



# OBJETO DIGITAL FAIR: ampliando as fronteiras da gestão de dados

## *Fair digital object: expanding the boundaries of data stewardship*

Luís Fernando Sayão<sup>1</sup>

 [0000-0002-6970-0553](https://orcid.org/0000-0002-6970-0553)

Luana Farias Sales<sup>2</sup>

 [0000-0002-3614-2356](https://orcid.org/0000-0002-3614-2356)

Washington Luís R. de Carvalho Segundo<sup>3</sup>

 [0000-0003-3635-9384](https://orcid.org/0000-0003-3635-9384)

### RESUMO

O exercício da pesquisa científica, especialmente na cauda longa que é povoada por pequenos laboratórios que trabalham numa infinidade de projetos, gera e consome uma vasta quantidade de objetos digitais que são heterogêneos em várias dimensões: em termos de volume, formatos, modelos, tecnologias, complexidade estrutural, configurando um genuíno big dado científico. Para que esses objetos digitais sejam apropriadamente gerenciados de acordo com suas peculiaridades, a automação se torna determinante, e a necessidade de identificação dos tipos de diferentes objetos digitais, seu contexto disciplinar, bem como a qualificação da relação entre eles, tornam-se de grande relevância. O problema que se coloca neste contexto é que as atuais práticas de gestão de dados têm se mostrado ineficientes relegando grande parte dos ativos digitais a uma condição de invisibilidade e sem condições de acesso e reuso por *stakolders* humanos e computacionais. Em busca de uma solução para esse dilema é proposto uma unidade abstrata de conhecimento, referenciada por identificadores persistentes e descrita por metadados capazes de interagir com sistemas automatizados, denominado Objeto Digital FAIR (FDO). Neste sentido, o presente artigo tem como metodologia uma pesquisa qualitativa, de cunho exploratório, apresentando um estudo conceitual, epistemológico, histórico e técnico, cujo objetivo é analisar o conceito de FDO, examinando as suas possíveis contribuições para as convergências tecnológicas e informacionais necessárias para se superar a atual ineficiência da gestão de dados. Como resultado é apresentada uma nova concepção de gestão de objetos digitais, que considera os princípios FAIR já na sua arquitetura e que através da acionabilidade por máquina pode nos levar a uma nova geração da Internet: Uma Internet FAIR de Dados e Serviços. Conclui-se que o conceito de FDO muda

---

Artigo submetido em 16/11/2023 e aceito para publicação em 29/12/2023.

<sup>1</sup>Comissão Nacional de Energia Nuclear - CNEN, Programa de Pós-graduação em Ciência da Informação – PPGCI IBICT-UFRJ – [luis.sayao@cnen.gov.br](mailto:luis.sayao@cnen.gov.br)

<sup>2</sup>Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT, Programa de Pós-graduação em Ciência da Informação – PPGCI IBICT-UFRJ – [luanasaes@ibict.br](mailto:luanasaes@ibict.br)

<sup>3</sup>Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT - [washingtonsegundo@ibict.br](mailto:washingtonsegundo@ibict.br)

significativamente a visão sobre o que um dataset deve ser, tornando necessária a decodificação e a disseminação de conhecimentos que habilitem esses dados a se tornarem fontes integradas de serviços, no escopo de um ecossistema informacional automatizado que possa atender às necessidades dos mais diversos setores.

**Palavras-Chave:** objeto digital FAIR; gestão de dados de pesquisa; framework de objetos digitais FAIR.

## **ABSTRACT**

*The exercise of scientific research, especially in the long tail which is populated by small laboratories working on a multitude of projects, generates and consumes a vast amount of digital objects that are heterogeneous in several dimensions: in terms of volume, formats, models, technologies, structural complexity, configuring genuine scientific big data. The automation becomes decisive for those digital objects to be appropriately managed according to their peculiarities. The need to identify the types of different digital objects, their disciplinary context, as well as the qualification of the relationship between them, also becomes of great importance. The problem that arises in this context is that current data management practices have proven to be inefficient, relegating a large part of digital assets to a condition of invisibility and without conditions for access and reuse by human and computer stakeholders. Looking for a solution to this dilemma is proposed an abstract unit of knowledge, referenced by persistent identifiers, and described by metadata capable of interacting with automated systems, named FAIR Digital Object (FDO). In this sense, the methodology adopted by the present study has qualitative approach and an exploratory nature, presenting a conceptual, epistemological, historical, and technical study, whose objective is to analyze the concept of FAIR Digital Object (FDO), examining its possible contributions to the technological and informational convergences necessary to overcome the current inefficiency of data management. As a result, a new conception of digital object management is presented, which considers FAIR principles already in its architecture and which, through machine actionability, can lead us to a new generation of the Internet: A FAIR Internet of Data and Services. It is concluded that the concept of FDO significantly changes the vision of what a dataset should be, making it necessary to decode and disseminate knowledge that enables these data to become integrated sources of services, within the scope of an automated information ecosystem that can meet the needs of the most diverse sectors.*

**Keywords:** FAIR digital object; research data management; FAIR digital object framework

## **1 INTRODUÇÃO**

No universo digital há a necessidade constante de interação com diferentes tipos de entidade ou objetos – códigos, *software*, *datasets*, metadados etc. Isto se torna ainda mais evidente no exercício da pesquisa científica, onde os objetos digitais gerados e consumidos são heterogêneos em várias dimensões, como em termos de volume, formatos, modelos, tecnologias, complexidade estrutural, configurando um genuíno **big dado científico**. Por conta dessa diversidade, cada tipo de objeto digital, para ser apropriadamente gerenciado, exige diferentes formas de interação. Nesse ambiente

crescentemente complexo, a automação se torna determinante, e a necessidade de identificação dos tipos de diferentes objetos digitais, seu contexto disciplinar, bem como a qualificação da relação entre eles, tornam-se de grande relevância. “Isto, exige requisitos adicionais que não estão apropriadamente cobertos pela Internet e pela *World Wide Web*”, problematizam Santos *et al.* (2023, p.3), posto que a Internet foi projetada e implementada como uma plataforma de múltiplo propósito cujo objetivo é subministrar conectividade global entre computadores, dispositivos e redes de todo tipo. Nessa concepção genérica, nenhuma aplicação específica foi destacada para tratamento especial.

Todavia, por volta da virada do milênio, consolida-se a percepção de que a ciência de dados se torna imprescindível para o endereçamento de muitos dos desafios da pesquisa contemporânea que impactam a sociedade como um todo, como por exemplo, a construção de modelos climáticos que equacione as questões de aquecimento global, o monitoramento de espécies ameaçadas ou a detecção de notícias falsas. Porém, de imediato, torna-se evidente também que a rápida aceleração na produção de dados de pesquisa não foi acompanhada por um desenvolvimento equivalente nas estratégias de acesso e análise do acúmulo de dados produzidos e consumidos pelos processos científicos.

Já em 2005, Jim Gray e colaboradores alertavam para esse gargalo quando testemunhavam que os instrumentos científicos e simulação computacional estavam criando um vasto acúmulo de dados que “requerem novos métodos científicos para analisar e organizar os dados” (Gray *et al.*, 2005, p.1). Esta situação tem como consequência direta uma descontinuidade no fluxo de aproveitamento dos dados para o reúso e geração de novos conhecimentos, e demanda que os algoritmos, ferramentas e métodos de análise sejam capazes de lidar com o imenso volume de dados e, simultaneamente, encontrar os efeitos sutis não detectados em estudos prévios. pois em muitos casos os avanços vêm de novas metodologias de análise aplicadas sobre os dados existentes, destacam Gray e Szalay (2004, p.3).

Nas análises sobre o fenômeno do big data científico é preciso considerar também a ampliação das fronteiras do que se entende por objeto de pesquisa, que acontece principalmente pela apropriação dos pressupostos mais colaborativos da ciência aberta, pelos fluxos de trabalho dos laboratórios, cujos

requisitos exigem também atenção às condições de reprodutibilidade/replicabilidade dos experimentos científicos. Para atender essas condições de contorno, além da gestão de publicações e dados, é preciso uma curadoria mais abrangente que inclua informações sobre códigos, *software*, *workflow*, metodologias, modelos, algoritmos, especificações de instrumentos científicos, *hardware* etc., que tornam mais formalizadas e transparentes as cadeias de proveniência e rastreabilidade e a confiança e autenticidade dos processos, produtos e dos contextos de todo o ciclo da pesquisa.

As dificuldades de gestão afetam ainda os fluxos de trabalho dos pesquisadores de áreas intensivas em dados. Estudos relevantes em projetos de *data analytic*, incluindo o levantamento do *Research Data Alliance* (RDA) de 2019, revelam que os pesquisadores que trabalham com dados em seus empreendimentos científicos consomem até 80% do seu tempo tornando os dados aptos para que sejam submetidos aos processos de análise (Schwardmann, 2020); pelo lado do reuso, o cientista, seja na academia ou na indústria, gasta de 70 a 80% do tempo em procedimentos tediosos e manuais para localizar, acessar e formatar os dados para reuso. Esses cenários podem nos indicar que as práticas correntes de gestão e curadoria de dados são profundamente ineficientes (Schultes; Wittenburg, 2018) e dispendiosas em termos de tempo, recursos materiais e infraestruturais, e de esforço intelectual.

Em face desses desafios, Schwardmann (2020) sugere que somente um alto grau de automação baseado em estruturas simples pode oferecer uma alternativa a essa gestão ineficiente e propensa a erros. Porém, ele nos alerta que o maior obstáculo para a automação é a heterogeneidade e a complexidade dos dados, e que a “abstração é um caminho geral para ocultar essas características dos dados de pesquisa por meio do encapsulamento e virtualização” (Schwardmann, 2020, p.1), que são métodos conhecidos da ciência da computação para contornar situações de alta complexidade, que podem ser aplicados, de forma geral, à natureza idiossincrática dos objetos digitais de pesquisa.

Para endereçar esses desafios, que constituem um obstáculo para a aceleração da pesquisa, várias organizações internacionais como a RDA, COODATA, GO FAIR Foundation e o *FDO Steering Committee* vêm trabalhando no desenvolvimento de uma unidade abstrata de conhecimento, acionável por

máquina, estável no tempo e no espaço, que registra informações suficientes que permitem a interpretação e processamento confiáveis dos dados que ela contém. Esse constructo abstrato é chamado de objeto digital FAIR ou *FAIR Digital Object* – FDO – como na sigla em inglês. Subjacente ao conceito de FDO está a ideia de virtualização e de encapsulamento de dados e de diferentes tipos de metadados por essa unidade de conhecimento que, dessa forma, eleva significativamente o nível de confiança com o qual ele pode ser analisado transversalmente por disciplinas e por modos de compartilhamento. O domínio do FDO é um ambiente virtual denominado *FAIR Digital Object Framework* (FDOF), que define um modelo para representar objetos em ambientes digitais e oferece suporte para uma Internet de Dados e Serviços FAIR (IFDS).

Assim sendo, o objetivo principal do presente estudo é analisar o conceito de Objeto Digital FAIR e os demais conceitos e abstrações relacionados de um ponto de vista epistemológico, histórico e técnico, e examinar as suas possíveis contribuições para as convergências tecnológicas e informacionais necessárias para se superar a profunda ineficiência da gestão de dados; elencamos como objetivos específicos: conhecer os componentes básicos da arquitetura de Objeto Digital FAIR na qualidade de uma unidade de conhecimento autocontida, tipificada e acionável por máquina; e mostrar o seu papel na implementação prática e escalável dos princípios FAIR na concepção de sistemas de informação para a pesquisa e para outras aplicações.

Os conceitos de FDO e FDOF estão ainda em construção, porém a sua concepção vem sendo acompanhada por inúmeros testes e práticas em diversas instâncias, que tramitam entre a academia e os setores industriais. Diante dessas circunstâncias, a presente pesquisa tem um percurso metodológico eminentemente exploratório: conhecer, aprofundar, decodificar e divulgar em linguagem clara esses novos movimentos globais em torno do FAIR. Para tal, optou-se por analisar os escritos, bibliográficos e documentais, como artigos, relatórios, palestras, projetos etc., e os produtos técnicos, como *test beds*, provas de conceitos, modelos, programas, especificações etc. dos pesquisadores que delineiam esse novo constructo, e as bases históricas e epistemológicas internalizadas na sua construção, que são alicerces teóricos e técnicos colocados pelos grupos do RDA afeitos a essa questão, e pelo *FDO Steering Committee*.

## 2 ALGUNS DESAFIOS NA GESTÃO DE OBJETOS DE PESQUISA

As atuais práticas de gestão de objetos digitais de pesquisa têm se mostrado ineficientes, relegando grande parte dos ativos digitais a uma condição de invisibilidade e sem condições de acesso e reuso por *stakolders* humanos e computacionais. São muitos os desafios pautados na agenda crítica da ciência de dados, especialmente no domínio da pesquisa científica, por exemplo: lidar com o crescente aumento do volume, da complexidade e da heterogeneidade dos objetos digitais de pesquisa; desenvolver metodologias para a interpretação interdisciplinar que proporcionem novas descobertas; promover a interoperabilidade de dados, metadados e sistemas; desenvolver ambientes que tenham níveis adequados de confiança; tornar os dados unidades de conhecimento acionáveis por computador disponíveis para provedores de serviço em ambientes automatizados; manter por tempo indeterminado as assertivas científicas com o seu contexto e proveniência.

Além do mais, num futuro em que já estamos envolvidos na sua gênese, não será mais viável, e sequer possível, para os pesquisadores encontrar, extrair, avaliar, integrar, processar e analisar dados digitais sem o apoio de métodos automatizados, tendo em vista a escala, velocidade e complexidade nas quais eles se apresentam. Isso significa que os dados e as informações sobre os dados precisam estar prontos para serem acionados, interpretados e consumidos por *stakeholders* artificiais que atuam em nome dos pesquisadores no provimento de serviços. “Vivemos cada vez mais na era da descoberta de conhecimento a partir de dados e orientados por agentes”, qualificam Batista *et al.* (2022, p.1), a dinâmica da pesquisa contemporânea. Nesse contexto, os ecossistemas informacionais precisam estar suficientemente interconectados e interoperáveis para permitir integração e interpretação significativas e escaláveis de evidências em diferentes domínios da ciência por meio de interfaces sociotécnicas interdisciplinares (De Smedt; Koureas; Wittenburg, 2020).

Outro problema que se instala nos ambientes de pesquisa é que os cientistas se confrontam com uma pluralidade de ferramentas e padrões que eles precisam selecionar para lidar com novas questões. “Os pesquisadores prefeririam, de forma geral, lidar com objetos de dados de num nível mais

abstrato onde são mais estáveis e independentes de tecnologias particulares” (De Smedt; Koureas; Wittenburg, 2020, p.6). Nesses ambientes, as mudanças tecnológicas devem ser transparentes para os usuários e implementadores; a camada de serviços deveria fazer parte de uma etapa da virtualização que evitaria a necessidade, por parte do pesquisador, de interagir com o emaranhado das tecnologias subjacentes aos objetos de pesquisa.

Para superar esses desafios, é definida uma unidade abstrata de conhecimento, referenciada por **identificadores persistentes e descritas por metadados**, que é capaz de interagir com sistemas automatizados. Ela se fundamenta no conceito pioneiro de Objeto Digital de Robert Khan e Robert Wilensky (1995) e no alinhamento aos aspectos destacados atualmente pelos Princípios FAIR. Esse construto abstrato é chamado de Objeto Digital FAIR (FDO). O seu percurso epistemológico, histórico e técnico será abordado em seguida.

### **3 DO OBJETO DIGITAL ATÉ O FDO**

O meio intensivo em mídias digitais em que estamos imersos, compreende no mundo científico a ecossistemas sociotécnicos informacionais, composto por dados, formalizados por esquemas ou ontologias que podem ser generalizados por metadados e consubstanciados por ofertas de serviços consumidos por seres humanos e máquinas (Hui, 2012). Este cenário corporifica-se pela intensa informatização da geração de conhecimento protagonizada pela eScience, compreendida como uma construção discursiva que se localiza na interface das práticas de pesquisa, tecnologia computacional e política científica, como destaca Paul Wouters (2006). Esta nova configuração, onde várias práticas disciplinares e tecnologias convergem, reivindica métodos inéditos de organização dos objetos digitais produzidos e consumidos por seus fluxos. Isto se torna possível pelo desenvolvimento, em larga escala, de sistemas sociotécnicos, de concepções inovadoras para a gestão desses objetos, e pelo aprofundamento de uma nova compreensão epistemológica para os objetos digitais, que vai além da sua natureza binária. Neste trabalho, essa discussão se restringirá às assertivas necessárias à construção do conceito de FDO.

A ideia de objeto está associada a uma materialidade tocável que restringe a sua inserção no universo digital. Entretanto, essa concepção vem

incorporando um novo conceito de materialidade “não tocável”, que congrega também os objetos digitais numa fisicalidade específica, mas que muitas vezes remete a características comuns aos objetos não digitais, como consumo de energia, desdobramentos ambientais e controvérsias jurídicas (Lischer-Katz, 2017). Assim sendo, o conceito de “objeto” pode também distinguir entidades abstratas ao invés de somente entidades concretas. Além do mais, pode incluir também a **representação digital** dessas entidades cuja natureza binária pressupõe propriedades e atributos específicos (Wittenburg *et al*, 2018).

Podemos, assim, pensar que os objetos digitais são unidades significativas que existem no mundo dos *bits*, que incluem todos os tipos de entidades que precisam ser identificadas univocamente por humanos e por máquinas, tais como dados, metadados, *software*, questões, configurações e outras entidades disponíveis em formatos digitais. “Isto inclui ainda representações digitais de entidades físicas tais como pessoas, instituições e entidades abstratas, tais como conceitos e relações”, enfatizam Peter Wittenburg e colaboradores (2018, p.3).

Os objetos digitais, na esfera científica, podem ser conceitualizados ainda como “coisa” epistêmica, ou seja, objetos que guardam todas as respostas, mas que precisam de processamento de muitos tipos – por exemplo, técnicos e informacionais - para se tornarem objetos técnicos e assim revelarem suas respostas em distintos contextos. “Muitos objetos da ciência [...] foram criados para gerar conhecimento. Podem ser instrumentos de observação ou medição; eles próprios podem ser objetos de estudo, como amostras ou espécimes; ou podem ser representações” (Tybjerg, 2017, p. 269). Todos esses objetos são chamados de “**objetos epistêmicos**” no sentido de que possuem um grande potencial para gerar conhecimento. O conceito de “objeto epistêmico” baseia-se no trabalho de Hans-Jörg Rheinberger (1977) e em seu conceito de “coisas epistêmicas” e montagens experimentais, anunciado em seu notável livro *Toward a History of Epistemic Things: Synthesizing proteins in the test tube*.

Resumindo em poucas palavras: o objeto epistêmico incorpora aquilo que ainda não é conhecido. O segundo elemento – denominado “objeto técnico” – é o conjunto de condições experimentais nas quais os objetos de pesquisa estão inseridos. É por meio desse arranjo “que os objetos de investigação se tornam contextualizados e articulados em um campo mais amplo de prática epistêmica

e cultura material, incluindo instrumentos, dispositivos de inscrição, modelos e os teoremas flutuantes ou conceitos de fronteira ligados a eles” (Rheinberger, 1977, p.29). É por meio dessas condições técnicas que o contexto institucional passa para o trabalho de bancada, enfatiza o autor. Isso acontece em termos de instalações de medição locais, fornecimento de materiais, tradições de pesquisa, fluxo de trabalho de laboratório, convenções e tradições disciplinares e habilidades acumuladas pelo longo caminho percorrido pelo pessoal técnico.

Rheinberger (1977) também nos esclarece sobre o papel que os objetos epistêmicos desempenham no espaço de representação criado nas atividades científicas, trazendo à tona a ideia de descontextualização que se torna um conceito relevante no âmbito da curadoria de dados de pesquisa. “O que há de significativo na representação enquanto inscrição é que as coisas podem ser representadas fora de seu contexto original e local e inseridas em outros contextos. É o tipo de representação que importa”, afirma Rheinberger (1977, p.106). Wouters (2006, p. 11) destaca o interesse especial de muitos projetos de eScience, cujo cerne visa a descontextualização de objetos e posterior contextualização em tempo real e em qualquer contexto. Ele pergunta: “Como isso é possível?” (Wouters, 2006, p.11). Imediatamente, ele mesmo responde: Isso é possível por meio de metadados que devem descrever o significado do objeto de pesquisa para que outras máquinas e humanos possam fazer (re)uso desses objetos em contextos impensáveis no momento da produção do objeto. “Metadados são representações do contexto original de objetos epistêmicos em tais termos que novos contextos podem ser criados para esses objetos, para gerar novas questões”, observa Wouters (2006, p.11).

“As condições técnicas determinam o âmbito das representações possíveis de uma coisa epistêmica, e a coisa epistêmica suficientemente estabilizada se transforma em repertório técnico de arranjo experimental”, assim Rheinberger resume seus argumentos (Rheinberger, 1977, p.29). Portanto, um objeto epistêmico é antes de tudo uma máquina geradora de perguntas; enquanto o objeto técnico é uma máquina de responder perguntas.

No âmbito do presente ensaio, o objeto epistêmico ou objeto de conhecimento, como representação abstrata, é o nosso foco de atenção, que se estabiliza no constructo FDO. Isto porque Rheinberger (1977) coloca a **representação** no coração do empreendimento científico, na qualidade de um

sistema de significados, interpretação e de contextualização/descontextualização. Essa abordagem vai nos ajudar a compreender a natureza abstrata do Objeto Digital e o papel dos metadados como um elemento essencial da sua arquitetura no domínio da ciência e, mais especificamente, da eScience. Vamos compreender, a seguir, como evoluiu essa arquitetura.

### 3.1 O MODELO PIONEIRO DE KHAN E WILENSKY

O conceito de objeto digital foi introduzido por Robert Kahn e Robert Wilensky, em um artigo clássico publicado em 1995 – *A framework for distributed digital objects service* – e que foi atualizado em 2006. Nesses trabalhos, os autores descrevem o aspecto fundamental de uma “infraestrutura que é aberta em sua arquitetura e que suporta uma grande e extensível classe de **serviços distribuídos de informação digital** (Kahn; Wilensky, 2006, p.1). Bibliotecas digitais são um exemplo de tais serviços, enfatizam os autores.

Numerosos outros exemplos podem ser encontrados em aplicações emergentes de comércio eletrônico, completam os autores. O modelo proposto fornece um método para nomear, identificar, e/ou evocar objetos digitais em um sistema distribuído de repositórios que oferecem grande flexibilidade e são adequados para empreendimentos de alcance nacional.

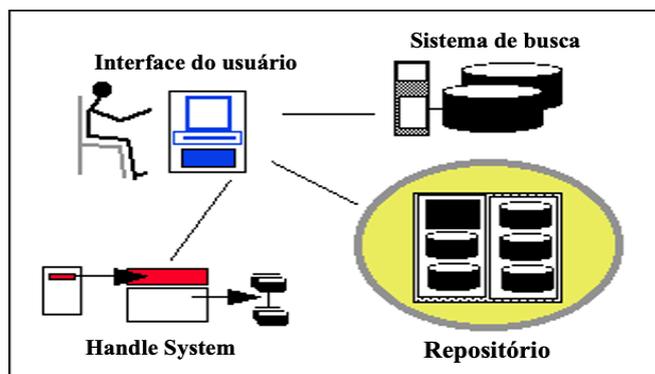
Khan e Wilensky (1995, 2006) definem as entidades básicas que devem estar presentes nesse sistema, no qual as informações, na forma de objeto digital, são armazenadas, acessadas, disseminadas e gerenciadas. Porém, os autores ressaltam que o termo “objeto digital” é aplicado em um sentido técnico sobre um tipo específico de estrutura de dados, e não no sentido geral para qualquer objeto que tenha formato digital. “Talvez um termo como objeto de infraestrutura digital captasse melhor essa intenção”, especificam os autores (Para distinguir esses conceitos, vamos grafar, neste artigo, esse tipo de objeto digital específico com iniciais em caixa alta, como já fizemos). O modelo estabelece ainda convenções de **nomes para identificar e localizar objetos digitais**, bem como descrevem um serviço para usar nomes de objetos para localizá-los e disseminá-los, e, ainda, um **protocolo de acesso** (Khan; Wilensky, 1995; 2006).

No modelo proposto, um Objeto Digital é uma estrutura de dados cujos principais componentes são material digital, ou dados, além de um **identificador**

**único** para esse material denominado *handle*. Assim, um usuário pode então depositar o objeto digital em um ou mais **repositórios**, a partir do qual o objeto estará disponível para outros interessados – sujeitos, é claro, aos **termos e condições** do objeto em particular. Ao depositar um objeto digital em um repositório, seu identificador (*handle*) e o nome do repositório ou endereço IP são registrados em um **sistema de servidores de identificador disponível globalmente** que indica os **endereços em que o objeto digital está armazenado**. Interações como depositar objetos digitais ou acessar objetos digitais em repositórios são realizadas usando uma estrutura para serviços distribuídos de objetos digitais, formalizada por um **protocolo de acesso a repositórios** (RAP, sigla para *Repository Access Protocol*).

O Modelo foi elaborado com o propósito de fundamentar uma rede de bibliotecas digitais de Relatórios Técnicos de Ciência da Computação (CS-TR), iniciado em 1992. O projeto incorporava uma visão de acesso contínuo a uma rede de informações distribuídas e contava com o apoio da Agência de Projetos de Investigação Avançada (ARPA) (Sayão, 2009). O foco inicial deste esforço, organizado e liderado pelo CNRI - *Corporation for National Research Initiative* -, foi desenvolver um *corpus* de material digitalizado a partir das coleções de relatórios técnicos de ciência da computação mantidos por cinco universidades com programas líderes em ciência da computação: Universidade Carnegie Mellon, Universidade Cornell, Instituto de Tecnologia de Massachusetts, Universidade de Stanford e Universidade da Califórnia, em Berkeley (Khan, 1995). A estrutura da biblioteca digital permite a coexistência de muitos sistemas diferentes. Os principais componentes são mostrados na **figura 1** abaixo.

Figura 1 – Principais componentes da Biblioteca Digital



Fonte: Arms; Bianchi; Overly (1997).

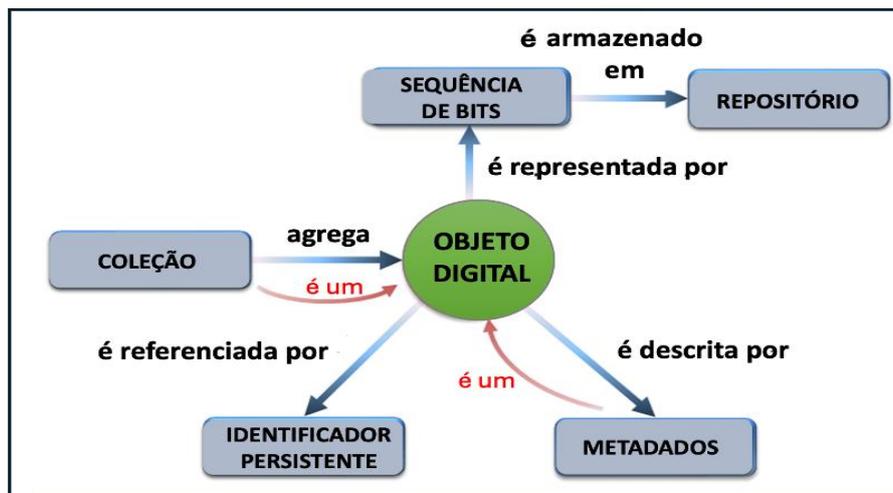
O modelo pioneiro de arquitetura de objetos digitais proposto por Robert Khan e Robert Wilensky (1995, 2006) era agnóstico no sentido de não considerar os conteúdos que transportava. Nos artigos, os autores enfocam os aspectos de rede da infraestrutura proposta, “ou seja, aqueles para os quais o conhecimento do **conteúdo do objeto digital não é necessário**. A definição dos aspectos baseados em conteúdo da infraestrutura não é propositalmente abordada [...]”, confirmam os autores (Khan; Wilensky, 1995, p.118).

Parece claro que os elementos estruturais que permeiam os atuais sistemas já estavam presentes no modelo, mas era preciso avançar na comunicação semântica dos Objetos Digitais. Estes serão os próximos passos.

### 3.2 CORE DATA MODEL

“Em resposta à imensa e crescente fragmentação e ineficiência do trabalho no domínio dos dados” (Wittenburg *et al.*, 2018, p.12), em 2014 o RDA’s *Data Foundation & Terminology Group* (DFT) apresentou seus resultados sobre a definição de seu **core data model** centrado nos Objetos Digitais e na sua organização, além da base terminológica correspondente. O modelo está no escopo de atuação do DFT, que entre os seus objetivos está: deslocar a discussão que se desenrola no âmbito da comunidade de dados, para um conjunto consensual de termos relevantes para os Grupos de RDA, com foco em um modelo básico (core model) de termos e suas relações, juntamente com alguns princípios básicos que harmonizam as soluções de organização de dados; e promover uma cultura comunitária RDA eficaz, pela convergência na direção de uma terminologia essencial que surge a partir de modelos de referência consensuais (Berg-Cross; Ritz; Wittenburg, 2015). O modelo foi baseado numa ampla análise de mais de vinte casos documentados provenientes de vários domínios científicos e explica a natureza pervasiva dos DO’s no contexto do *DFT Core Model* e na estrutura onde o Objeto Digital está integrado/contextualizado com outros importantes elementos de dados e conceitos, conforme sintetizado pela figura 2.

Figura 2 – Relação entre os elementos associados ao Objeto Digital



Fonte: Schwarzmann (2020).

O modelo estabelece que:

- **Um DO é representado por uma sequência de bits**, ou seja, o seu conteúdo é codificado por uma cadeia estruturada de *bits*;
- **O DO é referenciado** por um **identificador único**, resolvível e persistente (PID), emitido por um sistema confiável e disponível globalmente.  
**Definição:** Um identificador persistente é um ID de longa duração representado por uma *string* que identifica univocamente um DO e que se destina a ser persistentemente resolvido para fornecer informações significativas sobre o DO identificado.  
Associado ao PID temos os seguintes elementos:  
**PID Record:** Um registro PID contém um conjunto de atributos armazenados com um PID que descreve propriedades DO.  
**PID Resolver** – O resolvidor de PID é um sistema de resolução baseado em uma infraestrutura disponível globalmente que tem a capacidade de resolver um PID em informações úteis e atuais, que descrevem as propriedades de um DO.
- **O DO é descrito por vários tipos de metadados ricos que também são DOs.**  
**Definição:** Os metadados contêm afirmações descritivas, contextuais e de proveniência sobre as propriedades de um DO. Esses metadados tornarão um DO, por exemplo, encontrável, acessível e utilizável/interpretável. Para tornar os metadados referenciáveis, eles precisam estar associados a um PID e, portanto, ser um DO. Os metadados precisam conter minimamente o PID do DO.
- **Os DOs são agregados em coleções**, que também são DOs – ou seja, têm um PID e são descritos por metadados, cujos conteúdos fazem referência aos seus componentes. Uma **agregação digital** é um pacote de entidades digitais; uma **coleção digital** é uma agregação que contém DOs e outras Entidades Digitais (Des; uma coleção digital é um DO complexo;
- **O DO é armazenado, gerenciado e servido em vários repositórios.**  
**Definição:** um repositório digital é um componente da infraestrutura capaz de armazenar, gerenciar e curar DOs e retornar suas cadeias de *bits* quando uma solicitação for emitida.

“Essa definição [...] difere levemente da formulação original de Khan & Wilensky (1995), que estabelece que um DO tem uma sequência de *bits*

estruturada, um PID único e, no mínimo, metadados-chave que incluem o PID”, ressaltam Peter Wittenburg e colaboradores (2018, p.12) atestando a conexão histórica. Ela também torna o DO um conceito genérico, na medida em que se abstrai de inúmeros possíveis tipos de conteúdo de um DO, que permeiam inteiramente o domínio de entidades de dados digitais (De Smedt; Koureas; Wittenburg, 2020, p. 10).

### 3.3 DIGITAL OBJECT ARCHITECTURE (DOA)

O conceito de Objeto Digital foi aprimorado pela definição da **Digital Object Architecture** (DOA), que se constitui em uma extensão lógica da arquitetura da Internet para atender à necessidade de apoiar o gerenciamento de informações de forma mais geral do que apenas transmitir informações em formato digital, de um local para outro, como permitido pela Internet. A DOA foi estruturada pelo *DONA Foundation*, que tem como mantra “promover as tecnologias que permitem a interoperabilidade entre sistemas de informação”, conforme registra a sua página Web. (DONA, 2020)

Mantido esse propósito, o modelo do DOA introduz o conceito de **objeto digital**, que é o elemento-chave da arquitetura; define também os três sistemas principais e dois protocolos: o sistema de **identificação/resolução**, o **sistema de repositório** e o **sistema de registro**; um **protocolo de interface** é definido para o **sistema de identificação/resolução** e um segundo protocolo de interface para uso pelo **repositório** e pelo **sistema de registro**. Na prática, os componentes do repositório e do registro podem ser combinados. Os elementos da arquitetura são os seguintes:

- **Digital Object** – Objeto digital, uma sequência de *bits* ou um conjunto de sequências de *bits* interpretável por algum sistema computacional e associado a um único identificador.
- **Digital Object Interface Protocol (DOIP)** – Protocolo de Interface do Objeto Digital é um protocolo conceitual simples, porém poderoso, voltado para *softwares* aplicativos interagirem com “serviços” que podem ser objetos digitais ou sistemas que gerenciam esses objetos digitais.
- **Digital Object Identifier Resolution Protocol (DO-IRP)** – Protocolo de Resolução de Identificadores Persistente de Objetos Digitais é um protocolo de resolução rápida para criar, atualizar e resolver identificadores globalmente únicos. Cada identificador está associado a um registro de identificador que os clientes podem resolver usando este protocolo.
- **Identifier/ Resolution System** – Sistema de Resolução de Identificadores permite a atribuição de identificadores persistentes à informação em formato digital estruturada como objeto digital, independentemente da localização dessas informações ou da tecnologia subjacentes a elas; permite também a resolução do

identificador em informações sobre o estado atual do objeto: localização, acesso, política de acesso e uso, permissões, selo de tempo e chaves públicas. As informações de estado são armazenadas na forma de um objeto digital, e a resolução rápida é habilitada usando o DO-IRP.

- **Repository System** – Sistema de Repositório gerencia os objetos digitais incluindo o acesso aos objetos com base no uso de identificadores e com segurança integrada. O acesso ao sistema é via protocolo DOIP. Por meio do uso de identificadores no protocolo de acesso, o sistema de repositório abstrai os detalhes das tecnologias de armazenamento dos clientes, permitindo um mecanismo de longa duração para depositar e acessar objetos digitais.
- **Registry System** – Sistema de Registro é um repositório especializado que armazena metadados sobre os tipos de objetos digitais e as operações possíveis e permitidas sobre os objetos. Tipicamente, armazena metadados de objetos digitais que são gerenciados por um ou mais sistemas de repositório. O acesso a esses sistemas é realizado via DOIP.

### 3.3 FAIR DIGITAL OBJECT

A abordagem FAIR é expressa por 15 princípios orientadores, que podem ser sintetizados pela afirmativa de Wilkinson e colaboradores (2016) como dados e serviços que são encontráveis, acessíveis, interoperáveis, com o propósito de serem reusáveis tanto por máquinas como por pessoas. Essa articulação de orientações se estende ainda a outros ativos informacionais de pesquisa como códigos, algoritmos e modelos (Batista *et al.*, 2022). Após a publicação dos Princípios FAIR, em março de 2016, a ideia em torno de Objetos Digitais se desenvolve em direção ao conceito de FAIR Digital Object, com o propósito de alinhar as características dos DOs com os aspectos destacados pelos princípios FAIR. Nessa perspectiva, a importância do papel dos metadados se torna chave para ambas as concepções. Os princípios FAIR refletem forte relação entre metadado, dado ou objeto digital propriamente dito e identificador persistente.

Uma vez que os princípios FAIR atribuem uma importância significativa aos metadados, um dos principais acréscimos ao conceito original de DO foi a introdução de **metadados como um tipo particular de objeto digital** usado para descrever outros objetos (Santos, 2022, p. 5, grifo nosso).

No plano histórico, a primeira menção ao termo *FAIR Digital Object* apareceu no relatório *Turning FAIR Into Reality*, publicado em novembro de 2018 pela Comissão Europeia, certificando que os princípios FAIR já faziam parte do *roadmap* do *European Open Science Cloud* (EOSC). O relatório destaca que para tornar os princípios FAIR parte dos fluxos de pesquisa é central a criação de uma cultura FAIR e, simultaneamente, a conceitualização e construção de um

ecossistema técnico que habilite um ambiente **FAIR de dados e serviços**. Nesse ecossistema, Objetos Digitais representam dados, software ou outro recurso de pesquisa e devem estar acompanhados por identificador persistente, metadados e documentação contextual suficiente para permitir a descoberta, citação e reuso. Adicionalmente, o dado deve estar representado por formatos de arquivos comuns – idealmente abertos – e documentados ricamente por meio de padrões de metadados e de vocabulários adotados por uma dada comunidade de pesquisa viabilizando a interoperabilidade e o reuso. Além do mais, o compartilhamento dos códigos subjacentes aos dados se torna essencial. Essa ação não se restringe somente ao código fonte, mas também à documentação pertinente, que deve incluir declarações acionáveis por máquina sobre licenças e restrições de acesso (European Commission, 2018).

Os princípios FAIR e os conceitos que a eles estão relacionados oferecem diretrizes importantes para as implementações técnicas, especificamente, em relação ao Objeto Digital FAIR e ao Ecossistema Técnico do FAIR Data (European Commission, 2018). É isso que veremos a seguir, com um grau a mais de profundidade.

#### **4 OBJETO DIGITAL ENCONTRA OS PRINCÍPIOS FAIR**

Partindo do projeto de Robert Khan (1995), que estabelecia os princípios básicos de uma Internet onde datagramas sem significado científico eram roteados num ciclo de desconstrução e posterior reconstrução, logo se compreendeu que os objetos para serem intercambiados de maneira efetiva entre emissores e consumidores deveriam comportar algum significado. Um pouco depois, o projeto da *World Wide Web* idealizado por Berners-Lee, Hendler e Lassila (2001) delineava os primeiros passos nessa direção.

Mais tarde, em 2014 o *Lorentz Centre*, localizado em Leiden, Holanda, organizou um workshop onde foram formulados os Princípios Orientadores FAIR que visam tornar dados localizáveis, acessíveis e interoperáveis com o propósito de potencializar o seu reuso. Este conjunto de diretrizes foi posteriormente publicado, na forma de artigo, no *Nature Scientific Data* (Wilkson *et al.*, 2016; Mons *et al.*, 2017). Neste momento, os princípios FAIR não se referiam explicitamente a Objetos Digitais, mas havia uma intercessão conceitual clara entre esses constructos, na medida em que o FAIR requisitava identificadores

persistentes e que metadados ricos fossem associados às entidades digitais. Nesse contexto, revelava-se um crescente interesse na fusão das abordagens do Digital Object e do FAIR, tanto no nível conceitual quanto no técnico. Assim sendo, nos movimentos seguintes o RDA DFT Group começa a equacionar a ampliação do escopo do DO para o contexto dos Princípios FAIR, definindo o *FAIR Digital Object* (FDO) na qualidade de uma unidade de dados que é capaz de interagir com os sistemas automatizados de processamento de dados.

Ratificando as fronteiras conceituais do FDO, Luiz Bonino da Silva Santos (2022) define-o mais formalmente como uma **sequência de bits** que representa uma **unidade de informação** acionável por máquina, **identificada** por um identificador globalmente único, persistente e resolvível com comportamento de **resolução previsível**, descrito por registros de **metadados** – que também são Objetos Digitais FAIR -, e classificados por um **sistema de tipagem**. Nessa perspectiva, um FDO é uma unidade estável e acionável que agrupa informações suficientes para permitir a **interpretação** e **o processamento confiáveis** dos dados contidos nele (De Smedt; Koureas; Wittenburg, 2020), tendo como característica primordial a essencialidade do papel dos identificadores persistentes como parte da sua arquitetura e uma âncora essencial para a futura organização do domínio digital, não importando a sua natureza. A introdução do FDO estável baseado em identificadores persistentes não mudará a atual característica da Web, que é informação desestruturada, mas oferecerá uma solução mais estável e duradoura para *datasets* num domínio registrado e que precisam ser preservados por longo tempo (De Smedt, Koureas e Wittenburg, 2020).

Assim sendo, o conceito de FDO, que viabiliza um cenário inovador para os sistemas de conteúdos digitais, sejam eles científicos, sejam de outra natureza, está ancorado nas bases estabelecidas por Khan e Wilensky (1995) na última década do século passado. Na trajetória histórica, conceitual e técnica iniciada pelos autores, os elementos estruturais por eles delineados são atualmente colocados no contexto dos Princípios Orientadores FAIR. Esses princípios assumem globalmente um papel proeminente como um arcabouço para a sustentabilidade dos dados de pesquisa e para a construção apropriada de sistemas de curadoria. Além disso, a abordagem FAIR sempre considera a ideia de acionabilidade por máquina - entendida como a capacidade de sistemas

computacionais de realizarem serviços sobre os dados sem intervenção humana (De Smedt; Koureas; Wittenburg, 2020; Schwardmann, 2020). Para isso, as máquinas precisam agir de **forma autônoma** e adequada diante da ampla gama de tipos de dados encontrados durante a exploração do ecossistema global de dados (Wilkinson *et al.*, 2016). Assim, o termo “**acionável por máquina**” é usado atualmente para indicar um *continuum* de estados possíveis em que um **objeto digital fornece uma informação** cada vez mais detalhada para um explorador de dados computacional de ação autônoma (Wilkinson *et al.*, 2016). Essa concepção de acionalidade se estende para os metadados, configurando o chamado Metadado para Máquinas (Batista *et al.*, 2022).

Este cenário parece possibilitar a realização de uma Internet de Dados e Serviços FAIR (IFDS), cujo ponto central é o conceito de FDO, um tipo de Objeto Digital que está no contexto *FAIR Digital Object Framework* (FDOF), que como o próprio nome diz, é um arcabouço que define um modelo para **representar objetos** em um ambiente digital e um conjunto de recursos para fornecer suporte fundamental para os princípios FAIR (Santos, 2021). É o que veremos a seguir.

## 5 FAIR DIGITAL OBJECT FRAMEWORK (FDOF)

Os Objetos Digitais podem estar localizados numa cartografia abstrata delineada no final de 2022, por Santos, em um documento de trabalho. O autor denomina esse território virtual de *FAIR Digital Object Framework*, e o define da seguinte forma:

O FAIR Digital Object Framework (FDOF) é um arcabouço que define um modelo para representar objetos em um ambiente digital e um mecanismo para criar, manter e (re)usar esses objetos. O arcabouço consiste em um identificador de resolução de comportamento previsível, um mecanismo para recuperar os metadados do objeto e um sistema de tipagem de objetos. O principal objetivo do FDOF é definir um conjunto de recursos que ofereça um apoio fundacional para os princípios FAIR (Santos, 2022, p.1, grifo nosso).

De uma forma sucinta, no contexto do FDOF, um objeto digital é uma sequência de *bits* que representa uma unidade informacional – um *dataset*, um documento, uma foto, um serviço etc. – e que é apresentado de acordo com os princípios FAIR. O FDOF tem como objetivo central providenciar recursos que permitam

responder as perguntas seguintes, destacando que elas precisam ser interpretadas por seres humanos e por máquinas:

- **Qual é o objeto que está identificado** por esse identificador?
- Como eu posso **conseguir mais informações** sobre esse objeto?
- **O que pode ser feito** com esse objeto? (autorizações gerais)
- **O que eu posso fazer** com esse objeto particular? (minhas autorizações)

Santos (2022) relembra que, por projeto, os princípios FAIR não prescrevem nenhuma tecnologia e deixam em aberto as opções de implementação. Entretanto, para concretizar os comportamentos dos princípios FAIR é necessário definir opções de implementações. O FDOF, como uma construção de uma estrutura, naturalmente, faz algumas escolhas de implementação, embora mínimas, elas são selecionadas baseadas em níveis de interoperabilidade que o arcabouço pretende garantir aos participantes do ecossistema. O passo seguinte é compreender a estrutura dos Objetos Digitais FAIR e as suas diversas camadas de abstração.

## 6 OBJETO DIGITAL FAIR

Os Objetos Digitais FAIR desempenham um papel crucial na ciência aberta, na pesquisa colaborativa e na preservação de recursos digitais. Subjacente a este conceito está a ideia de garantir que o conhecimento e os dados científicos sejam acessíveis e utilizáveis por um público amplo, promovendo assim avanços na pesquisa e no desenvolvimento de soluções inovadoras. Para que isto seja possível, os Objetos Digitais FAIR são estruturados em diversas camadas de abstração que filtram sua complexidade.

### 6.1 ABSTRAÇÃO COMO MEIO DE OCULTAR A COMPLEXIDADE

Um grande obstáculo para a automação dos processos de gestão de dados e dos demais objetos de pesquisa é a complexidade e heterogeneidade intrínsecas desses objetos. Um meio de superar esses obstáculos é ocultá-los por meio de encapsulamento e virtualização, cujos métodos têm se mostrado extremamente poderosos quando aplicados a projetos de outros sistemas complexos e podem beneficiar o ecossistema de infraestrutura de dados que se tornam altamente complexos (Wittenburg *et al.*, 2018). Encapsulamento, nesse contexto, significa que usuários humanos não precisam saber qual é a estrutura

interna e como as operações são implementadas, dando margem para que detalhes que não são necessários numa camada específica possam ser ocultados. Por exemplo: na camada de infraestrutura de dados não há diferença considerável entre dado e metadado, *software*, assertivas semânticas entre outros objetos, posto que todos podem ser vistos como uma instância de dados; enquanto na camada de gestão de dados e de reuso é necessário distingui-los, visto que os metadados têm inúmeras funcionalidades que vão ser usados para governar a gestão de dados. Na **virtualização**, o objeto é substituído por sua representação lógica. Nessa direção, um enfoque clássico da ciência da computação, que é considerado o modo mais abstrato de representação lógica, é o ponteiro que conduz ao objeto, ocultando, dessa forma, toda a complexidade por trás de uma referência pura ao objeto (Schwardmann, 2020).

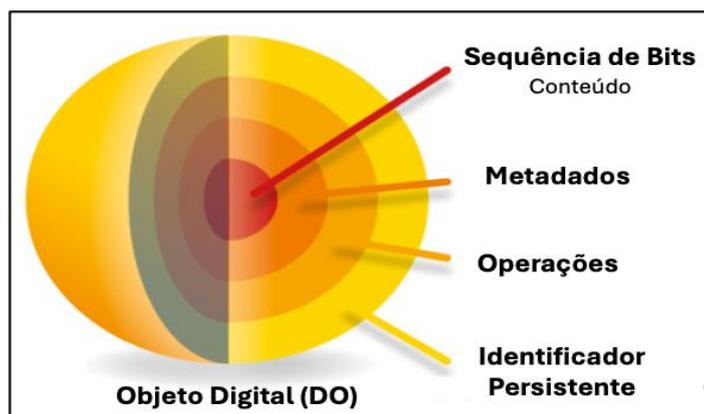
O primeiro passo da abstração – virtualização e encapsulamento – dos dados é a identificação dos elementos mínimos que são, em algum grau, atômicos para a perspectiva da gestão de dados e do reuso. Conforme visto na seção 3.1, esses elementos foram chamados de objetos digitais por Khan e Wilensky, já em 1995, e foram reusados e adaptados pelo grupo de trabalho do RDA em Data Foundation and Terminology (DFT), e estão relacionados a outros importantes conceitos representados pela figura 2 (Berg-Cross; Ritz; Wittenburg, 2015; Schwardmann, 2020).

Ainda segundo Ulrich Schwardmann (2020), identificar como representar a estrutura lógica de um objeto digital no nível apropriado de abstração ainda é um desafio importante, posto que isso depende de muitas variáveis, como o dado propriamente dito, *workflows* específicos e casos de uso da gestão e do reuso dos dados. Mas em todos os casos, o ponteiro – como a mais abstrata representação lógica – tem papel proeminente. Além do mais, considerando que o objeto deve estar disponível transversalmente por domínios e sites, o ponteiro deve ser uma referência globalmente única.

## 6.2 CAMADAS DE ENCAPSULAMENTO DE INFORMAÇÃO NO FDO

A figura 3 ilustra as várias camadas de abstração, encapsulamento e ligações que compõem a arquitetura proposta para um Objeto Digital FAIR e as requisições de operações possíveis sobre eles, elementos que são centrais para a construção do conceito.

Figura 3 – Camadas de encapsulamento no FDO



Fonte: De Smedt; Koureas; Wittenburg (2020).

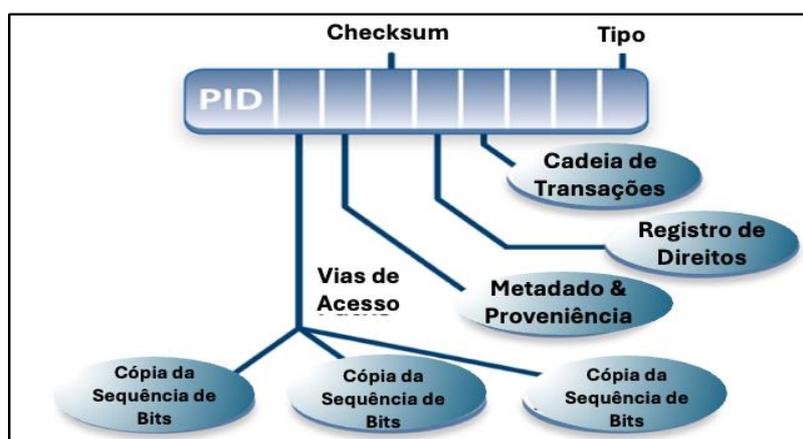
- Os objetos são acessados por **meio do seu PID** que os identificam univocamente; os atributos do PID apontam para as localizações da sequência de *bits*, seus metadados e outras propriedades relevantes, ou seja, os registros de PID permitem incluir ponteiros – também PIDs – para todas as entidades relevantes e tipos de metadados associados a um DO.
- O coração do DO é uma **sequência de bits** que codifica algum conteúdo, que pode ser de qualquer natureza: dados, metadados, *software* etc.
- A sequência de *bits* é descrita por **metadados de diferentes tipos** – descritivos, técnicos, contextuais, de proveniência, direitos de acesso etc. - para permitir o acesso e a correta interpretação do conteúdo DO ao longo do tempo.
- Os DOs podem receber **requisições para operações**, as quais são herdadas dos seus tipos, como conhecido da programação orientada a objeto; o tipo de DO permite ao usuário definir o tipo específico de operação e ao fazer isso aplica o encapsulamento. Isso implica na necessidade de um **registro de tipos**.
- **Por meio das operações** os metadados de um DO podem ser acessados; os FDOs permitem/habilitam abstração na medida em que, no nível da gestão de objetos, não importa se o conteúdo do FDO é dado, metadado, *software*, afirmação, *knowlets* etc.

No FDO, uma sequência de *bits* de dados é vinculada a todos os componentes de metadados necessários (descritivos, científicos, sistema, direitos de acesso, transações etc.) durante todo o tempo. Enquanto o PID e metadados do FDO são geralmente abertos, o acesso a sua sequência de *bits* pode ser sujeito a autenticação e autorização. Por exemplo, para proteger dados pessoais ou dados protegidos de outra forma. Todos os atores no ciberespaço que lidam com os FDOs estão conectados usando um protocolo unificado que garante interoperabilidade. Para este propósito, é proposto o DOIP – *DO Interface Protocol* – cujo papel pode ser comparado ao HTTP na *Web*.

## 7 IDENTIFICADOR PERSISTENTE

A figura 4 indica os papéis do PID como um concentrador dos componentes e ponteiros num Objeto Digital. Esta proposta de configuração estabelecida para o DO define e amplia os papéis do PID como um ponto de ancoragem para a identificação, acesso e reuso do DO. “Assumindo que PID é genuinamente persistente, [...] faz sentido vincular as informações essenciais ao registro de PID que serão retornadas para o usuário quando o PID for resolvido” (Wittenburg *et al.*, 2018, p.10). Faz sentido também armazenar informações críticas ou ponteiros que permitam que as máquinas encontrem as informações relevantes sobre o DO nesse registro.

Figura 4 – Uso do PID para apontar informações relevantes sobre o DO



Fonte: Wittenburg *et al.* (2018).

Nesse contexto, se revela a relevância do identificador persistente para conectar entidades proximalmente relacionadas e que são constitutivas para os DOs, tais como os diversos tipos de metadados que são necessários para encontrar, acessar, interpretar e controlar o reuso da sequência de *bits*, que é presença física do conteúdo. As informações podem conter caminhos para acesso à sequência de bits, metadados, os registros de direitos contendo as especificações de permissão, um ponteiro para entrada de *blockchain* associados com um DO, baseado no seu tipo.

Assim, “**PID não é somente um identificador** [posto que] ele deve incluir ou apontar para todas as informações relevantes que são importantes para trabalhar com o DO, e nesse sentido ser resolvido para formar instancias de

significado e/ou localização”, destacam Wittenburg *et al.* (2018, p. 13, grifo nosso). Entretanto, para cumprir essas novas funções, os identificadores persistentes precisam ter um comportamento previsível de forma que os agentes artificiais possam saber o que esperar – o objeto, uma *landing page*, metadados - quando o identificador for resolvido. Isto coloca em pauta a necessidade de identificadores previsíveis que sejam capazes de viabilizar a acionalidade por máquina.

Para permitir as máquinas interpretarem os atributos de um registro PID, existe a necessidade de associar **tipos** a eles. Isto pode ser realizado utilizando-se o **Sistema de Tipificação de Dados** (*Object Typing System*).

## 8 SISTEMA DE TIPIFICAÇÃO DE OBJETOS

As operações que podem ser aplicadas sobre um objeto devem estar registradas em um sistema, na forma de metadados, que caracteriza o tipo do objeto, mas também o contexto que envolve um **objeto individual**. Isto inclui, por exemplo, além do elenco de operações dadas pelo tipo de objetos, as permissões para manipular o objeto. Portanto, o sistema, que é um repositório especializado que armazena metadados sobre o objeto, define as operações que o objeto deve performar no **nível do tipo**, que informa que tipo de operação é aplicável para um tipo particular de objeto; e no **nível individual**, que lida com as permissões sobre quem pode realizar essas operações. Santos (2022) alinha as informações que o sistema pode disponibilizar:

- **Formato de codificação** do objeto;
- **Tipo do objeto** digital em relação a sua função informacional, por exemplo, *dataset*, imagem, serviço etc.;
- As **entidades que estão representadas** pelo objeto digital, por exemplo, uma montanha numa foto, uma dada proteína num *dataset* de proteínas;
- As **operações que são aplicáveis** para um dado tipo de objeto;
- As **operações que são permitidas** de serem efetuadas sobre um objeto digital específico.

Desde que diferentes tipos de objetos digitais podem ser descritos com diferentes conjuntos de propriedades, o FDOF *typing system* pode determinar esquemas de metadados específicos para Objetos Digitais FAIR específicos. Nesse contexto, o esquema de metadados define que propriedades devem ou podem estar presentes num dado registro de metadados. Quanto mais propriedades, mais ricos são os metadados.

## 9 À GUIA DE CONCLUSÃO

O desafio colocado para a ciência de dados, especialmente no âmbito científico, é que o ritmo extraordinário de geração de dados não tem sido acompanhado por uma dinâmica mais produtiva da gestão aplicada a esses dados que os habilite a se tornarem fontes integradas de serviços, no escopo de um ecossistema informacional automatizado. Esse desafio precisa de soluções disruptivas, mas que ao mesmo tempo estejam ancoradas em conhecimentos longamente consolidados das áreas de Ciência da Informação e da Ciência da Computação. O FDO se localiza nessa intercessão, porém alinhado também aos saberes e tradições disciplinares.

Além dos limites da Ciência da Computação e da Informação, esse constructo abstrato tem sido objeto de estudos e pesquisa por grupos de domínios científicos específicos e comunidades de práticas como biodiversidade, física de partículas, geoquímica. Observa-se também que o setor produtivo demonstra interesse na aplicação do FDO para aprimorar os processos de gestão dos seus dados, posto que as soluções adotadas pelo setor são fragmentadas (*ad hoc*), despadronizadas e complexas (WITTENBURG *et al.*, 2018). Essa aproximação contribuiu para a formação de um ecossistema de dados e serviços mais ricos e abrangentes. O que os estudos e os testes que estão sendo realizados revelam é que o conceito de FDO muda significativamente a visão sobre o que um *dataset* deve ser. Este fato torna relevante que os conhecimentos gerados sobre o tema sejam decodificados para a ampla audiência responsável pelos ciclos de vida dos dados, nos mais diversos setores.

O forte acoplamento entre objetos digitais e princípios FAIR, não obstante tenham enfoques conceituais bem distintos, na medida em que os princípios são diretrizes e política, e os objetos digitais são abstrações técnicas, configura uma solução extremamente frutífera, posto que a implementação concreta do FDO levará as estruturas de dados a cumprirem implicitamente, pelo menos, parte das orientações preconizadas pelo FAIR (Schardmann, 2020), o que nos leva a concluir que a implementação do conceito trará impactos importantes na composição de uma nova gestão de dados que supere os desafios críticos de hoje. O conceito de FDO viabiliza um cenário inovador para os sistemas de

conteúdos digitais. sejam eles científicos, sejam eles de outra natureza e abrem perspectivas importante para pesquisa interdisciplinar.

O presente estudo apresenta uma abordagem que toca mais na construção conceitual e abstrata do FDO e do IFDO, no entanto, para a implementação desses constructos e para os ensaios e *testbeds*, apesar da máxima independência de tecnologias específicas preconizadas, são necessárias algumas opções nessa direção. Assim sendo, como parte da atuação de um dos autores no FDO *Steering Committee*, a pesquisa prossegue pelos territórios mais técnicos do FDO, que serão abordados na próxima publicação.

## REFERÊNCIAS

ARMS, William Y.; BLANCHI, Christophe; OVERLY, Edward A. An Architecture for Information in Digital Libraries. **D-Lib Magazine**, Feb. 1997. Disponível em: <https://www.dlib.org/dlib/february97/cnri/02arms1.html>. Acesso em: 01 set. 2023.

BATISTA, Dominique *et al.* Machine actionable metadata models. **Scientific Data**, v. 9, 2022. Disponível em: <https://www.nature.com/articles/s41597-022-01707-6>. Acesso em: 07 set. 2023.

BERG-CROSS, Gary; RITZ, Raphael; WITTENBURG, Peter. **RDA Data Foundation and Terminology - DFT: Results RFC**. State: 29.6.2015. Version 1.5. Disponível em: <https://www.rd-alliance.org/system/files/DFT%20Core%20Terms-and%20model-v1-6.pdf>. Acesso em: 10 set. 2023.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 284, n. 5, p. 34-43, May 2001. Disponível em: <http://www.krchowdhary.com/ai/ai16/sematic%20web-sci-am.pdf>. Acesso em: 07 set. 2023.

DE SMEDT, Koenraad; KOUREAS, Dimitris; WITTENBURG, Peter. FAIR Digital Objects for Science: From data pieces to actionable knowledge units. **Publications**, v. 8, n. 2, 2020. Disponível em: [https://www.mdpi.com/2304-6775/8/2/21?type=check\\_update&version=2](https://www.mdpi.com/2304-6775/8/2/21?type=check_update&version=2). Acesso em: 06 jan. 2023.

DONA Foundation. **DONA**. 2020. Disponível em: <https://www.dona.net/> Acesso em: 02 nov. 2023

EUROPEAN COMISSION. **Turning FAIR into reality**: Final report and action plan from the Eropean Comission expert group on FAIR data. 2018. Disponível em: <https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8->

9982-01aa75ed71a1/language-en/format-PDF/source-80611283. Acesso em: 28 out. 2023.

GRAY, Jim *et al.* **Scientific data management in the coming decade.** Redmont, WA: Microsoft Corporation, Jan. 2005. Disponível em: <https://arxiv.org/ftp/cs/papers/0502/0502008.pdf>. Acesso em: 19 jul. 2022.

GRAY, Jim; SZALAY, Alexander S. **Where the rubber meets the sky:** Bridging the gap between database and science. Redmont, WA: Microsoft Corporation, Oct. 2004. Disponível em: <https://arxiv.org/abs/cs/0502011>. Acesso em: 22 mar. 2023.

HUI, Yuk. What is a digital object? **Metaphilosophy**, v. 43, n. 04, 2012. Disponível em: [https://www.digitalmilieu.net/documents/HUI\\_what%20is%20a%20digital%20object%20Metaphilosophy.pdf](https://www.digitalmilieu.net/documents/HUI_what%20is%20a%20digital%20object%20Metaphilosophy.pdf). Acesso em: 07 set. 2023.

KHAN, Robert. **An Introduction to the CS-TR Project.** Corporation for National Research Initiatives (CNRI), 1995. Disponível em: [http://www.cnri.reston.va.us/tmp\\_hp/describe.html](http://www.cnri.reston.va.us/tmp_hp/describe.html). Acesso em: 01 set. 2023.

KHAN, Robert; WILENSKY, Robert. **A framework for distributed digital objects service.** May 1995. Disponível em: <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>. Acesso em: 06 dez. 2022.

KHAN, Robert; WILENSKY, Robert. A framework for distributed digital objects service. **International Journal on Digital Libraries**, v. 6, n. 2, p. 115–123, 2006.

LISCHER-KATZ, Zack. Studying the materiality of media archives in the age of digitization: Forensics, infrastructures and ecologies. **First Monday**, v. 22, n. 1-2, 2017. Disponível em: <https://firstmonday.org/ojs/index.php/fm/article/view/7263/5769>. Acesso em: 06 dez. 2022.

MONS, Barend *et al.* Cloudy, increasingly FAIR: revisiting the FAIR Data guiding principle for the European Open Science. **Information Service & Use**, v. 37, n. 1, p. 49-56, 2017. Disponível em: <https://content.iospress.com/articles/information-services-and-use/isu824>. Acesso em: 22 mar. 2023.

RHEINBERGER, Hans-Jörg. **Toward a History of Epistemic Things:** Synthesizing proteins in the test tube. California: Stanford University Press, 1977.

SANTOS, Luiz Olavo Bonino da Silva. **FAIR Digital Object Framework Documentation.** 2022. Disponível em: <https://fairdigitalobjectframework.org/>. Acesso em: 20 jun. 2023.

SANTOS, Luiz Olavo Bonino da Silva *et al.* **Towards a conceptual model for the FAIR Digital Object Framework.** 23 Feb. 2023. Disponível em: <https://arxiv.org/pdf/2302.11894.pdf>. Acesso em: 20 jun. 2023.

SAYÃO, Luis Fernando. Afinal, o que é biblioteca digital. **Revista da USP**, São Paulo, n. 80, p. 6-17, 2009. Disponível em: <https://www.revistas.usp.br/revusp/article/view/13709>. Acesso em: 07 set. 2023.

SCHULTES, Erick; WITTENBURG, Peter. FAIR Principles and digital objects: Accelerating convergence on data infrastructure. **Communications in Computer and Information Science**, v. 1003, p. 3-16, 2018.

SCHWARDMANN, Ulrich. Digital objects – FAIR Digital Objects: Which services are required? **Data Science Journal**, v.19, n.1, 2020. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2020-015/>. Acesso em: 20 mar. 2023.

TYBJERG, Karin. Exhibiting Epistemic Objects. **Museum & Society**, v.15, n. 3, p. 269-286, 2017. Disponível em: <https://pdfs.semanticscholar.org/a152/1b29555cb7982794a6c80a1e3504ba2d8782.pdf>. Acesso em: 06 dez. 2022.

WILKINSON, Mark D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 1, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 22 mar. 2023.

WITTENBURG, Peter *et al.* **Digital objects as drivers towards convergence in data infrastructures**. Dec. 2018. Disponível em: [https://www.rd-alliance.org/sites/default/files/Digital\\_Objects\\_as\\_Drivers\\_towards\\_Convergence\\_in\\_Data.pdf](https://www.rd-alliance.org/sites/default/files/Digital_Objects_as_Drivers_towards_Convergence_in_Data.pdf). Acesso em: 10 jul. 2023.

WOUTERS, Paul. What is the matter with e-science? – thinking aloud about informatisation in knowledge creation. **The Pantaneto Forum**, United Kingdom, n. 23, July 2006. Disponível em: <https://pure.knaw.nl/ws/portalfiles/portal/489419/21643.pdf>. Acesso em: 27 out. 2023

## **AGRADECIMENTO AOS FINANCIADORES**

CNPq, FAPERJ e IBICT



Este obra está licenciado com uma Licença [Creative Commons Atribuição-NãoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).