Tendências da Pesquisa Brasileira em Ciência da Informação

REPRESENTAÇÃO SEMÂNTICA DE CADERNOS ABERTOS DE PESQUISA: APLICAÇÃO DO PADRÃO DE METADADOS DATA DOCUMENTATION INITIATIVE (DDI)[1]

SEMANTIC REPRESENTATION OF RESEARCH NOTEBOOKS: APPLYING THE DATA DOCUMENTATION INITIATIVE (DDI) METADATA STANDAR

REPRESENTACIÓN SEMÁNTICA DE CUADERNOS ABIERTOS DE INVESTIGACIÓN: APLICACIÓN DEL ESTÁNDAR DE METADATOS DATA DOCUMENTATION INITIATIVE (DDI)

Luciana Candida da Silva[2]

Submetido em: 24/04/2026
Aprovado em: 10/07/2026
Publicado em:31/07/2026

Artigo submetido ao sistema de similaridade turnitin

[1] Texto ampliado a partir do trabalho submetido, avaliado, aprovado, apresentado e premiado no XXV Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação – ENANCIB 2025 (GT 8 – Dados, Informação e Tecnologia)

[2] Doutora em Ciência da Informação pela UNESP-Marília. Universidade Federal de Goiás. E-mail: luciana_candida@ufg.br. ORCID: https://orcid.org/0000-0003-0729-3339

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional.

Resumo: investiga a representação de cadernos abertos de pesquisa a partir dos elementos de descrição do Data Documentation Initiative (DDI) e as relações semânticas associadas a esses elementos. Trata-se de uma pesquisa qualitativa, de caráter descritivo, exploratória e documental. A literatura referente aos cadernos abertos de pesquisa e ao DDI foi levantada nas basesde dados Biblioteca Digital de Teses e Dissertações (BDTD), da Base de Dados em Ciência da Informação (BRAPCI) e da Web of Science. Esta última baseou-se principalmente nos documentos oficiais da DDI Alliance. O resultado alcançado foi que os elementos do DDI descrevem os dados dos cadernos de pesquisa de forma compatível com os princípios Findable, Accessible, Interoperable e Reusable (FAIR), promovendo sua encontrabilidade, acessibilidade, interoperabilidade e reutilização; no entanto, é necessário complementá-los com elementos de outros padrões de metadados para uma descrição mais abrangente. Neste estudo, sugere-se a utilização integrada dos vocabulários DDI, Dublin Core Terms, Schema.org e SKOS, de modo a ampliar a capacidade de descrição semântica e promover maior interoperabilidade entre sistemas de informação científica.

Palavras-Chave: cadernos de pesquisa; cadernos abertos de pesquisa; Data Documentation Initiative (DDI); metadados; dados de pesquisa científica.

Abstract: *it investigates the representation of open research notebooks based on the descriptive elements of the Data Documentation Initiative (DDI) and the semantic relationships associated with these elements. This is a qualitative study of a descriptive, exploratory, and documentary nature. The literature on open research notebooks and DDI was surveyed in the Biblioteca Digital de Teses e Dissertações (BDTD), the Base de Dados em Ciência da Informação (BRAPCI), and the Web of Science databases. The latter relied primarily on official documents from the DDI Alliance. The results showed that the DDI elements describe research notebook data in a manner compatible with the Findable, Accessible, Interoperable, and Reusable (FAIR) principles, promoting their findability, accessibility, interoperability, and reusability; however, it is necessary to complement them with elements from other metadata standards to achieve a more comprehensive description. This study suggests the integrated use of the DDI, Dublin Core Terms, Schema.org, and SKOS vocabularies in order to expand semantic description capabilities and promote greater interoperability among scientific information systems.*

Keywords: *notebook science; open notebooks science; Data Documentation Initiative (DDI); metadata; scientific research data.*

***Resumen**: investiga la representación de cuadernos abiertos de investigación a partir de los elementos de descripción del Data Documentation Initiative (DDI) y las relaciones semánticas asociadas a estos elementos. Se trata de una investigación cualitativa, de carácter descriptivo, exploratorio y documental. La literatura referente a los cuadernos abiertos de investigación y al DDI fue recopilada en las bases Biblioteca Digital de Teses e Dissertações (BDTD), Base de Dados em Ciência da Informação (BRAPCI) y Web of Science. Esta última se basó principalmente en los documentos oficiales de la DDI Alliance. Los*

resultados obtenidos muestran que los elementos del DDI describen los datos de los cuadernos de investigación de forma compatible con los principios Findable, Accessible, Interoperable y Reusable (FAIR), promoviendo su encontrabilidad, accesibilidad, interoperabilidad y reutilización; sin embargo, es necesario complementarlos con elementos de otros estándares de metadatos para lograr una descripción más amplia. En este estudio, se sugiere el uso integrado de los vocabularios DDI, Dublin Core Terms, Schema.org y SKOS, con el fin de ampliar la capacidad de descripción semántica y promover una mayor interoperabilidad entre los sistemas de información científica.

***Palabras clave**: cuadernos de investigación; cuadernos abiertos de investigación; Data Documentation Initiative (DDI); metadatos; datos de investigación científica.*

1 INTRODUÇÃO

O ecossistema da pesquisa contempla a divulgação de dados primários, preferencialmente em tempo real, à medida que são gerados, e não apenas nos casos de sucesso ou de resultados consolidados. A tendência é permitir a colaboração simultânea, a distância e, cada vez mais, de modo visível e aberto à ampla contribuição. Nesse contexto, Curty (2015, p. 5) complementa que os dados podem ser úteis para outras pessoas, dentro e fora de domínios disciplinares, pois melhoram as chances de novos resultados e conhecimentos científicos decorrentes dos mesmos dados já disponíveis.

Os cadernos de pesquisa, ou notebook science, também denominados de cadernos de laboratório eletrônicos (ELNs) são instrumentos de anotações de dados, principalmente experimentais, gerados em laboratórios, para fundamentar as publicações científicas, que são divulgadas normalmente em suas configurações finais, seja no formato de teses ou dissertações, publicações de livros, artigos de periódicos e comunicações de congressos. Para Schnell (2019), o caderno de laboratório registra as hipóteses, experimentos e análises iniciais ou interpretações

dos experimentos; e serve como um registro legal de propriedades das ideias e resultados obtidos por um cientista. Apesar da importância atribuída aos dados registrados em cadernos de pesquisa, tais dados, em sua maioria, não são divulgados juntamente às publicações finais, ficando inacessíveis aos demais pesquisadores e à sociedade.

Nesse cenário, Silva, Prado e Araújo (2024) apontaram fatores que podem influenciar no processo de não compartilhamento e reuso de dados de pesquisa científica. Entre esses fatores, encontra-se a falta de infraestrutura e de pessoal para orientar. Assim, destacam-se os metadados como a infraestrutura capaz de atribuir significado aos dados. Para Riley (2017), metadados são informações estruturadas que descrevem, explicam, localizam e facilitam a recuperação, o uso e o gerenciamento de um recurso de informação. É nesse viés que desperta o interesse por estudar padrões de metadados desenvolvidos para a representação e publicação semântica de dados de pesquisa científica, especificamente a tipologia caderno de laboratório.

De acordo com Sanchez, Silva e Vechiato (2018) e Menezes (2025), o Data Documentation Initiative (DDI) está entre os três padrões de metadados mais utilizados para a representação de dados de pesquisa nos repositórios submetidos ao Registry of Research Data Repositories (Re3data). Para os autores, o DDI se destaca na descrição de pesquisa como coleta, normalização, análise, compartilhamento e arquivamento de dados. Sendo assim, este estudo pretendeu responder à seguinte questão-problema: os

elementos de metadados disponibilizados pelo DDI descrevem adequadamente os cadernos abertos de pesquisa?

O objetivo deste estudo, portanto, é analisar as possibilidades de representar cadernos abertos de pesquisa a partir dos elementos de descrição do DDI, bem como compreender as relações semânticas inseridas nesses elementos.

Para tanto, este artigo apresenta uma organização em seções que contemplam esta introdução, os aspectos metodológicos, a descrição dos cadernos abertos de pesquisa, assim como a representação semântica e o DDI. Na sequência, realizou-se o mapeamento dos elementos de metadados do DDI e de outros vocabulários complementares à proposta do estudo. Por fim, são apresentadas as considerações finais, a conclusão e as referências.

2 CADERNOS ABERTOS DE PESQUISA

O caderno aberto de pesquisa ou *open notebook science* (ONS), versão original em inglês, é a técnica de disponibilizar, em formato aberto, os dados de pesquisa científica registrados em cadernos de pesquisa ou cadernos de laboratório para que outros pesquisadores possam acessar, usar e reutilizar sem restrição, estando sujeitos, no máximo, a atender exigências que visem preservar a proveniência e a abertura dos dados. Bradley (2010) destaca que o objetivo dos cadernos abertos de pesquisa é tornar os detalhes dos experimentos feitos em laboratório disponíveis gratuitamente na *web*, o que não se restringe apenas aos dados bem-sucedidos.

De acordo com Silva (2020), a proposta de abertura de conjuntos de dados de pesquisa registrados em cadernos de laboratório faz parte de um movimento maior da Ciência Aberta denominado *e-Science,* caracterizado pelo uso intensivo de

tecnologias e esforços colaborativos, os quais trazem a oportunidade de se pensar os novos contextos e práticas científicas. Para Clinio e Albagli (2017), a proposta trata de uma inovação no modo de produzir e comunicar a ciência, tendo sido desenvolvida por diversos cientistas, entre os quais se destaca Jean-Claude Bradley como o principal entusiasta no recrutamento de apoiadores para a definição do conceito e práticas de cadernos abertos em química.

Para Schapira e Harding (2019), a abertura e o compartilhamento de dados de pesquisa científica registrados em cadernos de pesquisa são uma maneira eficiente e rápida de disseminar dados antes de serem publicados em periódicos revisados por pares, e apresentam vantagens em relação ao tradicional.

Nessa perspectiva, o caderno de pesquisa científica produzido e publicado em formato aberto permite a disponibilidade de acesso, uso e reuso em tempo real, na íntegra, dos dados brutos produzidos em laboratórios. Assim, Clinio (2016) ressalta que o laboratório demanda o desenvolvimento de infraestruturas que preservem a proveniência dos dados e promovam a circulação do conhecimento científico, percorrendo desde o registro de dados brutos de pesquisa até a descrição no formato de experimentos e a alimentação de aplicativos e bancos de dados especializados. Em se tratando de bancos de dados, Clinio (2016) reforça que os cadernos abertos devem ser plataformas amigáveis que facilitam a sua adoção por pesquisadores e integram sistemas de informação projetados para a gestão aberta do conhecimento, lidando com questões como a interoperabilidade entre plataformas, a

automação de tarefas, o desenvolvimento de tecnologias da Web Semântica e aplicações.

Nesse sentido, Silva (2020) destaca os metadados como uma infraestrutura necessária para a descrição e atribuição de significado aos dados registrados nos cadernos de pesquisa. Os metadados são vistos como fundamentais para descrever, recuperar e acessar dados digitais na *web*, pois, segundo Taylor (2003, p. 139), "o termo metadado designa —[...] a informação estruturada que descreve atributos de recursos informacionais com o propósito de identificação, descoberta e, às vezes, administração".

Além disso, destaca-se a necessidade de um conjunto de infraestrutura composto por tecnologias e boas práticas voltadas à organização e representação dessas informações, denominadas, neste estudo, como representação semântica, conforme apresentado na seção 3.

3 REPRESENTAÇÃO SEMÂNTICA

A representação semântica é a maneira de descrever dados estruturados que possam ser legíveis por usuários humanos e máquinas, a partir do uso de tecnologias da Web Semântica e um conjunto de melhores práticas para publicação de dados, de modo que sejam facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis.

A Web Semântica é responsável por apresentar uma estrutura ao conteúdo significativo das páginas da *web* por meio de tecnologias indispensáveis, como a *eXtensible Markup Language* (XML) e o *Resource Description Framework* (RDF). A linguagem XML permite

descrever o conteúdo e sua estrutura semântica de forma legível por máquinas e humanos, mas não é utilizada para exibir dados. Para atribuir significado, adota-se o modelo RDF, que codifica as informações em conjuntos de triplas, cada uma delas representando, de forma simples, um sujeito, um verbo e um objeto. O terceiro componente são as ontologias, que definem formalmente as relações entre os termos (Berners-Lee; Hendler; Lassila, 2001).

No entanto, apenas atribuir semântica às informações da *web* não é suficiente para criar um banco de dados global; faz-se necessário estabelecer conexões entre os dados. Nesse contexto, surge o conceito de *Linked Data* como uma forma de publicar dados estruturados e conectados na *web*.

Para Berners-Lee (2006), o conceito de *Linked Data* refere-se a um conjunto de princípios voltados à criação de ligações entre recursos de diferentes fontes, promovendo a reutilização e o enriquecimento de dados na *web*. Para isso, adota-se o uso de *Uniform Resource Identifiers* (URIs) para nomear itens; URIs pelo protocolo *Hypertext Transfer Protocol* (HTTP) para que pessoas possam localizar esses nomes; e tecnologias como RDF e SPARQL *Protocol and RDF Query Language* (SPARQL), entre outras, para que, ao consultar uma URI, sejam fornecidas informações em formatos padronizados e úteis. Além disso, essas informações devem incluir sentenças com *links* para outras URIs, permitindo a descoberta de itens relacionados.

Bizer, Heath e Berners-Lee (2009, p. 2, tradução nossa) definem o *Linked Data* como "um conjunto de boas práticas para publicar e conectar dados estruturados na Web, com o intuito de criar uma Web de dados". Para esse fim, as boas práticas adotam padrões

reconhecidos internacionalmente e recomendados pelo World Wide Web Consortium (W3C).

Entre as iniciativas de maior destaque na literatura, encontram-se as Melhores Práticas para Publicação de Dados na Web, recomendadas pelo W3C, e os princípios *Findable*, *Accessible*, *Interoperable* e *Reusable* (FAIR), voltados à publicação de dados de pesquisa. Conforme estudo de Silva, Santarem Segundo e Silva (2018), essas recomendações se complementam em termos de orientações e tecnologias.

Os princípios FAIR estabelecem que os recursos, ferramentas, vocabulários e infraestrutura de dados contemporâneos devem ser apresentados de forma a favorecer sua descoberta e reutilização por terceiros (Force11, 2014). Na interpretação de Veiga (2024, p. 269-270), os princípios possuem os seguintes significados:

> ***Findable*** (localizável), destacam a necessidade de adotar identificadores persistentes e metadados ricos, incluindo um conjunto abrangente de atributos; ***accessible*** (acessível), por meio dos identificadores, dados e metadados possam ser acessados por meio de protocolos de comunicação padronizados, abertos e gratuitos, com mecanismos de autenticação e autorização quando necessário; ***interoperable*** (interoperável), uso de linguagens de representação do conhecimento, vocabulários e/ou ontologias que adotem os princípios FAIR, além de dados e metadados interligados; e ***reusable*** (reutilizável), os dados e metadados devem ser descritos por múltiplos atributos, usar licenças apropriadas, indicar a proveniência e seguir padrões específicos da sua comunidade.

Esses princípios enfatizam a importância do uso de dados, metadados e identificadores persistentes como elementos fundamentais para garantir a interoperabilidade entre os repositórios de dados de pesquisa, permitindo que dados sejam facilmente localizáveis, acessíveis e reutilizáveis em investigações futuras.

Sendo assim, destaca-se a necessidade de fornecer metadados para usuários humanos e máquinas, considerando os metadados descritivos e estruturados. Para este estudo, será analisada a adequação dos metadados do DDI para a descrição dos cadernos abertos de pesquisa.

4 DATA DOCUMENTATION INITIATIVE - DDI

O *Data Documentation Initiative* é uma iniciativa internacional, mantida pelo Consórcio DDI Alliance, para descrever e documentar dados de pesquisa nas áreas de ciências sociais, econômicas, comportamentais e da saúde. Conforme o DDI Alliance (2024), o DDI oferece um conjunto de especificações para a criação de metadados que descrevem o conteúdo, a estrutura e o contexto dos dados coletados e utilizados nessas áreas.

Segundo relatos de Block e Thomas (2003), a busca por um padrão internacional capaz de descrever e documentar dados de ciências sociais teve início em 1994, a partir da iniciativa do Inter-university Consortium for Political and Social Research (ICPSR) em reunir representantes de diversos ambientes de informação nos Estados Unidos, Canadá e Europa para substituir o padrão conhecido como OSIRIS, considerado já obsoleto, por um mais moderno e tecnologicamente capaz. Nessa perspectiva, com o apoio da National Science Foundation, o ICPSR formou o Comitê DDI para desenvolver um novo padrão. Esse comitê reconheceu que metadados processáveis por máquinas poderiam oferecer um padrão compatível com capacidade de buscas *online* em seus e outros acervos. Em última análise, ferramentas de *software* poderiam ser desenvolvidas

para localizar dados e metadados na *internet*, interpretá-los e realizar análises de dados *online*.

O Comitê DDI publicou a versão 1.0 de uma *Document Type Definition* (DTD) XML para *codebook* de ciências sociais em março de 2000. O W3C (2024) define a linguagem XML como um formato de texto simples e flexível, originalmente projetado para atender aos desafios da publicação eletrônica em larga escala, além de permitir a troca de uma ampla variedade de dados na *web*. A DTD é um arquivo de especificação que contém um conjunto de declarações que definem um tipo de documento para uma linguagem de marcação, como o XML. Assim, uma DTD estabelece as regras pelas quais um determinado tipo de documento XML deve ser criado. A soma de esforços e investimentos tecnológicos proporcionou uma estrutura compreensível por máquina e permitiu o processamento automatizado por *software* de acesso a dados, apontando perspectivas em direção à Web Semântica. De acordo com Vardigan (2013), era importante tornar-se uma aliança de adesão autossustentável, seguindo os moldes bem-sucedidos do Consórcio World Wide Web. Posteriormente, o DDI passou a ser formado por uma aliança, compondo uma nova estrutura de membros e suporte financeiro para possibilitar que a iniciativa pudesse continuar com seus desenvolvimentos tecnológicos e buscar ser expresso como Schema XML, RDF, DTD ou, possivelmente, em outros formatos (Vardigan, 2013).

Os objetivos do DDI consistem em facilitar a criação de metadados, reduzindo a complexidade da descrição de dados de pesquisa, além de estruturar metadados para tornar a documentação de dados utilizáveis para usuários humanos e máquinas. De acordo

com a estrutura com que se apresenta, o DDI possui características que possibilitam a flexibilidade e a usabilidade ao possibilitar a harmonização de dados e torná-los comparáveis e legíveis por máquinas. Esse padrão foi projetado para trabalhar com outros padrões, por exemplo, International Organization for Standardization (ISO) 19115, ISO 8601, *Data Catalog Vocabulary* (DCAT), *Provenance Ontology* (PROV), entre outros. Conforme o DDI Alliance (2024b), ele independe de plataforma tecnológica, mas requer configurações adicionais.

O DDI apresenta um conjunto de produtos para documentação e gestão de dados de pesquisa. Entre os principais produtos, encontram-se o *codebook* e o *lifecycle.* De acordo com a Alliance (2025), o DDI-Codebook (DDI-C) apresenta uma documentação estruturada e descritiva do conteúdo, significado, procedência e acesso para um único conjunto de dados. A estrutura do padrão DDI-C aplica-se ao conteúdo básico para variáveis, arquivos, material de origem e informações do nível de estudo. Atende a questões relacionadas à descoberta, preservação e uso de informações dos dados (DDI Alliance, 2025).

O *lifecycle* surge com a proposta de expandir a ideia do *codebook* em termos de cobertura de conteúdo, profundidade e gerenciamento de diferentes estágios do ciclo de vida de dados de pesquisa, incluindo a preservação de metadados ao longo do tempo, metadados reutilizáveis e suporte para planejamento, captura, processamento, armazenamento, descoberta e disseminação dos dados. Além disso, permite agrupar e comparar estudos relacionados ou séries de estudos (DDI Alliance, 2025).

Os padrões DDI-C e DDI-Lifecycle estão disponíveis como um XML Schema que, segundo o W3C (2020), fornece um meio de definir a estrutura, o conteúdo e a semântica de documentos XML. No momento, o DDI está no processo de preparação de uma nova versão do DDI-C para a publicação da versão 2.6 e também do DDI-Lifecycle para a versão 4.0. O que se observa, segundo Thomas (2024), é que a mudança da versão do *lifecycle* é apenas estrutural, e que suportará o modelo RDF, o formato *JavaScript Object Notation* (JSON) e outras implementações, além de XML. O DDI-Codebook foi atualizado em resposta a pedidos por novos elementos e atributos DDI e para facilitar a transformação de metadados do DDI-*Lifecycle* para o DDI-C e vice-versa.

Além dos padrões *codebook* (modelo descritivo original) e *lifecycle* (suporta o reuso e metadados operacionais), foram adicionados outros produtos relacionados, como um conjunto de vocabulários controlados, comumente usados em ciências sociais e outras disciplinas. Esses vocabulários podem ser "uma lista de termos claramente definidos, mutuamente exclusivos e exaustivos, que são as únicas opções para uso em um contexto específico por meio de uma classificação para algo tão complexo quanto um dicionário de sinônimos com milhares de termos e relacionamentos de termos" (DDI Alliance, 2024d). Os vocabulários controlados do DDI também são incorporados em ferramentas de edição e publicação que funcionam com a especificação do DDI. São representados em RDF:SKOS e XML:DDI *Lifecycle Code List Structure*.

O *Extended Knowledge Organization System* (XKOS), uma extensão do modelo *Simple Knowledge Organization System*

(SKOS), é utilizado para representar vocabulários de assuntos, como esquemas de classificações e tesauros de forma padronizada e legível por máquina. No contexto do DDI, o XKOS adiciona as extensões necessárias para atender aos requisitos da comunidade estatística.

Outro produto é a *Structured Data Transformation Language* (SDTL), uma "linguagem intermediária independente para representar comandos de transformação de dados" (DDI Alliance, 2025). A SDTL propõe-se a documentar as operações de transformação de dados, facilitando a replicação, interoperabilidade e transparência de processos de manipulação de dados. Esse processo permitirá que outros pesquisadores compreendam com transparência como os dados foram gerados e, assim, reproduzir a mesma operação. A SDTL apresenta representações em JSON Schema, XML Schema, RDF/Web Ontology Language (OWL) e Unified Modeling Language (UML)/XML Metadata Interchange (XMI).

Os elementos do DDI serão apresentados, na seção seguinte, no mapeamento para o recurso caderno de pesquisa.

5 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa é de natureza qualitativa, do tipo descritiva, exploratória e documental. A pesquisa documental foi baseada no estudo dos registros de memória das reuniões do DDI Alliance. Para tanto, contextualizou os cadernos abertos de pesquisa e, na sequência, descreveu a representação semântica contemplando seus conceitos e infraestrutura, além de analisar o *Data Documentation Initiative* em seus aspectos conceituais, evolutivos e semânticos. Logo, partiu para o mapeamento dos elementos de metadados do

DDI correspondentes aos elementos de descrição de um caderno de pesquisa.

A literatura consultada foi recuperada nas bases de dados da Biblioteca Digital de Teses e Dissertações (BDTD), da Base de Dados em Ciência da Informação (BRAPCI) e da Web of Science, sendo que a construção da pesquisa se baseou, principalmente, nos documentos oficiais encontrados no *site* do DDI Alliance. Os principais termos de busca foram "Data Documentation Initiative", "padrões de metadados", "dados de pesquisa" e "caderno de pesquisa", além de termos sinônimos e nos idiomas português, espanhol e inglês. Não foi estabelecido recorte temporal para os temas pesquisados.

6 RESULTADOS E DISCUSSÃO

O ecossistema de dados de pesquisa científica pode ser compreendido como um conjunto de elementos que se relacionam sobre um determinado contexto, como também de atores envolvidos no processo de produção dos dados e maneiras de publicar dados abertos. Um caderno de pesquisa normalmente é temático, isto é, realiza várias pesquisas sobre um determinado assunto. Cada pesquisa gera um conjunto de dados (*datasets*) que pode ser formado por um único elemento ou por um conjunto de elementos relacionados que se complementam com arquivos em planilhas, imagens, anotações técnicas, dentre outros. Esses elementos podem ser gravados em um ou vários arquivos de documentos.

Apresenta-se, a seguir, o resultado do mapeamento de metadados e de vocabulários que visam descrever e individualizar os objetos que compõem o ecossistema dos cadernos de pesquisa,

especialmente no que se refere a pesquisas experimentais para estruturação e publicação de cadernos abertos de pesquisa.

Os elementos de metadados do DDI para cadernos de pesquisa foram mapeados e organizados em seis classes: citação; descrição do conjunto de dados; descrição do assunto; materiais, técnicas e métodos; direitos, licenças, proveniência; e acesso e uso de recursos.

Quadro 1 - Mapeamento de propriedades de vocabulários para descrição de cadernos de pesquisa

Planilha de Metadados (Rótulos)	Propriedade dos Vocabulários Schema.org, Dublin Core (DC Terms) e SKOS		Uso
Identificador do registro	<IDNo>	DDI	**Essencial**
Data e horário do registro	<Date>	DDI	**Essencial**
Autor ▲	<authEnty>	DDI	**Essencial**
• **Identificador▲**	<IDNo>	DDI	**Essencial**
• **Ocupação**	schema:hasOccupation	Schema	**Complementar**
• **Afiliação▲**	<authEnty Affiliation>	DDI	**Essencial**
• **Departamento da Instituição**	<authEnty Affiliation>	DDI	**Complementar**
• **Identificador▲**	<IDNo>	DDI	**Essencial**
Colaborador▲	<authEnty>	DDI	**Essencial**
• **Identificador▲**	<IDNo>	DDI	**Essencial**
• **Profissão/Ocupação**	schema:hasOccupation	Schema	**Complementar**
• **Afiliação▲**	<authEnty Affiliation>	DDI	**Essencial**
• **Departamento da Instituição**	<authEnty Affiliation>	DDI	**Complementar**
• **Identificador▲**	<specificElements>	DDI	**Essencial**
Produtor▲	<Producer>	DDI	**Essencial**
• **Identificador▲**	<IDNo>	DDI	**Essencial**
• **Profissão/Ocupação**	schema:hasOccupation	Schema	**Complementar**
• **Afiliação▲**	<producer Affiliation>	DDI	**Essencial**
• **Departamento da Instituição**	<producer Affiliation>	DDI	**Complementar**
• **Identificador▲**	<IDNo>	DDI	**Essencial**
Distribuidor▲	<Distrbtr>	DDI	**Essencial**
• **Identificador do agente▲**	<IDNo>	DDI	**Essencial**
• **Afiliação▲**	<distrbtr Affiliation>	DDI	**Essencial**
• **Abreviação do distribuidor**	<Distrbtr>	DDI	**Complementar**
• **Data da**	dcterms:date	DDI	**Essencial**

Planilha de Metadados (Rótulos)	Propriedade dos Vocabulários Schema.org, Dublin Core (DC Terms) e SKOS		Uso
distribuição			
Agência de financiamento	<fundAg>	DDI	**Complementar**
Título	<Titl>	DDI	**Essencial**
Subtítulo	<subTitl>	DDI	**Complementar**
Título alternativo	<altTitl>	DDI	**Complementar**
Idioma ▲	dcterms:Language	Dublin Core	**Essencial**
Formato ▲	dct:format	Dublin Core	**Essencial**
Tipo▲	dct:type	Dublin Core	**Essencial**
Cobertura espacial▲	<Geomap>	DDI	**Essencial**
Período de execução da pesquisa	schema:startTime	Schema	**Essencial**
Público	schema:audience	Schema	**Essencial**
Resultados alcançados	schema:result	Schema	**Essencial**
Assunto▲	<Abstract>	DDI	**Essencial**
• **Identificador▲**	<IDNo>	DDI	**Complementar**
• **Ponto de acesso controlado**	skos:prefLabel	Skos	**Essencial**
• **Vocabulário**	<vocabURI>	DDI	**Essencial**
• **Assunto mais amplo**	skos:broader	Skos	**Essencial**
• **Vocabulário**	<vocabURI>	DDI	**Essencial**
• **Assunto mais específico**	skos:narrower	Skos	**Essencial**
• **Vocabulário**	<vocabURI>	DDI	**Essencial**
Descrição	<Absctract>	DDI	**Essencial**
Publicação relacionada	<OtherMaterial>	DDI	**Complementar**
• **Citação**	<relPub>	DDI	**Complementar**
• **Identificador▲**	<IDNo>	DDI	**Complementar**
Materiais	<Absctract>	DDI	**Complementar**
• **Reagentes▲**	schema:activeIngredient	Schema	**Complementar**
• **InChIKey▲**	schema:identifier	Schema	**Complementar**
• **Equipamentos**	schema:instrument	Schema	**Complementar**
• **Fórmula molecular▲**	schema:identifier	Schema	**Complementar**
• **Peso molecular▲**	schema:weight	Schema	**Complementar**
• **Técnica de medição**	schema:measurementTechique	Schema	**Complementar**
• **Nomes químicos▲**	skos:related	Skos	**Complementar**
• **Nome comercial▲**	skos:related	Skos	**Complementar**
• **Data de criação**	schema:dateCreated	Schema	**Complementar**
• **Data de modificação**	schema:dateModifield	Schema	**Complementar**
• **Período de encerramento da pesquisa**	schema:endTime	Schema	**Complementar**

Planilha de Metadados (Rótulos)	Propriedade dos Vocabulários Schema.org, Dublin Core (DC Terms) e SKOS		Uso
• ***Status* da ação**	schema:actionStatus	Schema	**Complementar**
• **Error**	schema:error	Schema	**Complementar**
Fonte de dados	schema:provider	Schema	**Complementar**
Declaração de proveniência	dct:provenance	Dublin Core	**Complementar**
Licença de uso	schema:license	Schema	**Complementar**
Declaração de direitos	dct:RightsStatement	Dublin Core	**Complementar**
Titular dos direitos	dct:rightsHolder	Dublin Core	**Complementar**
Tamanho do aplicativo	schema:fileSize	Schema	**Complementar**
***Software* necessário**	schema:availableOnDevice	Schema	**Complementar**
Registros de exibição	schema:RegisterAction	Schema	**Complementar**
Controle de uso e usuários	schema:userInteractionalCount	Schema	**Complementar**
Tipo de interação do usuário	schema:interactionType	Schema	**Complementar**
Tipo de dados	**<dataKind>**	**DDI**	**Essencial**

Fonte: Elaborado pela autora (2025).

Em uma publicação de dados de pesquisa científica de cadernos de laboratório, há o envolvimento de pessoas e organizações enquanto atores do ecossistema da pesquisa. A pessoa pode ser o pesquisador que desempenha as funções de autor, colaborador, contribuidor, dentre outras funções de uma pesquisa. A organização pode ser a instituição responsável pela pesquisa, bem como o laboratório que desenvolve a pesquisa.

Sendo assim, na primeira coluna, encontram-se os metadados adaptados da modelagem realizada por Silva (2020) e a inserção de atributos que descrevem as especificidades dos dados registrados no objeto protocolo de pesquisa, acrescidos de elementos que podem ser enriquecidos com vocabulários externos. A exemplo de Silva (2020), os elementos sinalizados com um triângulo (▲) indicam a importância do reuso de informações de *datasets* externos, por meio de URI, sempre que possível, no intuito de evitar ambiguidades, garantir padronização e carregar

informações adicionais àquelas requeridas pelos metadados. São exemplos de *tags* de uso de *datasets* externos para enriquecimento de dados dos cadernos de pesquisa:

a) Nas *tags* referentes aos autores e às instituições, podem ser reutilizadas as informações dos vocabulários Lattes, *Open Researcher and Contributor ID* (ORCID) e *Virtual International Authority File* (VIAF), entre outros, para representar as várias formas do nome dos agentes.
b) Na *tag* de identificação de idioma, recomenda-se o uso das orientações da norma ISO 639-1, que descreve a estrutura, conteúdo, construção e semântica de *tags* para idioma.
c) Na *tag* de cobertura espacial, recomenda-se o uso do GeoNames, que oferece o identificador do local e traz informações complementares do número de população, fuso horário, informações de latitude e longitude, entre outras.
d) Na *tag* de indicação do formato do arquivo, meio físico ou dimensões do objeto, a recomendação é adotar a lista de mídia da *internet*, a *Multipurpose Internet Mail Extensions* (MIME).
e) Na *tag* destinada à indicação do tipo de dados, o DC Terms orienta o uso de vocabulário do tipo Dublin Core Metadata Initiative (DCMI), disponível na página do Dublin Core.
f) Nas *tags* de assunto, é recomendado o uso do identificador do assunto principal fornecido pelos vocabulários semânticos PubChem, *Medical Subject Headings* (MeSH), LSHS, entre outros.
g) A *tag* de descrição é preenchida por meio de *string* textual, mas, para seu detalhamento, deve ser adotado um vocabulário semântico como o PubChem, que fornece todos

os nomes químicos relacionados, fórmulas moleculares, peso molecular, entre outras informações. Vale destacar, segundo Bird e Frey (2013), que as fórmulas moleculares, embora passíveis de interpretação por máquinas, não são necessariamente únicas e frequentemente são ambíguas. Sendo assim, é reforçado o uso do identificador único de química para as descrições químicas.

As *tags* referentes às duas primeiras linhas (identificador do registro e data/horário do registro) e às últimas linhas são preenchidas automaticamente, por se tratarem de metadados administrativos e metadados de uso.

A propriedade mapeada para representar os identificadores foi <*specificElements*>, do vocabulário DDI, o qual pode ser mapeado para <*dcterms:identifier*>; enquanto, para representar a data e horário do registro, mapeou a propriedade <*dcterms:date*>, do Dublin Core. Destaca-se que os metadados qualificadores do agente (identificador, ocupação, afiliação) se aplicam aos autores pessoais, colaboradores, produtores e distribuidores, podendo ser adicionado o metadado depositor <*depositr*>, instituição pela qual os autores são vinculados e agência de fomento. Na segunda coluna, encontram-se as propriedades dos vocabulários DDI, DC Terms, Schema.org e SKOS, correspondentes aos metadados da primeira coluna. As propriedades do DDI não foram capazes de contemplar todos os elementos mapeados para o caderno de pesquisa. Portanto, foram atribuídos vocabulários de outros padrões amplamente reconhecidos e utilizados para publicação de dados de pesquisa.

Para refinar os metadados, adotou-se as classes e propriedades do vocabulário SKOS, as quais estão declarando rótulos preferidos na indicação de agentes e conceitos dos conjuntos de dados descritos, sendo que a propriedade *skos:prefLabel* indica o ponto de acesso autorizado e a *skos:altLabel* indica o ponto de acesso alternativo. O vocabulário SKOS pode sinalizar relações de hierarquia entre os conceitos a partir do uso da propriedade *skos:broader* para indicar o conceito mais amplo e a *skos:narrower* para os conceitos mais específicos. A propriedade *skos:related* sinaliza os conceitos relacionados.

Na terceira coluna, são apresentados os tipos de valores desejados para os metadados. Os elementos sinalizados como essenciais são altamente recomendáveis, enquanto os complementares são opcionais. Contudo, caberá a cada instituição definir suas políticas de representação de cadernos de pesquisa.

5 CONSIDERAÇÕES FINAIS

O estudo realizado evidencia que os cadernos de pesquisa ou de laboratórios possuem características específicas voltadas para o registro de dados experimentais e observacionais, tais como materiais, reagentes, equipamentos, fórmulas moleculares, técnicas de medição, análises em andamento, entre outros elementos. Quando descritos de forma completa, esses registros favorecem a localização e a acessibilidade dos dados. O DDI oferece uma estrutura padronizada e consistente para a descrição de diversos tipos de dados de pesquisa, com ênfase nas áreas das ciências sociais e comportamentais, para as quais foi originalmente desenvolvido. O

DDI permite a descrição de elementos relevantes presentes nos cadernos de laboratório, possibilitando seu enriquecimento semântico por meio do uso de identificadores persistentes atribuídos a agentes, assuntos, publicações correlatas e cobertura espacial, entre outros.

Salienta-se que os identificadores são essenciais para a interoperabilidade entre os repositórios de dados, facilitando a integração, o reuso e a confiabilidade das informações científicas. Sua adoção contribui para a mitigação de problemas como ambiguidade de autoria, duplicidade de registros e perda de vínculos entre publicações e autores.

Apesar das vantagens oferecidas pelo DDI na representação de cadernos de laboratório, sua versão atual apresenta lacunas que demandam a complementação com outros padrões de metadados, a fim de assegurar uma descrição mais abrangente e precisa dos dados. Neste estudo, sugere-se a utilização integrada dos vocabulários DDI, *Dublin Core Terms, Schema.org* e SKOS, de modo a ampliar a capacidade de descrição semântica e promover maior interoperabilidade entre sistemas de informação científica.

Assim, sugere-se como perspectiva para pesquisas futuras a análise integrada entre o DDI e outros esquemas de metadados, visando a uma representação mais completa e interoperável dos cadernos de laboratório. Além disso, sugere-se ainda, a investigação do uso do padrão de metadados ISA (*Investigation*, *Study*, Assay), voltado à documentação do processo de geração de dados em experimentos científicos, sobretudo nas áreas das ciências da vida, como biologia, medicina, química e bioinformática. Por fim, analisar de que forma esses cadernos estão sendo registrados, por meio da

consulta a repositórios de dados de pesquisa, de modo a apresentar um panorama geral dos registros de cadernos de pesquisa.

REFERÊNCIAS

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, New York, 17 may 2001.

BERNERS-LEE, T. **Linked Data**. [*S.l.*], 2006. Disponível em: https://www.w3.org/DesignIssues/LinkedData.html. Acesso em: 13 jun. 2018.

BLOCK, W.; THOMAS, W. Implementing the Data Documentation Initiative at the Minnesota Population Center, Historical Methods. **A Journal of Quantitative and Interdisciplinary History**, [*s.l.*], v.36, n.2, p. 97-101. DOI: 10.1080/01615440309601219

BRADLEY, J. C. The impacto of open notebook Science. **Information Today**, [*s.l.*], v. 27, n.8, p. 50-51, set. 2010. Entrevista realizada por Richard Poyder. Disponível em: http://www.infotoday.com/it/sep10/Poynder.shtml#top. Acesso em: 02 set. 2019.

CLINIO, A. **Novos cadernos de laboratório e novas culturas epistêmicas**: entre a política do experimento e o experimento da política. 2016. 240 f. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

CLINIO, A.; ALBAGLI, S. Cadernos abertos de laboratório e publicações líquidas: novas tecnologias literárias para uma Ciência Aberta. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde (RECIIS)**, [*s.l.*], v. 11, sup., p. 1-17, 2017. Disponível em: https://doi.org/10.29397/reciis.v11i0.1427. Acesso em: 02 set. 2019.

CURTY, R. G. **Beyond "Data Thrifting"**: na investigation of factors influencing research data reuse in the social sciences. Orientador: Jian Qin. 2015. 266f. Tese (Doutorado em Filosofia) – School of Information Studies [Escola de Estudos da Informação],

Universidade de Syracuse (NY), 2015. Disponível em: https://surface.syr.edu/etd/266. Acesso em: 17 jan. 2025.

DATA DOCUMENTATION INITIATIVE – DDI. **Overview of current products**. [*S.l.*], 2024. Disponível em: https://ddialliance.org/products/overview-of-current-products. Acesso em: 10 jul. 2024.

FORCE11. **The Future of Research Communications and e-Scholarship. Guiding principles for findable, accessible, interoperable and reusable data publishing version B1.0**. [*S.l.*], 2014. Texto digital. Disponível em: https://www.force11.org/fairprinciples. Acesso em: 25 maio 2025.

RILEY, J. **Understanding metadata**: what is metadata, and what is it for? NISO Primer Series. Baltimore: NISO, 2017.

RODRIGUES, D. G. A. **Elementos de ciclos de vida dos dados no percurso metodológico das teses brasileiras da área da ciência da informação**: um estudo diagnóstico. 2021. 181 f. Tese (Doutorado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Paraíba, João Pessoa, 2021.

SANCHEZ, F. A.; SILVA, N. B. P.; VECHIATO, F. L. Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa. **Informação & Tecnologia**, Marília, v. 5, n. 1, 2018.

SCHAPIRA, M.; HARDING, R.J. J. Open laboratory notebooks: good for Science, good for society, good for scientists. **F1000Research Open for Science**, [*S.l.*], 2019. Disponível em: https://doi.org/10.12688/f1000research.17710.1. Acesso em: 21 set. 2025.

SCHELL, S. Tem Simple Rules for a computational biologist's laboratory notebook. **PloS Computational Biology**, São Francisco, v.11, n.9, 2015. Disponível em: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004385. Acesso em: 25 fev. 2024.

SILVA, L. C.; PRADO, C. B.; ARAÚJO, E. A. Fatores que influenciam o processo de compartilhamento e reuso de dados de pesquisa: revisão sistemática de literatura a partir da base de dados BRAPCI.

AtoZ: novas práticas em informação e conhecimento, [*s.l.*], v.13, p. 1-15, 2024.

SILVA, L. C. **Publicação de dados de pesquisa científica**: proposta de estruturação semântica de cadernos abertos de pesquisa frente às dimensões da e-Science. 2020. 243 f. Tese (Doutorado em Ciência da Informação) - Programa de Pós- Graduação em Ciência da Informação, Universidade Estadual Paulista Júlio Mesquita Filho (PPGCI/UNESP), Marília, 2020.

SILVA, L. C.; SANTAREM SEGUNDO, J. E.; SILVA, M. F. Princípios FAIR e melhores práticas do Linked Data na publicação de dados de pesquisa. **Informação & Tecnologia**, Marília, v.5, n.2, p.81-103, jul./dez. 2018. Disponível em: https://periodicos.ufpb.br/index.php/itec/article/view/44812/27746 . Acesso em: 20 maio 2025.

TAYLOR, A. G. **The organization of the information**. Westport: Libraries Unlimited, 2003. 280 p.

VEIGA, V.; HENNING, P.; LUIZ, P. P.; SANTOS, O. B. S. S. Um panorama dos princípios de dados fair: teoria, práticas e serviços. **Informação & Informação**, Londrina, v. 29, n. 4, 2024. Disponível em: https://brapci.inf.br/v/340002. Acesso em: 21 maio 2025.

WORLD WIDE WEB CONSORTIUM.W3C. **XML Schema**. 2000. Disponível em: https://www.w3.org/XML/Schema. Acesso em: 10 abr. 2025.

LICENÇA DE USO

Direitos autorais das pessoas autoras, 2025. Licenciado sob Licença Creative Commons Atribuição 4.0 Internacional (CC BY 4.0).

PUBLISHER

Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EQUIPE EDITORIAL

Martha Suzana Cabral Nunes, Maria Cleide Rodrigues Bernardino, Franciéle Carneiro Garcês da Silva.

COMO CITAR

SILVA, Luciana Candida da. Representação semântica de cadernos abertos de pesquisa: aplicação do padrão de metadados Data Documentation Initiative (Ddi). **Tendências da Pesquisa Brasileira e Ciência da Informação**, São Paulo, v. 19, p. 1-26, jan./jun. 2026.