

# **Organização e representação da informação e do conhecimento em contextos informacionais: uma proposta de um modelo teórico-conceitual para a qualidade de objetos culturais digitais**

---

*Daniela Lucas da Silva Lemos<sup>1</sup>*

*Dalton Lopes Martins<sup>2</sup>*

*Renato Rocha Souza<sup>3</sup>*

## **Resumo:**

Atualmente, nota-se o interesse de comunidades científicas e de prática, que lidam com tecnologias digitais e ao mesmo tempo com áreas pertencentes às Ciências Humanas, Sociais e Sociais Aplicadas, em fazer reuso de fontes de informação na web e tratar seus objetos culturais digitais como elementos preponderantes na geração de valor em suas circulações e reuso social. Nesse ensejo, o artigo objetiva apresentar uma proposta de um modelo teórico-conceitual capaz de reunir e fundamentar elementos seminais e interdisciplinares da Ciência da Informação, da Ciência de Dados e da Ciência da Computação e, logo, contribuir na representação de entidades de informação com qualidade para a produção de conhecimento cultural em rede. Trata-se de uma pesquisa teórica e qualitativa, de cunho exploratório e descritivo usando como técnica a pesquisa bibliográfica em bases de dados especializadas nos campos ora citados, juntamente com o uso de categorias analíticas oriundas do método de Análise de Conteúdo no auxílio a inferências e a interpretações acerca dos dados coletados. Os artefatos de representação ora evidenciados no modelo teórico-conceitual proposto podem ser considerados sínteses de um Sistema de Organização do Conhecimento que busca, primariamente, a organização de recursos de informação, produzindo relacionamentos; pontos de acesso; ordenação para itens em coleções; facetas genéricas e específicas para navegação em sistemas web, e objetos culturais digitais interoperáveis em rede. O profissional da informação, portanto, tem como papel preponderante o desenvolvimento de terminologias específicas para a

---

<sup>1</sup> Doutora em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG); mestre em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG); daniela.l.silva@ufes.br. Lattes (<http://lattes.cnpq.br/9280443047358807>). ORCID (<https://orcid.org/0000-0003-1565-7366>).

<sup>2</sup> Doutor em Ciência da Informação pela Universidade de São Paulo; mestre em Engenharia da Computação pela Universidade Estadual de Campinas (UNICAMP); daltonmartins@unb.br. Lattes (<http://lattes.cnpq.br/3774617443225038>). ORCID (<https://orcid.org/0000-0002-6244-6791>).

<sup>3</sup> Doutor em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG). Mestre em Engenharia de Produção pela Universidade Federal de Santa Catarina (UFSC). rsouza.fgv@gmail.com. Lattes (<http://lattes.cnpq.br/4726949697973381>). ORCID (<https://orcid.org/0000-0002-1895-3905>).

estruturação de informações com qualidade, seguindo modelos semânticos para o tratamento da informação, visando serviços de informações inteligentes, inovadores e de valor ao domínio da cultura e ao campo das Humanidades Digitais.

**Palavras-chave:** Organização e Representação da Informação e do Conhecimento; Sistema de Organização do Conhecimento; Objeto Cultural Digital; Qualidade de Dados; Profissional da Informação.

*Organization and representation of information and knowledge in informational contexts: a proposal for a theoretical-conceptual model for the quality of digital cultural objects*

**Abstract:**

Currently, there is interest from scientific and practical communities that deal with digital technologies and at the same time with areas belonging to the Human, Social and Applied Social Sciences in reusing information sources on the web and treating their digital cultural objects as preponderant elements in the generation of value in their circulations and social reuse. In this opportunity, the article at presenting a proposal for a theoretical-conceptual model capable of integrating contributions from Information Science, Data Science and Computer Science and, therefore, contribute to the representation of information entities with quality for the production of networked cultural knowledge. This is a theoretical and qualitative research, of an exploratory and descriptive nature, using bibliographical research in specialized databases in the fields cited above as a technique, together with the use of analytical categories derived from the method of Content Analysis in helping to inferences and interpretations about the collected data. The representation artifacts now evidenced in the proposed theoretical-conceptual model can be considered syntheses of a Knowledge Organization System that seeks, primarily, the organization of information resources, producing relationships; access points; sorting for items in collections; generic and specific facets for navigation in web systems, and interoperable digital cultural objects in a network. The information professional, therefore, has as a preponderant role the development of specific terminologies for the structuring of information with quality, following semantic models for the treatment of information, aiming at intelligent, innovative and valuable information services to the field of culture and the field of the Digital Humanities.

**Keywords:** Organization and Representation of Information and Knowledge; Knowledge Organization System; Digital Cultural Object; Data Quality; Information Professional.

## 1 INTRODUÇÃO

O fenômeno da “explosão da informação” (LE COADIC, 1996; SARACEVIC, 1999), evidenciado desde o último século, vem demandando da sociedade contemporânea, variadas formas criativas e adequadas de organização e representação dos volumosos estoques de informação gerados para a sua posterior recuperação. Vive-se uma era em que a informação se caracteriza como um objeto de consumo cobiçado, de valor criativo e indispensável para a realização de diversas atividades do cotidiano, demandadas por diversos tipos de instituições, governos e indivíduos.

A partir desse contexto, nota-se o interesse de comunidades científicas e de prática, que lidam com tecnologias digitais e ao mesmo tempo com áreas pertencentes às Ciências Humanas, Sociais e Sociais Aplicadas, em fazer reúso de fontes de informação na web e tratar seus objetos culturais digitais como elementos preponderantes na geração de valor em suas circulações e reúso social (MARTINS et al., 2022).

Contudo, para que se possa obter fontes de informação com qualidade visando reúso, os dados precisam se adequar a dimensões importantes como, por exemplo, precisão, integridade, consistência e completude, as quais corroborarão com a medida da adequação dos dados a um propósito específico (ECKERSON, 2002). Para tal, torna-se necessário considerar o significado dos dados e como eles são usados (CHAPMAN, 2005).

Em suma, o que se projeta para a obtenção da qualidade de dados em conjunto de dados, tanto no âmbito de pesquisas científicas quanto em práticas profissionais, é a criação e a modelagem apropriada de metadados pelo profissional da informação endereçando processos, sobre os quais incluem: análise, contextualização, cálculo, síntese e descrição dos dados de modo a transformá-los em informação (WANG, 2018). Geralmente, tais processos são orientados por soluções inteligentes de organização e tratamento da informação advindas de campos científicos interdisciplinares como a Ciência da Informação (CI), a Ciência de Dados (CD) e a Ciência da Computação (CC) (VIRKUS; GAROUFALLOU, 2019; 2020; MARTINS et al., 2022).

Segundo Wang (2018), o controle da qualidade de dados é um dos papéis mais importantes do profissional da informação para com a CD, cujo principal propósito deste campo é identificar e extrair conhecimento a partir de grandes volumes de dados existentes em diversos tipos de fontes disponíveis em ambientes digitais. Abrem-se, a partir disto, espaços para o crescimento de campos emergentes como as Humanidades Digitais (HD) (KOLTAY, 2016; POOLE, 2017; CLEMENT; CARTER, 2017), que buscam a extração de valor em objetos digitais disponíveis na rede por meio de ações de reúso para o desenvolvimento econômico, social e humano, a exemplo da cultura, indústria criativa, ciência aberta, turismo, educação, entre outros.

Para tal, o profissional da informação precisa estar alinhado com as principais tendências científicas e tecnológicas que permitam a estruturação de infraestruturas informacionais contemporâneas e adequadas para a produção, a preservação, a organização e o acesso ao conhecimento pela sociedade em rede, além de disponibilizar dados com qualidade em variados tipos de mídia e em variados contextos informacionais, de modo que possam ser úteis à tomada de decisões. Borgman (2010) entende infraestrutura informacional em ambientes abertos e distribuídos como um conjunto de elementos técnicos, sociais e políticos, incluindo pessoas, tecnologias, ferramentas e serviços, usados em favor da disseminação colaborativa de conteúdos em rede.

Percebe-se, assim, a necessidade de um entendimento sistêmico e aprofundado acerca de princípios teórico-metodológicos de campos considerados fundamentais a essas práticas científicas e profissionais, como os ora citados, e seus conceitos elementares. Logo, pergunta-se: *quais conteúdos interdisciplinares no âmbito da CI, da CD e da CC, em conjunto com seus princípios teórico-metodológicos, estão sendo discutidos e encaminhados como pesquisa na perspectiva da disponibilização qualificada de objetos culturais digitais na rede?*

O objetivo da presente pesquisa é, então, apresentar uma proposta de um modelo teórico-conceitual capaz de fundamentar e reunir sistematicamente elementos seminais e interdisciplinares da Ciência da Informação, da Ciência de Dados e da Ciência da Computação, contribuindo, logo, na representação qualificada de

entidades de informação para a produção de conhecimento cultural em rede. Tal conhecimento pode vir a ser útil aos cientistas de dados, humanistas digitais, cientistas da informação e da computação interessados no reuso de fontes de informação cultural para aplicações diversas.

Entende-se que a proposição de um modelo teórico-conceitual, fundamentado em princípios teórico-metodológicos de importantes campos interdisciplinares, pode vir a ser útil na idealização de programas de formação adequados para a capacitação do profissional da informação em lidar com objetos digitais em variados tipos de mídia e em variados contextos informacionais. O tratamento, neste caso, requer uma representação da informação e do conhecimento diferenciada, visando melhorar os sistemas de catalogação, indexação, busca, navegação, agregação e recuperação, de forma a obter resultados mais significativos a várias tipologias de usuários no âmbito da cultura digital. Adicionalmente, o profissional da informação tem como papel a implantação de padrões para a estruturação de base de dados com qualidade em repositórios e bibliotecas digitais, por exemplo, seguindo modelos contemporâneos para tratamento documental que visem normalização, qualidade e intercâmbio de descrições ou anotações em rede.

Um das contribuições deste artigo é, portanto, identificar potenciais elementos que possam fornecer subsídios conceituais e metodológicos para futuros programas de formação no país que capacitem profissionais da informação, a exemplo de bibliotecários, arquivistas e museólogos, a conduzirem projetos voltados à construção de infraestruturas informacionais envolvendo acervos culturais digitais para o reuso de informação do patrimônio cultural brasileiro.

O artigo está organizado como se segue. Nesta Introdução são apresentados o contexto da pesquisa, seu problema e objetivo, e a sua justificativa. O Capítulo 2 apresenta os procedimentos metodológicos usados na condução e concretização da pesquisa. Os Capítulos 3 e 4 se incumbem de apresentar os resultados da pesquisa, tecendo elucidações e discussões teórica-metodológicas acerca dos processos, instrumentos e produtos presentes no modelo teórico-conceitual proposto. E, por fim, o Capítulo 5 encerra o artigo com as considerações finais.

## 2 METODOLOGIA

O estudo aqui conduzido classifica-se como uma pesquisa teórica e qualitativa, de cunho exploratório e descritivo, usando-se de pesquisa bibliográfica para fundamentar conceitos e dar sustentação teórica e metodológica ao estudo, além de estabelecer uma visão geral do problema de pesquisa.

Foram elencadas e consultadas bases de dados nacionais e internacionais, considerando a recuperação de artigos científicos seminais ou, recentes, relevantes e com fator de impacto considerável aos propósitos de uma revisão de literatura. Os campos revisados foram: Ciência da Informação, Ciência de Dados e Ciência da Computação, incluindo também as Humanidades Digitais como campo interessado na qualidade de objetos culturais digitais, tema central da pesquisa.

As bases de dados consultadas são aqui relacionadas: SciELO – Scientific Electronic Library Online; Google Scholar; Library and Information Science Abstracts (LISA); Base de Dados em Ciência da Informação (BRAPCI); Web of Science; Computer and Information Systems; Computers and Applied Sciences Complete; Information Science & Technology Abstracts (ISTA); Library, Information Science & Technology Abstracts with Full Text; Scopus (Elsevier B.V); e Biblioteca Brasileira de Teses e Dissertações (BDTD).

Como complemento, especialmente para a revisão do campo da CD, uma das mais importantes revistas científicas da CI internacional, a *Journal of the Association for Information Science and Technology* (JASIST), abriu uma chamada de artigos para um número especial da revista para o ano de 2022 com foco em discutir as especificidades da CD em suas relações com a CI. A chamada disponibilizou uma extensa bibliografia especializada sobre tema, com a qual se fez uso na presente pesquisa.

O material coletado nas etapas de levantamento bibliográfico foi examinado a partir de técnicas oriundas do método de Análise de Conteúdo proposto por Bardin (2016), que adota como estratégia o emprego de um conjunto de métodos de análise das comunicações para fins de extração, análise e interpretação dos dados,

e assim criar categorias que permitam a inferência de conhecimentos presentes nos documentos.

As categorias analíticas determinadas apoiaram os recortes dos conteúdos de disciplinas de cada campo de conhecimento envolvido (CI, CD e CC) em conjunto com seus princípios teórico-metodológicos, o que permitiu a proposição do modelo teórico-conceitual (implementado na ferramenta CmapTools) com a reunião sistemática de elementos seminais úteis à sua discussão interdisciplinar.

### **3 A REPRESENTAÇÃO DE DOMÍNIOS DE CONHECIMENTO E SUA APLICAÇÃO NAS ATIVIDADES DE ORGANIZAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO EM CONTEXTOS INFORMACIONAIS**

O grande volume de informação resultante da produção e do consumo de ativos informacionais em rede vem sendo objeto de estudo por parte da CI, sendo, inclusive, parte integrante de sua gênese (BORKO, 1968), além de outros campos de pesquisa como as HD, a CD e a CC (CLEMENT; CARTER, 2017; VIRKUS; GAROUFALLOU, 2019; 2020; MARTINS et al., 2022). Nesse contexto, reforça-se cada vez mais a necessidade em desenvolver métodos e técnicas eficientes para organização, representação, armazenamento, recuperação e intercâmbio da informação por meio da internet, sejam na produção científica, documental ou quaisquer outros registros em suportes físicos ou digitais, em variados contextos e conjunturas.

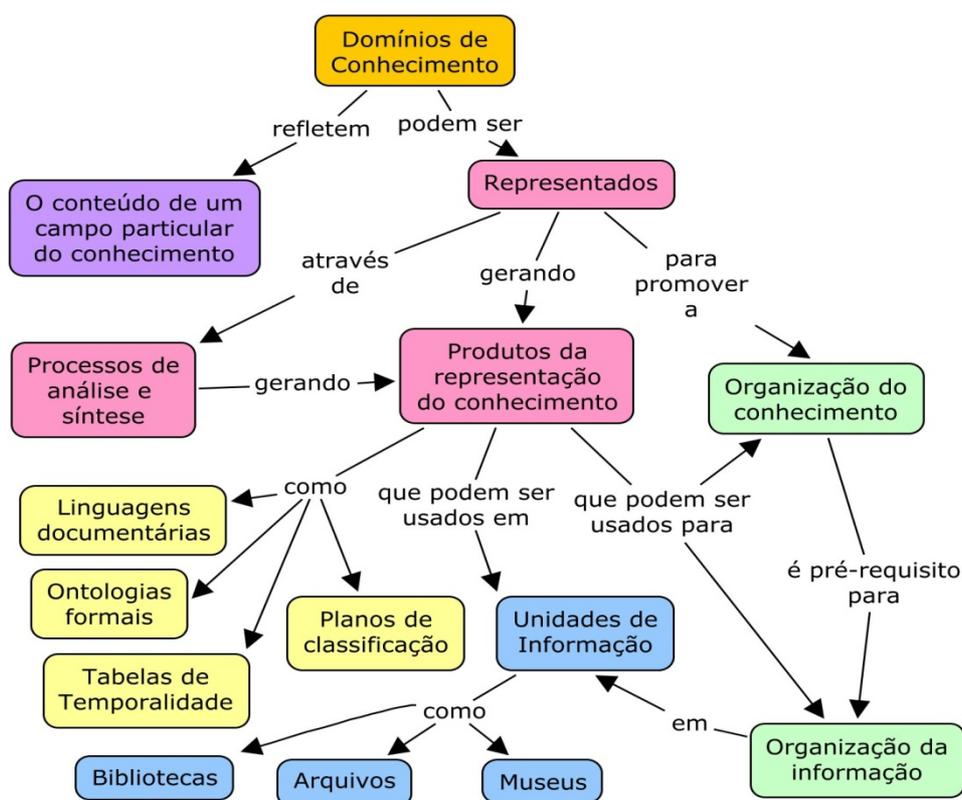
A área de Recuperação da Informação (RI) (LUHN, 1953; MOOERS, 1960; LANCASTER, 1986; ABADAL; CODINA, 2005; BAEZA-YATES; RIBEIRO-NETO, 2011; HJØRLAND, 2016) traz no seu bojo o aspecto do tratamento documental (catalogação, indexação e classificação) como uma de suas principais frentes de pesquisa, especialmente no que se refere a questões ligadas à natureza semântica da informação (RANGANATHAN, 1967; DAHLBERG, 1978; 1993; BARITE, 2000; JACOB, 2004; HJØRLAND, 2007; KHOO; NA, 2006; ALMEIDA et al., 2011; ALMEIDA, 2013). Nesse sentido, a RI pode ser considerada uma área que sempre buscou (e busca) empregar instrumentos tecnológicos que auxiliem as práticas de análise de conteúdo e categorização com soluções de apoio à compreensão, à interpretação e à sistematização de conteúdos constantes nos registros de variados domínios de

conhecimento subjacentes aos sistemas de recuperação da informação (SRIs) (LANCASTER, 1986; 2004; ABADAL; CODINA, 2005).

Os domínios de conhecimento podem ser organizados e representados, conforme se indica na Figura 1, por meio de processos de análise e síntese acerca de conceitos e relacionamentos de um domínio (DAHLBERG, 1978; 1993; HJØRLAND, 2007), gerando produtos ou instrumentos da representação do conhecimento, tais como linguagens documentárias, ontologias formais, tabelas de temporalidade e planos de classificação.

Assim, por exemplo, bibliotecários, museólogos e arquivistas empregam sistemas de classificação ou símbolos de classificação (descritores codificados) para descreverem o conteúdo do documento de caráter bibliográfico, museológico e arquivístico de forma mais abrangente a partir do uso de tais produtos ou sistemas de organização, que refletem o domínio de conhecimento.

Figura 1 – Representação de domínios de conhecimento



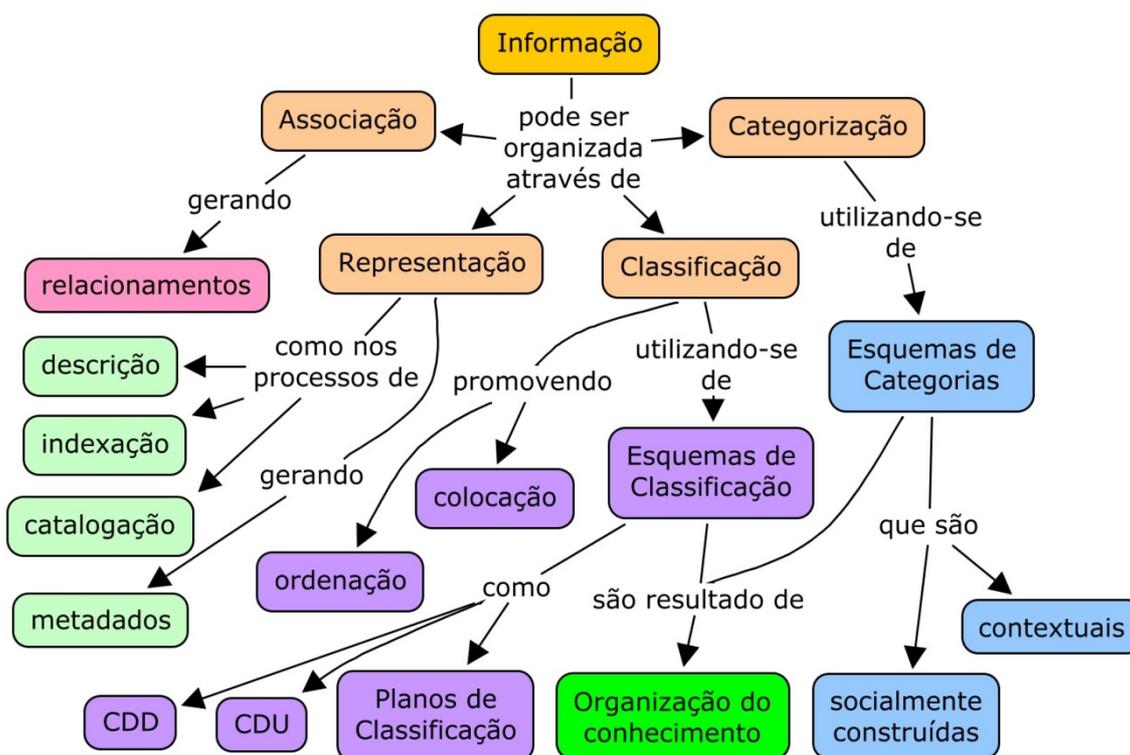
Fonte: elaborado pelos autores (2023)

Logo, torna-se oportuno salientar que organizar a informação difere-se de organizar o conhecimento, embora muitas vezes organizar o conhecimento seja um pré-requisito para organizar a informação, conforme é proposto na Figura 1. Bräscher e Café (2008), numa análise reflexiva sobre o assunto, compactuam com tal ponderação. Consideram que no contexto da organização da informação tem-se como objeto os registros de informação (exs.: catálogos de museus, livros, coleções, jornais, periódicos, arquivos públicos, fotografias, manuscritos, entre outros), distinto, portanto, do mundo da cognição e dos conceitos (leis, propriedades de objetos, regras, axiomas), do qual reflete o conhecimento e não a informação. Ou seja, quando nos referimos à organização do conhecimento, estamos no mundo dos conceitos e não naquele dos registros de informação (BRÄSCHER; CAFÉ, 2008, pgs. 5-6).

A Organização do Conhecimento, portanto, está ligada à análise dos conceitos e relacionamentos de um domínio de conhecimento, com a consequente síntese de um Sistema de Organização do Conhecimento (SOC) (KOS – *Knowledge Organization System*, do inglês) (HJØRLAND, 2007), que busca, conforme ilustrado na Figura 2, auxiliar a organização de registros de informação nos aspectos de: i) associação, gerando relacionamentos; ii) representação, gerando metadados, pontos de acesso e índices em processos de descrição, indexação e catalogação; iii) classificação, promovendo colocação e ordenação para os documentos a partir de esquemas de classificação; e iv) categorização, gerando esquemas de categorias socialmente construídas para variados contextos.

Nesse caminho, Hodge (2000) evidencia estudos sobre tipos de SOCs que consideram características distintas a partir de seus contextos e aplicações, podendo se desdobrar em três categorias classificatórias, a saber: i) lista de termos: arquivos de autoridade, glossários, dicionários e índices para mapas; ii) classificação e categorização: cabeçalhos de assuntos, esquemas de classificação, taxonomias e esquemas de categorização; e iii) modelos de relacionamentos: tesouros, redes semânticas e ontologias. Contudo, Hjørland (2007) acrescenta que os tipos de SOCs se caracterizam por seus níveis de complexidade estrutural, isto é, pelo formalismo promovido pela adequabilidade no uso de tecnologias semânticas, especialmente no que se refere aos relacionamentos.

Figura 2 – Formas de organização da informação



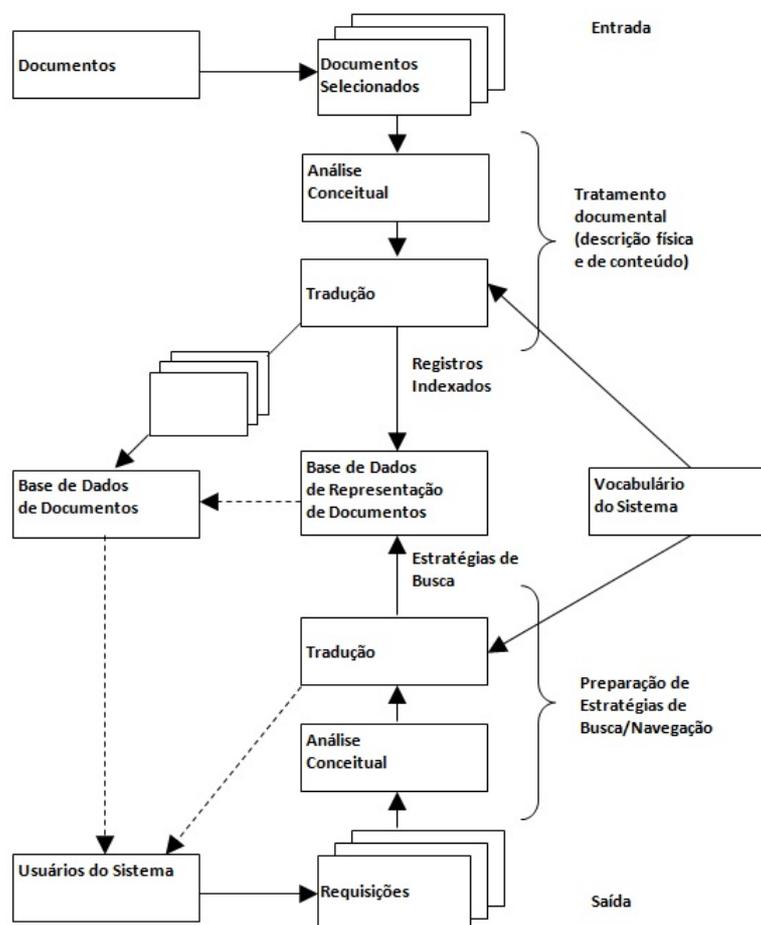
Fonte: elaborado pelos autores (2023)

Taxonomias, por exemplo, são coleções de termos classificados em uma estrutura hierárquica, na qual se emprega relacionamentos de generalização e especialização. Já nos tesouros, as relações semânticas se estendem em equivalência, hierárquicas e associativas entre os termos, em que são claramente mostradas e identificadas através de indicadores de relação padrão (ANSI, 2005). Ontologias possuem os mesmos princípios dos vocabulários controlados (SILVA; SOUZA; ALMEIDA, 2008), isto é, trabalham com linguagem natural e fazem a delimitação de termos e de relações. Entretanto, a semântica envolvida na terminologia da ontologia se difere dos vocabulários controlados (dentro do contexto das linguagens documentárias) por incluir axiomas formais (através de declarações lógicas) que restringem a utilização do vocabulário. Outra distinção está no fato de as especificações de relações no contexto das ontologias serem em número superior às dos tesouros, por exemplo.

No âmbito da presente pesquisa, os SRIs em contextos digitais são casos particulares de sistemas de informação (exs.: bibliotecas e repositórios digitais, agregadores culturais, catálogos online, sistemas de gestão de conteúdo, entre

outros) que lidam com tarefas de representação, organização, armazenamento e acesso a itens (documentos ou objetos digitais). Adicionalmente, um SRI é também considerado a interface entre uma coleção de recursos de informação e uma população de usuários, e se constituem, principalmente, por dois subsistemas interdependentes, quais sejam, o de entrada e o de saída (LANCASTER, 1986; 2004; ABADAL; CODINA, 2005; BAEZA-YATES; RIBEIRO-NETO, 2011), conforme se apresentam no modelo de organização, representação e recuperação da informação da Figura 3.

Figura 3 – Modelo de organização, representação e recuperação da informação



Fonte: adaptado de Lancaster (2004)

O subsistema de entrada refere-se ao processo de seleção de documentos (ou na formação de coleções) e descrição de suas características física e de conteúdo, por meio das quais as bases de dados são constituídas. Já o subsistema de saída refere-se à interface do usuário com o sistema, partindo da necessidade de informação, passando pela estratégia de busca, de modo a propiciar a recuperação de itens

relevantes (segundo medidas de revocação e precisão) e, finalmente, o acesso ao documento e/ou suas representações pelo usuário.

Finalmente, um SOC visto como vocabulário do sistema (Figura 3) congrega a função de viabilizar a recuperação de informação na medida em que busca padronizar a entrada de dados na constituição dos registros na base de dados de representação, facilitar a estratégia de busca e, conseqüentemente, melhorar a interação do usuário com o sistema de recuperação de informação. Logo, vislumbra-se que o uso de SOCs em processos de organização da informação em SRIs é uma boa prática em termos de se buscar a produção de bases de dados com qualidade para aplicações diversas.

As Seções a seguir buscam discorrer sobre encaminhamentos de pesquisas interdisciplinares oriundas especialmente de 3 campos do conhecimento, quais sejam Ciência da Informação, Ciência de Dados e Ciência da Computação, no esforço de se produzir metadados com qualidade a partir de princípios teóricos e metodológicos da área de Organização e Representação da Informação e do Conhecimento discutida no presente Capítulo. Para cada campo, buscou-se apresentar de forma sucinta algumas revisões de casos que refletem a produção e o uso (reúso) de metadados em contextos digitais.

### **3.1 Ciência da Informação e qualidade de fontes documentais**

Os fundamentos dos processos de catalogação, de classificação e de indexação são considerados abordagens do tratamento da informação e responsáveis pela formação da área de Organização da Informação em CI (GUIMARÃES, 2003; 2008; 2009). Esta área, por sua vez, é abordada na CI sob o prisma de pressupostos teóricos e metodológicos ao tratamento da informação, e como atividade operacional inerente ao fazer profissional relativo a soluções de problemas da comunicação efetiva de conhecimento e de seus registros à sociedade (SARACEVIC, 1999).

Desse modo, disciplinas como Representação Descritiva, Representação Temática e Modelagem Conceitual se fazem presentes na proposição de esquemas de dados (ou bases de dados) que redundam em informações para um determinado domínio

de conhecimento. A informação, neste caso, precisa ser organizada, e, para tal, precisa ser descrita, e o produto desse processo descritivo é a representação da informação (SVENONIUS, 2000), conforme se percebe no modelo da Figura 3 em que uma base de dados de representação de documentos é constituída por registros de metadados.

Metadados no âmbito da CI (GILLILAND, 2016; ZENG; QIN, 2016) é um assunto que tem sido tratado há décadas na disciplina de Catalogação, especialmente na área de Biblioteconomia, sendo considerada importante atividade na padronização e descrição de recursos de informação, promovendo interpretação uniforme e universal, em qualquer idioma e em qualquer tipo de unidade de informação; além de ser a forma mais comumente empregada para agregar semântica a informações com o propósito de facilitar a busca de recursos de informação.

Ressalta-se, nesse contexto, que alguns tipos de representação da informação são construídos através do uso de linguagens, as quais são subdivididas em linguagens que descrevem a informação (o conteúdo) e linguagens que descrevem o documento (o suporte), no todo ou em partes (BRÄSCHER; CAFÉ, 2008).

A linguagem usada para descrever o documento está relacionada à Representação Descritiva ou Catalogação Descritiva (ABBAS, 2010; GILLILAND, 2016; IFLA, 2009; 2016; ZENG; QIN, 2016), também tratada como processo de catalogação, que busca retratar aspectos específicos do documento que permitam a sua individualização e determinação dos pontos de acesso para proporcionar aos usuários a condição de encontrar, identificar, selecionar e obter o item por meio de um catálogo (WYNAR, 1985; MEY, 1995; IFLA, 2016). Já a linguagem destinada à descrição do conteúdo do documento está relacionada à Representação Temática ou Catalogação de Assunto (FOSKETT, 1973; LANCASTER, 1986; ANSI, 2005; GUIMARÃES, 2009) que vislumbra aspectos intelectuais e semânticos, logo, subjetivos, como a compreensão do assunto do documento para fins de tradução para uma linguagem de classificação, indexação ou resumo.

A classificação consiste em ações de ordenar, organizar, reunir, segundo características comuns a fim de compreensão e entendimento melhor do mundo (PIEDEDE, 1977). Em Ciência da Informação, a classificação é uma questão central

para a organização de assuntos, facilitando a recuperação da informação em ambiência digital, especialmente com o emprego de instrumentos de organização do conhecimento contemporâneos como as ontologias, aplicadas juntamente aos princípios teóricos e metodológicos seminais do campo da informação (ALMEIDA; TEIXEIRA, 2020).

A indexação, elemento preponderante às ações bem sucedidas de recuperação da informação, pode descrever o assunto dos documentos seguindo uma linguagem documentária em que se busca obter um vocabulário controlado de um assunto ou domínio específico (LANCASTER, 2004). Tal artefato de representação pode auxiliar nos processos de análise e descrição de documentos, permitindo a criação padronizada de metadados ao nomear, de forma consistente, os pontos de acesso aos documentos e a informação neles contida; e os usuários no momento da escolha dos filtros destinados à busca, expansão do vocabulário das consultas, bem como na navegação em sistemas de informação para a web. Exemplos de vocabulários controlados (ABBAS, 2010) incluem: esquemas de classificação, listas de cabeçalhos de assuntos, tesouros, taxonomias e ontologias.

As atividades envolvidas na catalogação descritiva e de assunto são geralmente orientadas por regras ou códigos de catalogação, linguagens documentárias e padrões de metadados visando a obtenção de informações consistentes para recuperação e acesso eficientes por parte da comunidade de usuários.

Regras de catalogação determinam como elaborar a descrição de um recurso de informação e os pontos de acesso, tornando-se práticas essenciais na padronização, na descrição e, portanto, na agregação semântica de recursos de informação, com o propósito de viabilizar a interoperabilidade, o compartilhamento de recursos, o intercâmbio contínuo e a reutilização de metadados (JOUNDREY; TAYLOR; MILLER, 2015; BACA et al., 2006; OLIVER, 2011; ZENG, 2019; RDA, 2022). Exemplos de regras de catalogação incluem: Anglo-American Cataloging Rules (AACR); Cataloging Cultural Objects (CCO); Resource Description and Access (RDA).

Linguagens documentárias são destinadas à descrição do conteúdo do documento e está relacionada à Catalogação de Assunto ou Representação Temática, conforme

elucidado anteriormente. Exemplos de linguagens documentárias (TRUST, 2023) incluem: Library of Congress Authority Files; Union List of Artist Names – Getty ULAN; Art & Architecture Thesaurus - Getty AAT; Thesaurus of Geographic Names – Getty TGN; Cultural Objects Name Authority – Getty CONA. Adicionalmente, sistemas de organização do conhecimento contemporâneos, incluindo o *Simple Knowledge Organization System* (SKOS), também podem ser considerados linguagens documentárias modeladas em linguagens da Web Semântica (MOREIRO GONZÁLEZ, 2011), conforme exemplos das linguagens documentárias supracitadas. As terminologias apresentam em suas estruturas um identificador único no formato *Uniform Resource Identifier* (URI) que estabelecem interligações e anotações sobre dados sob licença aberta, o que os confere possibilidades de reuso e interoperabilidade com outros conjuntos de dados disponíveis em rede.

Já os padrões de metadados têm a capacidade de prover um vocabulário comum para descrever uma variedade de estruturas de dados capazes de satisfazer a várias comunidades, e, geralmente, são estruturados seguindo modelos para tratamento dos dados (com regras de catalogação e linguagens documentárias), o que redundando em normalização, qualidade e intercâmbio de suas descrições. Exemplos de padrões de metadados (LE MOS et al., 2022) incluem: Dublin Core; VRA Core; LIDO; MPEG-7.

Logo, com objetivo de viabilizar a interoperabilidade, instituições vêm investindo na adoção de padrões e práticas recomendadas para a produção de metadados, sendo, portanto, uma tentativa de se obter um vocabulário comum e consistente para descrever uma variedade de estruturas de dados capazes de satisfazer a várias comunidades. Nesse sentido, Boughida (2005) e Gilliland (2016) promovem orientações acerca do uso de padrões para tratamento nos dados focando padronização, normalização, qualidade das fontes documentais e intercâmbio de metadados em ambiente digital, indicando o uso de regras de catalogação (conteúdo dos dados), linguagens documentárias (valores dos dados); padrões de metadados (estrutura de dados) e padrões para comunicação de dados (padrões de metadados expressados em uma linguagem de representação legível para a máquina – MARC21; BIBFRAME; Dublin Core RDF/XML, entre outros).

A disciplina de Modelagem Conceitual vem sendo empregada na Ciência da Informação para amenizar os problemas na comunicação entre sistemas de informação e seus usuários (SAYÃO, 2001; IFLA, 2009; CAMPOS, 2004), além de os modelos resultantes poderem impactar na qualidade de dados que, eventualmente, sejam utilizados para treinamento em aplicações e serviços usando recursos de aprendizagem de máquina e inteligência artificial (MARTINS et al., 2022). Logo, é perceptível a importância da área de modelos e modelagem levando-se em consideração que as organizações sociais e as pessoas têm exigências cada vez mais sofisticadas em relação aos sistemas de informação.

A atividade de modelagem requer do cientista da informação uma habilidade intelectual em abstrair o que de fato é de interesse em ser representado. Pode-se perceber o caráter subjetivo dessa atividade, uma vez que, a abstração implica em decisões de modelagem muitas vezes subjetiva e passível de discussão, ou seja, os elementos componentes do modelo podem ser construídos conforme perspectiva, pontos de vista ou dimensões de um domínio particular. Além disso, o estabelecimento dos elementos a serem usados no sistema envolve um conhecimento amplo e atual do domínio ou área de assunto na qual se está modelando. Essa escolha também se relaciona com a heteroglossia, isto é, com a linguagem própria dos integrantes da área.

O modelo, resultado da ação de modelar, é uma representação simplificada e inteligível de um domínio ou campo de estudo em que características essenciais são consideradas em sua composição para fins de comunicação e de entendimento sobre a realidade.

Alguns tipos de modelos são evidenciados na literatura de CI e campos correlatos como formas de representações consistentes para os sistemas de informação, a saber: modelo cognitivo, modelo de dados e modelo conceitual. Segundo Sayão (2001), os modelos que buscam descrever o sistema e sua estrutura são conhecidos como modelos de dados; os modelos cognitivos são os que mais se aproximam do modelo mental do usuário em relação ao sistema; finalmente, os modelos conceituais enquadram-se numa perspectiva de interpretar os usuários, o sistema e a interação entre eles.

Assim, a Ciência da Informação traz contribuições significativas para a modelagem de domínios em que teorias e metodologias já consolidadas no campo podem ajudar o modelizador em sua atividade de elaboração de modelos voltados a organização e a representação de linguagens documentárias, metadados, bases de dados, documentos multimídia e ontologias (HODGE, 2000; HJÖRLAND, 2007; LEMOS; SOUZA, 2020); e ainda podem proporcionar benefícios no acesso, disseminação, compartilhamento e uso do conhecimento pelos usuários. Para tal, destacam-se revisões teóricas e propostas sobre classificação e categorias que auxiliam a estruturação de conhecimento presente nesses domínios (RANGANATHAN, 1967; DAHLBERG, 1978; JACOB, 2004; BARITE, 2000); estudos sobre a natureza de relações conceituais e semânticas envolvidas em modelos de representação (DAHLBERG, 1978; HJÖRLAND, 2007; KHOO; NA, 2007); e modelos de referência enquanto recurso metodológico para auxiliar o processo de organização genérica de metadados para o campo do patrimônio cultural (IFLA, 2009; 2016; LE BOEUF et al., 2018; HARPRING, 2022).

Alguns modelos de referência no âmbito da catalogação (SPECTRUM, FRBROO, CIDOC CRM, EDM, CRMdig, Linked Art, CDWA, CCO, VRA Core, MPEG-7, DUBLIN CORE, LIDO, apenas para citar alguns) são recomendados no campo do patrimônio cultural, incluindo modelos de dados, modelos ontológicos, padrões de metadados, guias de gestão de acervos e guias de catalogação. Adicionalmente, instituições de patrimônio cultural (LEMOS et al., 2022; LEMOS; COELHO-JÚNIOR; CARMO, 2021; CHARLES et al., 2017; FINK, 2018; DIJKSHOORN et al., 2018; HARPRING, 2022) geralmente aderem a esses modelos para produzirem metadados de coleções de objetos culturais digitalizados ou nato digitais, padronizando aspectos únicos de coleções culturais tanto fisicamente quanto digitalmente, bem como fornecerem dados administrativos para descrever a digitalização, os direitos autorais e as disposições de uso dos objetos.

### **3.2 Ciência de Dados e qualidade de dados**

O campo da Ciência de Dados pode ser considerado relativamente novo, e cujo interesse se dá na agregação de conhecimentos, métodos e técnicas interdisciplinares, especialmente oriundas do campo da Ciência da Computação (VIRKUS; GAROUFALLOU, 2019), em que a automação é dada como elemento

central no processamento de informações, permitindo uma análise mais rápida e precisa de grandes conjuntos de dados, levando a uma tomada de decisão mais precisa e mais produtiva.

Desse modo, a CD aproveita de técnicas de Aprendizado de Máquina (AM) da Inteligência Artificial (IA) para aprender padrões em dados e inferir os mesmos padrões em novos conjuntos de dados (FACELI et al., 2021). Para tal, técnicas implementadas em algoritmos de AM possuem duas abordagens da IA, conhecidas como supervisionada (problemas de classificação) e não-supervisionada (problemas de associação e agrupamento). Tais abordagens buscam, em essência, facilitar a automação de processos documentais e melhorar a qualidade dos dados para fins de extração de seu valor, transformando-os em informação e conhecimento a partir de análise e mineração (WANG, 2018; VIRKUS; GAROUFALLOU, 2019). Torna-se válido, contudo, ressaltar a importância dessa visão ao destacar o processo de conversão de dados em informação, cujo tema é altamente central, relevante e estudado há anos pela área de CI.

Nesse sentido, a CI, conforme se observou em 3.1, traz um arcabouço teórico-metodológico em seu percurso científico bem consolidado e respeitado à produção de metadados visando dar significado e contexto ao recurso de informação por meio de modelos conceituais, regras de catalogação, linguagens documentárias e padrões de metadados. Como tal, os metadados ajudam na qualificação de bases de dados, tanto em termos de seu tratamento descritivo quando temático, tornando-as fontes de informação adequadas para a realização de experimentos e aplicações de interesse a CD, as HD e a comunidade de patrimônio cultural (HAGOOD, 2012; CLEMENT; CARTER, 2017; IFLA, 2020; EUROPEANA TECH, 2021).

A declaração da IFLA (IFLA, 2020), por exemplo, sobre bibliotecas e IA reforça que uma das principais preocupações ligadas à pesquisa de IA reside no fato de que muitos dados usados para o treinamento são incompletos, incorretos ou têm vieses sociais e culturais. A IFLA declara explicitamente que uma das contribuições das bibliotecas seria melhorar a qualidade dos dados para aplicações em CD. A curadoria digital (HIGGINS, 2011), nesse contexto, torna-se elemento central de representação e gestão ao longo de todo o ciclo de vida dos metadados em

ambiente informacional. Para tal, a adoção de uma política de qualidade de dados (SIQUEIRA et al., 2021) tem sido explorada para dar condições de geração de infraestrutura informacional adequada a aplicações e pesquisas em CD, especialmente em Aprendizado de Máquina (EUROPEANA TECH, 2021).

Pesquisas envolvendo técnicas de CD vêm sendo desenvolvidas ao longo da última década (CHARDONNENS et al., 2018; TARMIZI et al., 2019; DE CLERCQ et al., 2019; LIAO; ZHAO, 2020; PURWITASARI et al., 2020; LEMOS; COELHO JUNIOR, 2023) com temas dedicados a processamento de metadados, o que constitui também interesse para a CI, incluindo qualidade de dados e metadados; pré-processamento para melhoramento dos dados; automatização de tarefas repetitivas envolvendo catalogação; enriquecimento de metadados para a descrição semântica de conteúdos documentais; descobrimento de objetos e localização em imagens; e extração de textos relevantes. Também, nesse sentido, a CD contribui sobremaneira com a possibilidade da geração automática e semiautomática de metadados a partir de aplicações de processamento de linguagem natural em textos e processamento em documentos multimídia para identificação de padrões (EUROPEANA TECH, 2021; IFLA, 2020).

Como exemplos, há diversos algoritmos e estratégias de AM supervisionada que demandam dados bem descritos (valores de dados) para treinamento e teste de eficiência de seus resultados. A oferta de bases de dados de alta qualidade com dados classificados por meio de vocabulários controlados ou mesmo ontologias torna-se um atrativo fundamental para o desenvolvimento de aplicações e pesquisas em CD (MARTINS et al., 2022).

Ainda em se tratando de técnicas usadas em AM, muitos conjuntos de dados apresentam problemas de ruídos, dados inconsistentes, dados ausentes e dados redundantes. Entretanto, mesmo que os algoritmos envolvidos nessas técnicas já estejam preparados para tais realidades em relação aos problemas nos dados, dependendo de sua amplitude, esses problemas podem influenciar os resultados almejados, comprometendo, assim, o aprendizado do modelo empregado. Desse modo, técnicas de pré-processamento de dados são cada vez mais utilizadas nessas situações para identificar e minimizar a ocorrência desses problemas. Para tal,

ferramentas são empregadas na limpeza e transformação de dados, como OpenRefine (ABHILASH; MAHESH, 2021) e PDFBox (JIANG et al., 2018), visando normalização, limpeza, tratamento de problemas sintáticos, para que os dados (metadados) estejam em condições mínimas de qualidade para análise e mineração.

O projeto “Data Science in Libraries” financiado pelo Institute for Museum and Library Services (IMLS) nos Estados Unidos, por exemplo, publicou um relatório em 2018 chamado “Shifting to Data Savy: the future of data science in Libraries” (BURTON et al., 2018, p. 7) no qual menciona que as áreas de algoritmos, AM e técnicas estatísticas são muito importantes para os cientistas de dados; no entanto, o conjunto de conhecimentos mais importante para os cientistas de dados dominar são as técnicas de limpeza e preparação de dados. O relatório reforça que as bibliotecas poderiam se responsabilizar por essas ações, o que seria uma contribuição fundamental para a CD.

Por fim, instituições como bibliotecas, arquivos e museus e centros de informação em geral são instituições que tradicionalmente já possuem grandes conjuntos de dados em condições para aplicação em projetos e experimentos envolvendo IA e AM com seus conjuntos de dados. Falta, porém, uma atualização na formação do profissional da informação para lidar com novos serviços de inteligência no que diz respeito a conhecimentos em tecnologias relevantes para manipulação de dados na web (exs.: Python, SQL, PHP, Javascript, XML, entre outras); modelagem de bases de dados; APIs para coleta de dados em plataformas digitais; internet das coisas; big data, design de sistemas; apenas para citar algumas competências. Em síntese, faltam profissionais e pesquisas dedicadas aos temas da área (SOUZA; ALMEIDA; BARACHO, 2015; SEMELER; PINTO, 2019).

### **3.3 Ciência da Computação e qualidade de objetos digitais**

Na Ciência da Computação, a modelagem conceitual exerce um importante papel em áreas como banco de dados e projetos de sistemas de informação; engenharia de software e de domínio; projetos de sistemas baseados em conhecimento, incluindo técnicas de IA; processamento de linguagem natural; modelagem de empresas; e ontologias formais no contexto da Web Semântica (GUIZZARDI, 2005).

Os estudos sobre concepção de modelos conceituais destinados à construção de sistemas de informação mais interativos com o usuário têm sido cada vez mais intensificados na busca de uma representação mais próxima da realidade (GUIZZARDI, 2020; ALMEIDA; TEIXEIRA, 2020). Guizzardi (2005) assim como Mylopoulos (1992) assinalam que a modelagem está intrinsecamente associada à representação das coisas do mundo físico e social, independente de projeto específico ou escolhas tecnológicas, objetivando, portanto, entender (aprender), comunicar (entre os envolvidos no domínio) e raciocinar (resolver problemas) sobre uma dada porção da realidade, o que resultará numa conceituação consistente e coerente de um domínio de conhecimento.

Segundo Guizzardi (2005), o problema de interoperabilidade semântica existe porque as comunidades divergem em suas conceituações. A viabilização no processo de comunicação entre a conceituação e a linguagem tanto do ponto de vista do tratamento pela máquina quanto para a compreensão do usuário é denominada de compromisso ontológico (GUARINO, 1998; GUIZZARDI, 2005).

Entende-se, para a presente pesquisa, que a complexidade em se modelar conceituações envolvendo objetos culturais digitais, considerando a variedade e a riqueza de assuntos envolvidos em seus conteúdos, determina o emprego de SOCs mais sofisticados para o tratamento semântico da informação, o que inclui as ontologias formais.

Conforme já discorrido nesta pesquisa, o uso de metadados é a forma comumente empregada para agregar semântica em documentos (ZENG; QIN, 2016), entretanto a Web Semântica (BERNERS-LEE; HENDLER; LASSILA, 2001) propõe anotar conteúdo de documentos usando ontologias de domínio, o que é atestado por Shadbolt, Berners-Lee e Hall (2006) quando assinalam que anotação semântica é uma abordagem subjacente aos conceitos preconizados pela Web Semântica (web de dados) no que tange ao fornecimento de significado à organização da informação por meio de conexões lógicas entre os termos para promover interoperabilidade entre sistemas.

Nesse viés, ontologias são vistas nesta pesquisa como modelos de anotação mais sofisticados (ANDREWS; ZAIHRAYEU; PANE, 2012; LEMOS; SOUZA, 2020) em

termos de tratamento semântico da informação, o que permite um usuário descrever e interligar recursos existentes por meio de qualificadores como conceitos, instâncias, propriedades, relações e restrições mantidas entre tais recursos. Apresentam-se, portanto, como possibilidades de representação de conhecimento sobre esquemas de metadados na medida em que buscam organizar e padronizar conceitos, termos e definições aceitas por uma comunidade particular, permitindo que a máquina interprete as anotações com semânticas bem definidas, possibilitando compartilhamento de significado entre o anotador e o consumidor do documento.

Almeida (2013) destaca que pesquisas desenvolvidas em ontologias possuem similaridades com pesquisas seminais em CI, especialmente nos fundamentos da classificação para a concepção de terminologias consistentes, ao que se percebe o interesse pelo tema desde a década de 1990 pela comunidade *Library and Information Science* (VICKERY, 1997; ALMEIDA; TEIXEIRA, 2020) na busca de soluções para padronizar a linguagem envolvida no domínio visando facilitar a comunicação entre usuários e sistemas de informação.

Logo, ontologias têm sido usadas para anotação semântica em documentos em contextos das HD e na comunidade de patrimônio cultural. Por exemplo, na Arqueologia, cria-se um modelo de dados semântico para documentar fragmentos de objetos 3D em seus aspectos geométricos (CATALANO et al., 2020); na História, cria-se um modelo conceitual ontológico para a codificação e a difusão em rede de dados associados a acervos históricos fotográficos (ROBLEDANO-ARILLO, NAVARRO-BONILLA; CERDÁ-DÍAZ, 2020); no campo do Patrimônio Cultural Digital, modelos de dados semânticos são usados para a publicação online de coleções culturais (DIJKSHOORN et al., 2018; LEMOS et al., 2022); e na Arquitetura, cria-se uma ontologia de domínio endereçada a anotações semânticas de objetos 3D baseadas na realidade do estado de conservação de edifícios (MESSAOUDI et al., 2018).

O *Linked Open Data* (LOD) (BIZER; HEATH; BERNERS-LEE, 2009) destaca-se no contexto de construção de ontologias como uma técnica contemporânea de organização e tratamento documental na web que consiste numa proposta de uso

de padrões abertos concebidos pelo *World Wide Web Consortium* (W3C) a fim de interligar e anotar dados sob licença aberta por meio de um URI. A proposta possibilita que provedores de conteúdo enriqueçam seus esquemas de metadados com especificações estruturadas e bem definidas de conhecimento a partir de vocabulários controlados, ontologias e padrões de metadados, o que pode viabilizar o consumo e o reuso de informações de alta qualidade.

Numa perspectiva similar, os princípios norteadores do FAIR (WILKINSON et al., 2016; GUIZZARDI, 2020), acrônimo para *Findable* (possibilidade do dado ser encontrado), *Accessible* (acessibilidade), *Interoperable* (interoperabilidade), e *Reusable* (reutilização), buscam melhorar a capacidade das máquinas em localizar, por meio de uma identificação única e persistente, objetos digitais e seus metadados na rede, além de viabilizar autenticação e autorização para acessos, comunicação harmônica e eficaz com outras aplicações para diversos fins, e informações completas e descritas para compreensão e análise tanto por indivíduos quanto por agentes computacionais, visando reuso.

Nesse panorama, instituições de cultura inseridas em ambientes digitais (FINK, 2018; DIJKSHOORN et al., 2018; EUROPEANA TECH, 2021) vêm mantendo seus próprios vocabulários alinhados com modelos conceituais baseados em ontologias e vocabulários com menos rigor formal, como os SKOS. Um exemplo seria o uso de vocabulários do Getty (TRUST, 2023) que tipicamente descrevem pessoas, localizações, eventos e conceitos específicos relacionados à cultura, e são mantidos em conformidade com normas internacionais pelo The Getty Research Institute através de contribuições de comunidades interessadas em suas temáticas.

O cenário evidencia um panorama de uso e reuso de SOCs na organização de uma quantidade significativa de dados que tem sido gerada, conectada e distribuída na web, especialmente dados multimídia, os quais necessitam de metadados eficientes para gestão, organização e recuperação de conteúdos associados a acervos culturais distribuídos na rede. Apesar dos significativos avanços conceituais e práticos que têm sido realizados nas últimas décadas, existe a percepção de que poucos desses recursos estão de fato implementados nas instituições culturais brasileiras (MARTINS et al., 2022), o que se traduz no fato de

que seus sistemas de informação e projetos voltados para a web são ainda incipientes na adoção dessas práticas, dificultando, por exemplo, a disponibilização de infraestruturas informacionais adequadas para aplicações em CD que poderiam ser úteis às comunidades de HD, CI e outras, no reúso de objetos digitais.

#### **4 MODELO TEÓRICO-CONCEITUAL PARA A QUALIDADE DE OBJETOS CULTURAIS DIGITAIS**

Os fundamentos explorados e descritos no Capítulo 3 fornecem condições necessárias à proposição do modelo teórico-conceitual apresentado na Figura 4, em que se vislumbra uma representação sinóptica interdisciplinar envolvendo princípios teórico-metodológicos de 3 campos de conhecimento abordados na presente pesquisa, a CI, a CD e a CC. O modelo busca apontar explicitamente os encaminhamentos pontuais de pesquisa de cada campo envolvido visando à produção de metadados com qualidade; e as relações entre processos, instrumentos e produtos resultantes da ação de organizar e representar conteúdos em mídias diversas, dando como exemplo o domínio da cultura digital.

O modelo teórico-conceitual apresenta os elementos essenciais para a construção de uma política de qualidade de dados que dê condições necessárias às instituições de cultura produzirem bases de dados curadas que possam ser utilizadas para aplicações, pesquisas e análises de CD, por exemplo.

Em **CI**, as Representações Descritivas e Temáticas são evidenciadas, mostrando como as duas disciplinas empregam princípios teórico-metodológicos em processos que colaboram na geração de instrumentos de representação de conhecimento ou SOCs (modelos conceituais para a realidade documental, regras de catalogação, esquemas e padrões de metadados e linguagens documentais) endereçados à produção de metadados com qualidade para as fontes documentais.

Metadados quando produzidos a partir de padrões de documentação, incluindo modelos de referência, regras de catalogação, padrões de metadados e linguagens documentárias (GILLILAND, 2016; IFLA, 2016), tornam-se elementos centrais na produção de bases de dados com qualidade. Estas, por sua vez, são produtos de informação considerados cruciais para a sociedade quando adotados para realizar a mediação entre documentos e comunidade de usuários interessada na

preservação, busca, recuperação, acesso e reuso, pois referenciam e divulgam o conhecimento a partir do uso qualificado da informação.

Nesse ensejo, a comunidade das HD pode se beneficiar do uso dessas bases de dados curadas, de acesso aberto, acessíveis, de ampla representatividade e de interesse científico e cultural para a sociedade da informação, sendo um dos pontos fundamentais para a pesquisa científica e o desenvolvimento conceitual do campo.

Poole (2017, p. 100) destaca que bases de dados são instrumentos centrais para as HD em aspectos que vislumbram organização integrada de documentos, busca e recuperação, acrescentando que o tratamento adequado dessas bases, visando qualidade em seus dados, passa a ser perseguido pela comunidade, consciente da importância de atividades envolvendo descrição, anotação, coleta, agregação, armazenamento, migração e preservação de objetos digitais. Em suma, atividades inerentes a área de Curadoria Digital (HIGGINS, 2011).

A identificação de fontes documentais que possam ser utilizadas como base para pesquisas em HD, ou seja, que forneçam informações sobre seus acervos e ainda permitam reutilização a partir da disponibilização em formatos abertos e interoperáveis, torna-se um problema de grande interesse tanto para a CI quanto para as HD, evidenciando interesses de pesquisa em comum (KOLTAY, 2016).

Desse modo, fontes documentais em contextos digitais disponíveis, por exemplo, em bibliotecas, arquivos e museus são beneficiadas com informações bem descritas, classificadas e indexadas, o que proporciona aos usuários melhor condição de encontrar, identificar, selecionar e obter o item por meio de um catálogo interoperável na rede (IFLA, 2016).

Em **CD**, considerando que o uso de SOC nas práticas de organização da informação seja uma boa prática em termos de curadoria digital, o seu uso também traz benefícios diretos para a pesquisa em CD. O fato de poder utilizar metadados de boa qualidade oriundos de processos que refletem o uso de vocabulários controlados, regras de catalogação e modelos de metadados bem definidos amplia o potencial das aplicações de CD. Tais aplicações podem se dedicar menos às

etapas de pré-processamento de dados, em geral voltadas para melhorar a qualidade dos dados coletados, e mais às etapas de análise dos dados, modelagem algorítmica e identificação de padrões em textos e multimídia.

Nesse contexto, dados sem os preceitos de qualidade, ora apontados nesta pesquisa, dificilmente podem ser utilizados para aplicações mais elaboradas de CD, tais como algoritmos supervisionados, redes neurais, entre outros. Como já mencionado anteriormente, grande esforço de pré-processamento de dados será necessário para o desenvolvimento de pesquisas e aplicações. A depender do grau de problemas, dificilmente fluxos automatizados ou semi-automatizados poderiam ser propostos.

Tal perspectiva nos leva a compreender as contribuições de CD com a área de CI numa perspectiva de sua relação com os metadados, logo, possibilitando visualizar oportunidades interdisciplinares para os profissionais da informação em serviços de informações inovadores e de valor à sociedade. Podem-se citar técnicas de pré-processamento que visam melhorar a qualidade dos dados descritivos e temáticos nas bases de dados, sobretudo quando catalogados manualmente, incluindo normalização, limpeza, inclusão de valores ausentes, entre outros tratamentos (VIRKUS; GAROUFALLOU, 2020).

A CD também contribui nesse contexto com a possibilidade da geração automática e semiautomática de metadados a partir de aplicações de processamento de linguagem natural e de recursos audiovisuais. Têm-se também aplicações de CD que podem se valer de metadados de boa qualidade para aplicações de aprendizagem de máquina não-supervisionadas e supervisionadas fazendo bom reuso dos dados para clusterização, visualizações, inferências estatísticas, teste de hipóteses, reconhecimento de padrões em texto e multimídia, entre outros (GREENBERG, 2017; EUROPEANA TECH, 2021).

Finalmente, Greenberg (2017) reforça a compreensão do papel dos metadados na CD afirmando que se o campo não aprofundar a pesquisa sobre os metadados em seu repertório conceitual e operacional, o campo enfrentará problemas em seu progresso. Para o autor, os metadados são indispensáveis no desenvolvimento de

aplicações de CD, pois oferecem contexto para a interpretação, uma vez que descrevem os dados.

Em CC, destacam-se também as contribuições do campo numa perspectiva de exploração semântica de objetos digitais em rede, como é o caso dos dados abertos ligados da Web Semântica, dos princípios FAIR, como técnicas contemporâneas de organização e tratamento documental na web visando reutilização de dados para diversas finalidades na sociedade (BIZER; HEATH; BERNERS-LEE, 2009; GUIZZARDI, 2020).

Os modelos de anotação (ANDREWS; ZAIHRAYEU; PANE, 2012), uma frente de pesquisa em CC (embora não exclusivo para o campo), são adequados para descrever formalmente informações sobre recursos na web. Esses modelos são representados por tags, atributos, relações e ontologias, alguns dos quais, como as ontologias formais, são considerados estruturas sofisticadas para anotações em ambientes digitais, refletindo a evolução dos metadados para anotação semântica (SILVA; SOUZA, 2014; LEMOS; SOUZA, 2020), ou mesmo vocabulários controlados (ex.: The Art & Architecture Thesaurus) com menos rigor formal (SKOS), porém com o mesmo propósito em anotar conteúdo documental. Tais artefatos de representação podem ser utilizados como sistemas de organização do conhecimento, mantendo os princípios orientadores FAIR, a fim de melhor descrever os dados e garantir padronização, qualidade e troca de descrições, melhorando assim o escopo de interoperabilidade semântica (GUIZZARDI, 2020) entre diferentes esquemas de metadados e aplicações disponíveis na web.

Assim sendo, os artefatos de representação ora evidenciados no modelo teórico-conceitual proposto (Figura 4) podem ser considerados sínteses de um SOC que busca, primariamente, a organização de recursos de informação, produzindo relacionamentos; pontos de acesso; ordenação para os documentos em coleções; e facetas genéricas e específicas para navegação em sistemas web. Nesse sentido, tem-se que o emprego de SOCs para viabilizar a recuperação de informação nos mais diversos contextos informacionais justifica-se na medida em que buscam construir bons esquemas de dados, padronizar a entrada de dados nas bases de

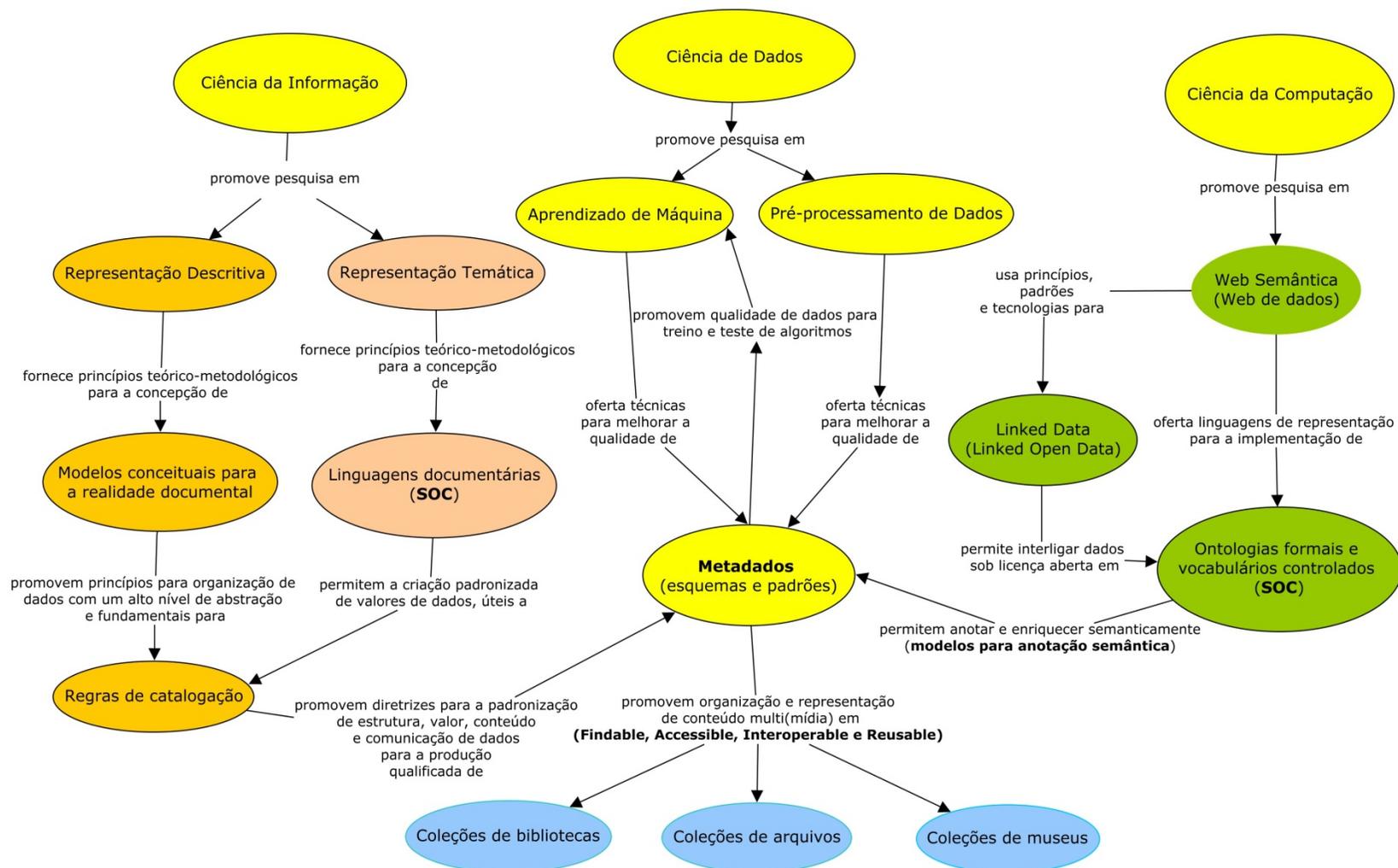
dados, facilitar a estratégia de busca e, conseqüentemente, melhorar a interação do usuário com o SRI.

O profissional da informação, nesse contexto, além de conhecer seus instrumentos de trabalho tradicionais, necessita dominar outros, tais como vocabulários controlados para plataformas digitais, ontologias formais, linguagens de representação de recursos de informação para a web, e também estar apto a lidar com as características multimídia dos novos documentos (LEMOS; SOUZA, 2020) a fim de representá-las a contento.

Adicionalmente, esse profissional tem como papel preponderante o desenvolvimento de terminologias específicas para a estruturação de bases de dados com qualidade, seguindo modelos para tratamento documental, ora apresentados e elucidados no presente artigo, visando serviços de informações inovadores e de valor à sociedade.

Finalmente, o modelo teórico-conceitual pode contribuir para o avanço de um paradigma que de fato vislumbra o uso dos dados não apenas para a gestão interna das instituições de cultura, mas que foca na socialização de suas informações, nas demandas do usuário final e na circulação em rede da informação cultural como instrumento educacional e de transformação social.

Figura 4 – Modelo teórico-conceitual para a qualidade de objetos culturais digitais



Fonte: elaborado pelos autores (2023)

## 5 CONSIDERAÇÕES FINAIS

A pesquisa aqui apresentada traz contribuições a várias comunidades que lidam com fontes de informação para as mais diversas finalidades de uso e reúso, ao identificar potenciais elementos interdisciplinares oriundos da Ciência da Informação, da Ciência de Dados e da Ciência da Computação, por meio dos quais esforços conjugados de pesquisas são endereçados à organização e à representação qualificada de entidades de informação (dados e metadados) para a produção de conhecimento cultural em rede.

Na CI, destacam-se os padrões de qualidade (modelos conceituais, regras de catalogação, padrões de metadados e linguagens documentárias) fundamentados nos aportes das catalogações descritivas e de assunto, se atendo, logo, à criação e ao uso de metadados (esquemas e padrões) pertinentes a variados perfis de aplicação.

Na CD, destacam-se as técnicas de aprendizado de máquina e de pré-processamento de dados para fins de produção automática de metadados qualificados que podem ser aproveitados para processos de catalogação, classificação e indexação de grandes bases de dados de itens de acervo; além do processo inverso de os metadados qualificados também servirem de produtos para aplicações de aprendizagem de máquina não-supervisionadas e supervisionadas, incluindo clusterização, visualizações, inferências estatísticas, teste de hipóteses, reconhecimento de padrões em texto e multimídia, entre outras.

E na CC, destaca-se o aspecto de organização e representação semântica da informação, em que se recomenda o uso de vocabulários semânticos, como ontologias formais e SKOS, seguindo os preceitos LOD e os princípios orientadores FAIR, na produção de metadados interoperáveis e enriquecidos para auxiliarem buscas, navegações, agregações, acessos e recuperações de objetos digitais na web.

O modelo teórico-conceitual, portanto, serve como um arcabouço teórico-metodológico capaz de apoiar a estruturação de sistemas de informação inteligentes, como agregadores culturais, repositórios e bibliotecas digitais, adequados ao acesso e a recuperação de objetos culturais qualificados na internet.

Nesse sentido, cientistas de dados, humanistas digitais, cientistas da informação e da computação interessados no reuso desses objetos qualificados, conseguem se assegurar das possibilidades de extração de valor para a geração de conhecimento útil à sociedade.

Desse modo, a pesquisa consegue responder a sua questão por meio da sistematização de conteúdos interdisciplinares no âmbito da CI, da CD e da CC, em conjunto com seus princípios teórico-metodológicos, discutidos e encaminhados como frentes de pesquisa na qualificação de objetos culturais digitais na rede. E, concomitante, cumpre com o objetivo de apresentar um modelo teórico-conceitual para organizar, representar e discutir elementos centrais advindos de campos científicos importantes que se articulam no sentido de unir esforços de pesquisa para a produção de serviços de informações mais interessantes e sofisticados à sociedade.

Os processos, instrumentos e produtos discutidos nos campos da CI, da CD e da CC permitem visualizar o avanço contínuo à área de Organização e Representação da Informação e do Conhecimento para ambientes de informações digitais, o que pode proporcionar melhorias aos sistemas de recuperação da informação, especialmente para o domínio da cultura digital.

Finalmente, o modelo teórico-conceitual proposto pode servir como um veículo pedagógico capaz de evidenciar habilidades e competências necessárias ao profissional da informação no planejamento de políticas de catalogação que almejem consistência descritiva aos objetos culturais de suas instituições para fins de integração de seus acervos e recuperação eficiente de objetos digitais neste ambiente. Espera-se, desse modo, estimular o debate em nível nacional sobre o tema, ampliando a interlocução entre os demais pesquisadores da CI que estão produzindo experiências similares e ainda em estágios iniciais de elaboração.

## REFERÊNCIAS

- ABADAL, E.; CODINA, L. *Bases de datos documentales: características, funciones y método*. Madrid: Síntesis, 2005.
- ABBAS, J. *Structures for organizing knowledge: exploring taxonomies, ontologies, and other schema*. New York: Neal-Schuman Publishers, 2010.
- ABHILASH, C.; MAHESH, K. Graph analytics applied to COVID19 karnataka state dataset. In: *4th International Conference on Information Science and Systems*. Edinburgh United Kingdom: ACM, 2021, p. 74–80. Disponível em: <https://dl.acm.org/doi/10.1145/3459955.3460603>. Acesso em: 17 abr. 2023.
- ALMEIDA, M.; SOUZA, R.; FONSECA, F. Semantics in the semantic web: A critical evaluation. *Knowledge Organization*, v. 38, n. 3, p. 187–203, 2011.
- ALMEIDA, M. B. Revisiting ontologies: A necessary clarification. *Journal of the American Society for Information Science and Technology*, v. 64, n. 8, p. 1682–1693. 2013.
- ALMEIDA, M. B.; TEIXEIRA, L. M.D. Revisitando os fundamentos da classificação: uma análise crítica sobre teorias do passado e do presente. *Perspectivas em Ciência da Informação*, p. 28-56, 2020.
- ANDREWS, P.; ZAIHRAYEU, I.; PANE, J. A classification of semantic annotation systems. *Semantic Web*, v. 3, n. 3, p. 223-248, 2012.
- ANSI/NISO Z39.19-2005 (R2010). *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Baltimore: NISO Press, p. 184, 2005.
- ANSI/NISO Z39.19-2005 (R2010). *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Baltimore: NISO Press, p. 184, 2005.
- BACA, M.; HARPRING, P.; LANZI, E.; MCRAE, L.; WHITESIDE, A. *Cataloging cultural objects: a guide to describing cultural works and their images*. Chicago: American Library Association, 2006.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. *Modern information retrieval: The concepts and technology behind search*. Addison Wesley, 2011.
- BARDIN, L. *Análise de conteúdo*. São Paulo: Edições 70, 2016.
- BARITE, M. G. The notion of “category”: Its implications in subject analysis and in the construction and evaluation of indexing languages. *Knowledge Organization*, v.27, n. 1–2, p. 4–10, 2000.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific american*, v. 284, n. 5, p. 34-43, 2001.

- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, v. 5, n. 3, p. 1–22, 2009.
- BORKO, H. Information science: what is it?. *American documentation*, v. 19, n. 1, p. 3-5, 1968.
- BORGMAN, C. L. *Scholarship in the digital age: Information, infrastructure, and the internet*. MIT Press, 2010.
- BOUGHIDA, K. CDWA Lite for Cataloguing Cultural Objects (CCO): A New XML Schema for the Cultural Heritage Community. In: *Proceedings of the XVI International Conference of the Association for History and Computing*. Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2005, p. 49–56. Disponível em: . Acesso em: 17 abr. 2023.
- BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento?. In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. *Anais...* São Paulo: Ancib, 2008.
- BURTON, M.; LYON, L.; ERDMANN, C.; TIJERINA, B. *Shifting to data savvy: The future of data science in libraries (Project report)*. University of Pittsburgh. 2018.
- CAMPOS, M. L. de A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ciência da Informação*, Brasília, v. 33, n. 1, p. 22- 32, jan./abr. 2004.
- CATALANO, C. E. et al. Representing quantitative documentation of 3D cultural heritage artefacts with CIDOC CRMdig. *International Journal on Digital Libraries*, v. 21, n. 4, p. 359-373, 2020.
- CHAPMAN, A. D. *Principles of Data Quality*. Copenhagen, 2005. Disponível em: <https://www.gbif.org/document/80509>. Acesso em: 17 abr. 2023.
- CHARDONNENS, A; RIZZA, E.; COECKELBERGS, M; et al. Mining user queries with information extraction methods and linked data. *Journal of Documentation*, v. 74, n. 5, p. 936–950, 2018.
- CHARLES, V.; CLAYPHAN, R.; ISAAC, A. *Definition of the Europeana Data Model v5.2.8*. 2017. Disponível em: [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation//EDM\\_Definition\\_v5.2.8\\_102017.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf). Acesso em: 17 abr. 2023.
- CLEMENT, T. E.; CARTER, D. Connecting theory and practice in digital humanities information work. *Journal of the Association for Information Science and Technology*, [S. l.], v. 68, n. 6, p. 1385–1396, 2017.
- DAHLBERG, I. A referent-oriented, analytical concept theory for INTERCONCEPT. *International Classification*, v.5, n. 3, p. 142–151, 1978.

DAHLBERG, I. Knowledge organization: Its scope and possibilities. *Knowledge Organization*, v.20, n. 4, p. 211–222, 1993.

DE CLERCQ, D.; DIOP, N.; JAIN, D.; et al. Multi-label classification and interactive NLP-based visualization of electric vehicle patent data. *World Patent Information*, v. 58, p. 101903, 2019.

DIJKSHOORN, C.; JONGMA, L.; OSSENBRUGGEN, J. V.; SCHREIBER, G.; WEELE, W. T.; WIELEMAKER, J. The Rijksmuseum collection as linked data. *Semantic Web*, v. 9, n. 2, p. 221-230, jan. 2018. Disponível em: <https://research.vu.nl/en/publications/the-rijksmuseum-collection-as-linkeddata>. Acesso em: 17 abr.2023.

ECKERSON, W. W. DATA QUALITY AND THE BOTTOM LINE: Achieving Business Success through a Commitment to High Quality Data. *The Data Warehouse Institute*, [S. l.], 2002. Disponível em: <http://download.101com.com/pub/tdwi/Files/DQReport.pdf>. Acesso em: 17 abr. 2023.

EUROPEANA TECH. *AI in relation to GLAMs Task Force: Report and recommendations*. 2021. Disponível em: [https://pro.europeana.eu/files/Europeana\\_Professional/Europeana\\_Network/Europeana\\_Network\\_Task\\_Forces/Final\\_reports/AI%20in%20relation%20to%20GLAMs%20Task%20Force%20Report.pdf](https://pro.europeana.eu/files/Europeana_Professional/Europeana_Network/Europeana_Network_Task_Forces/Final_reports/AI%20in%20relation%20to%20GLAMs%20Task%20Force%20Report.pdf). Acesso em: 17 abr.2023.

FACELI, K. et al. *Inteligência artificial: uma abordagem de aprendizado de máquina*. Rio de Janeiro: LTC, 2021.

FINK, E. E. *American Art Collaborative (AAC) Linked Open Data (LOD) Initiative Releases “Overview and Recommendations for Good Practices”*. 2018. 80 p. Disponível em: <https://repository.si.edu/bitstream/handle/10088/106410/OverviewandRecommendationsAccessible.pdf>. Acesso em: 17 abr. 2023.

FOSKETT, A. C. *A abordagem temática da informação*. Tradução de Antônio Agenor Briquet de Lemos. São Paulo: Polígono; Brasília: Ed.UnB, 1973.

GILLILAND, A. J. Setting the Stage. In: BACA, Murta. (ed.). *Introduction to metadata*. 3. ed. Los Angeles: Getty Research Institute, 2016. Disponível em: <https://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 17 abr. 2023.

GREENBERG, J. Big metadata, smart metadata, and metadata capital: Toward greater synergy between data science and metadata. *Journal of Data and Information Science*, v. 2, n. 3, p. 19, 2017.

GUARINO, N. (Ed.). *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98)*, June 6-8, Trento, Italy. IOS press, 1998.

GUIMARÃES, J. A. C. A análise documentária no âmbito do tratamento da informação: elementos históricos e conceituais. In: RODRIGUES, J.M.; LOPES, I.L.

(org.) *Organização e representação do conhecimento na perspectiva da ciência da informação*. Brasília: Thesaurus, 2003. (Estudos avançados em ciência da informação; 2). p. 100-117.

GUIMARÃES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-americana de Ciência da Informação*, 2008, janeiro/abril, v. [1], n. [1].

GUIMARÃES, J. A. C. Abordagens teóricas em tratamento temático da informação: catalogação de assunto, indexação e análise documental. In: García Marco, Francisco Javier (org.) *Avances y perspectivas en sistemas de información y de documentación*. Zaragoza: Prensas Universitarias de Zaragoza, 2009, pp. 105-117.

GUIZZARDI, G. *Ontological foundations for structural conceptual models*. 2005. 416 f. Tese (Doutorado em Ciência da Computação) - Universidade de Twente, Enschede, Holanda, 2005.

GUIZZARDI, G. *Ontological foundations for structural conceptual models*. 2005. 416 f. Tese (Doutorado em Ciência da Computação) - Universidade de Twente, Enschede, Holanda, 2005.

HAGOOD, J. A brief introduction to data mining projects in the humanities. *Bulletin of the American Society for Information Science and Technology*, v. 38, n. 4, p. 20-23, 2012.

HARPRING, P. *Metadata Standards Crosswalks*. 2022. Disponível em: [https://www.getty.edu/research/publications/electronic\\_publications/intrometadata/crosswalks.html#endnote1CCO](https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html#endnote1CCO). Acesso em: 17 abr. 2023.

HIGGINS, S. Digital curation: The emergence of a new discipline. *International Journal of Digital Curation*, v. 6, n. 2, p. 78-88, 2011.

HJØRLAND, B. Does the traditional thesaurus have a place in modern information retrieval? *Knowledge Organization*, v.43, n.3, p. 145-159, 2016. <https://doi.org/10.5771/0943-7444-2016-3-145>

HODGE, G. *Systems of knowledge organization for digital libraries: Beyond traditional authority files*. Digital Library Federation, Council on Library and Information Resources, 2000.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS (IFLA). *Declaração dos Princípios Internacionais de Catalogação*. Haia, 2016. Disponível em: [https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp\\_2016-pt.pdf](https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-pt.pdf). Acesso em: 17 abr. 2023.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS (IFLA). *Statement on libraries and artificial intelligence*. 2020. Disponível em: <https://repository.ifla.org/handle/123456789/1646>. Acesso em: 17 abr. 2023.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS (IFLA). *Statement on libraries and artificial intelligence*. 2020. Disponível em: <https://repository.ifla.org/handle/123456789/1646>. Acesso em: 17 abr. 2023.

JACOB, E. K. Classification and categorization: A difference that makes a difference. *Library Trends*, v. 52, n.3, p. 515–540, 2004.

JIANG, C.; LIU, J.; OU, Dongyang; et al. Implicit Semantics Based Metadata Extraction and Matching of Scholarly Documents. *Journal of Database Management*, v. 29, n. 2, p. 1–22, 2018.

JOUDREY, D. N.; TAYLOR, A. G.; MILLER, D. P. *Introduction to cataloging and classification*. 11 ed. Santa Barbara: ABC-CLIO, 2015.

KHOO, C. S. G.; NA, J.-C. Semantic relations in information science. *Annual Review of Information Science and Technology*, n.40, v.1, p. 157–228, 2006.

KOLTAY, T. Library and information science and the digital humanities: Perceived and real strengths and weaknesses. *Journal of Documentation*, v. 72, n. 4, p. 781–792, 2016.

LANCASTER, F. W. *Vocabulary control for information retrieval*. Information Resources Press, 1986.

LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 2004.

LE BOEUF, P. et al. (ed.). *Definition of the CIDOC Conceptual Reference Model: version 6.2.3*: International Council of Museums (ICOM); International Committee for Documentation (CIDOC), 2018. Documentation Standards Group. Disponível em: <http://www.cidoc-crm.org/Version/version-6.2.3-0>. Acesso em: 17 abr. 2023.

LE COADIC, Y-F. *A Ciência da Informação*. 2. ed. Brasília: Briquet de Lemos, 1996.

LEMOS, D. L. da S.; SOUZA, R. R. Knowledge Organization Systems for the Representation of Multimedia Resources on the Web: A Comparative Analysis. *KNOWLEDGE ORGANIZATION*, v. 47, n. 4, p. 300–319, 2020.

LEMOS, D. L. da S.; COELHO-JÚNIOR, A.; CARMO, D. do. Ontologias para anotação semântica em mídias: uma construção colaborativa de redes de conhecimento do patrimônio cultural. *Fronteiras da Representação do Conhecimento*, v. 1, n. 1, p. 94–125, 2021.

LEMOS, D. L. da S.; MARTINS, D. L. ; SA, A. M. E. ; MARTINS, L. C. ; CARMO, D. . A Proposal in Creating a Semantic Repository for Digital 3D Replicas: The Case of Modernist Sculptures in Public Spaces of Rio De Janeiro. *KNOWLEDGE ORGANIZATION*, v. 49, p. 151-171, 2022.

LEMOS, D. L. da S.; COELHO JUNIOR, A. Qualidade de dados em acervos do patrimônio cultural: uma avaliação diagnóstica semiautomática nos objetos

culturais sob gestão do Instituto Brasileiro de Museus. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 28, p. 1-22, 2023.

LIAO, X.; ZHAO, Z. Unsupervised Approaches for Textual Semantic Annotation, A Survey. *ACM Computing Surveys*, v. 52, n. 4, p. 1-45, 2020.

LUHN, H. P. A new method of recording and searching information. *American Documentation*, v. 4, n. 1, p. 14-16, 1953. DOI: <https://doi.org/10.1002/asi.5090040104>.

MARTINS, D. L.; LEMOS, D. L. da S.; OLIVEIRA, L. F. R.; SIQUEIRA, J.; CARMO, D.; MEDEIROS, V. N. Information organization and representation in digital cultural heritage in Brazil: Systematic mapping of information infrastructure in digital collections for data science applications. *Journal of the Association for Information Science and Technology*, [S. l.], p. asi.24650, 2022.

MESSAOUDI, T. et al. An ontological model for the reality-based 3D annotation of heritage building conservation state. *Journal of Cultural Heritage*, v. 29, p. 100-112, 2018.

MEY, E. S. A. *Introdução à catalogação*. Brasília: Briquet de Lemos Livros, 1995.

MOOERS, C. N. The next twenty years in information retrieval; some goals and predictions. *American Documentation*, v. 11, n. 3, p. 229-236, 1960. DOI: <https://doi.org/10.1002/asi.5090110306>.

MOREIRO GONZÁLEZ, J. A. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFBA, p. 128, 2011.

MYLOPOULOS, J. Conceptual modelling and Telos. *Conceptual modelling, databases, and CASE: An integrated view of information system development*, p. 49-68, 1992.

OLIVER, C. *Introdução à RDA: um guia básico*. Brasília: Briquet de Lemos 2011.

PIEDADE, M. A. R. *Introdução à teoria da classificação*. Rio de Janeiro: Interciência, 1977. 185 p.

POOLE, A. H. The conceptual ecology of digital humanities. *Journal of Documentation*, v. 73, n. 1, p. 91-122, 2017. DOI: <https://doi.org/10.1108/JD-05-2016-0065>.

RANGANATHAN, S. R. *Prolegomena to library classification*. Asia Publishing House. 1967.

RDA Steering Committee. *About RDA*. Online. 2022. Disponível em: <http://www.rda-rsc.org/content/about-rda>. Acesso em: 17 abr. 2023.

ROBLEDANO-ARILLO, J.; NAVARRO-BONILLA, D.; CERDÁ-DÍAZ, J. Application of Linked Open Data to the coding and dissemination of Spanish Civil War photographic archives. *Journal of Documentation*, v. 76, n. 1, p. 67-95, 2020.

SARACEVIC, T. Information science. *Journal of the American Society for Information Science*, v.50, n12, p. 1051–1063. 1999.

SAYÃO, L. F. Modelos teóricos em ciência da informação-abstração e método científico. *Ciência da Informação*, v. 30, p. 82-91, 2001.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. *Ciência da Informação*, v. 48, n. 1, 2019.

SHADBOLT, N.; BERNERS-LEE, T.; HALL, W. The semantic web revisited. *IEEE intelligent systems*, v. 21, n. 3, p. 96-101, 2006.

SILVA, D. L. da; SOUZA, R. R.; ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. *Ciência da Informação*, v. 37, n. 3, p. 60–75, 2008.

SILVA, D. L. da; SOUZA, R. R. Representação de documentos multimídia: dos metadados às anotações semânticas. *Tendências da Pesquisa Brasileira em Ciência da Informação*, [s. l.], v. 7, n. 1, 2014. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/119482>. Acesso em: 17 abr. 2023.

SIQUEIRA, J.; CARMO, D. do; MARTINS, D. L.; LEMOS, D. L. da S.; MEDEIROS, V. N.; OLIVEIRA, L. F. R. Elements for the construction of a data quality policy for the aggregation of digital cultural collections: the cases of the Digital Public Library of America.Inc and the Europeana Foundation. In: ÁLVAREZ, Edgar Bisset. (eds) *Data and Information in Online Environments: Second EAI International Conference-DIONE 2021*. Springer International Publishing, 2021.

SOUZA, R. R.; ALMEIDA, M. B.; BARACHO, R. M. A. Ciência da informação em transformação: Big Data, nuvens, redes sociais e Web Semântica. *Ciência da Informação*, v. 42, n. 2, 2015.

SVENONIUS, E. *The intellectual foundation of information organization*. Cambridge: The MIT Press, 2000.

TRUST, J. P. G. Getty Vocabularies. 2023. Disponível em: <https://www.getty.edu/research/tools/vocabularies>. Acesso em: 17 abr. 2023.

VICKERY, B. C. Ontologies. *Journal of Information Science*, [S.l.], v. 23, n. 4, p. 277-286, 1997.

VIRKUS, S.; GAROUFALLOU, E. Data science from a library and information science perspective. *Data Technologies and Applications*, v.53, n.4, p. 422–441. 2019. DOI: <https://doi.org/10.1108/DTA-05-2019-0076>. 2019.

WANG, L. Twinning data science with information science in schools of library and information science. *Journal of Documentation*, v.74, 2018.

WILKINSON, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, [S. l.], v. 3, n. 1, p. 160018, 2016.

WYNAR, B. S. *Introduction to cataloging and classification*. 7<sup>a</sup> ed. Colorado: Libraries Unlimited Inc., 1985.

ZENG, M. L.; QIN, J. *Metadata*. ALA Neal-Schuman, 2016.

ZENG, M. L. Interoperability. *Knowledge Organization*, v.46, n.2, p. 122-146, jan. 2019.