações de

derivação associadas à produção da afirmação (Lebo; Sahoo; McGuinness, 2013). As nanopublicações são estruturadas em Head, Assertion, Provenance e Publication Info e implementadas como RDF Dataset em TriG; a Publication Info pode registrar autoria, data e licença. Para identificação e integridade das nanopublicações, podem ser empregados Trusty URIs (Kuhn; Dumontier, 2014); em caso de alteração, a manutenção deve preservar o artefato anterior e documentar a nova versão, evitando sua sobrescrita. Por fim, os componentes são recuperados e avaliados quanto à rastreabilidade, citabilidade, verificabilidade e potencial de reuso.

3.2 CONTRIBUIÇÃO DO *WORKFLOW* PROPOSTO EM RELAÇÃO ÀS METODOLOGIAS DE REFERÊNCIA

O workflow proposto não pretende substituir as cinco abordagens metodológicas analisadas, mas articulá-las em um percurso integrado voltado à organização semântica de dados em contextos GLAM. A comparação evidencia que cada abordagem cobre dimensões relevantes, mas requer complementação para atender ao percurso metodológico integrado proposto.

Em relação à Ontology Development 101, a proposta preserva a importância da definição de domínio, escopo, classes, propriedades e instâncias (Noy; McGuinness, 2001). No entanto, essa abordagem tem caráter geral e introdutório, não sendo voltada especificamente a contextos GLAM, nem a problemas de limpeza, reconciliação, enriquecimento de metadados culturais, publicação de dados ligados ou nanopublicações. Por isso, no *workflow* proposto, sua contribuição é incorporada sobretudo às etapas de modelagem e formalização ontológica, mas articulada a outras abordagens.

Em relação à OntoForInfoScience, o workflow incorpora a aproximação entre ontologias, Ciência da Informação, Organização do Conhecimento e análise de domínio (Mendonça; Almeida, 2016). Essa abordagem é relevante para compreender a ontologia como artefato de representação conceitual e informacional. Contudo, ela não cobre plenamente as etapas técnicas e operacionais de publicação em dados ligados, alinhamento semântico com padrões GLAM, consulta por SPARQL e publicação de afirmações por meio de nanopublicações. Assim, sua contribuição é fundamental para a dimensão conceitual do *workflow*, mas precisa ser complementada por abordagens voltadas à implementação e publicação semântica.

Em relação à LOT, a proposta adota a preocupação com desenvolvimento leve, documentação, publicação e manutenção de ontologias e vocabulários em ambientes de dados ligados (Poveda-Villalón, 2018). Entretanto, a LOT não se concentra especificamente em dados GLAM, nem enfatiza problemas típicos de metadados culturais, como inconsistências terminológicas, reconciliação com autoridades, decisões curatoriais e limitações documentais. No workflow proposto, seus princípios são incorporados principalmente às camadas de organização terminológica, ontologia operacional e interoperabilidade, mas articulados a procedimentos de preparação e enriquecimento de dados culturais.

Em relação à abordagem Linked Data for Libraries, Archives and Museums, o workflow incorpora diretamente a limpeza, reconciliação, enriquecimento e publicação de metadados culturais (Hooland; Verborgh, 2014). Essa abordagem é a mais próxima dos problemas práticos enfrentados por instituições GLAM. Ainda assim, ela não oferece, por si só, uma estrutura completa para desenvolvimento de ontologias operacionais, diferenciação entre vocabulários SKOS, ontologia e grafo RDF, nem incorpora nanopublicações como mecanismo de publicação granular com proveniência. Por isso, no workflow proposto, suas contribuições são combinadas com metodologias de desenvolvimento ontológico e com o bloco final dedicado às nanopublicações.

Em relação à SABiOx, a proposta preserva a passagem entre conceitualização do domínio e implementação computacional. Na SABiOx, a construção da ontologia de referência prevê a definição de uma linguagem de modelagem e de uma ontologia de fundamentação (Aguiar; Souza, 2024). O workflow aqui proposto não torna esse percurso obrigatório: admite modelos conceituais com diferentes níveis de formalização e reserva o uso de ontologias de fundamentação e linguagens ontologicamente fundamentadas para situações em que esse rigor adicional seja requerido. A SABiOx sustenta, assim, a transição entre conceitualização e implementação, sendo complementada no workflow por preparação de dados GLAM, interoperabilidade, povoamento, recuperação por SPARQL e nanopublicações.

Dessa forma, a principal contribuição do workflow está em organizar essas contribuições parciais em uma arquitetura metodológica única e incremental, capaz de conduzir dados culturais heterogêneos a uma aplicação semântica documentada, consultável e interoperável, estendida à publicação de afirmações com proveniência.

A inclusão das nanopublicações como bloco final constitui um diferencial específico da proposta. Enquanto as metodologias de base se concentram em ontologias, vocabulários ou dados ligados, o *workflow* seleciona afirmações relevantes do grafo RDF e as publica com proveniência e informação de publicação. Essa etapa é particularmente pertinente em contextos GLAM, nos quais muitas afirmações envolvem fonte, atribuição, incerteza, interpretação ou responsabilidade curatorial.

Outro diferencial é a documentação distribuída das decisões metodológicas e curatoriais. Decisões de escopo, preparação, modelagem, URI, alinhamento, povoamento, consulta e nanopublicação são registradas ao longo do próprio *workflow*, favorecendo rastreabilidade e avaliação crítica. A proposta também diferencia ontologia operacional, vocabulários SKOS, grafo RDF e nanopublicações, evitando confusões entre artefatos conceituais, terminológicos, dados povoados e afirmações publicadas.

4 CONSIDERAÇÕES FINAIS

Este artigo apresentou um *workflow* metodológico para orientar a representação, conexão, recuperação e publicação semântica de dados em contextos GLAM. A proposta foi fundamentada em cinco abordagens de referência: *Ontology Development* 101, *OntoForInfoScience*, *Linked Open Terms*, *Linked Data for Libraries*, *Archives and Museums* e SABiOx.

O workflow foi organizado em cinco blocos analíticos, 11 camadas e 32 etapas, concebidos para apoiar aplicações semânticas incrementais, documentadas e reutilizáveis em contextos GLAM.

A principal contribuição da proposta está na integração entre abordagens que, em geral, aparecem separadas: desenvolvimento de ontologias, Organização do Conhecimento, publicação de dados ligados em GLAM, distinção entre ontologias de referência e ontologias operacionais e publicação granular de afirmações com proveniência. Como nenhuma das abordagens analisadas cobre isoladamente todo esse percurso, o *workflow* organiza suas contribuições em uma sequência metodológica integrada. A inclusão das nanopublicações como Bloco E amplia a proposta ao permitir que afirmações específicas sejam tratadas como unidades citáveis, verificáveis e reutilizáveis.

O *workflow* apresenta limites. Não pretende cobrir todos os requisitos de projetos institucionais de grande escala nem substituir metodologias completas de desenvolvimento

ontológico; etapas como alinhamento com CIDOC CRM, consultas federadas e nanopublicações podem exigir adaptações à maturidade dos dados e à experiência da equipe. Além disso, embora venha sendo experimentado em uma disciplina de mestrado, o presente estudo ainda não apresenta aplicação integral nem validação empírica sistemática do workflow, que permanecem como etapas de continuidade da pesquisa.

Como continuidade, recomenda-se aplicar o *workflow* a diferentes tipos de bases GLAM, comparar dificuldades de implementação, avaliar sua compreensão por estudantes e profissionais, desenvolver modelos de documentos para cada camada e produzir exemplos completos envolvendo CIDOC CRM/ISO 21127:2023, Linked Art, vocabulários Getty, consultas SPARQL e nanopublicações em TriG. Conclui-se que a proposta favorece interoperabilidade, rastreabilidade, recuperação e reuso de dados culturais, preservando a centralidade das decisões de representação, documentação e curadoria.

REFERÊNCIAS

AGUIAR, Camila Zachhé; SOUZA, Vítor E. Silva. SABiOx: the extended systematic approach for building ontologies. *In*: SEMINAR ON ONTOLOGY RESEARCH IN BRAZIL (ONTOBRAS), 17., 2024, Vitória. **Anais** [...]. Aachen: CEUR Workshop Proceedings, 2024. 13 p. (CEUR Workshop Proceedings, v. 3905). Disponível em: https://ceur-ws.org/Vol-3905/paper1.pdf. Acesso em: 29 maio 2026.

BAADER, Franz; MCGUINNESS, Deborah L.; NARDI, Daniele; PATEL-SCHNEIDER, Peter F. **The description logic handbook**: theory, implementation, and applications. Cambridge: Cambridge University Press, 2003. Disponível em: https://courses.cs.umbc.edu/graduate/691/fall17/01/papers/DescriptionLogicHandbook.pdf . Acesso em: 29 maio 2026.

BERNERS-LEE, Tim. **Linked data**. [*S. l.*]: W3C, 2006. Disponível em: https://www.w3.org/DesignIssues/LinkedData.html. Acesso em: 29 maio 2026.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. **Scientific American**, New York, v. 284, n. 5, p. 28-37, 2001.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data: the story so far. **International Journal on Semantic Web and Information Systems**, Hershey, v. 5, n. 3, p. 1-22, 2009. Disponível em: Disponível em: https://doi.org/10.4018/jswis.2009081901. Acesso em: 7 nov. 2026

CIDOC CRM SPECIAL INTEREST GROUP. **CIDOC Conceptual Reference Model (CIDOC CRM)**: version 7.1.3. [*S. l.*]: CIDOC CRM SIG, 2024. Disponível em: https://cidoc-crm.org/Version/version-7.1.3. Acesso em: 29 maio 2026.

DOERR, Martin. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. **AI Magazine**, Menlo Park, v. 24, n. 3, p. 75-92, 2003. Disponível em: https://dl.acm.org/doi/10.5555/958671.958678. Acesso 7 nov. 2026.

FALBO, Ricardo de Almeida. SABiO: systematic approach for building ontologies. *In*: JOINT WORKSHOP ONTO.COM/ODISE ON ONTOLOGIES IN CONCEPTUAL MODELING AND INFORMATION SYSTEMS ENGINEERING, 1., 2014, Rio de Janeiro. **Anais** [...]. Aachen: CEUR Workshop Proceedings, 2014. p. 1-14. Disponível em: https://ceur-ws.org/Vol-1301/ontocomodise2014_2.pdf. Acesso em: 7 nov. 2026.

GROTH, Paul; GIBSON, Andrew; VELTEROP, Jan. The anatomy of a nanopublication. **Information Services and Use**, Amsterdam, v. 30, n. 1-2, p. 51-56, 2010. Disponível em: https://doi.org/10.3233/ISU-2010-0613. Acesso em: 7 nov. 2026.

GRUBER, Thomas R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, London, v. 5, n. 2, p. 199-220, 1993. Disponível em: https://doi.org/10.1006/knac.1993.1008. Acesso em: 7 nov. 2026.

GUIZZARDI, Giancarlo. Ontology, ontologies and the "I" of FAIR. **Data Intelligence**, Cambridge, v. 2, n. 1-2, p. 181-191, 2020. Disponível em: https://doi.org/10.1162/dint_a_00040. Acesso em: 7 nov. 2026.

GUIZZARDI, Giancarlo; WAGNER, Gerd; ALMEIDA, João Paulo Andrade; GUIZZARDI, Renata S.S. Towards ontological foundations for conceptual modeling: the Unified Foundational Ontology (UFO) story. **Applied Ontology**, Amsterdam, v. 10, n. 3-4, p. 259-271, 2015. Disponível em: https://journals.sagepub.com/doi/abs/10.3233/AO-150157. Acesso em: 7 nov. 2026.

HJØRLAND, Birger. Domain analysis in information science: eleven approaches: traditional as well as innovative. **Journal of Documentation**, Bingley, v. 58, n. 4, p. 422-462, 2002. Disponível em: https://doi.org/10.1108/00220410210431136. Acesso em: 7 nov. 2026.

HOOLAND, Seth van; VERBORGH, Ruben. **Linked data for libraries, archives and museums**: how to clean, link and publish your metadata. London: Facet Publishing, 2014.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION; INTERNATIONAL ELECTROTECHNICAL COMMISSION. **ISO/IEC 21838-2:2021**: information technology: top-level ontologies (TLO): part 2: Basic Formal Ontology (BFO). Geneva: ISO, 2021. Disponível em: https://www.iso.org/standard/74572.html. Acesso em: 29 maio 2026.

ISAAC, Antoine; SUMMERS, Ed. **SKOS simple knowledge organization system primer**. [*S. l.*]: W3C, 2023. Disponível em: https://www.w3.org/TR/skos-primer/. Acesso em: 29 maio 2026.

KNUBLAUCH, Holger; KONTOKOSTAS, Dimitris. **Shapes Constraint Language (SHACL)**. [*S. l.*]: W3C, 2017. W3C Recommendation. Disponível em: https://www.w3.org/TR/shacl/. Acesso em: 29 maio 2026.

KUHN, Tobias; DUMONTIER, Michel. Trusty URIs: verifiable, immutable, and permanent digital artifacts for linked data. *In*: EXTENDED SEMANTIC WEB CONFERENCE (ESWC), 11.,

2014, [*S. l.*]. **The semantic web**: trends and challenges. Cham: Springer, 2014. p. 395-410. DOI: 10.1007/978-3-319-07443-6_27. Disponível em: https://doi.org/10.1007/978-3-319-07443-6_27. Acesso em: 7 nov. 2026.

LEBO, Timothy; SAHOO, Satya; MCGUINNESS, Deborah. **PROV-O**: the PROV ontology. [*S. l.*]: W3C, 2013. W3C Recommendation. Disponível em: https://www.w3.org/TR/prov-o/. Acesso em: 29 maio 2026.

MCCOMB, Dave. Data centric revolution: is a knowledge ontology the missing link? **TDAN.com**, [*s. l.*], 7 jun. 2023. Disponível em: https://tdan.com/data-centric-revolution-is-knowledge-ontology-the-missing-link/30790. Acesso em: 29 maio 2026.

MENDONÇA, Fabricio; ALMEIDA, Mauricio Barcellos. OntoForInfoScience: a detailed methodology for construction of ontologies and its application in the blood domain. **Brazilian Journal of Information Science**: research trends, Marília, v. 10, n. 1, p. 12-19, 2016. Disponível em: https://doi.org/10.36311/1981-1640.2016.v10n1.02.p12. Acesso em: 7 nov. 2026.

NOY, Natalya F.; MCGUINNESS, Deborah L. **Ontology development 101**: a guide to creating your first ontology. Stanford: Stanford Knowledge Systems Laboratory, 2001. Disponível em: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf. Acesso em: 29 maio 2026.

POVEDA-VILLALÓN, María. **Linked Open Terms (LOT)**. [*S. l.*]: [*s. n.*], 2018. Disponível em: https://lot.linkeddata.es/. Acesso em: 29 maio 2026.

RILEY, Jenn. **Understanding metadata**: what is metadata, and what is it for? A primer. Baltimore: NISO, 2017. Disponível em: https://www.niso.org/publications/understanding-metadata-2017. Acesso em: 29 maio 2026.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015.

SAYÃO, Luis Fernando; SALES, Luana Farias. Nanopublicações: abordagens de representação do conhecimento como inovação para a comunicação científica. **Informação & Informação**, Londrina, v. 29, n. 4, p. 247-266, 2024. DOI: 10.5433/1981-8920.2024v29n4p247. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/51883. Acesso em: 7 nov. 2026.

STUDER, Rudi; BENJAMINS, V. Richard; FENSEL, Dieter. Knowledge engineering: principles and methods. **Data & Knowledge Engineering**, Amsterdam, v. 25, n. 1-2, p. 161-197, 1998. Disponível em: https://dl.acm.org/doi/10.1016/S0169-023X(97)00056-6. Acesso em: 7 nov. 2026.

WANG, Richard Y.; STRONG, Diane M. Beyond accuracy: what data quality means to data consumers. **Journal of Management Information Systems**, Armonk, v. 12, n. 4, p. 5-33, 1996. Disponível em: Disponível em: https://doi.org/10.1080/07421222.1996.11518099. Acesso em: 11 nov. 2026.

WILKINSON, Mark D.; DUMONTIER, Michel; AALBEERSBERG, Ijsbrand; APPLETON, Gabrielle *et al*. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, art. 160018, 2016. Disponível em: https://doi.org/10.1038/sdata.2016.18. Acesso em: 7 nov. 2026.

ZENG, Marcia Lei. Interoperability. **Knowledge Organization**, Baden-Baden, v. 46, n. 2, p. 122-146, 2019. Disponível em: Disponível em: https://doi.org/10.5771/0943-7444-2019-2-122. Acesso em: 11 nov. 2026.

ZENG, Marcia Lei. Knowledge organization systems (KOS). **Knowledge Organization**, Baden-Baden, v. 35, n. 2-3, p. 160-182, 2008. Disponível em: https://www.imrpress.com/journal/ko/35/2-3/10.5771/0943-7444-2008-2-3-160. Acesso em: 7 nov. 2026.

FINANCIAMENTO

O presente trabalho recebeu apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES), por meio da Bolsa Produtividade Estadual, no âmbito do Termo de Cooperação CNPq/FAPES, Processo nº 1011/2025.