
Panorama das Abordagens de Integração de Dados: um estudo da produção científica brasileira em Ciência da Informação

*Overview of Data Integration Approaches: a study of Brazilian scientific production in
Information Science*

Sarah Rúbia de Oliveira Santos (1), Célia da Consolação Dias (2)

(1) Universidade Federal de Minas Gerais (UFMG), Brasil, sarahrubia22@gmail.com

(2) celiadias@gmail.com



Resumo

Em um cenário de crescente complexidade na gestão de informações, marcado pela diversidade de sistemas e demandas por interoperabilidade, este artigo propõe um panorama das abordagens de integração de dados na produção científica brasileira em Ciência da Informação, sugerindo uma categorização dessas abordagens. Este artigo tem como objetivo mapear e analisar as abordagens de integração de dados na produção científica brasileira em Ciência da Informação, propondo uma categorização dessas abordagens. A metodologia adotada é de natureza aplicada e exploratória, baseada em pesquisa bibliográfica, com análise de 33 artigos selecionados da Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci). Os resultados revelam três categorias principais: Integração de Sistemas e Tecnologias, Integração Semântica e Ontológica, e Integração de Dados e Informações Heterogêneas em Contextos Aplicados. A primeira categoria destaca a interoperabilidade técnica e a padronização de metadados, especialmente em bibliotecas digitais. A segunda enfoca o uso de ontologias e tecnologias da Web Semântica para superar barreiras semânticas. A terceira aborda estratégias para lidar com a heterogeneidade de dados em contextos específicos, como pesquisa científica e patrimônio cultural. Conclui-se que a produção científica brasileira tem evoluído de uma abordagem técnica para uma perspectiva mais semântica e contextualizada, visando superar desafios como a heterogeneidade tecnológica, a fragmentação dos dados e a complexidade da padronização.

Palavras-chave: Integração de dados; Interoperabilidade; Web Semântica; Linked Data; Ciência da Informação

Abstract

In a context marked by the growing complexity of information management, characterized by system diversity and interoperability demands, this article presents an overview of data integration approaches in Brazilian scientific production in Information Science, suggesting a categorization of these strategies. This article aims to map and analyze data integration approaches in Brazilian scientific production in Information Science, proposing a categorization of these approaches. The methodology is applied and exploratory, based on bibliographic research, with analysis of 33 articles selected from the Reference Database of Journal Articles in Information Science (Brapci). The results reveal three main categories: Integration of Systems and Technologies, Semantic and Ontological Integration, and Integration of Heterogeneous Data and Information. The first category highlights technical interoperability and metadata standardization, especially in digital libraries. The second focuses on the use of ontologies and Semantic Web technologies to overcome semantic barriers. The third addresses strategies for dealing with data heterogeneity in specific contexts, such as scientific research and cultural heritage. It is concluded that Brazilian scientific production has evolved from a technical approach to a more semantic and contextualized perspective, aiming to overcome challenges such as technological heterogeneity, data fragmentation, and standardization complexity.

Keywords: Data integration; Interoperability; Semantic Web; Linked Data; Information Science

1 Introdução

A integração de dados surgiu como forma de unificar sistemas de informação distribuídos e evoluiu para prover dados contextualizados de maneira conectada na web. No entanto, este processo de conexão entre os dados apresenta alguns desafios. Os desafios baseiam-se na heterogeneidade dos dados, na fragmentação e isolamento de repositórios digitais e acervos, na falta de coordenação entre instituições para prover acesso unificado e visualização de informações contextualizadas na web. Esses obstáculos dificultam não apenas a interoperabilidade entre sistemas, mas também o acesso e a reutilização de informações, especialmente em contextos multidisciplinares ou de grande escala, como os setores científico, cultural e governamental.

Um exemplo relevante da integração de dados pode ser ilustrado pela plataforma Conecta GOV.BR que unifica informações de sistemas do governo brasileiro visando reduzir a burocracia e facilitar os processos de preenchimento de formulários, emissão de certidões e envio de documentos para os cidadãos. A diversidade de sistemas e a falta de coordenação entre instituições podem resultar em duplicação de esforços e desperdício de recursos, no entanto, a partir dessa integração dos sistemas governamentais foi gerada uma economia de R\$1,4 bilhão em 2023 (Brasil 2023).

Ao pensar na importância de combinar dados de diferentes origens, percebe-se uma transformação substancial no tratamento da informação. A união efetiva de sistemas possibilita uma compreensão mais profunda dos recursos, gerando descobertas que podem orientar tanto decisões estratégicas quanto inovações. Na perspectiva da Ciência da Informação, esta integração transcende aspectos puramente técnicos: representa uma busca por otimizar como os dados fluem, assegurando que permaneçam acessíveis e adaptáveis a diversos contextos. Neste sentido, este artigo tem como objetivo mapear e analisar o panorama das abordagens de integração de dados na produção científica brasileira em Ciência da Informação, buscando identificar como a temática é abordada na área e de que forma as iniciativas se apresentam no cenário atual.

O artigo está dividido nas seguintes seções: inicialmente, discute-se a integração de dados, com foco em sua evolução conceitual, a relação com a interoperabilidade e os principais desafios para sua concretização (Seção 2). Em seguida, apresenta-se a metodologia utilizada no estudo (Seção 3) e analisa-se os resultados obtidos, destacando aspectos práticos e teóricos relevantes (Seção 4). Por fim, nas considerações finais são apresentados os principais achados do trabalho (Seção 5).

2 Fundamentação Teórica

Na presente seção são discutidos os principais aspectos da integração de dados. Apresenta-se brevemente a evolução do conceito de integração de dados, desde sua emergência até às abordagens contemporâneas, identificadas na literatura. Em seguida, são exploradas a integração de dados e a interoperabilidade, conceitos comumente utilizados como sinônimos, na literatura da Ciência da Informação. E, por fim, discute-se os principais desafios relacionados à integração de dados que precisam ser enfrentados e superados.

2.1 Evolução do Conceito de Integração de Dados

A discussão sobre integração de dados possui raízes interdisciplinares, conectando tanto a Ciência da Computação quanto a Ciência da Informação. Na década de 1960, ainda antes do surgimento formal do termo “integração de dados” na computação, estudos pioneiros no campo da Ciência da Informação já abordavam desafios análogos. Como destacam Campos, Campos e

Barbosa (2020), pesquisas sobre compatibilidade e conversão entre linguagens documentárias buscavam criar ferramentas para converter estruturas de dados heterogêneas, permitindo que usuários acessassem múltiplas fontes de informação. Esse esforço foi impulsionado pela explosão informacional da época e pela necessidade estratégica de compartilhamento de dados entre instituições governamentais. Embora focados em linguagens documentárias (como tesouros e classificações), esses trabalhos anteciparam questões centrais da integração, como a harmonização semântica e técnica entre sistemas distintos.

Na década de 1970, o conceito de integração de dados emergiu no contexto da Ciência da Computação e dos Sistemas de Informação. Um dos primeiros e mais importantes estudos que abordaram a integração de dados de forma acadêmica foi o de Chen (1976), na sua publicação sobre o modelo Entidade-Relacionamento (ER) no contexto de bancos de dados. Embora esse trabalho não trate especificamente da “integração de dados” como é entendida atualmente, ele introduz uma forma de representar dados de diferentes fontes e integrá-los por meio de modelos lógicos. Paralelamente, conforme Campos *et al.* (2020), os estudos sobre compatibilidade na Ciência da Informação entraram em declínio temporário, em parte devido à disseminação da automação de processos, que deslocou o foco para questões técnicas de armazenamento e recuperação.

Nos anos 1980, a integração de dados na computação consolidou-se em torno de desafios técnicos, como a integração de bancos de dados distribuídos, que exigiam coordenação entre sistemas com estruturas heterogêneas (relacionais, hierárquicas etc.) e localizados em ambientes físicos distintos (Sheth 1999). Na mesma década, a Ciência da Informação retomou os estudos sobre compatibilidade, agora adaptados ao ambiente computacional. Como observam Campos *et al.* (2020), as pesquisas passaram a abordar problemas sintáticos e semânticos em linguagens automatizadas, alinhando-se parcialmente aos esforços da computação para lidar com dados distribuídos.

A década de 1990 marcou uma mudança significativa no foco da integração de dados, que começa a mudar dos sistemas, da sintaxe e da estrutura para a semântica dos dados. O avanço da Web e a proliferação de formatos (*Application Programming Interfaces* - APIs, dados semi e não estruturados) exigiram abordagens que considerassem o significado e o contexto dos dados (Sheth

1999). Na Ciência da Computação, esse movimento se refletiu em sistemas voltados para interoperabilidade semântica, enquanto na Ciência da Informação, conforme Campos *et al.* (2020), os estudos sobre compatibilidade evoluíram para a elaboração de ontologias, ferramentas capazes de mediar a interoperabilidade entre sistemas por meio de representações conceituais compartilhadas. Essa convergência entre as áreas destacou a interdependência entre integração técnica e alinhamento semântico, preparando o terreno para discussões sobre interoperabilidade em escala global.

O surgimento da Web Semântica no final dos anos 1990, consolidado por iniciativas do World Wide Web Consortium (W3C) e impulsionado pelo artigo seminal de Berners-Lee, Hendler, e Lassila (2001), trouxe uma nova perspectiva para a integração de dados ao propor padrões e estruturas para conectar e interpretar informações de forma mais eficiente. Um marco inicial foi a especificação do *Resource Description Framework* (RDF), publicada pelo W3C em 1999, que estabeleceu as bases técnicas para esse movimento (W3C 1999). Posteriormente, os conceitos de *linked data*, discutidos por Heath e Bizer (2011), ampliaram sua aplicação. Conforme Santarém Segundo (2015), embora os conceitos de Web Semântica e dados abertos não tenham uma relação direta em sua origem, a expansão do movimento de dados abertos evidenciou a urgência de soluções de integração capazes de combinar fontes heterogêneas, demandando estruturas semânticas como as propostas pelo RDF e *linked data* para garantir acesso universal.

Um exemplo prático desses padrões e estruturas pode ser visto na área cultural. O CIDOC *Conceptual Reference Model* (CIDOC-CRM) é uma ontologia padrão ISO para o domínio cultural, desenvolvida para viabilizar a interoperabilidade semântica entre museus, bibliotecas, arquivos e outras instituições culturais. Sua relevância é atestada pela ampla adoção: centenas de instituições e projetos de pesquisa em andamento utilizam o modelo, consolidando-o como referência global (Tzitzikas *et al.* 2022). Mantido regularmente por um grupo dedicado, o CIDOC-CRM possui atualizações contínuas. O modelo oferece uma estrutura conceitual para a integração e compartilhamento de dados culturais. Essas iniciativas institucionais e acadêmicas, tanto na proposição quanto na aplicação de tecnologias, influenciam as práticas atuais, mantendo-se como referências centrais na busca por soluções semânticas e interoperáveis para a integração de dados em diferentes áreas.

2.2 Integração de Dados e Interoperabilidade

A integração pode ser entendida como o processo de unificar e padronizar os recursos de informação provenientes de diferentes fontes, a fim de promover a interoperabilidade e facilitar o acesso aos dados e informações de maneira coesa (Mello, Mesquita, e Vieira 2015; Zeng e Qin 2016). Nesse contexto, a integração ocorre tanto em um nível técnico, envolvendo a padronização e conexão dos recursos de informação, como em um nível conceitual, envolvendo a integração semântica e organizacional dos dados.

Os conceitos de integração e interoperabilidade são comumente utilizados como equivalentes na área de Tecnologias da Informação e Comunicação, no entanto, eles são diferentes. Mello, Mesquita e Vieira (2015) explicam que a integração é o processo de conectar vários sistemas e gerar uma dependência tecnológica entre eles, enquanto a interoperabilidade é o processo de comunicação entre sistemas sem gerar dependência tecnológica. Para Zeng e Qin (2016, p. 349, tradução nossa),

a interoperabilidade, não importa em que nível, está relacionada à capacidade de diferentes sistemas de informação se comunicarem entre si. Essa comunicação pode assumir várias formas, como transferência, troca, transformação, mediação, migração ou integração de informações.

Conforme Almeida (2002, p. 76), “os estudos relativos à integração de fontes heterogêneas têm demonstrado que se trata de um problema complexo, o qual deve ser abordado de diferentes formas. É importante, numa primeira etapa, definir que nível de interoperabilidade é desejado”, seja ele sistêmico, sintático, estrutural e/ou semântico (Sheth 1999). A Figura 1 ilustra as dimensões da interoperabilidade, apresentadas por Zeng (2019).

Figura 1 – Padrões e recomendações para interoperabilidade



Fonte: Traduzido de Zeng (2019, on-line).

A Interoperabilidade entre Sistemas, que está na base da pirâmide, está relacionada com troca de dados através de redes, computadores, aplicativos e serviços web, ou seja, demanda por uma infraestrutura tecnológica. Refere-se à capacidade de diferentes sistemas de computador e redes de comunicação trabalharem juntos (Sheth 1999). A interoperabilidade em nível de sistema (camada técnica) envolve a superação de incompatibilidades entre hardware, sistemas operacionais e infraestruturas de rede, garantindo a troca técnica de dados entre ambientes heterogêneos. (Zeng 2019).

A Interoperabilidade Sintática, construída sobre a base sistêmica, envolve a capacidade de diferentes sistemas entenderem e processarem os formatos de dados uns dos outros, o que inclui a compatibilidade de tipos de dados e formatos, permitindo que dados sejam trocados e compreendidos entre sistemas diferentes (Sheth 1999). Para isso, utiliza tecnologias da Web Semântica para estabelecer uma linguagem comum para representação dos dados e formatos padronizados de intercâmbio, tornando a codificação consistente e permitindo que esquemas de metadados sejam compartilhados conforme especificações técnicas. Se houver diferenças na codificação e representação, isto é, se os dados diferem em relação aos pontos mencionados, não há interoperabilidade sintática; portanto, os dados não podem ser trocados ou mesclados (Zeng 2019). Alguns dos mais importantes padrões, como o RDF, a *Ontology Web Language* (OWL) e

o *Simple Knowledge Organization System* (SKOS), seguem as recomendações oficiais do W3C para viabilizar uma web semântica.

A Interoperabilidade Estrutural, que compõe o terceiro nível, está apoiada na troca de dados por meio de modelos de dados, estruturas de dados e esquemas de dados predefinidos. Para que a troca de dados ocorra, é necessário ter uma arquitetura de dados consistente, fazer uso de modelos conceituais para uma modelagem padronizada que constitua relações estruturais bem definidas, promovendo a organização lógica da informação (Sheth 1999; Zeng 2019).

Os modelos conceituais desempenham um papel fundamental na facilitação da interoperabilidade entre acervos de diferentes unidades de informação (Marcondes 2015). Eles fornecem uma representação abstrata e genérica de um determinado domínio da realidade, permitindo compreender, raciocinar e estabelecer um entendimento comum sobre a realidade dentro de uma comunidade. Por terem uma estrutura pré-definida, esses modelos permitem a troca de dados sem perdas (Zeng 2019). Exemplos incluem o *Europeana Data Model* (EDM), o BibFrame, o CIDOC-CRM, o FRBR, o IFLA LRM e o RiC-CM, usados em bibliotecas, arquivos e museus para integrar a descrição de recursos de informação (Carrasco 2019; Carrasco e Vidotti 2019).

As dimensões sintática e estrutural fornecem as bases tecnológicas e as estruturas de representação necessárias para que a dimensão semântica seja possível. Com a aplicação das tecnologias da Web Semântica e com os metadados adicionados a uma estrutura de descrição, é possível combinar fontes de dados, de modo que os dados de uma fonte podem ser enriquecidos semanticamente com dados vinculados adicionais de outra fonte (Hyvönen 2012; Marcondes 2021).

A Interoperabilidade Semântica, topo da pirâmide, é responsável pela troca de dados por meio de contextos pré-definidos e entregues em Sistemas de Organização do Conhecimento. Essa dimensão trata das inconsistências na terminologia e nos significados dos metadados. Para atribuir significado ao que está sendo representado, é necessário considerar o contexto envolvido e suas variáveis. Segundo Zeng (2019, on-line, tradução nossa), “um contexto é comumente entendido como as circunstâncias que formam o cenário para um evento, declaração, processo ou ideia, e em termos dos quais o evento, declaração, processo ou ideia pode ser entendido e avaliado”. A

dimensão semântica está baseada em padrões como a ISO 25964, que asseguram a preservação do significado dos dados, a interpretação consistente entre sistemas, a manutenção do contexto cultural e a compreensão compartilhada dos conceitos (ISO 25964-1:2011). Iniciativas como os princípios FAIR (Wilkinson *et al.* 2016) reforçam a interdependência entre os diferentes tipos de interoperabilidade ao levantarem a necessidade de que os dados disponíveis na web sejam localizáveis, acessíveis, interoperáveis e reutilizáveis, o que demanda alinhamento entre metadados robustos, ontologias e infraestruturas abertas. A interoperabilidade semântica é crucial para lidar com a heterogeneidade intrínseca dos dados e para suportar consultas de alto nível e sensíveis ao contexto (Sheth 1999).

A visão dos padrões e recomendações da interoperabilidade em formato de pirâmide evidencia como cada nível depende dos anteriores e como a verdadeira integração só pode ser alcançada quando todas as camadas estão adequadamente implementadas e funcionando em harmonia. Para que a integração de dados ocorra efetivamente, é necessário enfrentar os desafios que surgem à medida que a escala aumenta.

2.3 Desafios para a Integração de Dados

Os desafios da integração de dados são discutidos com base, principalmente, em Sheth (1999), Doan, Halevy e Ives (2012), Kadadi *et al.* (2014) e Campos *et al.* (2020). Esses desafios podem ser organizados em três categorias principais: sistemas, semântica e organizacional, cada uma com especificidades que afetam a integração de dados em diferentes contextos.

No âmbito dos sistemas, a integração de dados enfrenta desafios técnicos significativos devido à natureza distribuída das informações, que podem estar espalhadas por sistemas isolados ou redes globais interconectadas. Essa distribuição geográfica, aliada à heterogeneidade de padrões de armazenamento e descrição dos dados, dificulta a comunicação entre sistemas e a execução eficiente de consultas distribuídas (Sheth 1999; Doan *et al.* 2012; Kadadi *et al.* 2014). O crescimento exponencial das fontes de dados na web impõe desafios adicionais para a otimização de consultas e gerenciamento de grandes volumes de dados.

A complexidade técnica é intensificada pelo *Big Data*, caracterizado por grandes volumes de dados heterogêneos gerados em alta velocidade, como os provenientes de sensores inteligentes e interações em redes sociais (Kadadi *et al.* 2014). A necessidade de soluções escaláveis torna-se

essencial, exigindo sistemas de processamento rápido e estratégias de armazenamento híbrido (HDD/SSD). Além disso, os processos de Extração, Transformação e Carregamento (*Extraction, Transformation, Load* - ETL) enfrentam desafios como a conversão de dados não estruturados para formatos padronizados e a otimização de recursos durante o carregamento. A infraestrutura de rede também afeta a integração, com limitações como largura de banda insuficiente e perda de pacotes comprometendo a eficiência na transferência de grandes volumes de dados.

A gestão de metadados é um fator crucial para garantir interoperabilidade. Sheth (1999) destaca a necessidade de capturar e estruturar informações independentemente de sua representação original, enquanto Campos *et al.* (2020) ressaltam a importância da padronização e da adoção de tecnologias como *linked data* e modelos conceituais para conectar sistemas heterogêneos. Esses desafios são particularmente relevantes em áreas como patrimônio cultural e pesquisa científica, onde a diversidade de fontes e formatos demanda soluções flexíveis para integração.

No campo semântico, a heterogeneidade persiste como uma barreira crítica. Diferenças em esquemas de organização, conflitos terminológicos (sinônimos, homônimos) e variações estruturais exigem alinhamento contextual para garantir interpretação consistente dos dados (Doan *et al.* 2012). Campos *et al.* (2020) posicionam a interoperabilidade semântica como um elemento essencial para ambientes de *e-Science* e ciência aberta, propondo metodologias de mapeamento entre vocabulários distintos, sem a imposição de uma linguagem universal. Exemplos incluem a matriz de compatibilidade de Dahlberg e o uso de ontologias como SKOS e OWL, que estruturam relações semânticas para harmonizar dados multidisciplinares.

Kadadi *et al.* (2014) destacam que a inconsistência dos dados em fontes não estruturadas demanda técnicas avançadas, como etiquetagem e filtragem, para garantir a qualidade dos dados. Além disso, Grunzke *et al.* (2019) enfatizam a necessidade de descrições detalhadas e armazenamento eficiente, enquanto Felicetti e Murano (2021) ressaltam a importância de modelos conceituais para representar entidades complexas e possibilitar a interoperabilidade entre dados de diferentes áreas, como arqueologia, história e linguística.

A escalabilidade dos sistemas de informação depende do uso de múltiplas ontologias independentes, que devem ser capazes de processar solicitações baseadas em uma ontologia

específica, mesmo em ambientes onde os dados utilizam ontologias distintas. Sheth (1999) aponta que um dos principais desafios é representar correlações de informação em nível semântico, independentemente do meio ou da localização das fontes. No entanto, a falta de padronização e compatibilidade entre ontologias e vocabulários de domínio compromete a portabilidade e a interoperabilidade entre sistemas e disciplinas (Chen e Ou 2020; Abgaz *et al.* 2021).

No contexto organizacional, a fragmentação de repositórios e a resistência institucional ao compartilhamento de dados, motivada por preocupações com segurança, privacidade ou controle, dificultam a colaboração (Sheth 1999; Doan *et al.* 2012). Kadadi *et al.* (2014) também apontam a escassez de recursos, como financiamento e profissionais qualificados, como obstáculo para a implementação de soluções robustas.

A diversidade cultural e linguística entre comunidades de pesquisa também impõe desafios à ciência aberta, exigindo metalinguagens intermediárias para evitar ambiguidades sem comprometer identidades locais. Para enfrentar essas questões, é necessário investir em governança de dados eficaz, com políticas claras de curadoria, preservação e descrição de metadados em repositórios abertos (Campos *et al.* 2020).

Resolver completamente os desafios da integração de dados é improvável (Doan *et al.* 2012). Como destacam Kadadi *et al.* (2014) e Campos *et al.* (2020), a integração de dados é um problema multidimensional que exige sinergia entre avanços técnicos, estruturas semânticas e cooperação institucional. No entanto, diversas iniciativas estão sendo desenvolvidas para mitigar esses desafios, indicando que a solução passa tanto por inovação tecnológica quanto por mudanças paradigmáticas na gestão do conhecimento e na colaboração global.

3 Metodologia

A pesquisa adotou uma abordagem exploratória e bibliográfica buscando fornecer um panorama das abordagens de integração de dados existentes no âmbito da Ciência da Informação brasileira. Como pesquisa exploratória, buscou-se levantar informações sobre o tema integração de dados, com vistas a mapear as condições de manifestação desse objeto na literatura (Severino 2017). Do ponto de vista dos procedimentos técnicos, trata-se de uma pesquisa bibliográfica,

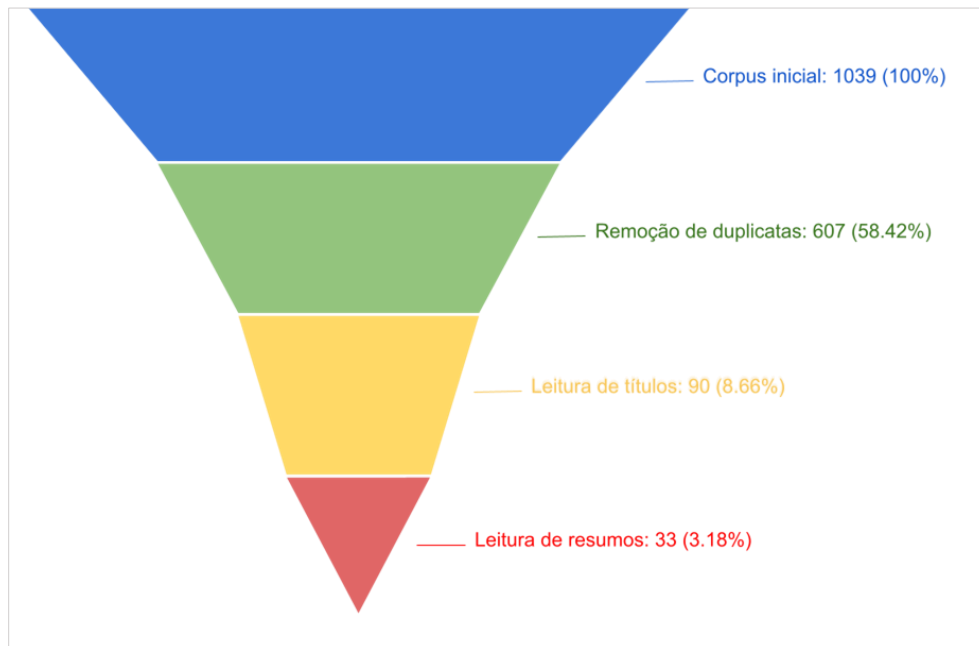
fundamentada na análise de materiais já publicados e disponíveis na literatura acadêmica da Ciência da Informação (Prodanov e Freitas 2013).

O ambiente de pesquisa é a Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci), repositório especializado, referência na compilação e organização de publicações científicas no campo da Ciência da Informação provenientes de fontes brasileiras e baseadas na América Latina. Este recurso concentra-se na reunião sistemática de artigos de periódicos, trabalhos de eventos, livros e outras fontes acadêmicas relevantes (Brapci 2025).

A estratégia de busca concentrou-se na utilização do descritor “integração” em todos os campos (título, resumo, palavras-chave). A opção pelo termo simples, sem associação com a palavra “dados”, se deu para ampliar o escopo da pesquisa e capturar estudos com perspectivas mais diversificadas sobre integração. A busca foi realizada na base de dados em dezembro de 2024 e resultou em um conjunto inicial de 1.039 estudos, compostos de artigos de periódicos e artigos apresentados em eventos. Não foi realizado recorte temporal, para que fosse possível observar a evolução das discussões sobre o conceito de integração na literatura da área.

Para organizar e analisar esses registros, os metadados foram exportados no formato de planilha (.csv) e ordenados por título, o que permitiu a identificação e remoção de duplicatas. Após essa etapa, restaram 607 estudos, que foram submetidos a uma análise preliminar dos títulos. O objetivo era selecionar apenas os trabalhos que abordavam, de forma direta ou indireta, a integração de dados e temas relacionados, como sistemas de informação, tecnologias semânticas (web semântica, *linked data* e ontologias), gestão e reúso de dados e contextos aplicados de integração. Essa triagem reduziu o número de publicações para 90, as quais foram então submetidas à leitura dos resumos e das palavras-chave. Após esta análise, foram identificados 33 estudos que efetivamente contribuíram para o mapeamento de como a temática de integração de dados é abordada na literatura brasileira de Ciência da Informação. O processo de busca e filtragem dos trabalhos é ilustrado na Figura 2.

Figura 2 – Estratégia de busca e seleção de estudos



Fonte: Elaborado pelas autoras.

Por fim, foi realizada uma análise de conteúdo dos artigos selecionados, visando a compreensão e interpretação dos dados coletados de forma sistemática e permitindo categorizar os trabalhos a partir da semelhança nas abordagens identificadas. Este processo foi dividido em três etapas: 1) levantamento inicial: extração de termos e expressões; 2) agrupamento por tipos de integração; e 3) contextualização em categorias.

Inicialmente, realizou-se um levantamento dos termos e expressões utilizados pelos autores para abordar a integração de dados, identificando-se uma diversidade vocabular que reflete distintas abordagens e desafios. Expressões como “integração de sistemas”, “integração de fontes heterogêneas”, “integração de ontologias”, “integração e interoperabilidade entre bibliotecas digitais” e “integração semântica dos dados” evidenciaram ênfase em problemas como heterogeneidade de formatos, conflitos de significado e necessidade de padronização, além de aplicações práticas em domínios como patrimônio cultural, acervos digitais e pesquisa científica. Nesta fase, priorizaram-se termos que explicitavam mecanismos de integração, excluindo descrições genéricas.

Em seguida, os termos foram agrupados em tipos de integração, com base em similaridades técnicas e objetivos comuns. Essa etapa revelou quatro eixos principais: (1) integração de dados

heterogêneos, abrangendo a combinação de fontes com formatos ou origens distintas; (2) integração semântica/ontológica, focada em resolver ambiguidades contextuais por meio de tecnologias semânticas; (3) integração de sistemas, relacionada à interoperabilidade técnica entre infraestruturas; e (4) integração institucional/contextual, que aborda colaborações entre instituições. O critério de agrupamento considerou tanto a funcionalidade (resolução de heterogeneidade técnica *vs.* semântica) quanto o nível de atuação (dados, sistemas, organizações). As categorias emergiram dos próprios dados e refletiram a evolução do conceito de integração de dados, bem como os níveis de interoperabilidade apresentados nas seções teóricas.

Por fim, os tipos de integração foram contextualizados em três categorias estratégicas, a saber:

- **Integração de Sistemas e Tecnologias**, que enfatiza a interoperabilidade técnica e padronização de metadados (ex: modelos como CIDOC-CRM para acervos culturais);
- **Integração Semântica e Ontológica**, centrada na resolução de conflitos contextuais por meio de ontologias e tecnologias da web semântica;
- **Integração de Dados e Informações Heterogêneas em Contextos Aplicados**, aplicada a contextos específicos como gestão organizacional, pesquisa científica e patrimônio cultural.

A seguir, as categorias são apresentadas e discutidas com base nos autores que as compõem.

4 Resultados e Discussão

A análise documental da tipologia revelou artigos publicados em periódicos e artigos apresentados em eventos acadêmicos, com um intervalo temporal abrangendo os anos de 2001 a 2024. Os resultados da pesquisa permitiram identificar três categorias principais relacionadas à integração de dados no campo da Ciência da Informação, com base no que foi abordado em cada estudo. São elas: Integração de Sistemas e Tecnologias; Integração Semântica e Ontológica; e Integração de Dados e Informações Heterogêneas em Contextos Aplicados. Essa estrutura em três partes permite mapear as camadas da integração (técnica, semântica e organizacional) destacando

como a Ciência da Informação enfrenta desafios como a fragmentação de dados, a diversidade contextual e a necessidade de colaboração institucional.

Observa-se que as categorias identificadas não são mutuamente excludentes, um artigo pode constar em mais de uma categoria. Conforme discutido na fundamentação teórica, a integração de dados envolve múltiplos níveis de interoperabilidade, os quais apresentam interdependências significativas entre si, algo que foi evidenciado também no agrupamento das categorias. No Quadro 1, os autores foram agrupados pelo tipo de integração de dados e contexto que abordam nos estudos analisados.

Quadro 1 – Abordagens de Integração de Dados na Ciência da Informação Brasileira

Integração de Sistemas e Tecnologias	
Foco na interoperabilidade e padronização de metadados na web	Alves e Souza (2007) Dias <i>et al.</i> (2022) Marcondes e Sayão (2001) Nóbrega <i>et al.</i> (2022) Serra <i>et al.</i> (2017) Simionato, Arakaki e Santos (2017)
Integração Semântica e Ontológica	
Aspecto Lógico e Contextual	Alves e Souza (2007) Campos (2007) Castro e Santos (2010) Lemos e Souza (2017) Lemos e Souza (2019) Marcondes (2017) Melo, Botega e Santarém Segundo (2017) Santos e Dias (2022) Santos e Dias (2024) Silva e Souza (2014) Sousa e Ribeiro (2021)
Integração de Dados e Informações Heterogêneas em Contextos Aplicados	
Contexto Organizacional	Almeida, Xavier e Barreto (2018) Almeida (2002) Alves e Souza (2007) Lemos e Souza (2019) Nóbrega <i>et al.</i> (2022) Schulz <i>et al.</i> (2009) Silva e Souza (2014)

Contexto da Pesquisa Científica	Alencar, Bochner e Amaral (2018) Carvalho Segundo <i>et al.</i> (2024) Dias <i>et al.</i> (2022) Jorge <i>et al.</i> (2020) Santos e Rockembach (2018)
Contexto do Patrimônio Cultural	Alves <i>et al.</i> (2017) Carrasco e Vidotti (2019) Carrasco e Vidotti (2020) Carrasco, Vidotti e Santarém Segundo (2017) Marcondes (2016) Marcondes (2019) Ribeiro, Borges e Sousa (2020) Santos e Dias (2024) Simionato e Santos (2017)

Fonte: Elaborado pelas autoras.

A categoria de **Integração de Sistemas e Tecnologias** engloba os trabalhos que focam na integração de sistemas, como bibliotecas digitais e repositórios, e apresentam tecnologias e padrões utilizados para este fim. Marcondes e Sayão (2001), em um dos primeiros trabalhos a abordar a integração de sistemas na Ciência da Informação, descrevem a implementação da Biblioteca Digital Brasileira em Ciência e Tecnologia (BDB), enfatizando a importância da interoperabilidade através de protocolos como Z39.50 e OAI-PMH. Em uma iniciativa mais recente, as Ferramentas de Apoio à Integração de Sistemas (FAIS) são apresentadas por Nóbrega *et al.* (2022). Essas ferramentas foram criadas para tratar especificamente os desafios do ecossistema brasileiro de Educação, Ciência, Tecnologia e Inovação, considerando desafios como a heterogeneidade dos sistemas de informação, a dispersão e a inconsistência dos dados, e a falta de padrões dos esquemas de metadados e ontologias para representação de domínios.

Nesta categoria, também é levantada a questão da padronização e compatibilidade entre sistemas e tecnologias. Alves e Souza (2007), em estudo conduzido na Embrapa, investigam a correspondência entre os padrões de metadados Dublin Core e MARC 21, demonstrando as relações de correspondências entre os dois esquemas e enfatizando como esse tipo de iniciativa pode otimizar a organização e o reuso de metadados. Serra *et al.* (2017), por sua vez, analisam a aderência do MARC 21 e do formato ONIX aos princípios de descrição bibliográfica, buscando identificar de que forma esses formatos se alinham aos princípios de conveniência do usuário,

acurácia e suficiência de Svenonius (2000 *apud* SERRA *et al.* 2017). Por fim, Simionato, Arakaki e Santos (2017) investigam de que forma padrões como Dublin Core e BIBFRAME podem otimizar a interoperabilidade e facilitar a navegação entre acervos digitais de bibliotecas, arquivos, museus e galerias de arte. Os estudos são aplicados em contextos distintos, mas compartilham a preocupação com a necessidade de conectar sistemas e tecnologias com vistas a fornecer uma navegação mais fluida no ambiente web, permitindo que o usuário tenha acesso ao que busca, mas também a recursos informacionais relacionados.

A categoria **Integração Semântica e Ontológica** abrange a utilização de ontologias, vocabulários controlados e tecnologias da Web Semântica para alinhar conceitualmente e enriquecer informações provenientes de fontes heterogêneas, promovendo uma compreensão mais rica e precisa dos dados. Santos e Dias (2022), ao fundamentar historicamente o conceito de “integração de dados”, explorando sua evolução desde os primeiros esforços para padronizar catálogos até as tecnologias modernas da Web Semântica e Dados Interligados, destacam a integração semântica como uma direção promissora para as pesquisas no campo da Ciência da Informação.

Entretanto, os estudos sobre integração de ontologias no Brasil datam da primeira década dos anos 2000, quando a Web Semântica estava se estruturando. Em 2007, Campos (2007) enfatizou a necessidade de definições conceituais robustas para a criação e integração de ontologias, especificamente no domínio da bioinformática. Castro e Santos (2010), por sua vez, analisaram a MarcOnt Initiative, que utiliza tecnologias da Web Semântica para superar as limitações do MARC21 na descrição de objetos disponíveis em bibliotecas digitais. Os autores demonstram, por meio da ferramenta JeromeDL, como uma ontologia bem estruturada pode descrever a forma e o conteúdo de recursos informacionais de maneira compreensível por máquinas.

Considerando o ambiente da Web Semântica, Marcondes (2017) destaca a perspectiva dos Dados Abertos Interligados (*Linked Open Data* - LOD) que propõe a utilização de padrões RDF e Internationalized Resource Identifier (IRI) para transformar catálogos de arquivos, bibliotecas e museus em redes de dados interconectados. Nesse contexto, Melo, Botega e Santarém Segundo

(2017) contribuem com uma metodologia para avaliar a qualidade dos dados conectados, utilizando os requisitos e princípios de qualidade propostos pelo projeto LOD.

No campo multimídia, Lemos e Souza (2017, 2019) propõem modelos ontológicos para anotar semanticamente documentos multimídia, integrando padrões como MPEG-7 e Dublin Core. Essas propostas visam superar a falta de padronização e permitir um tratamento semântico mais rico e interoperável, melhorando a recuperação de informações multimídia na web. Esses avanços são complementados pelas discussões de Sousa e Ribeiro (2021) sobre o uso do modelo conceitual Europeia Data Model (EDM) para integrar acervos relacionados a Oswaldo Cruz, conectando documentos diversos por meio de vocabulários como FOAF, BIO e CIDOC-CRM. A abordagem semântica tem um potencial significativo para a contextualização e recuperação de informação, mas sua implementação demanda esforços para garantir alinhamento conceitual e adoção de padrões reconhecidos internacionalmente.

Os artigos identificados como pertencentes à categoria **Integração de Dados e Informações Heterogêneas em Contextos Aplicados** abordam, em grande parte, o desafio da heterogeneidade dos dados (Sheth 1999; Doan *et al.* 2012; Santos e Dias, 2024) e propõem soluções em contextos específicos. Santos e Dias (2024), identificaram que a heterogeneidade e a fragmentação dos repositórios continuam como obstáculos centrais para a efetivação da integração de dados e ressaltam a importância de vocabulários controlados e tecnologias como *linked data* para superar essas barreiras. Nos estudos levantados, a heterogeneidade dos dados é abordada em três contextos, a saber: contexto organizacional, da pesquisa científica e do patrimônio cultural.

No contexto organizacional, Almeida (2002) destaca a relevância da integração de fontes heterogêneas de dados para a gestão do conhecimento, especialmente em intranets corporativas, propondo o uso de ontologias, metadados e softwares que intermedeiam a comunicação e a gestão de dados entre aplicações distintas. Nos setores de negócio, ciência, governo, mídia, radiodifusão e cultura, Silva e Souza (2014) exploram a integração de dados multimídia através de ontologias e padrões de metadados, com foco na recuperação semântica de informações. No âmbito governamental, Almeida, Xavier e Barreto (2018) abordam a integração de dados administrativos para pesquisa em saúde pública, enfatizando aspectos éticos e de segurança. Note-se neste estudo que os contextos se misturam, visto que os dados são administrativos, mas serão utilizados na

pesquisa da área de saúde pública. E, em aplicações na área biomédica, Schulz *et al.* (2009) analisam o papel das ontologias formais na organização e integração de dados, propondo uma tipologia de representação do conhecimento.

Na área de pesquisa científica, Alencar, Bochner e Amaral (2018) demonstram a importância da integração de múltiplas bases de dados bibliométricos, enquanto Santos e Rockembach (2018) investigam a integração de dados de pesquisa com publicações científicas, que os autores denominaram de “publicações ampliadas”, isto é, expande-se o escopo de uma publicação tradicional, conectando-a aos dados de pesquisa subjacentes e outros materiais digitais relacionados. O Projeto BrCris, descrito por Carvalho Segundo *et al.* (2024), propõe uma plataforma mais abrangente que integra dados do ecossistema nacional de pesquisa científica, agregando dados de diversas fontes, como dados curriculares, informações sobre organizações, programas de pós-graduação, publicações, patentes e revistas científicas. Jorge *et al.* (2020), por sua vez, apresentam uma solução para integração de dados de pesquisa dos cursos de pós-graduação da Universidade do Estado da Bahia (UNEB). Por fim, Dias *et al.* (2022) demonstram a aplicação prática da integração de dados a partir do cruzamento de dados da Plataforma Lattes com dados do DOAJ, fornecendo uma lista de publicações em acesso aberto, permitindo, assim, análises sobre os autores e suas publicações.

No âmbito do Patrimônio Cultural, a integração de acervos das instituições de memória, como as bibliotecas, arquivos, museus e galerias (*Galleries, Libraries, Archives, and Museums - GLAM*) tem se destacado como um tema de destaque na literatura brasileira de Ciência da Informação, com ênfase particular nas tecnologias e padrões que viabilizam a interoperabilidade entre diferentes tipos de coleções, visando a ampliação do acesso e a preservação do patrimônio cultural. A evolução das abordagens de integração é bem documentada por Marcondes (2016), que traça um panorama desde os modelos tradicionais baseados na troca de metadados até as tecnologias contemporâneas de dados abertos interligados. Esta perspectiva é expandida em trabalho posterior, onde Marcondes (2017) propõe o uso da Web Semântica e do padrão RDF para criar uma rede de dados interconectados, transformando registros de catálogos em Objetos Digitais de Patrimônio (ODP). Marcondes (2019) contribui com a temática, ainda, ao propor uma classificação abrangente de tipos de objetos de patrimônio, visando melhorar a integração de acervos digitais heterogêneos.

A integração de acervos de GLAM levanta também a questão da harmonização, ou integração, de modelos conceituais. Esta harmonização é particularmente relevante considerando que as instituições de memória tradicionalmente operam com modelos conceituais distintos. Carrasco, Vidotti e Santarém Segundo (2017) exploram o uso da ontologia CIDOC-CRM como ferramenta para integrar dados heterogêneos por meio da Web Semântica, facilitando a conexão semântica entre informações de diferentes instituições culturais. Carrasco e Vidotti (2019) apresentam o FRBR-object oriented (FRBRoo) como uma solução para harmonizar a descrição e interoperabilidade entre acervos de museus e bibliotecas. Em trabalho subsequente, Carrasco e Vidotti (2020) analisam a implementação prática do EDM na plataforma Europeia, demonstrando como o EDM permite a integração efetiva de informações de diversas instituições europeias. Uma abordagem inovadora é apresentada por Simionato e Santos (2017) com o Modelo Conceitual *Digital Images for Libraries, Archives and Museums* (DILAM), que busca integrar princípios descritivos de diferentes instituições de memória especificamente para recursos imagéticos digitais. Esta proposta demonstra como a modelagem entidade-relacionamento pode ser utilizada para criar sistemas de metadados que permitam a integração efetiva de dados. No âmbito das instituições brasileiras, Alves *et al.* (2017) abordam a aplicação prática de padronização de metadados baseado no Dublin Core, que foi utilizada para integrar diferentes tipos de acervo em seu estudo de caso na Fundação Oswaldo Cruz (Fiocruz). Também na instituição, Ribeiro, Borges e Sousa (2020) demonstram a aplicação do modelo EDM e dos princípios de *linked data* na integração de recursos de bibliotecas, arquivos e museus relacionados à descoberta da Doença de Chagas.

A partir dos estudos analisados é possível observar que as pesquisas sobre a integração de dados na Ciência da Informação brasileira iniciaram em meados dos anos 2000 e acompanham a própria evolução histórica da integração de dados, deslocando do foco inicial dos sistemas para uma perspectiva mais semântica e interconectada, principalmente nos contextos científicos, do patrimônio cultural e das entidades governamentais. Embora progressos significativos tenham sido alcançados, persiste a necessidade de aprimorar a integração em suas múltiplas dimensões. Como propostas, os estudos sugerem a aplicação de tecnologias como Web Semântica, ontologias e padrões de metadados, visando desenvolver ambientes informacionais com maior interoperabilidade e acessibilidade. Os estudos analisados corroboram os desafios já identificados

na fundamentação teórica, destacando a heterogeneidade tecnológica e semântica, a fragmentação dos dados e a complexidade dos processos de padronização como questões centrais.

5 Conclusões

Por meio do mapeamento e categorização das abordagens de integração de dados presentes na produção científica brasileira em Ciência da Informação, foi possível identificar três categorias de estudos que englobam a Integração de Sistemas e Tecnologias; a Integração Semântica e Ontológica; e a Integração de Dados e Informações Heterogêneas em Contextos Aplicados. As categorias não são mutuamente exclusivas e demonstram como as abordagens são interdependentes e se retroalimentam para enfrentar os desafios da integração de conteúdos no ambiente digital.

Entre os principais desafios da integração de dados estão a heterogeneidade tecnológica e dos dados, a fragmentação dos dados, que se encontram dispersos em diferentes sistemas e instituições, e a complexidade da padronização dos esquemas de metadados. A Ciência da Informação brasileira tem se mostrado a par desses desafios e os aborda em diferentes contextos. Em suma, a Integração de Sistemas e Tecnologias foca na interoperabilidade técnica entre sistemas de informação, como bibliotecas digitais, e a integração entre padrões de metadados, como MARC 21, Dublin Core, OAI-PMH e BIBFRAME, embora os desafios para torná-los interoperáveis ainda persistam devido a suas singularidades. A Integração Semântica e Ontológica explora o uso de ontologias, vocabulários controlados e tecnologias da Web Semântica, como *linked data* e modelos conceituais (CIDOC-CRM e EDM), para superar barreiras semânticas em domínios como bioinformática e patrimônio cultural. Já a Integração de Dados e Informações Heterogêneas em Contextos Aplicados engloba as duas categorias anteriores e aborda, também, estratégias para harmonizar dados em contextos específicos, como organizacional, de pesquisa e do patrimônio cultural, utilizando metadados, ontologias e padrões para viabilizar a recuperação e reutilização eficiente desses dados.

O mapeamento das abordagens de integração de dados contribuiu para entender o panorama das iniciativas de integração de dados na área de Ciência da Informação e como são tratados os desafios complexos da integração. Embora o termo “interoperabilidade” não tenha sido

incluído na estratégia de busca inicial, a análise revelou que essa temática emerge como um eixo central das discussões mapeadas, evidenciando sua relevância intrínseca aos processos de integração. Além disso, evidenciou possíveis caminhos de pesquisa em diferentes frentes. A necessidade de integração eficiente de dados continua sendo um aspecto central para garantir o acesso à informação no ambiente digital, e cada iniciativa representa um passo em direção a medidas que garantam a interoperabilidade entre sistemas, tecnologias e a preservação do significado dos dados em diferentes contextos.

Referências

- Abgaz, Y., *et al.* “A Methodology for Semantic Enrichment of Cultural Heritage Images Using Artificial Intelligence Technologies.” *Journal of Imaging*, vol. 7, no. 8, 2021, p. 121.
DOI: <https://doi.org/10.3390/jimaging7080121>. Acesso em 4 Jan. 2025.
- Alencar, M. S. M., R. Bochner, e D. Giacometti. “Integração de Bases de Dados em Estudos Bibliométricos: A Produção Científica Nacional em Zika Vírus.” *Em Questão*, 2018, pp. 13–28.
<https://seer.ufrgs.br/emquestao/article/view/86923>. Acesso em 20 Nov. 2024.
- Almeida, B., P. Xavier, e M. Barreto. “Dados Governamentais na Perspectiva da Ciência Aberta: Potencialidades e Desafios para Saúde Pública a Partir de um Estudo de Caso.” *Cadernos BAD*, no. 1, 2018, pp. 172–179. <https://core.ac.uk/download/pdf/276550153.pdf>. Acesso em 20 Nov. 2024.
- Almeida, M. B. “A Necessidade de Integração de Fontes Heterogêneas de Dados em Projetos de Gestão de Conhecimento.” *Perspectivas em Ciência da Informação*, vol. 7, no. 1, 2002.
<https://periodicos.ufmg.br/index.php/pci/article/view/23413>. Acesso em 3 Jan. 2025.
- Alves, A. S., *et al.* “Descrição de Metadados Mínimos para Integração e Preservação Digital dos Acervos Científicos e Culturais da Fundação Oswaldo Cruz: Um Estudo de Caso.” *Reciis – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, vol. 11, no. sup, 2017.
<https://www.arca.fiocruz.br/handle/icict/23719>. Acesso em 20 Nov. 2024.
- Alves, M. das D. R., e M. I. F. Souza. “Estudo de Correspondência de Elementos Metadados: Dublin Core e MARC 21.” *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, vol. 5, no. 1, 2007, pp. 20–38. <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2019>. Acesso em 20 Nov. 2024.
- Berners-Lee, T., J. Hendler, e O. Lassila. “The Semantic Web.” *Scientific American*, vol. 284, no. 5, 2001, pp. 34–43. <https://lassila.org/publications/2001/SciAm.html>. Acesso em 5 Jan. 2025.
- Brapci. Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação. 2025.
<https://brapci.inf.br/>. Acesso em 3 Fev. 2025.

- Brasil. Ministério da Gestão e da Inovação em Serviços Públicos. “Gestão Economiza R\$ 1,4 Bilhão com Integração das Bases de Dados do Governo Federal em 2023.” 28 Dec. 2023.
<https://www.gov.br/gestao/pt-br/assuntos/noticias/2023/dezembro/gestao-economiza-r-1-4-bilhao-com-integracao-das-bases-de-dados-do-governo-federal-em-2023>. Acesso em 5 Jan. 2025.
- Campos, M. L. A. “Integração de Ontologias: O Domínio da Bioinformática.” *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, vol. 1, no. 1, 2007.
<https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/896>. Acesso em 27 Ago. 2023.
- Campos, M. L. de A, Campos, L. M. e Barbosa, N .T. “The Challenges of Semantic Interoperability in the Era of eScience on the Web.” *KO Knowledge Organization*, vol. 47, no. 8, 2020, pp. 680-695.
DOI: <https://doi.org/10.5771/0943-7444-2020-8-680>.
- Carrasco, L. B. Integração de Conteúdos Culturais Heterogêneos em Ambientes Digitais do Patrimônio Cultural: Harmonização de Modelos Conceituais. 2019, Unesp, Tese de Doutorado.
<https://repositorio.unesp.br/handle/11449/182113>. Acesso em 10 Mar. 2022.
- Carrasco, L. B., e S. A. B. G. Vidotti. “Patrimônio Cultural: Integração de Acervos Heterogêneos de Museus e Bibliotecas.” *Anais do XX Encontro Nacional de Pesquisa em Ciência da Informação*, 2019, pp. 1–14. <https://www.brapci.inf.br/index.php/res/v/121792>. Acesso em 27 Jan. 2022.
- Carrasco, L. B., e S. A. B. G. Vidotti. “Bibliotecas, Arquivos e Museus: Um Panorama sobre a Curadoria Digital da Europeia.” *Memória e Informação*, vol. 4, no. 1, 2020, pp. 37–55.
<http://hdl.handle.net/20.500.11997/18256>. Acesso em 20 Nov. 2024.
- Carrasco, L. B., S. A. B. G. Vidotti, e J. E. Santarém Segundo. “Convergência e Ampliação de Dados Culturais em Ambientes Semânticos.” *Anais do XVIII Encontro Nacional de Pesquisa em Ciência da Informação*, 2017, pp. 1–21.
http://enancib.marilia.unesp.br/index.php/XVIII_ENANCIB/ENANCIB/paper/view/523. Acesso em 20 Nov. 2024.
- Carvalho Segundo, W. L. R., *et al.* “Inovação e Conectividade: Uma Perspectiva sobre o Projeto BrCris e suas Tecnologias para Tratamento de Dados Científicos.” *Anais do VII Workshop de Informação, Dados e Tecnologia (WIDaT)*, 2024.
<https://labcotec.ibict.br/widat/index.php/widat2024/article/view/182>. Acesso em 20 Nov. 2024.
- Castro, F. F., e P. L. V. A. C. Santos. “Representação e Descrição de Recursos Informacionais: Aspectos Estruturantes no Delineamento de Ambientes Informacionais Digitais.” *Tendências da Pesquisa Brasileira em Ciência da Informação*, 2010, pp. 155–168.
<https://www.brapci.inf.br/index.php/res/v/119342>. Acesso em 1 Fev. 2022.
- Chen, P. P. “The Entity-Relationship Model.” *ACM Transactions on Database Systems*, vol. 1, no. 1, 1976, pp. 9–36. <https://dl.acm.org/doi/10.1145/320434.320440>. Acesso em 5 Jan. 2025.
- Chen, J., e S. Ou. “Research on the Construction of the Semantic Model for Chinese Ancient Architectures Based on Architectural Narratives.” *The Electronic Library*, vol. 38, no. 4, 2020, pp. 769–784. <https://doi.org/10.1108/EL-02-2020-0039>. Acesso em 20 Ago. 2020.

- Dias, P. M., *et al.* “Extração e Validação de Dados Científicos para Identificação de Publicações em Periódicos de Acesso Aberto.” *Transinformação*, vol. 34, 2022. <https://doi.org/10.1590/2318-0889202234e210040>. Acesso em 20 Nov. 2024.
- Doan, A., A. Halevy, e Z. Ives. *Principles of Data Integration*. Elsevier, 2012. ISBN 9780124160446.
- Felicetti, A., e F. Murano. “Ce qui est écrit et ce qui est parlé. CRMtex for Modelling Textual Entities on the Semantic Web.” *Semantic Web*, vol. 12, no. 2, 2021, pp. 169–180. <https://content.iospress.com/articles/semantic-web/sw200418>. Acesso em 27 Ago. 2023.
- Grunzke, R., *et al.* “The MASI Repository Service—Comprehensive, Metadata-Driven and Multi-Community Research Data Management.” *Future Generation Computer Systems*, vol. 94, 2019, pp. 879–894. <https://doi.org/10.1016/j.future.2017.12.023>. Acesso em 27 Ago. 2023.
- Heath, T., e C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.
- Hyvönen, E. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Morgan & Claypool, 2012.
- ISO 25964-1:2011. *Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*. International Organization for Standardization, 2011.
- Jorge, E. M. F., *et al.* “Arquitetura da Informação Analítica para Integração de Dados da Pesquisa e Pós-Graduação: Um Estudo de Caso da Universidade do Estado da Bahia.” *Informação & Informação*, vol. 25, no. 1, 2020, pp. 115–140. <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/36009>. Acesso em 20 Nov. 2024.
- Kadadi, A. *et al.* “Challenges of data integration and interoperability in big data.” 2014 IEEE international conference on big data (big data). IEEE, 2014. <https://doi.org/10.1109/BigData.2014.7004486>. Acesso em 24 mar. 2025.
- Lemos, D. L. S., e R. R. Souza. “Modelos de Anotação para Representação de Documentos: Uma Proposta Ontológica para o Domínio Multimídia.” *Informação & Tecnologia*, vol. 4, no. 2, 2017, pp. 102–123. <https://www.academia.edu/download/73419968/21039.pdf>. Acesso em 3 Jan. 2025.
- Lemos, D. L. S., e R. R. Souza. “Ontologias na Representação de Documentos: Um Panorama Atual para Descrição de Conteúdo Multimídia em Rede.” *Informação & Sociedade*, vol. 29, no. 4, 2019, p. 103. <https://www.proquest.com/openview/9d160e3cf7d20135882b3071cfdb3549/1?pq-origsite=gscholar&cbl=2030753>. Acesso em 20 Nov. 2024.
- Marcondes, C. H. *Dados Abertos Interligados: Publicação, Recuperação e Integração de Acervos de Arquivos, Bibliotecas e Museus na Web*. Oficina Universitária, 2021.
- Marcondes, C. H. “Interoperabilidade entre Acervos Digitais de Arquivos, Bibliotecas e Museus: Potencialidades das Tecnologias de Dados Abertos Interligados.” *Perspectivas em Ciência da Informação*, vol. 21, 2016, pp. 61–83. <https://periodicos.ufmg.br/index.php/pci/article/view/23065>. Acesso em 4 Fev. 2022.

- Marcondes, C. H. “O Papel dos Modelos Conceituais para Interoperabilidade entre Acervos Digitais de Arquivos, Bibliotecas e Museus.” Anais do 7º Encontro Ibérico EDICIC, EDICIC, 2015, pp. 1–14, E-Prints Complutense, <https://eprints.ucm.es/id/eprint/34550/>. Acesso em 8 abr. 2022.
- Marcondes, C. H. “Publicando e Interligando Acervos Digitais na Web através das Tecnologias de Dados Abertos Interligados.” *Revista Brasileira de Biblioteconomia e Documentação*, vol. 13, 2017, pp. 2135–2163. <https://www.brapci.inf.br/index.php/res/v/3085>. Acesso em 3 Fev. 2022.
- Marcondes, C. H. “Uma Classificação de Tipos de Objetos de Patrimônio para Integração de Acervos Digitais de Arquivos, Bibliotecas e Museus.” *Scire: Representación y Organización del Conocimiento*, vol. 25, no. 2, 2019, pp. 45–52. <https://www.ibersid.eu/ojs/index.php/scire/article/view/4618>. Acesso em 18 Fev. 2022.
- Marcondes, C. H., e L. F. Sayão. “Integração e Interoperabilidade no Acesso a Recursos Informacionais Eletrônicos em C&T: A Proposta da Biblioteca Digital Brasileira.” *Ciência da Informação*, vol. 30, 2001, pp. 24–33. <https://www.scielo.br/j/ci/a/GqvpM45Xd6S8Jd8KbqVhvf/?lang=pt>. Acesso em 20 Nov. 2024.
- Mello, A. P. P., H. Mesquita, e C. E. Vieira. Introdução à Interoperabilidade. Escola Nacional de Administração Pública (Enap), 2015. <http://repositorio.enap.gov.br/handle/1/2398>. Acesso em 27 Ago. 2023.
- Melo, J. O. S. F., L. C. Botega, e J. E. Santarém Segundo. “Metodologia de Avaliação de Qualidade para Dados Conectados.” *Informação & Tecnologia*, vol. 4, no. 2, 2017, pp. 80–101. <https://periodicos.ufpb.br/index.php/itec/article/view/40539>. Acesso em 20 Nov. 2024.
- Nóbrega, H. G. M., *et al.* “Integração de Dados de ECTI Entre Diferentes Sistemas de Informação: Proposta de Solução.” Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 22, 2022, Porto Alegre, Anais, ANCIB, 2022, pp. 1–11. <https://enancib.ancib.org/index.php/enancib/xxiiencib/paper/view/1241>. Acesso em 20 Nov. 2024.
- Prodanov, C. C., e E. C. Freitas. Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico. 2nd ed., Editora Feevale, 2013.
- Ribeiro, C. J. S., M. M. Borges, e A. M. C. Sousa. “Proposta de Integração de Recursos do Patrimônio Cultural Científico no Modelo EDM: Uma Perspectiva e Representação de um Domínio Temático.” *Biblos: Revista do Instituto de Ciências Humanas e da Informação*, vol. 34, no. 2, 2020, pp. 247–264. <https://periodicos.furg.br/biblos/article/view/11864>. Acesso em 2 Jan. 2022.
- Santarém Segundo, J. E. “Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais.” *Tendências da Pesquisa Brasileira em Ciência da Informação*, vol. 8, no. 2, 2015. <https://revistas.ancib.org/index.php/tpbci/article/view/359>. Acesso em 24 mar. 2025.

- Santos, S. R. O., e C. C. Dias. “Definindo Integração de Dados a Partir do Desenvolvimento dos Catálogos.” Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 22, 2022, Porto Alegre, Anais, ANCIB, 2022. <https://repositorio.ufmg.br/handle/1843/59666>. Acesso em 20 Nov. 2024.
- Santos, S. R. O., e C. C. Dias. “Desafios para a Organização do Conhecimento no Âmbito das Humanidades Digitais e do Patrimônio Cultural: Um Olhar a Partir da Área de Informação e Ciência da Computação na Dimensions.” *Encontros Bibli*, vol. 29, 2024, pp. 1–21. <https://doi.org/10.5007/1518-2924.2024.e98058>. Acesso em 20 Nov. 2024.
- Santos, D. B., e M. Rockembach. “Publicações Ampliadas: Aspectos da Integração de Dados de Pesquisa.” *Informação & Sociedade*, vol. 28, no. 2, 2018, pp. 257–269. <https://lume.ufrgs.br/handle/10183/182012>. Acesso em 20 Nov. 2024.
- Schulz, S., *et al.* “Vantagens e Limitações das Ontologias Formais na Área Biomédica.” *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, vol. 3, no. 1, 2009. <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/818>. Acesso em 20 Nov. 2024.
- Serra, L. G., *et al.* “Os Princípios da Descrição e os Formatos MARC 21 e ONIX.” *Ciência da Informação*, vol. 46, no. 2, 2017. <https://revista.ibict.br/ciinf/article/view/2327>. Acesso em 20 Nov. 2024.
- Severino, A. J. *Metodologia do trabalho científico*. Cortez editora, 2017.
- Sheth, A. P. “Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics.” *Interoperating Geographic Information Systems*, edited by Goodchild *et al.*, Springer US, 1999, pp. 5–29. https://link.springer.com/chapter/10.1007/978-1-4615-5189-8_2. Acesso em 5 Jan. 2025.
- Silva, D. L., e R. R. Souza. “Representação de Documentos Multimídia: Dos Metadados às Anotações Semânticas.” *Tendências da Pesquisa Brasileira em Ciência da Informação*, vol. 7, no. 1, 2014. <https://ancib.org/revistas/index.php/tpbci/article/view/346>. Acesso em 20 Nov. 2024.
- Simionato, A. C., F. A. Arakaki, e P. L. V. A. C. Santos. “Descrição em Bibliotecas, Arquivos, Museus e Galerias de Arte: Linkando Recursos e Comunidades.” *Informação & Informação*, vol. 22, no. 2, 2017, pp. 449–466. <https://www.uel.br/revistas/uel/index.php/informacao/article/view/31464>. Acesso em 20 Nov. 2024.
- Simionato, A. C., e P. L. V. A. da Santos. “Modelo Conceitual DILAM: Integração Entre Arquivos, Bibliotecas e Museus.” *Informação & Sociedade*, vol. 27, no. 2, 2017. <https://biblat.unam.mx/es/revista/informacao-sociedade/articulo/modelo-conceitual-dilam-integracao-entre-arquivos-bibliotecas-e-museus>. Acesso em 20 Nov. 2024.
- Sousa, A. M. C., e C. J. S. Ribeiro. “Relações Biográficas do Patrimônio Cultural no Modelo EDM: Uma Proposta de Integração de Acervos Sobre Oswaldo Cruz.” *Revista Ibero-Americana de Ciência da Informação*, vol. 14, no. 2, 2021, pp. 367–381. <https://www.arca.fiocruz.br/handle/icict/50782>. Acesso em 21 Fev. 2022.

- Wilkinson, M. D., *et al.* “The FAIR Guiding Principles for scientific data management and stewardship.” *Scientific data* 3.1 (2016): 1-9. <http://dx.doi.org/10.25607/OBP-800>. Acesso em: 24 Mar. 2025.
- World Wide Web Consortium (W3C). Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation, 22 Feb. 1999. www.w3.org/TR/1999/REC-rdf-syntax-19990222/. Acesso em 31 Mar. 2025..
- Tzitzikas, Y., *et al.* “CIDOC-CRM and machine learning: a survey and future research.” *Heritage* vol. 5, no. 3, 2022, pp. 1612-1636. <https://doi.org/10.3390/heritage5030084>. Acesso em 24 Mar. 2025.
- Zeng, M. L. “Interoperability.” *KO Knowledge Organization*, vol. 46, no. 2, 2019. <https://www.isko.org/cyclo/interoperability>. Acesso em 5 Jan. 2025.
- Zeng, M. L., e J. Qin. *Metadata*. 2nd ed., American Library Association, 2016.

Copyright: © 2025 SANTOS, Sarah Rúbia de Oliveira; DIAS, Célia da Consolação. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

Submetido: 08/02/2025

Aceito: 31/03/2025