

XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - XXV ENANCIB

GT 8 - Dados, Informação e Tecnologia

INTEROPERABILIDADE DE METADADOS EM REPOSITÓRIOS DE DADOS DE PESQUISA BRASILEIROS INDEXADOS NO RE3DATA

METADATA INTEROPERABILITY IN BRAZILIAN RESEARCH DATA REPOSITORIES INDEXED IN RE3DATA

Andréa da Silva Barboza - Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT),
Universidade Federal do Rio de Janeiro (UFRJ)/ Fundação Biblioteca Nacional (FBN)

Renata Braz Gonçalves - Universidade Federal do Rio Grande (FURG)

Naira Christofolletti Silveira - Instituto Brasileiro de Informação em Ciência e Tecnologia
(IBICT)/Fundação Biblioteca Nacional (FBN)/Universidade Federal do Estado do Rio de
Janeiro (UNIRIO)

Modalidade: Trabalho Completo

Resumo: o compartilhamento de dados de pesquisa representa um avanço significativo para a ciência, ao viabilizar o reaproveitamento de dados já coletados, otimizando recursos e potencializando novas investigações, eliminando a fase de coleta de dados. A padronização desse compartilhamento é fundamental para ampliar sua eficácia e promover maior integração entre sistemas. Esta pesquisa tem como objetivo analisar a compatibilização dos metadados descritivos utilizados em bases de dados de pesquisa brasileiras cadastradas no *Registry of Research Data Repositories*, com foco na interoperabilidade entre os padrões adotados e sua aderência aos princípios FAIR (*Findable, Accessible, Interoperable, Reusable*). A análise, realizada por meio do método *Crosswalk*, abrange 19 bases brasileiras e revela a adoção de diferentes padrões de metadados, frequentemente condicionados pelo software utilizado. O padrão mais recorrente é o *Metadata Schema*, embora o DDI (*Data Documentation Initiative*) se destaque como o mais compatível com os princípios FAIR. A pesquisa evidencia incompatibilidades entre campos e aponta para a necessidade de revisão e harmonização dos padrões de metadados empregados, visando o uso mais eficaz das informações e a melhoria da interoperabilidade entre as bases, beneficiando tanto os responsáveis pelo seu gerenciamento quanto os usuários.

Palavras-chave: metadados descritivos; dados científicos; interoperabilidade.

Abstract: the sharing of research data represents a significant advancement for science, enabling the reuse of previously collected data, optimizing resources, and enhancing new investigations by eliminating the data collection phase. Standardization of this sharing is essential to increase its effectiveness and promote greater integration between systems. This study aims to analyze the compatibility of descriptive metadata used in Brazilian research data repositories registered in the *Registry of Research Data Repositories*, with a focus on the interoperability of the adopted standards and their adherence to the FAIR principles (*Findable, Accessible, Interoperable, Reusable*). The analysis, conducted using the *Crosswalk* method, covers 19 Brazilian repositories and reveals the adoption of different metadata standards, often influenced by the software in use. The most frequently adopted standard is the *Metadata Schema*, although DDI (*Data Documentation Initiative*) stands out as the most compatible with the FAIR principles. The study highlights field incompatibilities and points to the need for reviewing and harmonizing the metadata standards employed, aiming at

more effective use of information and improved interoperability between repositories, benefiting both managers and users.

Keywords: descriptive metadata; scientific data; interoperability.

1 INTRODUÇÃO

A evolução dos paradigmas científicos demonstra a importância da coleta e do manuseio de dados. De acordo com Gray (2009) a ciência progrediu do paradigma empírico, baseado na observação, para o teórico, focado na estruturação do método, e, em seguida, para o computacional, que integrou o uso de computadores. Atualmente, a transição para o quarto paradigma é marcada pela predominância de *Big Data* e pela consolidação da *e-Science*.

O compartilhamento padronizado de dados de pesquisa possui grande relevância para o progresso científico, uma vez que permite o reaproveitamento de dados previamente coletados, otimiza recursos e impulsiona novas investigações ao eliminar a fase de coleta. O compartilhamento de dados é particularmente relevante no contexto da *e-Science*. A padronização é, portanto, fundamental para a integração eficaz entre sistemas. Nesse sentido, o reuso de dados de pesquisa vai além de quem os gerou, não apenas otimizando a coleta e aprofundando a análise de dados, mas também promovendo a interdisciplinaridade e potencializando o avanço científico de maneira mais ampla.

Em 2016 foram estabelecidos os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) através da publicação de um artigo que propõe diretrizes para o compartilhamento de dados. Eles visam não apenas orientar a gestão dos dados para uso humano, mas também garantir que sejam legíveis e processáveis por máquinas. Esses princípios são fundamentais para favorecer a reutilização da informação científica. O Brasil possui um escritório do GFISCO (*GO FAIR International Support and Coordination Office*) que tem a “responsabilidade de difundir, apoiar e coordenar as atividades relacionadas à adoção da estratégia de implementação dos princípios FAIR definida pela iniciativa GO FAIR em todo o território brasileiro” (Go FAIR Brasil, [202-?])

Essa pesquisa tem por objetivo analisar a compatibilização de metadados descritivos utilizados em bases de dados de pesquisa brasileiras disponíveis no *Registry of Research Data Repositories* (Re3Data), um repositório que reúne bases de dados de pesquisa do mundo todo. Objetiva-se realizar a análise com foco na interoperabilidade entre os padrões adotados e sua aderência aos princípios FAIR, sugerindo mudanças para o uso eficaz dos dados, bem como a

troca de dados entre diferentes sistemas, visando a otimização das bases de dados tanto para quem as alimenta quanto para seus utilizadores.

Ao filtrar as bases brasileiras, a plataforma afirma que existem 23 bases, no entanto, 4 delas não possuem sede no Brasil e são alimentadas por diversos países. Por esse motivo, consideramos 19 bases nessa pesquisa.

2 DADOS DE PESQUISA

Para que uma pesquisa produza resultados válidos, é imprescindível a realização de etapas metodológicas fundamentais, como a coleta e a análise de dados, que podem abranger distintos tipos, textuais, numéricos, observacionais, experimentais, entre outros. Uma vez coletados, esses dados são submetidos a tratamento analítico, gerando informações que fundamentam as conclusões do estudo. A e-Science introduz uma mudança significativa nesse processo ao tratar grandes volumes de dados computacionais. A prática de compartilhar dados acelera as fases iniciais da investigação, reduz custos operacionais, evita redundâncias e amplia o potencial de exploração dos dados, promovendo maior eficiência e colaboratividade na produção científica.

Segundo a *University of Toronto* “Dados de pesquisa são quaisquer informações coletadas, observadas, geradas ou criadas e usadas como evidência no processo de pesquisa para validar descobertas e resultados de pesquisa” (University of Toronto, 2023, tradução nossa). Esse conceito considera o dado de pesquisa em seu sentido mais amplo, sob a ótica do pesquisador e não do compartilhamento. No que tange ao compartilhamento dos dados de pesquisa a percepção do que é considerado dado é mais criteriosa, Sayão e Salles alertam que “para que o registro se configure como dado de pesquisa, ele precisa ser tratado e aceito pela comunidade. Isso coloca em pauta a obrigação de uma gestão mínima, com atribuição de metadados que tornem o registro compreensível para a comunidade científica” (Sayão; Sales, 2020, p.[39]).

Os tipos de dados de pesquisa sofrem variações por área do conhecimento e pelo tipo de pesquisa. A *National Science Foundation* (NSF) apresenta uma lista com tipos de dados de pesquisa para as ciências sociais comportais e econômicas, dos quais apresentamos aqui os dados utilizados na antropologia por sua considerável diversidade

Dados quantitativos
Fontes secundárias
Meios de comunicação
Texto interativo baseado na web
Dados que foram analisados usando software de análise de dados qualitativos
(National Science Foundation, 2016).

Os dados são gerados dentro do contexto da pesquisa, mas sua análise poderá ser realizada sob diferentes perspectivas em pesquisas diversas. O tratamento e análise de dados realizados em uma pesquisa geram informações que produzem um novo conhecimento. Sayão e Salles (2020) apresentam uma classificação para os dados que os subdivide nas categorias demonstradas na Figura 1.

Figura 1 - Tipologia de dados de pesquisa
Estrutura de Classificação de Dados



Fonte: elaborada pelas autoras através da plataforma Napkin com base em Sayão e Sales (2020).

A classificação apresentada por Sayão e Sales (2020) é utilizada na gestão de dados, e deve ter início no planejamento da pesquisa para um melhor resultado, auxiliando não apenas na gestão dos dados para o compartilhamento, mas também oferecendo ao pesquisador uma estruturação dos dados que facilita a análise, garantindo um melhor aproveitamento.

2.1 Gestão de dados

Para se falar em compartilhamento de dados é preciso compreender que antes de compartilhar se faz necessária a gestão dos dados, para que possam ser compreendidos e reutilizados por outros pesquisadores. Cunha e Costa destacam que

Na atualidade existe uma ênfase em abertura de dados, requisitos de planejamento e gerenciamento de dados, assim como potenciais oportunidades de pesquisa com o denominado “Big Data”. É preciso contextualizar que o processo de gestão de dados é decorrente da necessidade de compartilhamento dos mesmos (Cunha; Costa, 2020, p. 2).

Assim como a gestão deve ser pensada para o compartilhamento, é importante que se pense o compartilhamento para o reuso, todavia destacamos que essa não é uma obrigatoriedade. É possível compartilhar dados sem autorizar sua utilização ou pode ser definindo um prazo a partir do qual os dados poderão ser reutilizados. O reuso de dados de pesquisa possibilita a otimização de recursos através do melhor aproveitamento de dados que já foram coletados e que passam a não precisar de uma nova coleta; promove a interdisciplinaridade permitindo que áreas distintas utilizem um mesmo dado sob diferentes perspectivas e, conseqüentemente, promove a aceleração do progresso científico diminuindo o tempo de realização das pesquisas.

Para que tudo isso ocorra é preciso realizar uma boa gestão de dados. Idealmente a gestão deve ter início no planejamento da pesquisa, o que irá contribuir para uma melhor organização da coleta e contribuirá para que os dados sejam mais bem aproveitados na própria pesquisa. Para tanto, Sales e Sayão (2020) apresentam na Figura 2 algumas questões que precisam constar no plano de gestão de dados (PGD).

Figura 2 - Questões pertinentes ao PGD



Fonte: Sales e Sayão (2020).

Observando a Figura 2, é possível perceber a importância do PGD desde o planejamento da pesquisa. O *Digital Curation Center* (DCC) disponibiliza uma lista de verificação que auxilia o pesquisador a produzir o PGD através das seguintes questões:

- Quais dados você coletará ou criará?
- Como serão os dados coletados ou criados?
- Qual documentação e metadados acompanharão os dados?
- Como você administrará as questões de direitos autorais e direitos de propriedade intelectual?
- Como os dados serão armazenados e copiados durante a pesquisa?
- Como você gerenciará o acesso e a segurança?
- Quais dados devem ser retidos, compartilhados e/ou preservados?
- Qual é o plano de preservação de longo prazo para o conjunto de dados?
- Como você compartilhará os dados?
- São necessárias restrições ao compartilhamento de dados?
- Quem será o responsável pela gestão dos dados?
- Quais recursos você precisará para entregar seu plano? (Digital Curation Centre, 2013).

Observando esses questionamentos o pesquisador poderá estruturar melhor sua coleta e gestão de dados, aumentando a qualidade e conseqüentemente o aproveitamento dos dados e permitindo o compartilhamento e reuso.

2.2 Compartilhamento de dados

O compartilhamento de dados de pesquisa surge da necessidade de otimizar tanto os dados coletados quanto os recursos tecnológicos disponíveis, permitindo que os pesquisadores concentrem seus esforços na análise e, conseqüentemente, elevem a qualidade dos resultados obtidos. Esta prática contribui significativamente para o uso mais eficiente de recursos humanos e financeiros. É importante reconhecer, contudo, que o compartilhamento e a manutenção de dados em plataformas também envolvem custos operacionais. Por essa razão, torna-se fundamental implementar uma gestão adequada de dados seguida de processos rigorosos de curadoria.

Como vimos anteriormente, a gestão de dados começa pelo PGD que deve ser desenvolvido pelo pesquisador no início de sua pesquisa, devendo conter o

ciclo de vida de gestão para todos os dados que serão coletados, processados ou gerados por um projeto de pesquisa. De uma forma abreviada, ele se constitui em um documento formal que estabelece um compromisso de como esses dados serão tratados durante todo o desenvolvimento do projeto, e também após a sua conclusão (Sayão; Sales, 2015, p. 15).

De acordo com os autores, ele deve estar alinhado aos templates disponíveis da instituição a que o pesquisador é vinculado e deve levar em consideração a plataforma onde os dados ficarão hospedados. Para Sayão e Salles (2015) o PGD deve conter informações como: a descrição dos dados (tipo de dados, formatos, tipo de armazenamento, etc.); os

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

metadados (descrição dos dados); a política de acesso, compartilhamento e reuso; a gestão de arquivamento de longo prazo; e o orçamento envolvido na gestão dos dados.

Para que o compartilhamento de dados cumpra sua função e possa ser reutilizado o uso do PGD é essencial. A estruturação de dados em um plano prescinde de regras ou princípios que garantam sua reutilização. Pensando nessa estruturação, Eaves (2009) propõem três leis para dados governamentais abertos

- 1) se o dado não pode ser encontrado e indexado na web, ele não existe;
- 2) se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
- 3) se algum dispositivo legal não permitir sua replicação, ele não é útil (Eaves, 2009, tradução nossa).

As leis propostas por Eaves em 2009 vão ao encontro dos princípios FAIR criados em 2016. De acordo com Sales (2022) FAIR são 15 princípios organizados em quatro categorias que dão origem ao acrônimo FAIR

Na categoria F – Findable temos: F1. Os (meta) dados devem ter identificadores globais, persistentes e identificáveis; F2. Os dados devem ser descritos com (Meta) dados ricos; F3. Os (meta) dados devem ser registrados ou indexados em recursos que ofereçam capacidades de busca; F4. Metadados devem especificar o identificador dos dados.

Na categoria A – Accessible temos: A1. (Meta) dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado; A1.1 O protocolo deve ser aberto, gratuito e universalmente implementável; A1.2 O protocolo deve permitir procedimentos de autenticação e autorização, quando necessário; A2. (Meta) dados devem ser acessíveis, mesmo quando os dados não estão mais disponíveis.

Na categoria I – Interoperable temos: I1. (Meta) dados devem usar uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento; I2. (Meta) dados devem usar vocabulários que seguem os princípios FAIR; I3. (Meta) dados devem incluir referências qualificadas para outros (Meta) dados;

E por fim na categoria R – Reusable, temos R1. (Meta) dados devem ter atributos com pluralidade de precisão e serem relevantes, R1.1. (Meta) dados devem ser liberados com licenças de uso de dados claras e acessíveis, R1.2. (Meta) dados devem estar associados à sua proveniência e R1.3. (Meta) dados devem estar alinhados com padrões relevantes ao seu domínio (Sales, 2022)

Buscando garantir a implantação desses princípios foi criado o Go FAIR, uma iniciativa mundial que visa esclarecer sobre os princípios tornando os dados globalmente encontráveis e reutilizáveis. No Brasil o Go FAIR fica sediado no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e tem a professora Luana Sales como sua atual coordenadora.

Os principais autores sobre dados abertos se preocupam em salientar a importância e vantagens no compartilhamento. Sayão e Sales (2015) apontam um estudo feito no Reino Unido que elenca as seguintes vantagens

- Aumento na eficiência da pesquisa, por exemplo, evitando a duplicação de esforços através da pronta disponibilidade de ferramentas de pesquisa,

protocolos e exemplos de boas práticas, pela redução do custo de formação de coleção de dados e pela promoção e adoção de padrões abertos;

- Incentiva a investigação e o debate científico;
- Promove a inovação e potencializa novos usos para os dados;
- Possibilita novas formas de colaboração entre usuários e criadores de dados;
- Maximiza a transparência e a prestação de contas;
- Permite uma avaliação mais precisa das descobertas científicas;
- Colabora para o aperfeiçoamento e a validação dos métodos científicos;
- Evita o custo da duplicação de coleções de dados;
- Aumenta a visibilidade do impacto e a visibilidade dos resultados de pesquisa;
- Contribui para dar os devidos créditos ao criador dos dados;
- Oferece um recurso importante para a educação e a capacitação (Sayão; Sales, 2015).

Conclui-se, portanto, que o PGD desempenha um papel estratégico no aprimoramento da própria pesquisa, uma vez que o compartilhamento de dados contribui significativamente para o avanço do conhecimento científico e amplia a visibilidade e o reconhecimento do pesquisador na comunidade acadêmica. O uso do PGD proporciona a padronização dos dados que facilitará seu posterior compartilhamento em bases de dados, o que demonstra também a importância na padronização de metadados, o que é um passo crucial para facilitar o seu compartilhamento e a interoperabilidade entre diferentes bases de dados.

3 PADRÕES DE METADADOS PARA COMPARTILHAMENTO

Metadados são elementos que têm a função de dar contexto ao dado, agregando significado. Os metadados dos dados de pesquisa são fundamentais para que os dados se tornem inteligíveis e possam ser reusados, caso contrário seriam apenas um emaranhado de dados muitas vezes impossíveis de decifrar. Para Sayão e Sales (2015, p. 20) os metadados “são a documentação dos dados”. Sayão divide os metadados em três tipos: descritivos, estruturais e administrativos. De acordo com o autor

Metadados descritivos: é a face mais conhecida dos metadados, são eles que descrevem um recurso com o propósito de descoberta e identificação; podem incluir elementos tais como título, autor, resumo, palavras-chave e identificador persistente.

Metadados estruturais: são informações que documentam como os recursos complexos, compostos por vários elementos, devem ser recompostos e ordenados. Por exemplo, como as páginas de um livro, digitalizadas separadamente, são vinculadas entre si e ordenadas para formar um capítulo.

Metadados administrativos: fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação. Nessa categoria, estão metadados técnicos que explicitam as especificidades e dependências técnicas do recurso; inclui também os metadados voltados para apoio à gestão dos direitos relacionados ao recurso (Sayão, 2010, p. 5)

Além dos metadados utilizados para organizar os dados de pesquisa, existem também os metadados utilizados pelas plataformas de compartilhamento para que os dados

compartilhados sejam descritos, chamados aqui de metadados descritivos. Essas plataformas funcionam como catálogos e utilizam padrões de metadados para a descrição do acervo de dados, assim como ocorre nas bibliotecas. Essa pesquisa fará uma análise sobre os quatro padrões de metadados descritivos mais usados nas bases de dados brasileiras.

4 METODOLOGIA

Essa é uma pesquisa documental exploratória de caráter qualitativo. Primeiramente foi realizado o levantamento das bases de dados brasileiras presentes na *Re3Data*. Para tanto foi utilizado o filtro de 'países'. Foram identificadas 23 bases utilizando este filtro, das quais 19 possuem sede no Brasil. As demais são bases internacionais, alimentadas por diversos países incluindo o Brasil, estas foram desconsideradas. A coleta de dados foi concluída em maio de 2025.

Após a identificação das bases de dados foi realizado um levantamento dos padrões de metadados utilizados em cada base, objetivando identificar os quatro mais frequentes. A identificação dos padrões foi realizada dentro da própria *Re3Data*, abrindo os detalhes de cada base e selecionando a aba 'padrões'. Identificamos também os softwares utilizados para que se possa relacionar aos padrões de metadados utilizados. Levantados os padrões de metadados mais frequentes nas bases brasileiras, foi aplicado o método *CrossWalk* para observar as relações entre os padrões de metadados e relacioná-los aos princípios *FAIR*, apontando a falta de correspondências e proporcionando uma análise sobre a interoperabilidade.

4.1 *CrossWalk*

O método *CrossWalk* consiste na elaboração de uma planilha comparativa em que os metadados de um mesmo padrão são agrupados por coluna e se busca correspondência entre os metadados de outro padrão nas colunas ao lado, organizando-os por linha conforme correspondência. O objetivo da aplicação desse método é visualizar a interoperabilidade entre diferentes padrões de metadados. Através do *CrossWalk* também é possível realizar uma análise comparativa visando a escolha do padrão de metadados que melhor atende às necessidades de uma base de dados.

A aplicação desse método exige um conhecimento profundo dos padrões comparados, por este motivo se optou pela utilização de planilhas previamente elaboradas por instituições confiáveis. Para tabelar os campos *Dublin Core* e *DDI* foi utilizado o documento *Data Documentation Initiative Alliance (2023)* que apresenta o mapeamento comparativo dos dois

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

padrões. Para tabelar o *DataCite Metadata Schema* foi utilizado um documento do *DataCite Metadata Working Group (2021)* e o *Data Documentation Initiative Alliance (2014)*. Os metadados *FAIR data point* foram utilizados na análise, obtidos através do artigo de Felipe e Santos (2021, p. 13)

5 RESULTADOS

Entre as bases de dados identificadas, três não apresentaram descrição do padrão de metadados utilizado nos detalhes da *Re3Data*. O quadro 1 apresenta os resultados encontrados quanto às bases, os padrões de metadados e os softwares adotados.

Quadro 1 - Bases de dados brasileiras disponíveis no Re3data

Repositório	Instituição Mantenedora	Acesso aos Dados	Software	Padrões de Metadados
SciELO Data	Rede SciELO	Acesso aberto	Dataverse	Dublin Core, DDI, Metadata Schema
Base de Dados Científicos da UFPR	Universidade Federal do Paraná	Acesso aberto	DSpace	Dublin Core
IBICT Cariniana Dataverse Network	Instituto Brasileiro de Informação em Ciência e Tecnologia	Acesso aberto	Dataverse	DDI
Repositório de Dados de Pesquisa do CEDAP	Universidade Federal do Rio Grande do Sul	Acesso aberto	DSpace	Dublin Core
Repositório Institucional da UNESP	Universidade Estadual Paulista	Acesso aberto	Não especificado	Não especificado
Arca Dados	Fundação Oswaldo Cruz	Acesso aberto	Dataverse	Dublin Core, DDI, Metadata Schema, Open Archives
REDU - Repositório de Dados de Pesquisas Institucionais da Unicamp	Universidade Estadual de Campinas	Acesso aberto	Dataverse	Dublin Core, DDI
Maenduar	Laboratório de Humanidades Digitais	Acesso aberto	Não especificado	Dublin Core, Schema
FAPESP COVID-19 Data Sharing/BR	Fundação de Amparo à Pesquisa do Estado de São Paulo	Acesso aberto	DSpace	Dublin Core
Deposita Dados	Instituto Brasileiro de Informação em Ciência e Tecnologia	Acesso aberto	Dataverse	Dublin Core, DDI, Metadata Schema, Open Archives
Redape	Empresa Brasileira de Pesquisa Agropecuária	Acesso aberto	Dataverse	Dublin Core, Metadata Schema, Open Archives
Dados Abertos de Pesquisas	Instituto Federal Goiano - Campus Urutaí	Acesso aberto	Dataverse	Dublin Core, DDI, Metadata Schema, Open Archives
PPBio Data Repository	Instituto Nacional de Pesquisas da Amazônia	Acesso aberto	Não especificado	EML (Ecological Metadata Language)
Repositório Institucional da UFSCar	Universidade Federal de São Carlos	Acesso aberto	DSpace	Dublin Core
ATTO Data Portal	Instituto Nacional de Pesquisas da Amazônia	Acesso aberto	Não especificado	Metadata Schema
Domus Dados: Repositório de Dados de Pesquisa da Unifesp	Universidade Federal de São Paulo	Acesso aberto	Dataverse	Não especificado
Open Research Data @PUC-Rio	Pontifícia Universidade Católica do Rio de Janeiro	Acesso aberto	Não especificado	Não especificado
Aleia	Instituto Brasileiro de Informação em Ciência e Tecnologia	Acesso aberto	Dataverse	Dublin Core, DDI, Metadata Schema, Open Archives

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

Fonte: elaborado pelas autoras (2025).

Conforme podemos observar no quadro acima, o padrão *Dublin Core* é mais frequente, sendo utilizado em 13 bases de dados, seguido do padrão *DataCite Metadata Schema*, utilizado em nove bases, e, por fim o padrão DDI - *Data Documentation Initiative*, utilizado oito vezes. Percebemos que o padrão Dublin Core é o único utilizado nas bases que o software *DSpace*. O software *Dataverse* possibilita o uso de diferentes padrões.

A análise comparativa através do método *CrossWalk* foi realizada sobre os quatro padrões mais utilizados, conforme quadro 2, no qual o campo preenchido com “x” significa ausência de equivalência, ou seja, não há correspondência entre os padrões de metadados analisados.

Quadro 2 - CrossWalk dos campos utilizados em cada padrão

Dublin Core	DDI	Metadata Schema	FAIR data point	Descrição
Title	<titl> 2.1.1.1	3 Title	Title*	Título pelo qual um recurso é conhecido
Creator	<AuthEnty> 2.1.2.1	2 Creator	ContactPoint	Autores da coleta de dados
Subject	<keyword> 2.2.1.1	6 Subject	Keyword	Assuntos ou palavras-chave associadas aos dados
X	<topcClas> 2.2.1.2	X	Theme*	Descrição da classificação do tópico / Lista de termos que descrevem o assunto do conjunto de dados
Description	<abstract> 2.2.2	17 Description	Description	Descrição / resumo. Pode conter informações adicionais que não se encaixam nos outros tópicos
Publisher	<producer> 2.1.3.1	4 Publisher	Publisher*	Pessoa ou entidade com responsabilidade financeira e/ou administrativa
Contributor	<othId> 2.1.2.2	7 Contributor		Outras responsabilidades
Date	<prodDate> 2.1.3.3	8 Date	MetadataModified	Data relevantes para o trabalho / coleta de dados
Type	<dataKind> 2.2.3.10	10 Resource Type	Type*	Tipo de dados
X	X	13 Size	X	Tamanho, duração ou extensão
Format	<fileType> 3.1.5	14 Format	X	Formato do arquivo
Identifier	<IDNo> 2.1.1.5	1 Identifier	MetadataIdentifier*	Número de identificação / Identificador alternativo
		11 Alternate Identifier		
X	<holdings location="" callno="" URI=""> 2.1.8	X	Distribution*	Informações sobre Holdings / Lista de URLs de distribuição
Source	<sources> 2.3.1.8	X	X	Fontes usadas para a coleta de dados

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

Language	X	9 Language	Language	Idioma principal do recurso
Relation	<othrStdyMat> 2.5	12 RelatedIdentifier	X	Outros materiais relacionados ao recurso
		20 RelatedItem	X	
Coverage	<timePrd> 2.2.3.1	18 GeoLocation	X	Período de tempo coberto / Localização geográfica da coleta
X	<collDate> 2.2.3.2	X	X	Data da coleta de dados
X	<prodDate> 1.1.3.3	5 PublicationYear	MetadataIssued	Data de publicação dos dados
X	<nation> 2.2.3.3	X	X	Descrição do(s) país(es) abrangido(s) no arquivo em forma codificada
X	<geogCover> 2.2.3.4	X	X	Descrição da cobertura geográfica dos dados
X	<version> 1.1.6.1	15 Version	HasVersion*	Número da versão do recurso
Rights	<copyright> 2.1.3.2	16 Rights	Rights	Informações sobre direitos autorais
X	<fundAg> 1.1.3.6	19 FundingReference	X	Informações sobre financiamento
X	X	X	ConformsTo	A especificação do esquema de metadados do conjunto de dados (ShEx)
X	X	16.b RightsIdentifier	License	Licença associada ao conjunto de dados
X	<restrctn> 2.4.2.3	X	AccessRights*	Descrições dos direitos de acesso

Fonte: elaborado pelas autoras (2025).

Observa-se que entre os padrões analisados, o *Dublin Core* não contempla 12 campos de metadados em comparação aos demais padrões. Os padrões DDI e *Metadada Schema* são os que possuem maior compatibilidade entre si, todavia ainda possuem nove campos de metadados não correspondentes. Em comparação aos campos propostos no *FAIR data point*, o padrão DDI é o que possui maior correspondência, ainda assim três campos não possuem correspondência.

6 DISCUSSÃO E CONSIDERAÇÕES

Com base no Quadro 1, apresentado na seção 5, observa-se que as bases de dados de uma mesma área utilizam diferentes padrões de metadados e algumas vezes utilizam mais de um padrão. O *Dublin Core* é o padrão mais frequente, estando presente em 68,4% das bases de dados consultadas. As bases que utilizam o software *DSpace* utilizam apenas o *Dublin Core*. Em contrapartida, as bases que utilizam o software *Dataverse* costumam utilizar mais de um padrão de metadados.

Verifica-se que o padrão *Dublin Core* é o que possui menor correspondência com o *FAIR data point*, com sete campos faltantes em relação ao proposto no modelo FAIR. O padrão

DDI é o que possui maior correspondência com o *FAIR data point*, todavia, ainda faltam três campos de metadados para que o DDI tenha 100% de correspondência com o FAIR. O *Metadada Schema* não se distâcia muito, com quatro campos de metadados não correspondendo ao modelo *FAIR data point*.

Conclui-se que, embora os padrões apresentados possuam campos correspondentes, existem alguns campos que não estabelecem relações adequadas entre si, evidenciando a necessidade de uma revisão que promova a completa interoperabilidade entre eles. Como por exemplo o campo de tema/classificação que é indicado pelos princípios FAIR mas apenas o padrão DDI adota.

Acredita-se que a conscientização sobre a importância do Plano de Gestão de Dados (PGD), a compreensão dos princípios FAIR entre os pesquisadores e o conhecimento dos padrões de metadados podem facilitar significativamente esse processo. Essa facilitação ocorre porque é fundamental, em todos os padrões analisados, o aprimoramento dos campos de metadados para que estejam em conformidade com o que é proposto no *FAIR data point*.

REFERÊNCIAS

CUNHA, Murilo Bastos da; COSTA, Maria Murrieta. Fontes de informação sobre Gestão de Dados de Pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v. 30, n. 4, 2020. Disponível em: <https://periodicos.ufpb.br/index.php/ies/article/view/57183>. Acesso em: 18 maio 2023.

DATA CITE METADATA WORKING GROUP. **DataCite Metadata Schema Documentation for the Publication and Citation of research data and other research outputs**. Version 4.4, 2021. Disponível em: https://schema.datacite.org/meta/kernel-4.4/doc/DataCite-MetadataKernel_v4.4.pdf. Acesso em: 12 ago. 2023.

DATA DOCUMENTATION INITIATIVE ALLIANCE. **Linked Field-Level documentation**. 2014. Disponível em: https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation.html. Acesso em: 12 ago. 2023.

DATA DOCUMENTATION INITIATIVE ALLIANCE. **Mapping to Dublin Core (DDI Version 2)**. 2023. Disponível em: <https://ddialliance.org/mapping-to-dublin-core>. Acesso em: 22 maio 2025.

DIGITAL CURATION CENTRE (DCC). **Checklist for a Data Management Plan**. v.4.0. Edinburgh: Digital Curation Centre, 2013. Disponível em: <http://www.dcc.ac.uk/resources/data-management-plans>. Acesso em: 11 ago. 2023.

EAVES, David. **The Three Laws of Open Government Data**. [S. l.: s. n.], 2009. Disponível em: <https://eaves.ca/2009/09/30/three-law-of-open-government-data/>. Acesso em: 22 maio 2025.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

FELIPE, Carla Beatriz Marques; SANTOS, Raimunda Fernanda dos. Avaliação de metadados em repositórios de dados de pesquisa sobre biodiversidade. **Em Questão**, Porto Alegre, v. 27, n. 3, p. 1–25, 2021. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/117591/85109>. Acesso em: 10 maio 2025.

GO FAIR Brasil. **GO FAIR Brasil**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT, [202?]. Disponível em: <http://go-fair-brasil.ibict.br/>. Acesso em: 8 ago. 2025.

GRAY, Jim. Jim Gray on eScience: a transformed scientific method. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Ed.). **The Fourth Paradigm: Data-Intensive Scientific Discovery**. Redmond, WA: Microsoft Research, 2009. p. xiii–xviii.

NATIONAL SCIENCE FOUNDATION (NSF). **Public Access to NSF-Funded Research Data for the Social, Behavioral, and Economic Sciences**: Workshop Report. [S. l.: s. n.], 2016. Disponível em: <https://nsf-gov-resources.nsf.gov/files/SBE-Public-Access-Workshop-Report.pdf>. Acesso em: 14 ago. 2023.

SALES, Luana. Princípios FAIR. *In*: INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Princípios GO FAIR**. Brasília: IBICT, 2022. Disponível em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2022/marco-2022/principios-go-fair>. Acesso em: 14 ago. 2023.

SAYÃO, Luis Fernando; SALES, Luana Farias. Afinal, o que é dado de pesquisa? **Biblos**, Lima, v. 34, n. 02, p. 32-51, jul./dez. 2020.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para pesquisadores e bibliotecários**. Rio de Janeiro: CNEN, 2015. Disponível em: <https://www.aben.com.br/Arquivos/420/420.pdf>. Acesso em: 20 jul. 2023.

SAYÃO, Luís Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 15, n. 30, p.1-31, 2010. Disponível em: <https://www.redalyc.org/pdf/147/14716919002.pdf>. Acesso em: 17 jun. 2023.

UNIVERSITY OF TORONTO. **What is Research Data?** 2023. Disponível em: https://cris.utoronto.ca/dri_portal/intro/. Acesso em 25 jul. 2023.