



# Objetos Digitais FAIR: ampliando as possibilidades de uso e reuso de dados de pesquisa no contexto da Ciência dos Dados

**Guilherme Ataíde Dias**

Doutorado, Universidade de São Paulo (USP), São Paulo, SP, Brasil

Programa de Pós-Graduação em Ciência da Informação/UFPB

<http://lattes.cnpq.br/9553707435669429>

[guilhermeataide@ccsa.ufpb.br](mailto:guilhermeataide@ccsa.ufpb.br)



**Fernando de Assis Rodrigues**

Doutorado, Universidade Estadual Paulista (UNESP), Marília, SP, Brasil

Programa de Pós-Graduação em Ciência da Informação/UFPB

<http://lattes.cnpq.br/5556499513805582>

[deassis@ufpa.br](mailto:deassis@ufpa.br)

Submetido em: 23/07/2023. Aprovado em: 20/08/2024. Publicado em: dd/mm/aaaa.

## RESUMO

A Ciência da Informação tem desempenhado um papel importante na instrumentalização de diversas atividades relacionadas à pesquisa científica, oferecendo contribuições valiosas como o desenvolvimento de tesouros, sistemas de automação para revisão por pares, bibliotecas digitais e esquemas de metadados. Recentemente, a Ciência da Informação tem focado na gestão de dados de pesquisa, com pesquisadores empenhados no desenvolvimento dessa área, por meio da criação de ciclo de vida de dados, planos de gestão de dados e repositórios de dados de pesquisa. Uma inovação recente que tem impulsionado o uso e o reuso de dados de pesquisa é a adoção dos Princípios FAIR (Findable, Accessible, Interoperable, Reusable), que se tornaram um padrão para a disponibilização de dados de pesquisa. As tecnologias digitais que apoiam a aplicação dos princípios FAIR estão em constante evolução. A especificação dos Objetos Digitais FAIR é uma dessas iniciativas. Os Objetos Digitais FAIR facilitam a gestão de grandes volumes de dados e aumentam a confiabilidade e interoperabilidade dos dados. Este artigo investiga o atual estado de desenvolvimento dos Objetos Digitais FAIR e como essa iniciativa pode contribuir para o uso e reuso de dados de pesquisa no contexto da Ciência dos Dados. A pesquisa é de natureza exploratória, qualitativa e baseada em uma revisão bibliográfica. Após a análise dos documentos obtidos na revisão bibliográfica acerca dos Objetos Digitais FAIR ressalta-se que o desenvolvimento dessa tecnologia é recente e ainda está em um processo contínuo de evolução. A aplicação dos Objetos Digitais FAIR em contextos reais, por meio de estudos de caso, será fundamental para validar sua contribuição efetiva no campo da Ciência dos Dados.

**Palavras-chave:** Objetos digitais; objetos digitais FAIR; ciência dos dados; ciência da informação; tecnologias digitais de informação e comunicação.

## INTRODUÇÃO

A Ciência da Informação (CI), ao longo de sua trajetória, tem fornecido subsídios para a instrumentalização de diversas atividades relacionadas ao trabalho científico. Para esse contexto, pode-se mencionar: o desenvolvimento de tesouros que contribuem para o processo de desambiguação de termos em áreas específicas do campo científico; a aplicação de sistemas de automação do processo de revisão por pares de artigos científicos; o aprimoramento de sistemas de bibliotecas digitais; a produção de esquemas de metadados para uma variedade de sistemas de informação automatizados; entre diversas outras formas de instrumentalização. De forma contemporânea, observa-se a contribuição de pesquisadores e pesquisadoras da área da CI no desenvolvimento de modelos e na construção de artefatos que oferecem suporte teórico e prático para atividades relacionadas ao processo de gestão de dados de pesquisa.

Especificamente na gestão de dados de pesquisa, há um esforço nacional e internacional de pesquisa em CI para alavancar este tema, no qual destaca-se, mesmo que em uma seleção de maneira não exaustiva, o desenvolvimento de ciclos de vida de dados (Sant'ana, 2016), de planos de gestão de dados (Silva, 2021) e a construção de diversos repositórios específicos para o depósito de dados e de atividades de treinamento relacionadas com todos os aspectos relacionados com a gestão de dados pesquisa. No contexto brasileiro, Rodrigues (2021) menciona que as primeiras discussões acerca de dados científicos na Ciência da Informação deram-se através da publicação dos artigos científicos de Sayão e Sales (2012) e de Medeiros e Caregnato (2012), intitulados respectivamente de *Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa* e *Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica*.

Correntemente, um desenvolvimento que tem contribuído para ampliar o uso e reúso dos dados de pesquisa pela comunidade científica é a adoção dos Princípios FAIR. As letras do acrônimo FAIR significam, respectivamente, **F**indable (Encontrável), **A**ccessible (Acessível), **I**nteroperable (Interoperável) e **R**eusable (Reutilizável) (Wilkinson *et al.*, 2016; Rodriguez-Iglesias *et al.*, 2016; Dias *et al.*, 2019). Os princípios FAIR se tornaram praticamente um padrão no que se refere à padronização para a disponibilização de dados de pesquisas em repositórios de dados, servindo para facilitar a recuperação de dados tanto por agentes biológicos ou máquinas, ou seja, por seres humanos ou por algoritmos.

A criação e a evolução de tecnologias digitais que contribuem para o uso dos princípios FAIR são uma constante. Dentro dessas iniciativas, destaca-se o desenvolvimento da especificação dos Objetos Digitais FAIR (*FAIR Digital Objects*). Essa tecnologia dá provimento a uma abordagem que possibilita o tratamento de volumes de dados crescentes, considerando a complexidade envolvida no processo, assim como o desenvolvimento de ferramentas que ampliam a confiança nos dados, além de incrementar a interoperabilidade dos dados com outros recursos (FAIR Digital Objects Forum, 2023a).

Esse artigo apresenta uma pesquisa com fito a desvendar como se configura atualmente a situação de desenvolvimento dos Objetos Digitais FAIR e como essa iniciativa contribui para ampliar o uso e reuso de dados de pesquisa no contexto da Ciência dos Dados.

Com relação ao método empregado, trata-se de uma pesquisa com o objetivo exploratório, de abordagem qualitativa, de natureza básica, adotando como procedimento metodológico a pesquisa bibliográfica (Richardson, 2017). Para a realização da pesquisa bibliográfica utilizou-se o Portal de Periódicos da CAPES e o *Google Scholar*. Não foram estabelecidos critérios temporais na elaboração da estratégia de busca. A palavras-chave utilizadas foram “FAIR”, “Objetos Digitais FAIR” e “Ciência dos Dados”. Considerou-se também a versão das palavras-chave na língua inglesa.

## Objetos digitais

O termo objeto digital é consabido na área da Ciência da Computação, estando relacionado inicialmente ao domínio dos tipos abstratos de dados. Os tipos abstratos de dados evoluíram ao longo do tempo e contribuíram para o desenvolvimento das linguagens de programação orientadas a objetos. A programação nesta modalidade é correntemente bastante utilizada e é um padrão de mercado, estando presente no desenvolvimento de uma gama diversificada de produtos de software, visto a possibilidade de as linguagens tornarem os sistemas de informação mais robustos e escaláveis (Anders *et al.*, 2023).

A SIMULA, Actor, Eiffel, Smalltalk, C++ e Objective-C são exemplos das primeiras linguagens que implementaram o conceito de orientação a objetos. Atualmente, uma série de linguagens de programação tem como padrão ou tem a possibilidade de desenvolvimento de algoritmos orientados a objetos, tais como: Go, Java, Kotlin, PHP, Python, R e Ruby<sup>1</sup>.

Na literatura disponível da Ciência da Computação e da CI encontram-se algumas definições para o termo objeto digital. Uma definição para objeto digital que se considera bastante compreensiva é a que consta na patente desenvolvida para sistemas identificarem, gerenciarem e rastreamos objetos digitais de forma única e persistente, sob a numeração US6135646A (Kahn; Ely, 2000). Neste documento, compreende-se como objeto digital,

[...] qualquer conjunto de sequências de bits ou dígitos e um identificador único associado, que chamamos de “identificador” (ou “*handle*”). Um objeto digital pode incorporar informações ou material em que direitos (por exemplo, direitos autorais) ou outros interesses sejam ou possam ser reivindicados. Também podem haver direitos associados ao próprio objeto digital. Portanto, objetos digitais podem incluir representações digitais convencionais de obras (livros, artigos, imagens, sons, software) e, de forma mais abrangente, qualquer material digital que seja capaz de produzir manifestações desejadas para um usuário de computador. Portanto, um objeto digital pode incluir programas e dados que, embora não representem diretamente

---

1 A Enciclopédia Wikipedia para língua inglesa mantém uma lista em constante atualização das linguagens de programação que suportam a programação em orientação a objetos. Disponível em: [https://en.wikipedia.org/wiki/List\\_of\\_object-oriented\\_programming\\_languages](https://en.wikipedia.org/wiki/List_of_object-oriented_programming_languages). Acesso em 14 jul. 2023.

o texto de uma obra, permitem a entrega através de uma rede e a subsequente reprodução em uma tela de computador de partes selecionadas do texto da obra. (Kahn; Ely, 2000, p. 28, tradução nossa<sup>2</sup>).

Sob outra perspectiva, Kahn e Wilensky (2006, p. 117, tradução nossa<sup>3</sup>) propuseram uma definição sintética sobre objeto digital, publicada no periódico *International Journal on Digital Libraries* (Revista Internacional sobre Bibliotecas Digitais), compreendendo-os como “[...] uma instância de um tipo de dado abstrato que possui dois componentes, dados e metadados-chave”.

Outra definição para objeto digital, não associada a um artigo publicado em periódico no contexto da Ciência da Computação ou da Ciência da Informação, mas sim no contexto da Filosofia, é a de Hui (2012), em que os objetos digitais são

[...] simplesmente objetos na Web, como vídeos do YouTube, perfis do Facebook, imagens do Flickr e assim por diante, que são compostos de dados e formalizados por esquemas ou ontologias que se pode generalizar como metadados (Hui, 2012, p. 381, tradução nossa<sup>4</sup>).

As três definições trazem características distintas para definir o objeto digital, mas com algumas convergências. Khan e Ely (2000) estabelecem que o objeto digital é formado por critérios técnicos-computacionais (conjunto de sequências de *bits* atrelados a identificadores), as informações sobre direito autoral ou de propriedade e podem formar quaisquer tipos de materiais digitais. Khan e Wilensky (2006) compreendem sob uma perspectiva mais generalista estas características. Para os autores, o objeto digital é uma associação de dados e metadados (que descrevem os dados). No caso dos metadados, não ficam restritos ao direito autoral ou de propriedade, podendo descrever quaisquer características do objeto digital. Hui (2012) aborda uma mesma conceitualização de Khan e Wilensky (2006) para o objeto digital, porém adicionando a possibilidade de agregar esquemas ou ontologias aos dados do objeto digital.

Todavia, Khan e Ely (2000) identificam que um objeto digital deve ter a possibilidade de ser armazenado e recuperado, mesmo que estes elementos sejam tratados de forma indireta, pois compreendem que os objetos digitais podem ser transmitidos por meio de uma rede. Hui (2012) também aborda o tema, ao exemplificar os objetos digitais como produtos audiovisuais em aplicativos, *websites* ou perfis de Serviços de Redes Sociais *Online*, além de ficar entendido de forma implícita que os objetos digitais são passíveis de serem trafegados em uma rede, como a infraestrutura de rede da internet.

---

2 Original: “[...] any set of sequences of bits or digits and an associated unique identifier which we call a “handle”. A digital object may incorporate information or material in which rights (e.g., copyright rights) or other interests are or may be claimed. There may also be rights associated with the digital object itself. Thus digital objects may include conventional digital representations of works (books, papers, images, sounds, software), and more broadly any digital material which is capable of producing desired manifestations for a computer user. Thus, a digital object could include programs and data which, though not directly a representation of the text of a work, enable the delivery over a network and the subsequent reproduction on a computer screen of selected portions of the text of the work” (Kahn; Ely, 2000, p. 28).

3 Original: “[...] is an instance of an abstract data type that has two components, data and key-metadata” (Kahn; Wilensky, 2002, p. 117).

4 Original: “[...] simply objects on the Web, such as YouTube videos, Facebook profiles, Flickr images, and so forth, that are composed of data and formalized by schemes or ontologies that one can generalize as metadata” (Hui, 2012, p. 381).

## Objetos digitais fair no contexto da ciência dos dados

Sob a ótica de uma perspectiva histórica, Cao (2017) contextualiza que provavelmente a primeira menção ao termo Ciência dos Dados na literatura deu-se no livro intitulado *Concise Survey of Computer Methods* do autor Peter Naur no ano de 1974.

Para Cao (2017, p. 43, tradução nossa<sup>5</sup>), a Ciência de Dados como disciplina,

[...] é um novo campo interdisciplinar que sintetiza e se baseia em estatísticas, informática, computação, comunicação, gestão e sociologia para estudar dados e seus ambientes (incluindo domínios e outros aspectos contextuais, como aspectos organizacionais e sociais) a fim de transformar dados em percepções e decisões, seguindo um pensamento e metodologia de dados-para-conhecimento-para-sabedoria.

Em relação à definição da Ciência dos Dados, Moura Jr. (2020) explica que diversas abordagens podem ser empregadas para atingir esse objetivo, tais como: a abordagem computacional, a abordagem baseada em princípios, a abordagem multidisciplinar integradora<sup>6</sup> ou uma abordagem baseada em um paradigma.

Ao investigar na literatura as diversas acepções para o termo Ciência dos Dados, verifica-se, frequentemente, que o termo não é empregado explicitamente para a área da CI, aparecendo regularmente associado com a Estatística, a Matemática, a Informática, a Comunicação Social, a Administração e a Sociologia (Cao, 2017; Blei; Smyth, 2017; Moura Jr, 2020).

Em relação à evidência apresentada, entende-se que a CI possui relações com a Ciência dos Dados, evidenciadas pela sua interdisciplinaridade com a Informática (Marchionini, 2016; Reis; Carvalho, 2022). Todavia, concluir uma percepção que a conexão da CI com a Ciência dos Dados se, e somente se, por meio da Informática, é limitante. Esta relação pode ser mais robusta do que se limitar apenas à interdisciplinaridade via Informática.

De fato, a CI fornece abordagens teóricas e práticas que propiciam uma gestão adequada dos dados de pesquisa (e.g. pesquisas e a própria instrumentalização das políticas de informação em repositórios de dados), em todas as fases do ciclo de vida dos dados – desde a coleta, armazenamento e recuperação, até um eventual descarte. Esse suporte da CI permite aos pesquisadores de todas as áreas do conhecimento o acesso a conjuntos de dados alinhados às melhores práticas da área, possibilitando, assim, maximizar seu uso e reúso.

De Smedt *et al.* (2020) explicam que em um futuro próximo será inviável para os pesquisadores lidarem manualmente com dados digitais e que as iniciativas bem-sucedidas na área da Ciência dos Dados dependerão de métodos automatizados para identificar e recuperar dados de repositórios, bem como agregá-los para os fins necessários. A dificuldade de

---

5 Original: "From the disciplinary perspective, data science is a new interdisciplinary field that synthesizes and builds on statistics, informatics, computing, communication, management, and sociology to study data and its environments (including domains and other contextual aspects, such as organizational and social aspects) in order to transform data to insights and decisions by following a data-to-knowledge-to-wisdom thinking and methodology" (Cao, 2017, p. 43).

6 Ver: Cao (2017).



processar dados manualmente está relacionada ao volume crescente dos conjuntos de dados disponibilizados, resultado direto das iniciativas de incentivo à publicação de dados abertos, bem como pela sua heterogeneidade, em função da sua variedade e visto a oferta desses insumos por diversas áreas da ciência. Aqui reside um dos principais desafios da atualidade: como estabelecer mecanismos práticos para auxiliar pesquisadores a disponibilizarem dados heterogêneos, sem *de facto* limitar a criatividade ou os pré-requisitos para o ciclo de vida de dados de cada área do conhecimento, seja *small science* ou *big science* (Borgman, 2015).

A partir desse cenário apresentado, verifica-se que os pesquisadores enfrentam desafios cada vez mais sofisticados relacionados à interoperabilidade dos conjuntos de dados. Essa complexidade reflete-se no uso cada vez mais frequente de algoritmos baseados em técnicas de inteligência artificial, que podem contribuir em todas as etapas necessárias para a identificação, recuperação e processamento dos conjuntos de dados. Um exemplo deste tipo de sofisticação é a necessidade cada vez mais latente de resultados de pesquisas baseados em grandes quantidades de dados em áreas do conhecimento que, *a priori*, não os tinham, seja por quaisquer motivos (*e.g.* custos de coleta, dificuldade de aquisição por barreiras geográficas ou linguísticas).

Para que agentes automatizados possam contribuir de forma mais efetiva com os pesquisadores em todos os processos associados com a manipulação dos dados de pesquisa, é importante que os conjuntos de dados sejam publicados mediante as seguintes ações: depósito em um repositório de dados de pesquisa, descrição dos dados através de metadados, atribuição de um identificador persistente e revisão final dos conjuntos de dados (Dias *et al.*, 2020), ainda que, como anteriormente mencionado, a adoção dos princípios FAIR também se constitui de uma ação fundamental para o uso e reúso dos dados de pesquisa, facilitando significativamente a atuação dos agentes não humanos.

Mesmo levando em consideração a publicação de conjuntos de dados em repositórios adequados e o uso dos princípios FAIR, o crescimento no volume de dados disponíveis, associado à produção desses dados em domínios científicos variados, implica na necessidade de maximização da interoperabilidade. Ademais, a possibilidade do cientista de averiguar a relevância, a proveniência, a completude e a adequação dos dados aos seus respectivos casos concretos no processo de investigação científica demanda o desenvolvimento de novas soluções tecnológicas (De Smedt *et al.*, 2020).

Uma possibilidade de aumentar a interoperabilidade de conjuntos de dados, permitindo que sejam utilizados em diversos domínios do conhecimento científico (com certa similaridade a iniciativas anteriores, com o uso de uma política de *cross-domain solution*<sup>7</sup>, mas de forma ampliada), é a adoção de critérios voltados para a padronização por meio do uso do conceito de objetos digitais.

---

7 Ver Department of Defense (2015).

Existem soluções para isso, que são utilizadas por comunidades de pesquisadores envolvidos em investigações que aplicam métodos e técnicas relacionadas à Ciência de Dados. Um exemplo de solução que possibilita o encapsulamento de dados e metadados em um objeto digital é a tecnologia *Network Common Data Form* (netCDF).

Essa tecnologia se constitui em um padrão aberto que contribui para o acesso a dados científicos e começou a ser desenvolvida no ano de 1988, tendo sua primeira versão em formato binário lançada no ano de 1990 (Rew; Davis, 1990). A partir dessas considerações, podemos observar que a ideia de encapsular dados e metadados em uma única instância para facilitar seu acesso e manipulação não é uma ideia nova, remontando à segunda metade do século XX.

O manual do usuário do netCDF explica que:

O objetivo da interface *Network Common Data Form* (netCDF) é permitir que você crie, acesse e compartilhe dados orientados a arrays de forma auto-descritiva e portátil. “Auto-descritiva” significa que **um conjunto de dados inclui informações que definem os dados nele contidos**. “Portátil” significa que os dados em um **conjunto de dados são representados de uma forma que pode ser acessada por computadores com diferentes formas de armazenar inteiros, caracteres e números de ponto flutuante**. Utilizar a interface netCDF para criar novos conjuntos de dados torna os dados portáteis. Utilizar a interface netCDF em produtos de software para acesso, gestão, análise e exibição de dados pode tornar o produto de software mais útil de forma geral (Unidata, 2023, *online*, tradução nossa<sup>8</sup>, grifo nosso).

A netCDF é amplamente conhecida e utilizada pelas comunidades científicas que trabalham com geociências, meteorologia, oceanografia e outros campos correlatos. No entanto, sua utilidade não se limita a essas áreas específicas, pois pode ser aplicada em uma ampla gama de disciplinas científicas. Isso se deve à sua capacidade de descrever e de acessar dados de forma consistente e portátil, tornando-a útil para diferentes áreas de pesquisa científica.

Considerando o contexto dos objetos digitais, uma outra solução tecnológica que merece ser destacada são os Objetos Digitais FAIR (ODF), ou em Língua Inglesa, *Fair Digital Object* (FDO). Conforme explicado por Santos (2022), a primeira vez que o termo Objeto Digital FAIR surgiu em uma publicação foi em novembro do ano de 2018, no relatório intitulado de *Turning FAIR into reality*<sup>9</sup>.

A ideia que fundamenta o conceito de um ODF está relacionada à necessidade de reunir os atributos ((meta)dados FAIR) de uma determinada entidade digital em um único objeto, uma vez que esses atributos associados a uma entidade, podem se apresentar dispersos.

---

8 Original: “The purpose of the Network Common Data Form (netCDF) interface is to allow you to create, access, and share array-oriented data in a form that is self-describing and portable. “Self-describing” means that a dataset includes information defining the data it contains. “Portable” means that the data in a dataset is represented in a form that can be accessed by computers with different ways of storing integers, characters, and floating-point numbers. Using the netCDF interface for creating new datasets makes the data portable. Using the netCDF interface in software for data access, management, analysis, and display can make the software more generally useful.” (Unidata, 2023, *online*).

9 Disponível em: <https://fairdigitalobjectframework.org/>. Acesso em 17 jul. 2023.

A utilização de ODF tem o potencial de simplificar a manipulação dessas entidades nos variados domínios. O acesso a um ODF é simplificado e padronizado, podendo ser realizado por meio de um protocolo desenvolvido pela *DONA Foundation*<sup>10</sup> conhecido como *Digital Object Interface Protocol* (DOIP), o qual é independente da tecnologia adotada pelos repositórios de dados e da forma como os dados estão estruturados (FDO2022, 2022).

Santos (2022, não paginado, tradução nossa<sup>11</sup>) define um ODF como “[...] uma sequência de bits que representa uma unidade informacional e é apresentado de acordo com os princípios FAIR”. Portanto, pode-se concluir que para um objeto digital ser considerado um ODF, este deve ser apresentado, obrigatoriamente, de acordo com os princípios FAIR.

Ainda com relação ao que seria um ODF, traz-se o entendimento de De Smedt *et al.* (2020, p. 7, tradução nossa<sup>12</sup>) sobre o assunto:

Na perspectiva de um cientista de dados, um ODF é uma unidade acionável estável que reúne informações suficientes para permitir a interpretação e o processamento confiáveis dos dados contidos nela.

Uma definição mais elaborada do que vem a ser um ODF e que explicitamente menciona os princípios FAIR é apresentada por Anders *et al.* (2023, p. 9, tradução nossa<sup>13</sup>, grifo nosso):

Um Objeto Digital FAIR é uma unidade composta por dados e/ou metadados regulados por estruturas ou esquemas, e com um identificador único e persistente (PID) atribuído globalmente, que é **localizável**, **acessível**, **interoperável** e **reutilizável** tanto por humanos quanto por computadores para uma interpretação confiável e processamento dos dados representados pelo objeto.

Na **FIGURA 1**, adaptação de De Smedt *et al.* (2020) e Anders *et al.* (2023), ilustra a estrutura de quatro camadas que compõe um ODF e ainda a ordem com que diferentes tipos de (meta)dados estão associadas ao objeto.

10 Disponível em: <https://www.dona.net/>. Acesso em 19 jul. 2023.

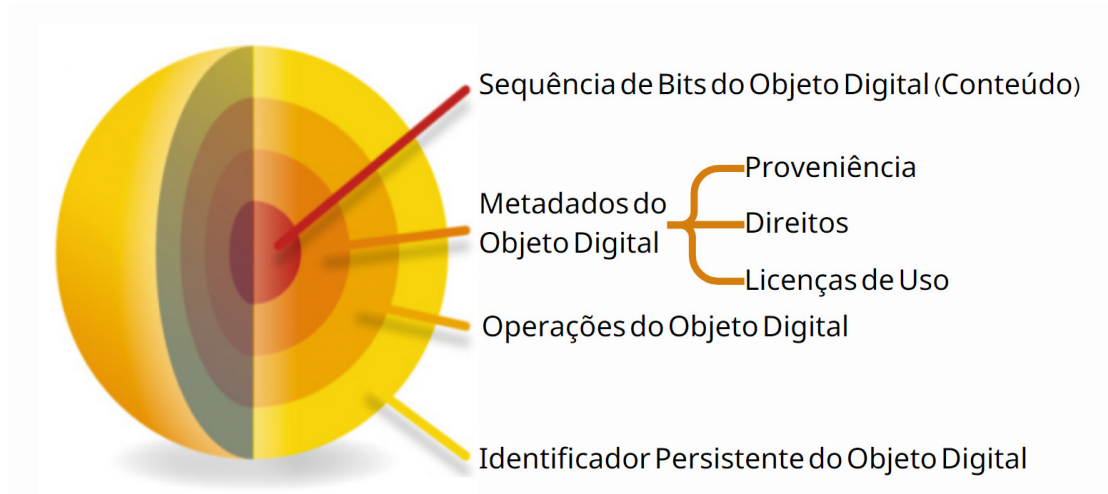
11 Original: “[...] a sequence of bits that represents an informational unit and is presented according to the FAIR principles” (Santos, 2022, n. p.).

12 Original: “From the perspective of a data scientist, an fdo is a stable actionable unit that bundles sufficient information to allow the reliable interpretation and processing of the data contained in it” (De Smedt, 2020, p. 7).

13 Original: “A FAIR Digital Object is a unit composed of data and/or metadata regulated by structures or schemas, and with an assigned globally unique and persistent identifier (PID), which is findable, accessible, interoperable and reusable both by humans and computers for the reliable interpretation and processing of the data represented by the object.” (Anders et al., 2023, p. 9, tradução nossa).



**FIGURA 1** – Estrutura de um Objeto Digital FAIR



Fonte: Adaptado pelos autores, a partir de De Smedt et al. (2020) e Anders et al. (2023).

A camada mais externa do ODF é representada pelo identificador persistente do objeto digital, possibilitando a sua identificação e localização de maneira inequívoca. A segunda camada ilustrada é composta pelas operações que podem ser realizadas sob o objeto digital, sendo seguida por uma camada de metadados e dos *bits* que compõem em si o conteúdo do objeto digital.

O **QUADRO 1** apresenta uma lista de documentos que versam sobre as especificações ODF, podendo ser utilizada como uma fonte de referência para uma compreensão detalhada do *status quo* de desenvolvimento dessa tecnologia.

**QUADRO 1 – Especificações FDO: lista de documentos**

<b>Título do documento</b>	<b>Descrição</b>	<b>Digital Object Identifier (DOI)</b>
PR Requirement Specification	Este documento de especificação de requisitos dos Objetos Digitais FAIR especifica os critérios que as entidades digitais precisam cumprir para serem aceitas como Objetos Digitais FAIR e devem ser seguidos por todas as implementações que afirmam estar em conformidade com os Objetos Digitais FAIR. Os requisitos são escritos em um nível que permite diferentes implementações.	<a href="https://doi.org/10.5281/zenodo.7781925">https://doi.org/10.5281/zenodo.7781925</a>
Implementation of Attributes, Types, Profiles and Registries	Este documento detalha o rascunho de trabalho sobre tipagem do ponto de vista de implementação. Destaca os papéis dos Objetos Digitais FAIR, seus Persistent Identifiers (PIDs), atributos, serviços e registros que precisam ser estabelecidos e operam em objetos. Portanto, é o documento central para construir infraestruturas Objetos Digitais FAIR.	<a href="https://doi.org/10.5281/zenodo.7825572">https://doi.org/10.5281/zenodo.7825572</a>
FDO Overview	Este documento é a primeira visão geral completa das especificações do Objeto Digital FAIR. Aborda vários aspectos discutidos no Fórum Objetos Digitais FAIR nos seus dois primeiros anos, mas contém apenas o essencial.	<a href="https://doi.org/10.5281/zenodo.7824713">https://doi.org/10.5281/zenodo.7824713</a>
DOIP Endorsement	A versão 2.0 do DOIP, lançada pela Fundação DONA, é endossada como uma contribuição importante para a pilha de especificações do Fórum Objetos Digitais FAIR. O protocolo de interface DOIP especifica como os clientes podem interagir com os Objetos Digitais, independentemente da tecnologia usada por um servidor DOIP e da organização de dados escolhida.	<a href="https://doi.org/10.5281/zenodo.7824795">https://doi.org/10.5281/zenodo.7824795</a>
Upload of FDO	Este documento ilustra as operações e verificações necessárias ao enviar Objetos Digitais FAIR para o domínio dos Objetos Digitais FAIR. Também descreve os tipos de serviços possíveis e o escopo das verificações que podem ser realizadas em diferentes níveis de Objetos Digitais FAIR.	<a href="https://doi.org/10.5281/zenodo.7825548">https://doi.org/10.5281/zenodo.7825548</a>
Typing FDOs	Os Objetos Digitais FAIR promovem a capacidade de ação por máquina ao exigir que cada objeto digital seja caracterizado por sinais que indiquem claramente as possíveis etapas de processamento que qualquer máquina acessando possa realizar. Chamamos esses sinais de tipos e este documento descreve a necessidade de tipos e os princípios que a comunidade Objetos Digitais FAIR segue no desenvolvimento de uma estrutura para criar e aplicar esses tipos.	<a href="https://doi.org/10.5281/zenodo.7825598">https://doi.org/10.5281/zenodo.7825598</a>
FDO PID Profile & Attributes	O Fórum Objetos Digitais FAIR precisa de uma definição abrangente do que é um Objeto Digital FAIR e quais são seus componentes relevantes. Os Perfis Persistent Identifier (PID) e o conjunto de atributos kernel que podem ser usados nos Perfis PID são pilares chave. O objetivo deste documento é especificar esses dois aspectos para incorporá-los no documento de especificação abrangente do Objeto Digital FAIR a ser escrito.	<a href="https://doi.org/10.5281/zenodo.7825629">https://doi.org/10.5281/zenodo.7825629</a>
FDO Machine Actionability	Para o trabalho com Objetos Digitais FAIR, precisamos de uma definição clara do que se entende por “capacidade de ação por máquina”. Este documento inclui uma descrição de termos relevantes e apresenta conclusões na forma de uma definição para uso no Fórum Objetos Digitais FAIR.	<a href="https://doi.org/10.5281/zenodo.7825649">https://doi.org/10.5281/zenodo.7825649</a>

Título do documento	Descrição	Digital Object Identifier (DOI)
FDO Granularity	Desde o início do desenvolvimento do conceito Objetos Digitais FAIR, houve uma discussão sobre quais atributos são necessários para o correto funcionamento dos Objetos Digitais FAIR em diferentes contextos e qual seria a diferença e as sobreposições com os metadados do objeto, como atualmente usado na gestão de dados de pesquisa. Este documento baseia-se em discussões anteriores com relação a atributos armazenados com PIDs do tipo Handle para facilitar a gestão de dados e discussões para derivar um esquema de metadados centrais para trabalho interdisciplinar.	<a href="https://doi.org/10.5281/zenodo.7825685">https://doi.org/10.5281/zenodo.7825685</a>
FDO Kernel Attributes and Metadata	Desde o início do desenvolvimento do conceito FDO, houve uma discussão sobre quais atributos são necessários para o correto funcionamento dos Objetos Digitais FAIR em diferentes contextos e qual seria a diferença e as sobreposições com os metadados do objeto, como atualmente usado na gestão de dados de pesquisa. Este documento baseia-se em discussões anteriores com relação a atributos armazenados com PIDs do tipo Handle para facilitar a gestão de dados e discussões para derivar um esquema de metadados centrais para trabalho interdisciplinar.	<a href="https://doi.org/10.5281/zenodo.7825692">https://doi.org/10.5281/zenodo.7825692</a>
FDO Configuration Types	Discussões intensivas no Fórum Objetos Digitais FAIR revelaram que existem muitas maneiras diferentes de organizar Objetos Digitais FAIR e muitas delas listadas neste documento são usadas na prática diária.	<a href="https://doi.org/10.5281/zenodo.7825702">https://doi.org/10.5281/zenodo.7825702</a>
FAIR Digital Objects Roadmap	Este documento não é uma especificação, mas um roteiro vivo fornecido pelos copresidentes do BIG & TSIG, baseado nas discussões sobre a especificação Objetos Digitais FAIR até agora e focando em aspectos tecnológicos.	<a href="https://doi.org/10.5281/zenodo.7824672">https://doi.org/10.5281/zenodo.7824672</a>

Fonte: Adaptado pelos autores, a partir de FAIR Digital Objects Forum (2023b).

Após a análise dos documentos acerca dos ODF obtidos na revisão bibliográfica, ressalta-se que o desenvolvimento dessa tecnologia é recente e ainda está em um processo contínuo de evolução. A aplicação dos ODF em contextos reais, através de estudos de caso, será fundamental para validar sua contribuição efetiva na Ciência dos Dados.

## CONCLUSÕES

Os desafios impostos aos pesquisadores pelo aumento contínuo na produção de dados de pesquisa no ambiente digital estimulam o aprimoramento e o desenvolvimento de novas tecnologias. Essas tecnologias possibilitam o tratamento desses dados em todas as instâncias dos ciclos de vida de dados. Nesse diapasão, pode-se fazer uma conexão das reflexões preliminares provocadas neste texto com o pensamento de Mayer-Schönberger e Cukier (2013), ao sugerirem que o aparato tecnológico responsável pela produção massiva de dados é a mesma estrutura tecnológica que fornece as ferramentas que possibilitam a sua análise.

Assim, aponta-se que o desenvolvimento tecnológico associado com a produção de dados serviu como catalisador para a proposição dos Princípios FAIR que trouxeram um

contributo para maximizar o uso e o reuso de dados de pesquisa. Os ODF são uma evolução necessária que prometem expandir ainda mais os benefícios trazidos pelo princípio FAIR, destacando-se em uma ambiência cada vez mais dependente da recuperação de conjuntos de dados por processos automatizados.

Indica-se que o binômio formado pelo avanço tecnológico e pela produção de dados atuou como catalisador para a proposição dos Princípios FAIR, que trouxeram contribuições significativas para maximizar o uso e o reuso de dados de pesquisa. Os ODF representam uma evolução necessária, que promete ampliar ainda mais os benefícios proporcionados pelo princípio FAIR, destacando-se em um ambiente cada vez mais dependente da recuperação de conjuntos de dados através de processos automatizados.

Os princípios FAIR foram bem recebidos e adotados por diversas comunidades científicas. Contudo, não há uma compreensão generalizada sobre a sua importância ou sobre como aplicá-los de forma efetiva. Esse fato tem implicações na popularização e adoção dos ODF, que ainda estão nos estágios iniciais do seu ciclo de vida tecnológico. Ambas as tecnologias exigem ações educativas sobre seus respectivos usos nas comunidades que lidam com a Ciência de Dados. Nesse aspecto, a Ciência da Informação desempenha um papel muito importante para o sucesso dessas iniciativas. Esse papel pode se concretizar através de treinamentos e suporte técnico provido pelos(as) Bibliotecários(as) de Dados que atuam nas bibliotecas universitárias (de pesquisa), além do papel fundamental em disseminar este tipo de avanço tecnológico por meio das mais variadas profissões que atuam no campo da informação.

## REFERÊNCIAS

IVONNE, A.; CHRISTOPHE, B.; DAAN, B.; MAGGIE, H.; SHARIF, I.; THOMAS, J.; LARRY, L.; KARSTEN, P. V. G.; ROBERT, Q.; ALEXANDER, S.; ULRICH, S.; STIAN, S. R.; GEORGE, S.; DIETER, V. U.; CLAUS, W.; PETER, W.; CARLO, Z. FAIR Digital Object Technical Overview Version PEN 2.0. **Zenodo**, United Kingdom. Apr. 2023. Disponível em: <https://doi.org/10.5281/zenodo.7824714>. Acesso em: 7 jul. 2023.

BLEI, D. M.; SMYTH, P. Science and data science. *In*: Proceedings of the National Academy of Sciences, 33., 2017, United States. **Proceedings** [...]. United States: PNAS, 2017. p. 8689-8692. v. 114, n. 33. Disponível em: <https://www.pnas.org/doi/10.1073/pnas.1702076114>. Acesso em: 16 jul. 2023.

BORGMAN, C. L. **Big data, little data, no data**: scholarship in the networked world. Cambridge: MIT Press, 2015.

CAO, L. Data Science: A Comprehensive Overview. **ACM Computing Surveys**, [s. l.], v. 50, n. 3, 2017, p. 1-42. Disponível em: <https://doi-org.ez15.periodicos.capes.gov.br/10.1145/3076253>. Acesso em: 11 jul. 2023.

DE SMEDT, K.; KOUREAS, D.; WITTENBURG, P. FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. **Publications**, [s. l.], v. 8, n. 2, Apr. 2020. Disponível em: <https://doi.org/10.3390/publications8020021>. Acesso em: 5 jul. 2023.

DIAS, G. A.; ANJOS, R. L.; RODRIGUES, A. A. Os princípios fair: viabilizando o reuso dos dados científicos. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados Científicos**: perspectivas e desafios. João Pessoa: UFPB, 2019. p. 177-187. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/359>. Acesso em: 12 jul. 2023.

DIAS, G. A.; SIEBRA, S. A.; SOUSA, R. P. M.; SOUSA, M. R. S. Publicando dados de pesquisa: contextualizando as principais etapas e elementos envolvidos no processo. **Ciência da Informação**, [s. l.], v. 49, n. 3, 2020. DOI: 10.18225/ci.inf.v49i3.5501. Disponível em: <https://revista.ibict.br/ciinf/article/view/5501>. Acesso em: 16 jul. 2023.

FAIR DIGITAL OBJECTS FORUM. **Homepage**. [S. l.], 2023a. Disponível em: <https://fairdo.org/>. Acesso em: 12 jul. 2023.

FAIR DIGITAL OBJECTS FORUM. **FDO Specifications**. [S. l.], 2023b. Disponível em: <https://fairdo.org/specifications/>. Acesso em: 21 jul. 2023.

FDO2022. Homepage. **1st International Conference on FAIR Digital Objects**: turning the internet into a meaningful data space. [S. l.], 2022. Disponível em: <https://www.fdo2022.org/>. Acesso em: 17 jul. 2022.

HUI, Y. What Is a Digital Object? **Metaphilosophy**, [s. l.], v. 43, n. 4, July 2012, p. 380–395.

KAHN, R.; ELY, D. K. **System for uniquely and persistently identifying, managing, and tracking digital objects**. Depositante: Corporation for National Research Initiatives, Reston, Va. US6135646A, Depósito: 24 Out. 2000.

KAHN, R.; WILENSKY, R. A framework for distributed digital object services. **International Journal on Digital Libraries**, [s. l.], v. 6, n. 2, p. 115-123, Apr. 2006.

MARCHIONINI, G. Information science roles in the emerging field of data science. **Journal of Data and Information Science**, [s. l.], v. 1, n. 2, Sept. 2016, p. 1-6. Disponível em: <https://doi.org/10.20309/jdis.201609>. Acesso em: 16 jul. 2023.

MAYER-SCHÖNBERGER, V.; CUKIER, K; **Big Data: a revolution that will transform how we live, work, and think**. Boston: Houghton Mifflin Harcourt, 2013. 242 p.

MEDEIROS, J. S.; CAREGNATO, S. E. Compartilhamento de dados e e-science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, set. 2012. p. 311-322. Disponível em: <https://revista.ibict.br/liinc/article/view/3368/2969>. Acesso em: 12 jul. 2023.

MOURA JR, P. J. Ciência dos Dados e implicações teóricas e práticas em administração. **Teoria e Prática em Administração**, [s. l.], v. 10, n. 2, 2020. Disponível em: <https://periodicos.ufpb.br/index.php/tpa/article/view/54385/>. Acesso em: 16 jul. 2023.

REIS, M. J.; CARVALHO, T. Ciência de Dados e Ciência da Informação. **Informação em Pauta**, Fortaleza, v. 7, n. 00, p. 1-21, 28 dez. 2022. Disponível em: <http://www.periodicos.ufc.br/informacaoempauta/article/view/78490>. Acesso em: 16 jul. 2023.

REW, R.; DAVIS, G. NetCDF: an interface for scientific data access. **IEEE Computer Graphics and Applications**, [s. l.], v. 10, n. 4, p. 76-82, July 1990. DOI: 10.1109/38.56302. Disponível em: <https://ieeexplore.ieee.org/abstract/document/56302>. Acesso em: 19 jul. 2023.

RICHARDSON, R. J. **Pesquisa social: métodos e técnicas**. 4. ed. rev. atual. ampl. São Paulo: Atlas, 2017, 424 p.

RODRÍGUEZ-IGLESIAS, A.; RODRÍGUEZ-GONZÁLEZ, A.; IRVINE, A. G.; SESMA, A.; URBAN, M.; HAMMOND-KOSACK, K. E.; WILKINSON, M. D. Publishing FAIR Data: an exemplar methodology utilizing PHI-Base. **Frontiers in plant science**, [s. l.], v. 7, n. 641. 2016. Disponível em: <http://doi.org/10.3389/fpls.2016.00641>. Acesso em: 12 jul. 2023.



RODRIGUES, D. G. A. **Elementos de ciclos de vida dos dados no percurso metodológico das teses brasileiras da área de Ciência da Informação**: um estudo diagnóstico. 2021. 183 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2021.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116–142, maio/ago. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 12 jul. 2023.

SANTOS, L. O. B. S. **FAIR Digital Object Framework Documentation**. [S. l.], 2022. Disponível em: <https://fairdigitalobjectframework.org/>. Acesso em: 17 jul. 2023.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. **Informação & Sociedade: estudos**, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012. Disponível em: <https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/curadoria-digital—sayao.pdf>. Acesso em: 12 jul. 2023.

SILVA, F. C. C. **Gestão de Dados Científicos**. Rio de Janeiro: INTERCIÊNCIA, 2021. 130 p.

UNIDATA. **The NetCDF User's Guide v.1.1**. [S. l.], 2023. Disponível em: <https://docs.unidata.ucar.edu/nug/current/>. Acesso em: 18 jul. 2023.

UNITED STATES. **Cross Domain (CD) Policy**. Washington: Department of Defense, [2015]. Disponível em: <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/854001p.pdf>. Acesso em: 23 jul. 2023.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, A. B.; BLOMBERG, N.; BOITEN, J.-W.; SANTOS, L. B. S.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; HOEN, P. A. C.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; SCHAİK, R. V.; SANSONE, S. A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; LEI, J. V D.; MULLIGEN, E. V.; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, [s. l.] v. 3, n. 160018, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 12 jul. 2023.