



## GT-8 – Informação e Tecnologia

ISSN 2177-3688

### PRINCÍPIOS FAIR E A COMPREENSÃO DE DADOS POR MÁQUINAS

#### *FAIR PRINCIPLES AND DATA UNDERSTANDING BY MACHINES*

**Emanuelle Torino - Universidade Tecnológica Federal do Paraná (UTFPR)**  
**Caio Saraiva Coneglian - Universidade de Marília (UNIMAR); Universidade Estadual Paulista (UNESP); Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)**  
**Silvana Aparecida Borsetti Gregório Vidotti - Universidade Estadual Paulista (UNESP)**

#### **Modalidade: Trabalho Completo**

**Resumo: Introdução:** As discussões sobre os Princípios FAIR enfatizam a relevância desses princípios no tratamento de dados, metadados, objetos digitais e infraestruturas de armazenamento. Em especial, a discussão acerca da capacidade das máquinas compreenderem os dados se entrelaça com as pesquisas de Web Semântica e da semântica dos dados. **Objetivo:** Objetivamos discutir as relações entre os Princípios FAIR e a semântica dos dados, visando contribuir para a aplicação de tais Princípios aos dados, metadados, objetos digitais e infraestruturas de disponibilização. **Método:** Por meio de uma pesquisa teórica exploratória, o estudo procura compreender as intersecções entre os Princípios FAIR e a semântica dos dados, de modo a auxiliar na sua aplicação de forma ampla. **Resultados:** Os resultados enfatizam a necessidade de considerar a compreensão dos dados por máquinas na implementação dos Princípios FAIR, destacando a aproximação entre o FAIR e a semântica dos dados. O estudo reforça a importância da compreensão de dados por máquinas na discussão e aplicação dos Princípios FAIR, ressaltando a necessidade da compreensão humana desses princípios e da semântica dos dados para uma implementação eficaz de uma infraestrutura de dados e serviços FAIR. **Conclusão:** A compreensão e domínio dos Princípios FAIR e da semântica dos dados possibilitará que haja uma infraestrutura de dados e serviços FAIR, preparada por humanos para que seja processável automaticamente por máquinas em benefício das ações requeridas por humanos, com o uso adequado de técnicas e tecnologias necessárias para a compreensão e o processamento de dados, metadados e objetos digitais por máquinas.

**Palavras-chave:** princípios FAIR; semântica dos dados; compreensão de dados por máquinas.

**Abstract: Introduction:** Discussions on the FAIR Principles emphasize the relevance of these principles in the handling of data, metadata, digital objects and storage infrastructures. In particular, the discussion about the ability of machines to understand data is intertwined with research on the Semantic Web and data semantics. **Objective:** We aim to discuss the relationships between the FAIR Principles and data semantics, aiming to contribute to the application of such Principles to data, metadata, digital objects and availability infrastructures. **Method:** Through an exploratory theoretical research, the study seeks to understand the intersections between the FAIR Principles and the semantics of data, in order to help its application in a broad way. **Results:** The results emphasize the need to consider the understanding of data by machines in the implementation of the FAIR Principles, highlighting the approximation between FAIR and data semantics. The study reinforces the importance of machine understanding of data in the discussion and application of the FAIR Principles, highlighting the need for human understanding of these principles and data semantics for

an effective implementation of a FAIR data infrastructure and services. **Conclusion:** Understanding and mastering the FAIR Principles and data semantics will enable an infrastructure of FAIR data and services, prepared by humans to be automatically processable by machines for the benefit of the actions required by humans, with the appropriate use of techniques and technologies necessary for the understanding and processing of data, metadata and digital objects by machines.

**Keywords:** FAIR principles; data semantics; machine data comprehension.

## **1 INTRODUÇÃO**

As discussões acerca dos Princípios FAIR têm se avolumado nos últimos anos. Isso mostra que os aspectos relacionados ao tratamento computacional dos dados para que sejam localizáveis, acessíveis, interoperáveis e reutilizáveis por máquinas e humanos é tema relevante no contexto científico, sobretudo na Ciência da Informação. Vale destacar que essa discussão teve início na aplicação aos dados de pesquisa, mas, aos poucos, foi sendo ampliada para o contexto dos dados, objetos digitais e suas infraestruturas de forma mais genérica.

É importante considerar que os Princípios FAIR estão diretamente vinculados à capacidade das máquinas de localizar, acessar e processar dados, visando o reuso, o que requer compreensão de tais dados. Por outro lado, a semântica dos dados, discutida de forma exaustiva pela Web Semântica, visa facilitar a comunicação e a aplicação de técnicas e tecnologias em diferentes fontes de dados.

No Brasil, ainda são incipientes as discussões acerca da aplicação e da maturidade dos Princípios FAIR, seja em dados, objetos digitais ou ambientes informacionais. A mesma situação é observada no que se refere à semântica dos dados e à aplicação de tecnologias da Web Semântica em conjunto com os Princípios FAIR. Esse contexto demonstra lacunas para a discussão.

Nesse sentido, objetivamos discutir as relações entre os Princípios FAIR e a semântica dos dados, visando contribuir para a aplicação de tais Princípios aos dados, objetos digitais e suas infraestruturas de disponibilização.

Para tanto, foi realizada uma pesquisa exploratória de cunho teórico, visando compreender as intersecções entre os Princípios FAIR e a semântica dos dados, de modo a auxiliar sua aplicação aos dados, objetos digitais e suas infraestruturas de armazenamento, além de favorecer a aplicação e a compreensão de dados por máquinas. A pesquisa teórica

teve como fonte as bases *Web of Science* e *Scopus* e, como ponto de partida, as publicações de autoria dos precursores dos Princípios FAIR.

Tal discussão reforça, ainda, a necessidade da compreensão dos Princípios FAIR e da semântica dos dados por humanos, para que possam ser aplicados e, com isso, possibilitar que haja uma infraestrutura de dados e serviços FAIR.

## **2 PRINCÍPIOS FAIR**

Os Princípios FAIR foram definidos por Wilkinson *et al.* (2016) como orientações independentes que visam favorecer a localização, o acesso, a interoperabilidade e o reuso de dados e metadados, orientados por máquinas e que trazem benefícios para humanos. Para Wilkinson *et al.* (2016, tradução nossa), “[...] os Princípios FAIR enfatizam especificamente o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente, além de apoiar sua reutilização por indivíduos”.

Contudo, as formas de implementação dos Princípios FAIR ainda não são consensuais (MONS *et al.*, 2017; JACOBSEN *et al.*, 2020), o que favorece que estudos sejam realizados na perspectiva de compreender como tais princípios podem ser aplicados em situações práticas (HEINRICHS; POLITZE; YAZDI, 2022).

Jacobsen *et al.* (2020, tradução nossa) refletem que muitas iniciativas relacionadas à capacidade de ação das máquinas já foram realizadas e que os Princípios FAIR consistem em uma “[...] visão multissetorial de uma infraestrutura que suporta a reutilização de dados acionáveis por máquina [...]”. Por outro lado, esclarecem que o fato de os Princípios FAIR serem orientadores e independentes entre si “[...] resultou em interpretações inconsistentes que carregam o risco de levar a implementações incompatíveis” (JACOBSEN *et al.*, 2020, tradução nossa), o que os levou a encorajar e fornecer considerações acerca da implementação de cada um dos Princípios.

Essa problemática exposta pelos autores se afasta do que seria adequado para o processamento das máquinas, ou seja, a interpretação comum e exaustiva dos Princípios FAIR e todas as suas facetas. Desta forma, os precursores dos Princípios FAIR (JACOBSEN *et al.*, 2020), nas perspectivas teórica e prática, admitem que os conceitos utilizados para estabelecer os Princípios FAIR podem ter gerado interpretações inadequadas no momento

da implantação, dentre os quais destacam a capacidade de ação da máquina e os metadados.

No que tange à capacidade de ação da máquina, “Isso implica (na verdade, exige) que os recursos que desejam cumprir ao máximo as diretrizes FAIR devem utilizar uma estrutura legível por máquina amplamente aceita para representação e troca de dados e conhecimento” (MONS *et al.*, 2017, tradução nossa). Destacam os autores que, atualmente, o *Resource Description Framework* (RDF)<sup>1</sup> e as ontologias formais são as soluções mais viáveis e amplamente adotadas para o FAIR, o qual não se limita a estas tecnologias e não deve ser confundido com Web Semântica.

Os Princípios FAIR traçam orientações genéricas e suas facetas as tornam escalonáveis. Enquanto orientações, há a necessidade de que todos os elementos constituintes dos Princípios e facetas FAIR sejam compreensíveis por humanos para que sejam aplicados aos dados, metadados, objetos digitais e ambientes informacionais digitais. Isso requer domínio e disponibilidade de uma infraestrutura favorável e compatível, ambos apoiados em técnicas e tecnologias.

Destaca-se que o objetivo principal dos Princípios FAIR não consiste na abertura dos dados, objetos e infraestruturas digitais, mas sim no desenvolvimento de uma infraestrutura de metadados e de dados que seja facilmente acionável e processável por máquinas. Desta forma, uma vez disponível em formato digital, atendendo às orientações dos Princípios FAIR, ambientes informacionais digitais, objetos digitais, dados e metadados são facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis por agentes e aplicações computacionais o que pode, inclusive, favorecer a disponibilização de uma internet de dados e serviços FAIR (MONS *et al.*, 2017; JACOBSEN *et al.*, 2020). Assim, os sistemas que objetivam gerenciar dados FAIR devem incorporar interface adequada à troca de dados e metadados (HEINRICHS; POLITZE; YAZDI, 2022).

Para Mons *et al.* (2017), os Princípios FAIR buscam assegurar que os objetos resultantes da pesquisa sejam reutilizados. Assim, são compostos por um conjunto de princípios orientadores para aumentar a reutilização, considerando a localização, o acesso e a interoperabilidade – com créditos apropriados e de modo a gerar benefícios ao autor e ao usuário. “O FAIR simplesmente descreve as qualidades ou comportamentos exigidos dos

---

<sup>1</sup> Disponível em: <https://www.w3.org/RDF/>.

recursos de dados para alcançar – possivelmente incrementalmente – sua descoberta ideal e reutilização acadêmica” (MONS *et al.*, 2017, tradução nossa).

Tais afirmações justificam a necessidade de que a comunidade se aprofunde na compreensão dos conceitos relacionados aos Princípios FAIR e das condições necessárias para que dados, metadados, objetos digitais e ambientes informacionais digitais otimizem, por exemplo, suas capacidades de visibilidade, de processamento e de enriquecimento. Destaca-se nesse sentido, sobretudo no caso dos objetos digitais, a necessidade de que haja o tratamento dos dados e metadados intrínsecos a eles, fator de extrema relevância no tratamento dos dados enquanto objeto digital, quer seja dado governamental, dado de pesquisa ou qualquer outro conjunto de dados materializado como objeto digital.

### **3 SEMÂNTICA DOS DADOS**

A semântica é, em linhas gerais, entendida como a área responsável pela significação, muito embora o significado seja dependente da língua e de sua estrutura, que se alicerça pela sintaxe (ILARI; GERALDI, 1992).

No contexto das máquinas, a significação, ou seja, a semântica tem sido estudada sobretudo no contexto da Web Semântica, considerando a necessidade de que os dados tenham uma estrutura que dê a eles significado e, com isso possibilite que sejam compreendidos por máquinas (BERNERS-LEE; HENDLER; LASSILA, 2001), a partir do uso de linguagens e ontologias.

A Web Semântica, apresentada por Berners-Lee, Hendler e Lassila (2001), previa a estruturação dos dados disponíveis na Web de modo a expressar significado e favorecer a “compreensão” e a navegação dos agentes computacionais para a realização de tarefas, em uma forma colaborativa de ação entre humanos e máquinas.

Nesse contexto, os humanos estruturam os dados de forma semântica para que as máquinas possam compreendê-los e acioná-los para processamentos que permitam otimizar as requisições feitas por humanos. Isso gera a necessidade de preparar os dados disponíveis na Web para que sejam legíveis por humanos e por máquinas de forma indiscriminada, utilizando interfaces e linguagens apropriadas.

A Web Semântica, cujo surgimento se relaciona com o advento e a expansão da própria Web, tem um contexto diferente na atualidade, em que a Web tem se tornado o

ambiente padrão de armazenamento, disponibilização e compartilhamento de dados – a exemplo dos softwares e aplicativos que são desenvolvidos para Web e do armazenamento em nuvem, já bastante comuns.

Esse contexto possibilita expandir os preceitos da Web Semântica para o tratamento de dados, considerando que o ambiente de armazenamento, disponibilização, localização e acesso é indiscutivelmente a Web.

Os avanços da Web Semântica e suas tecnologias, tanto conceituais quanto implementáveis, possibilitam entendê-la como área imprescindível ao tratamento dos dados para que sejam compreensíveis e acionáveis por máquinas, o que se aplica a diferentes contextos. Dentre eles, neste estudo, destaca-se o reuso de dados, metadados, objetos digitais e ambientes informacionais digitais, considerando a necessidade de que sejam localizáveis, acessíveis, interoperáveis e reutilizáveis, de acordo com os Princípios FAIR.

Inclusive, pode-se verificar um desafio apresentado por Berners-Lee, Hendler e Lassila (2001, tradução nossa) à Web Semântica que se aplica ao FAIR: “[...] fornecer uma linguagem que expresse dados e regras para raciocinar sobre os dados e que permita que as regras de qualquer sistema de representação de conhecimento existente sejam exportadas para a Web”. A estrutura da Web Semântica possibilita que haja compreensão de dados por máquinas e seus relacionamentos, de forma enriquecida.

Cumprido destacar que as tecnologias da Web Semântica estão em expansão e que são atreladas às tecnologias computacionais disponíveis, de forma que especificar uma ou um conjunto delas para o FAIR pode tornar o tratamento temporal. Desta forma, é importante enfatizar que o desenvolvimento das tecnologias trará aprimoramento ao tratamento dos dados, metadados, objetos digitais e ambientes informacionais digitais FAIR, e que a comunidade precisa acompanhá-los para assegurar que se mantenham FAIR. Tal afirmação elucida a necessidade de adotar tecnologias emergentes, abertas e que tenham foco no desenvolvimento coletivo.

Assim, ao entender as premissas dos princípios FAIR, em especial de tornar os dados localizáveis, acessíveis, interoperáveis e reutilizáveis com nenhuma ou mínima intervenção humana, verifica-se que há uma grande ligação com as propostas da Web Semântica e suas tecnologias. Pode-se, inclusive, tratar dessas tecnologias como ferramentas que buscam ampliar o nível de semântica dos dados.

Torino (2023, p. 231) afirma que “É por meio da semântica dos dados que os dados armazenados em diferentes fontes serão compatibilizados para o modelo de dados utilizado, o que requer que haja documentação acionável por humanos e máquinas visando estabelecer a compreensão”.

Ademais, Coneglian (2020, p. 104), aponta que a semântica “[...] aprofunda as relações entre as palavras, além de apontar os seus significados”. Assim, “Quando pensado sob a perspectiva humana, a semântica está na interpretação do sentido dos termos, de forma individual e relacionada a outros termos”. Por outro lado, “No nível de sistema, o processamento semântico irá determinar os possíveis significados de uma determinada frase ou conjunto de frases, focando principalmente na interação entre os termos dentro da sentença” (CONEGLIAN, 2020, p. 104).

De acordo com os apontamentos dos autores supracitados, há uma relação intrínseca entre a capacidade de ser lido por máquinas e a semântica dos dados e objetos digitais. Dessa forma, a discussão acerca da semântica dos dados e das ferramentas semânticas no âmbito do FAIR se mostra como um caminho para efetivar a aplicação de tais Princípios em dados, metadados, objetos digitais e suas infraestruturas de armazenamento e disponibilização.

#### **4 DISCUSSÃO**

Nos Princípios FAIR (WILKINSON *et al.*, 2016) algumas das facetas indicam a necessidade de uso de uma linguagem comum que possibilite a compreensão dos dados por máquinas. Jacobsen *et al.* (2020) apontam que os Princípios FAIR podem ser, por vezes, mal interpretados, e destacam esse fato relacionado especificamente a dois conceitos, com destaque à “capacidade de ação da máquina”, cuja relação é direta aos objetivos dos Princípios FAIR.

Nesse sentido, os autores apontam que a interoperabilidade – um dos princípios fundamentais – se refere à possibilidade de mesclar e/ou unificar dados de recursos digitais relacionados, bem como à capacidade de identificar, processar e relacionar dados de forma automática, o que “[...] requer que o significado (semântica) de cada recurso participante – sejam eles serviços de dados e/ou serviços – seja claro” (JACOBSEN *et al.*, 2020, tradução nossa).

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

---

No FAIR, Interoperável considera que “Os dados geralmente precisam ser integrados com outros dados. Além disso, os dados precisam interagir com aplicativos ou fluxos de trabalho para análise, armazenamento e processamento” (GOFAIR, 2023, tradução nossa), e é composto por três facetas: I1. (Meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento; I2. (Meta)dados usam vocabulários que seguem aos princípios FAIR; I3. (Meta)dados incluem referências qualificadas a outros (meta)dados.

Jacobsen *et al.* (2020) afirmam que os formatos de troca de dados estabelecidos por comunidades funcionam apenas neste escopo, tornando complexa a tarefa de interoperar e integrar dados, tanto para humanos quanto para máquinas, sendo a última, parte dos objetivos dos Princípios FAIR.

É evidente que o processo de localizar, acessar, interoperar e, finalmente, reusar dados requer compreensão por máquinas de modo inequívoco, o que pode ocorrer a partir de um entendimento comum decorrente de uma linguagem amplamente compreendida por máquinas. Nesse ponto, a interoperabilidade requerida no FAIR encontra a semântica dos dados.

Com o intuito de aprofundar a discussão acerca de como os princípios FAIR se relacionam com a semântica para a compreensão de dados por máquinas, foram identificadas as facetas que estão mais vinculadas a essa característica. Assim, apresenta-se, no Quadro 1, como cada faceta apoia, de forma efetiva, a ampliação do nível de semântica dos dados para possibilitar a compreensão por máquinas.

**Quadro 1 - Princípios FAIR e a Compreensão de Dados por Máquinas**

Princípio	Faceta	Semântica para compreensão por máquinas
Localizável	Identificadores persistentes	A utilização de identificadores persistentes é essencial para que os agentes computacionais obtenham informações adicionais do contexto e do significado das informações, além de permitir a obtenção de outros dados. Também permite que a recuperação desses dados no futuro aconteça de forma eficiente. Os identificadores persistentes devem ser utilizados como referência duradoura para os dados, metadados e objetos digitais na web. Para tanto, é necessário que qualquer alteração realizada no registro seja registrada no Identificador Persistente (PID). Destaca-se que, caso haja alteração de URL no registro fonte, o PID pode apontar página não localizada ou para conteúdo inadequado, o que reforça a necessidade de manter o registro do PID atualizado.
	Metadados	Ao apresentar metadados completos, descritos de forma exhaustiva, a

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB  
Aracaju-SE – 06 a 10 de novembro de 2023**

	enriquecidos	<p>capacidade da máquina compreender com um nível elevado de semântica formal aumenta significativamente. Dessa forma, a máquina é capaz de ampliar a quantidade de dados disponíveis, o que pode aumentar a visão que o agente possui de um determinado conjunto de dados, além de favorecer o processo de localização dos dados, metadados e objetos digitais.</p> <p>É importante destacar que tanto os metadados representacionais quanto os metadados intrínsecos ao objeto digital devem ser enriquecidos, o que auxilia, além da capacidade de localização, a interoperabilidade por meio de processamento automático de dados por máquinas.</p>
	Identificadores persistentes em metadados	<p>A utilização de identificadores persistentes em metadados amplia a capacidade da máquina de compreender qual é o contexto na qual aqueles metadados foram utilizados. Ao possuir PID, a ligação existente entre os objetos digitais, seus dados e metadados é formalmente explicitada, facilitando o processo da compreensão semântica de tais conjuntos.</p>
Interoperável	Vocabulários	<p>A utilização de vocabulários é que possibilita a compreensão dos dados por humanos e máquinas sem a necessidade de mediação, como o uso de tradutores ou mapeamentos. Essa compreensão é possível por meio da adoção de uma linguagem que expresse de forma precisa sua sintaxe e semântica.</p> <p>Tecnologias que possuem um nível maior de compreensão semântica, como o RDF, favorecem a interoperabilidade dos dados em primeira instância, mas também permitem que os dados disponíveis tenham estruturas que otimizem a capacidade de compreender o significado dos dados e possibilitem que agentes computacionais distintos possam trocar informações e interoperar/integrar dados e objetos digitais.</p>
	Vocabulários são FAIR	<p>Os vocabulários precisam apresentar características mencionadas, como possuir interligações, ser enriquecidos, apresentar identificadores persistentes, entre outras características, para que haja um nível de semântica no processo de interoperabilidade dos dados. Destaca-se a necessidade de que tais vocabulários apresentem características semânticas e possibilitem a compreensão e processamento por máquinas para que estes favoreçam que o objetivo digital e os seus metadados também tenham tais características.</p>
	Metadados ligados	<p>No contexto dos princípios do <i>Linked Data</i>, é essencial que os padrões de metadados utilizados, bem como os próprios metadados, apresentem ligações com outros conjuntos e outros padrões. Tal característica favorece que os agentes computacionais explorem os dados e obtenham uma maior compreensão da semântica do objeto e de seus metadados.</p> <p>Nesse contexto, o uso de metadados ligados efetiva o relacionamento entre objetos digitais, possibilitando, inclusive, que haja compreensão do processo de comunicação científica ampliada.</p>
Reutilizável	Metadados possuem múltiplos atributos	<p>No contexto da reutilização dos dados, é essencial que os metadados sejam completos e abundantes. Tal característica está fortemente vinculada à capacidade da máquina compreender o contexto dos dados, em especial quando tais metadados forem ricos e interligados com outras bases, pois permite a compreensão do contexto e da proveniência dos dados, possibilitando averiguar sua utilidade ao propósito de (re)uso. Quanto mais exaustiva for a representação dos objetos digitais, maior a possibilidade de localização e, por consequência, de reuso.</p>

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

---

	Padrões da comunidade	A utilização de padrões de comunidade favorece significativamente a compreensão por máquinas, em especial, pois as máquinas necessitam ser programadas para entender o que os diversos vocabulários, padrões e metadados significam. Desta forma, ao utilizar padrões estabelecidos, há uma maior probabilidade de uma máquina compreender o que aquele dado significa e poder, assim, classificá-lo e utilizá-lo de forma mais eficiente.
--	-----------------------	--

**Fonte:** Elaborado pelos autores.

A partir dos elementos apresentados no Quadro 1, destaca-se que os princípios FAIR buscam favorecer o reuso – por máquinas – de dados, metadados e objetos digitais, o que requer a compreensão por máquinas com um nível de semântica elevado. No entanto, há a necessidade de se demonstrar como isso ocorre, o que é apresentado na terceira coluna do quadro.

A discussão apresentada no Quadro 1 demonstra a necessidade de buscar, nas discussões e aplicações dos Princípios FAIR, tornar os dados, metadados e objetos digitais aptos a serem compreendidos por máquinas.

Jacobsen *et al.* (2020) alertam que os criadores de objetos digitais devem utilizar uma linguagem com mecanismo que permita a interpretação por máquinas, a fim de possibilitar o consumo e a compreensão básica dos dados. Considerando ainda que os Princípios FAIR buscam aplicações a agentes genéricos, é imprescindível que os dados sejam interpretados de maneira inequívoca por agentes humanos e máquinas, o que requer o uso de formatos genéricos e amplos. Nesse sentido, Jacobsen *et al.* (2020) destacam o uso do RDF, um formato compreensível por máquinas e amplamente utilizado na Web e da Web Semântica, a partir de uma recomendação da *World Wide Web Consortium* (W3C), cuja primeira versão data de 1999.

Mons *et al.* (2017) afirmam que o RDF, as ontologias formais e os vocabulários são eficazes para a interoperabilidade e o compartilhamento de dados e metadados, mas alertam que são tecnologias em uso na atualidade, que podem ser úteis desde que aplicadas adequadamente, mas que não podem ser consideradas as únicas.

De igual maneira, a interoperabilidade presente no FAIR necessita de vocabulários que auxiliem na compreensão inequívoca dos dados. Estão aqui contidos tesouros, modelos de dados e ontologias que auxiliam na distinção entre os termos utilizados por máquinas.

Destaca-se que, ainda que os dados sejam localizáveis e acessíveis, é necessário que as condições estabelecidas para que sejam interoperáveis sejam cumpridas para que possam

ser reutilizáveis. Esclarece-se assim que, apesar das facetas dos Princípios FAIR serem independentes e que haja a possibilidade de implementá-las de forma escalonada, sua interdependência leva à necessidade de que todos os princípios sejam cumpridos para que os dados, os metadados, os objetos digitais e as infraestruturas de armazenamento possibilitem que sejam localizáveis, acessíveis, interoperáveis e reutilizáveis. Além disso, a compreensão de dados por máquinas é que possibilita que os dados sejam conectados e enriquecidos, seguindo, por exemplo a *Linked Data Platform (LDP)*<sup>2</sup> – que estabelece regras para realizar operações e modelar interações entre recursos da Web e fornecer uma arquitetura para *Linked Data*.

No que tange à aplicação de licenças de uso, quer seja relacionada ao objeto digital ou aos metadados, é importante que haja compreensão das legislações aplicáveis para que o estabelecido surta efeitos legais. Ademais, é necessário aos humanos conhecer profundamente as especificações e requisitos para o licenciamento para adotá-los, e buscar condições para que tais licenças sejam aplicadas adequadamente e possam ser compreensíveis por humanos e máquinas. Nesse contexto, cumpre destacar os limites jurisdicionais das leis aplicáveis, bem como a forma de estruturação deste dado nos metadados e objeto digital.

De acordo com os precursores dos Princípios FAIR (JACOBSEN *et al.*, 2020, tradução nossa), “A interoperabilidade (indiscutivelmente o aspecto mais desafiador do FAIR) é obviamente muito dependente da convergência de soluções e padrões [...]”, mas que há a necessidade de orientar e auxiliar a comunidade no que tange à sua implementação. Ao que esclarecem: “Ninguém nunca disse que o FAIR era fácil, mas temos que passar pela dificuldade de tornar nossos recursos FAIR para possibilitar uma ciência melhor juntos” (JACOBSEN *et al.*, 2020, tradução nossa), e sabemos que isso requer que os dados sejam compreensíveis por máquinas.

Considerando as afirmações de Jacobsen *et al.* (2020, tradução nossa), a reutilização de dados requer convergência, tanto de intenções em seguir os Princípios FAIR e suas facetas, mas também de convergência tecnológica. Reforçamos que isso passa pela compreensão conceitual dos Princípios e facetas FAIR por humanos, quer sejam os responsáveis pelos dados e metadados ou ainda os detentores dos ciclos de vida de dados,

---

<sup>2</sup> Disponível em: <https://www.w3.org/TR/ldp/>.

para que haja condições de aplicação dos referidos princípios do ponto de vista tecnológico, de forma a favorecer a compreensão dos dados por máquinas.

Destaca-se, contudo, que os Princípios FAIR objetivam poupar tempo e esforços humanos na localização, acesso, interoperabilidade de dados, metadados e objetos digitais, ao incluir a colaboração das máquinas nesses processos. Para Mons *et al.* (2017), otimizar o tempo é o primeiro retorno proveniente da adequada gestão de dados: “Para atender a essa redução de custo potencialmente enorme, (meta)dados e serviços compatíveis com FAIR devem ser acionáveis por máquinas sem supervisão humana sempre e sempre que possível.” (MONS *et al.*, 2017, tradução nossa).

Tais discussões apontam que a característica escalonável e independente dos Princípios FAIR possibilita que dados, metadados, objetos digitais e infraestruturas tenham diferentes “graus” de localização, acesso, interoperabilidade e reuso. Contudo, apontam que, caso haja algum impedimento em qualquer dos princípios ou facetas, podem deixar de receber o benefício gerado pelo princípio. Assim, não se pode esperar que sejam reutilizados se não forem sequer localizáveis; caso o sejam, ainda dependem das condições de acesso; e se estas forem possíveis, precisam das características da interoperabilidade.

Diante do exposto, vale reforçar que caso os dados, os metadados, os objetos digitais e as infraestruturas não sejam FAIR e, portanto, não possam ser reutilizáveis, elas não devem ser consideradas “inutilizáveis”, uma vez que podem estar cumprindo adequadamente seu papel (por exemplo, ao não permitir acesso a dados pessoais, confidenciais e/ou sigilosos, tornando-os indisponíveis para reuso). Porém, Mons *et al.* (2017) apontam que há sim dados que podem ser considerados inúteis, citando como exemplo os que são publicados em links ou como arquivos suplementares de artigos, sem que haja um *Persistent identifier* (PID) exclusivo e associado a metadados ricos, legíveis por máquinas.

Outro aspecto importante que merece destaque é que a responsabilidade por implementar as facetas FAIR também é escalonável e, além disso, compartilhada.

Por um lado, o responsável pela criação do dado, metadado e/ou objeto digital necessita compreender e aplicar as facetas relacionadas, principalmente, a F2, F3, I1, I2, I3, R1, R1.2. Sua ação está diretamente atrelada ao que está contido no objeto – como metadados e dados intrínsecos –, bem como à representação desse objeto por meio dos

dados utilizados como “valor” nos campos de metadados disponíveis no ambiente informacional digital de armazenamento.

Uma ação importante do responsável pela criação do objeto digital é a definição do ambiente informacional de armazenamento do objeto digital, uma vez que, após o armazenamento, as intervenções no objeto digital e seus dados representacionais são restritas às possibilitadas pelo detentor do ciclo de vida dos dados. Desta forma, há uma transferência de responsabilidade do indivíduo responsável inicialmente pelos dados para o detentor do ciclo de vida de dados do ambiente no qual os dados estão ou serão armazenados.

Apesar disso, cabe ao responsável pela criação do objeto digital e pelos dados representacionais a ele relacionado atuar para que os dados e objetos digitais possam ser localizados e, a partir do acesso possibilitado pelo ambiente informacional digital, serem interoperados e reutilizados facilmente por processamento de máquina.

Por outro lado, o detentor do ciclo de vida de armazenamento dos dados e objetos digitais (TORINO, 2023) necessita compreender e aplicar todas as facetas FAIR<sup>3</sup> (F1, F2, F3, F4, A1, A1.1, A1.2, A2, I1, I2, I3, R1, R1.2 e R1.3), uma vez que os metadados, dados e objetos digitais dependem do ambiente de armazenamento para serem localizáveis, acessíveis, interoperáveis, o que permite que sejam reutilizados – objetivo principal do FAIR.

No que tange ao reuso, a exaustividade dos metadados auxilia o consumidor dos dados a definir se os dados são adequados ao propósito, além de identificar as instruções necessárias para consumo e reuso.

As condições técnicas de reuso estão atreladas às facetas do FAIR, enquanto as condições legais são expressas também em metadados – de representação e intrínsecos ao objeto digital. Esta é uma questão que requer atenção, uma vez que, em um ambiente informacional digital, pode haver condições diferentes de uso da ferramenta, dos metadados e dos objetos digitais, o que requer que haja expressão adequada dos direitos e permissões. No que tange aos metadados, são responsáveis por indicar as condições de uso do objeto digital, que deve expressar como metadado intrínseco a licença apropriada. Além disso, deve ser expressa a condição de uso dos próprios metadados, o que pode viabilizar o reuso dos dados representacionais. Expressar adequadamente as licenças de uso do objeto,

---

<sup>3</sup> Os Princípios FAIR e suas facetas podem ser acessados na página do GoFAIR, disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 10 jul. 2023.

dos metadados e da ferramenta de software utilizada e assegurar sua compreensão por humanos e máquinas é um desafio para o reuso ético e legal de dados, metadados e objetos digitais.

Ademais, o reuso está condicionado ao detalhamento da proveniência dos dados, incluindo as fontes, as condições e as formas de coleta, tratamento e armazenamento, autoria, financiamento e demais dados que explicitem o que foi aplicado aos dados, para que se possa subsidiar a decisão pelo reuso. Para tanto, recomenda-se o uso da Família PROV<sup>4</sup>.

## **5 CONSIDERAÇÕES FINAIS**

As discussões acerca dos Princípios FAIR e da relevância da localização, acesso, interoperabilidade e reuso de dados, metadados, objetos digitais e ambientes informacionais digitais se avoluma e acelera.

No mesmo compasso, é imprescindível que haja compreensão dos sujeitos envolvidos – na criação de objetos digitais, na disponibilização de infraestruturas para armazenamento de tais objetos, bem como na definição dos metadados e na criação dos dados representacionais – da influência que suas ações exercem na capacidade de compreensão por máquinas.

A compreensão e o domínio dos Princípios FAIR e da semântica dos dados por tais sujeitos é que possibilitará que haja uma infraestrutura de dados e serviços FAIR, preparada por humanos para que seja processável automaticamente por máquinas em benefício das ações requeridas por humanos, com o uso adequado de técnicas e tecnologias necessárias para a compreensão e o processamento de dados, metadados e objetos digitais por máquinas.

Isso requer que haja atenção aos processos de criação de objetos digitais e ao seu tratamento, incluindo metadados intrínsecos com vocabulário adequado e a definição do ambiente informacional de armazenamento, considerando a relevância da adoção dos Princípios FAIR aplicáveis e do tratamento adequado dos dados.

É sabido que atender aos Princípios FAIR requer domínio e infraestrutura compatível e que há a necessidade de que competências sejam desenvolvidas para que se faça de modo adequado. Nesse contexto é que as discussões acerca da compreensão da semântica dos

---

<sup>4</sup> Disponível em: <https://www.w3.org/2001/sw/wiki/PROV>.

dados e do impacto que esse tratamento exerce sobre o processo de FAIRificação necessitam de espaço, sobretudo na área de Ciência da Informação.

Merecem destaque, nesse sentido, os metadados que estão envolvidos em algum grau com os Princípios FAIR e com a semântica dos dados. Aos metadados, está imputada a tarefa de conectar as demandas de consumidores com as infraestruturas de armazenamento que fornecem dados representacionais acerca dos objetos a eles vinculados.

## REFERÊNCIAS

BERNERS-LEE T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, New York, v. 5, may 2001. Disponível em:

[https://www.researchgate.net/publication/225070375\\_The\\_Semantic\\_Web\\_A\\_New\\_Form\\_of\\_Web\\_Content\\_That\\_is\\_Meaningful\\_to\\_Computers\\_Will\\_Unleash\\_a\\_Revolution\\_of\\_New\\_Possibilities](https://www.researchgate.net/publication/225070375_The_Semantic_Web_A_New_Form_of_Web_Content_That_is_Meaningful_to_Computers_Will_Unleash_a_Revolution_of_New_Possibilities). Acesso em: 07 fev. 2023.

CONEGLIAN, C. S. **Recuperação da informação com abordagem semântica utilizando linguagem natural**: a inteligência artificial na Ciência da Informação. 2020. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2020. Disponível em: <http://hdl.handle.net/11449/193051>. Acesso em: 12 maio 2023.

GOFAIR. **FAIR principles**. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 07 fev. 2023.

HEINRICHS, B.; POLITZE, M.; YAZDI, M. A. Evaluation of architectures for FAIR data management in a research data management use case. *In*: INTERNATIONAL CONFERENCE ON DATA SCIENCE, TECHNOLOGY AND APPLICATIONS, 11., 2022. **Proceedings of the [...]**. Lisbon: SciTePress, 2022. p. 476-483. DOI: <https://doi.org/10.5220/0011302700003269>. Disponível em: <https://www.scitepress.org/Link.aspx?doi=10.5220/0011302700003269>. Acesso em: 15 mar. 2023.

ILARI, R.; GERALDI, J. W. **Semântica**. São Paulo: Ática, 1992.

JACOBSEN, A.; MIRANDA AZEVEDO, R. de; JUTY, N. *et al.* FAIR Principles: interpretations and implementation considerations. **Data Intelligence**, v. 2, n. 1-2, p. 10-29, 2020. DOI: [https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024). Disponível em: <https://direct.mit.edu/dint/article/2/1-2/10/10017/FAIR-Principles-Interpretations-and-Implementation>. Acesso em: 15 mar. 2023.

MONS, B.; NEYLON, C.; VELTEROP, J. *et al.* Cloudy, increasingly FAIR; revisiting the FAIR Data Guiding Principles for the European Open Science Cloud. **Information Services & Use**, v. 37, n. 1, p. 49-56, mar. 2017. DOI: <https://doi.org/10.3233/ISU-170824>. Disponível em: <https://content.iospress.com/articles/information-services-and-use/isu824>. Acesso em: 06 fev. 2023.

**XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**  
**Aracaju-SE – 06 a 10 de novembro de 2023**

---

TORINO, E. **Arquitetura de dados no contexto da Ciência da Informação**. Universidade Estadual Paulista (Unesp), 2023. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2023. Disponível em: <http://hdl.handle.net/11449/238875>. Acesso em: 06 abr. 2023.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J. *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, 160018, mar. 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 06 fev. 2023.