

XIX
ENANCIB encontro nacional
de pesquisa em
ciência da informação

22-26
OUTUBRO
2018
LONDRINA/PR

// SUJEITO INFORMACIONAL E AS
PERSPECTIVAS ATUAIS EM CIÊNCIA
DA INFORMAÇÃO. //



XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018

GT-8 – Informação e Tecnologia

DESMISTIFICANDO OS PRINCÍPIOS FAIR: CONCEITOS, MÉTRICAS, TECNOLOGIAS E APLICAÇÕES INSERIDAS NO ECOSISTEMA DOS DADOS FAIR

Patricia Henning (Universidade Federal do Estado do Rio de Janeiro - UNIRIO)

Claudio José Silva Ribeiro (Universidade Federal do Estado do Rio de Janeiro - UNIRIO)

Luana Sales (Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT)

João Moreira (University of Twente - UT)

Luiz Olavo Bonino da Silva Santos (GO FAIR - International Support and Coordination Office)

***DEMISTIFYING THE FAIR PRINCIPLES: CONCEPTS, METRICS, TECHNOLOGIES AND
APPLICATIONS IN THE FAIR DATA ECOSYSTEM***

Modalidade da Apresentação: Comunicação Oral

Resumo: Os princípios FAIR, um acrônimo para Findable, Accessible, Interoperable e Reusable, estão presentes nas discussões e práticas contemporâneas da ciência de dados, desde o início de 2014, e tiveram sua aplicação consolidada em 2017, quando a Comissão Europeia passou a exigir a adoção de plano de gestão de dados, com base nesses princípios, por projetos financiados por seus recursos. Desde então, tais princípios passaram a ser norteadores da descoberta, do acesso, da interoperabilidade, do compartilhamento e da reutilização dos dados de pesquisa. No entanto, quando colocados em prática levantam dúvidas e imprecisões, gerando diferentes interpretações, o que dificulta sua aplicação. Por essa razão buscou-se elucidar seu entendimento, utilizando-se de conceitos esclarecedores, apresentando métricas específicas que medem o nível de FAIRness dos objetos digitais; disseminando a proposta do ecossistema dos dados FAIR e as tecnologias Data FAIRPort e FAIR Data Point. Apresentamos, ainda, estudos realizados na Europa que comprovam o impacto e o potencial desses princípios, em diferentes áreas disciplinares, dando destaque às necessidades e aos exemplos de aplicação. A abordagem metodológica desta pesquisa é de natureza bibliográfica e de caráter qualitativo dando ênfase na descrição conceitual dos elementos necessários para a compreensão do ecossistema FAIR, o que permitiu, neste estudo, trabalhar a fundamentação teórica

e conceitual, bem como o uso das práticas do FAIR em diferentes contextos e dimensões. As considerações finais corroboram as mudanças culturais e tecnológicas que vêm ocorrendo, no mundo informacional, relacionadas às novas práticas de gestão de dados e às interações e parcerias necessárias para a sua complexa implementação.

Palavras-chave: Princípios FAIR; Dados FAIR; Métricas; Ecossistema de dados; Tecnologia FAIR.

Abstract: The FAIR Principles, an acronym for Findable, Accessible, Interoperable and Reusable, has been present in the contemporary discussions and practices of data science since the beginning of 2014 and had its application consolidated in 2017, when the European Commission began to demand the adoption of data management plans, based on these principles, by projects funded by its resources. Since then, these principles have been guiding the discovery, access, interoperability, sharing and reuse of research data. However, when it is put into practice it raises doubts and inaccuracies, leading to different interpretations, which hinders its application. For this reason, in this paper we aim at contributing to improve this understanding by clarifying related concepts, presenting specific metrics that measure the level of FAIRness of the digital objects; motivating the FAIR Data ecosystem proposal and related technologies such as the FAIR Data Point. We also present studies carried out in Europe that demonstrate the impact and potential, of these principles, in different disciplinary areas, highlighting needs and examples of application. The methodological approach used is qualitative and bibliographical in nature, emphasizing the conceptual description of the elements needed to understand the FAIR Ecosystem, which allowed, in this study, to work on the theoretical and conceptual basis, as well as the use of FAIR practices in different contexts and dimensions. The final considerations corroborate the cultural and technological changes that have been taking place in the information world related to new data management practices and the interactions and partnerships necessary for their complex implementation.

Keywords: FAIR principles; FAIR data; Metrics; Data ecosystem; FAIR technology.

1 INTRODUÇÃO

As novas formas de fazer ciência, pautadas fortemente no compartilhamento e no reuso de dados de pesquisa, vêm colocando em evidência a necessidade de deixar para trás a ideia dos dados apenas como insumos intermediários das atividades científicas. Este novo fazer científico está postulando novos processos e métodos que elevam os dados para um patamar de destaque nas áreas das ciências exatas, sociais e humanas, na arte ou na literatura.

Assim, o crescente desejo pelos dados gerados pelas atividades científicas, incentivou, tanto por parte das instituições de pesquisa e da comunidade acadêmica, quanto pelas agências de fomento, um aumento do pleito por políticas organizacionais, para a gestão de pessoas e infraestrutura tecnológica propiciando uma maior sustentabilidade e análise desses novos recursos informacionais.

Dada a complexidade e a heterogeneidade dos dados de pesquisa, condicionadas tanto pelas idiossincrasias inerentes aos diferentes domínios disciplinares quanto pelas características intrínsecas dos objetos digitais, cumpre destacar a necessidade desses objetos manterem uma contextualização precisa para transmitirem informação e conhecimento ao

longo do tempo. Isto possibilitará a sua interpretação e reuso coerente por outros pesquisadores.

Estes fatores implicam na necessidade de uma gestão ativa para além do simples armazenamento e da disponibilização na web, demandando um sistema confiável para a publicação de dados baseada em uma potente infraestrutura digital. Neste sentido, o potencial de reuso de dados de pesquisa compartilhados está fortemente relacionado à adoção de melhores práticas na gestão, na estruturação dos dados para interoperabilidade, no assinalamento de metadados de qualidade, no licenciamento apropriado e na acessibilidade (SAYÃO; SALES, 2018).

Para o encaminhamento das atividades científicas, na atualidade, a gestão de dados de pesquisa em um ambiente distribuído e em rede é considerada fundamental e torna-se um desafio crescente para o mundo da pesquisa da Ciência da Informação e da Computação. Esta gestão deve promover a identificação de princípios teóricos que apoiem os processos de organização, significação e recuperação, dando coerência, com o desenvolvimento de normas e métricas de apoio. Tudo isto apoiado por plataformas tecnológicas que impulsionem a curadoria dos dados e viabilizem o compartilhamento de dados de forma eficaz, para que esses recursos atinjam todo seu potencial de reuso.

Neste contexto, os princípios FAIR, um acrônimo para *Findable, Accessible, Interoperable e Reusable*, apresentam-se como elementos norteadores ao processo de gestão de dados de pesquisa. Estes princípios vêm sendo amplamente difundidos por comunidades internacionais nos últimos anos e, recentemente chegando ao Brasil como mais uma vertente de estudos e aplicação por gestores de dados. As discussões têm sido encaminhadas com o intuito de proporcionar o grau máximo de reuso de dados científicos, a partir da adoção de padrões, metadados, vocabulários controlados, ontologias e identificadores persistentes que proporcionam significado preciso aos dados e aos demais objetos a eles vinculados.

O presente trabalho vem apresentar os princípios FAIR com o intuito de esclarecer o seu arcabouço conceitual e colocar em discussão, pela comunidade de Ciência da Informação, a aplicabilidade destes elementos.

No Brasil, os debates acerca deste tema ainda estão em caráter inicial, restritos a alguns grupos de investigação específicos de universidades e institutos de pesquisa localizados na Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Universidade de São Paulo (USP), Universidade Estadual Paulista (UNESP), Universidade Federal da Paraíba (UFPB),

Universidade Federal do Espírito Santo (UFES), Fundação Oswaldo Cruz (Fiocruz), Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) entre possíveis outros.

A abordagem metodológica da presente pesquisa é de natureza bibliográfica e de natureza qualitativa dando ênfase na descrição conceitual dos elementos necessários para a compreensão do denominado Ecossistema FAIR. Como conduta metodológica, a investigação parte de um estudo que define as características que tornam os dados FAIR. Apresenta os elementos que compõem o ecossistema FAIR e das métricas específicas que medem o nível de compatibilidade com os princípios, denominadas *FAIRness*. A partir disso, apresenta estudos sobre o tema e discute proposta de implementação dos princípios FAIR, por meio de ferramentas já desenvolvidas para auxiliar a compreensão e aplicação desses princípios.

2 CONCEITUAÇÃO DOS PRINCÍPIOS FAIR

A criação dos princípios FAIR é datada do início de 2014, fruto de uma conferência internacional denominada “*Jointly designing a data FAIRPORT*”¹, que reuniu especialistas de diversas áreas do conhecimento envolvidos com a utilização, o tratamento e o reuso dos dados de pesquisa no âmbito *e-Science*.

O resultado dessa conferência gerou os princípios FAIR, que vêm sendo adotados por instituições de pesquisa, universidades, editores, gestores de dados, infraestruturas de dados e agências financeiras de pesquisa. Esses princípios foram considerados o instrumento norteador para a descoberta, o acesso, a interoperabilidade global, o compartilhamento e a reutilização dos dados de pesquisa, pois se propõe estabelecer orientações estruturais para a publicação de recursos digitais, conjunto de dados, códigos e objetos de pesquisa de maneira FAIR (WILKINSON *et al.*, 2017).

O quadro 1 apresenta os princípios propostos por Wilkinson *et al.* (2017), bem como suas utilidades e aplicações.

Quadro 1: Guia dos Princípios FAIR.

Princípios	Como aplicar
F –Findable	Para serem encontrados

¹ Disponível em: <<https://www.lorentzcenter.nl/lc/web/2014/602/info.php3?wsid=602>> Acesso em: 06, Ago, 2018

F1. Os (meta)dados devem ter identificadores globais, persistentes e identificáveis	Metadados com identificador único - <i>Unique and persistent identifier</i> (PID) ou equivalente, para cada conjunto de dados Ex: DOI, ARK, RRID, PID
F2. Os dados devem ser descritos com metadados enriquecidos (impacta diretamente R1)	Conjuntos de dados devem possuir metadados ricos o suficiente para que, uma vez indexados para um mecanismo de busca, esses metadados possam ajudar o usuário dos dados a encontrá-los mesmo que não possua o seu identificador.
F3. Os metadados devem incluir claramente e explicitamente os identificadores dos dados que descrevem	Como não podemos prever que os dados e seus metadados estejam sempre juntos (por exemplo quando os metadados são indexados por um mecanismo de busca e, portanto, está em uma plataforma diferente dos dados), a associação entre eles deve ser feita pela inclusão do identificador dos dados pelos metadados.
F4. Os (meta)dados devem ser registrados ou indexados em recursos que ofereçam capacidades de busca	Para que contribuam para que os dados sejam encontrados, seus metadados devem ser indexados por mecanismos de busca que, por sua vez, permitem aos usuários encontrá-los por meio de elementos desses metadados.
A – Accessible	Para serem acessíveis
A1. (Meta) dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado	Com o identificador do conjunto de dados ou de seus metadados, o usuário deve poder recuperá-los por meio de um protocolo de comunicação padronizado.
A1.1 O protocolo deve ser aberto, gratuito e universalmente implementável	Independente de licenciamento dos dados e dos metadados, o protocolo de comunicação usado para dar acesso a eles deve ser aberto, gratuito e passível de ser implementado por qualquer interessado.
A1.2 O protocolo deve permitir procedimentos de autenticação e autorização, quando necessário	Quando necessário, dependendo das restrições ao acesso aos dados e/ou metadados, um mecanismo para autenticação e autorização para o acesso deve ser permitido pelo protocolo de comunicação.
A2. Metadados devem ser acessíveis, mesmo quando os dados não estão mais disponíveis	É preciso existir um conjunto de estratégias de preservação para os dados e metadados. Minimamente, os metadados devem ser sempre acessíveis, possibilitando a criação de índices para conjuntos de dados atuais e passados.
I – Interoperable	Para serem interoperáveis
I1. (Meta) dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento	Para que possamos representar dados e metadados devemos utilizar linguagens de representação do conhecimento que sejam formais, acessíveis e amplamente aplicáveis. Por exemplo, RDF, XML, DICOM, etc.
I2. (Meta) dados devem usar vocabulários que seguem os princípios FAIR	Como os dados e metadados devem possuir referências aos vocabulários que contenham os conceitos utilizados, devemos garantir que sejam utilizados vocabulários que também sigam os princípios FAIR .
I3. (Meta) dados devem incluir referências qualificadas para outros (Meta) dados	Referenciar os conjuntos de dados devidamente, possibilitando que conjuntos de dados gerados, a partir de outros conjuntos de dados, sejam ligados. Possibilitar a

	ligação semântica entre eles por meio dos padrões adotados.
R – Reusable	Para serem reutilizáveis
R1. (Meta) dados são descritos com uma pluralidade de atributos precisos e relevantes.	Prover metadados que permitam aos potenciais usuários avaliarem se é possível reusar os dados ou metadados e também, se eles se adequam às suas necessidades.
R1.1. (Meta) dados devem ser disponibilizados com licenças de uso claras e acessíveis	É fundamental que o responsável pelos dados e metadados defina explicitamente quem pode ter acesso a eles, para que e sob quais condições. Essas informações são definidas por meio de suas licenças de uso.
R1.2. (Meta) dados devem estar associados à sua proveniência	Especificando a proveniência (linhagem) dos dados é importante não só para que o usuário possa avaliar a utilidade dos dados ou metadados, mas também para que possamos atribuir o devido crédito a quem produziu, manteve ou editou esses dados. Dentre as informações relevantes à proveniência destaca-se: (a) A linhagem dos dados, ou seja, o processo de obtenção do dado (gerado ou coletado). (b) Particularidades ou limitações sobre os dados que outros usuários devem conhecer. (c) Data da geração do conjunto de dados, condições de laboratório, quem preparou os dados, configurações de parâmetros, nome e versão do software utilizado. (d) Explicitar se são dados brutos ou processados. (e) A versão dos dados arquivados e/ou reutilizados deve ser claramente especificada e documentada.
R1.3. (Meta) dados devem estar alinhados com padrões relevantes ao seu domínio	Atender os padrões específicos da comunidade da área. Atender às boas práticas de arquivamento e ao compartilhamento do campo de pesquisa específico.

Fonte: WILKINSON *et.al*, 2016; traduzido e adaptado pelos autores.

Percebe-se que cada letra do FAIR endereça para um conjunto de especificações que demandam domínios de conhecimentos técnicos e estruturais distintos. Tais princípios devem ser entendidos não como um padrão, mas, como “[...] um guia de preparação dos dados de pesquisa para o reuso, sob condições claramente descritas por pessoas e máquinas”² (SURF, 2018, p.4, tradução dos autores).

Mons (2018, p.3, 57) convalida esta afirmação quando observa que esses princípios não têm o objetivo de constituir padrões ou especificar formatos ou tecnologias, mas, sim, contextualizar e apontar para direção de maior utilidade e melhores serviços de dados, oferecendo suporte à sua reutilização e, assim, facilitar a escolha de quais padrões podem ser utilizados para esses fins. Embora inspirados no contexto da ciência aberta, não tratam de questões morais e éticas relativas à sua abertura. Isso significa que para ser FAIR, não

² “[...] The principles serve as a guideline for preparing research data for reuse under clearly described conditions by both people and machines” (SURF, 2018, p.4.)

necessariamente os dados precisam estar abertos. “[...] O ‘A’ do FAIR representa ser acessível sob premissas bem definidas, enquanto as condições de reutilização são cobertas pelo requisito de ter uma licença legível por máquina, de acordo com o princípio R1.1” (MONS, 2018, p.57, tradução dos autores).³

Ainda assim, é possível considerar que esses princípios podem conduzir a diferentes interpretações e uso por áreas distintas, o que pode levar a um distanciamento dos significados fundamentais, pois muitas vezes não se coadunam com a proposta original (ALLEN; HARTLAND, 2018).

Dependendo da área, em alguns casos, é difícil aplicar todos os 15 elementos do FAIR, como por exemplo na área física, onde o grupo Nikhef’s,⁴ do *Dutch National Institute for Subatomic Physics*, encontrou uma enorme dificuldade em adotar o princípio (I.1) que trata do uso de linguagem formal para metadados, concluindo então que:

“[...] Esses princípios podem ser bem aplicados se forem trabalhados com conjuntos de dados comuns nas ciências da vida, por exemplo, conjuntos de genes. No entanto, não há como funcionar com nossos conjuntos de dados complexos - o risco de resultados falsos positivos seria extremamente alto.”⁵ (SURF, 2018, p.26).

Outro aspecto importante a destacar diz respeito ao processo de implantação. Para a ampla adoção desses princípios algumas iniciativas internacionais vêm elaborando métricas específicas que medem o grau de *FAIRness* dos objetos digitais que serão apresentadas na subseção 2.1.

2.1 Métricas Fair

As métricas FAIR devem definir, com certa precisão, um conjunto mensurável de propriedades e comportamentos que avaliam o nível de *FAIRness* dos objetos digitais. O termo *FAIRness*, significa o grau em que um recurso digital adere aos princípios FAIR. No entanto, segundo WILKINSON, *et al.*, (2017, p.1, tradução dos autores) tais princípios são considerados “[...] aspiracionais, na medida em que não definem estritamente como alcançar um estado de *FAIRness*, mas descrevem um conjunto de características, atributos e comportamentos que

³ “[...] The ‘A’ in FAIR stands for accessible under well-defined conditions, while reusability conditions are covered in the requirement to have a clear, machine-readable license as per the R of FAIR” (MONS, 2018, p. 57).

⁴ Disponível em: <<https://www.nikhef.nl/en/nikhef>>. Acesso em: 06 ago. 2018.

⁵ “[...] this principle could be applied well if you work with datasets that are common in the life sciences, for example gene sets. However, there is no way that will work with our complex datasets – the risk of false positive results alone would be enormously high.” (SURF, 2018, p.26).

moverão um recurso digital para mais perto desse objetivo.”⁶ Elas surgem fruto de discussões de diversas comunidades e grupos de interesse, que entendem que boas métricas devem ser acima de tudo, claras, realísticas, discriminadas, mensuráveis e universais. (WILKINSON, et al., 2017, p.3).

Embora já estabelecidos conceitualmente, os princípios são relativamente difíceis de serem diretamente utilizados porque não foram devidamente detalhados e explicados na sua publicação original (WILKINSON et al., 2016). Isto, consequentemente, dificulta a medição do nível de *FAIRness* dos objetos digitais. Para amenizar interpretações ambíguas sobre o entendimento dos princípios FAIR e monitorar o nível de *FAIRness* dos recursos digitais, algumas iniciativas vêm desenvolvendo, de forma colaborativa, métricas que descrevem características, atributos e comportamentos. Estas devem permitir avaliar de forma qualitativa e quantitativa o nível em que os recursos digitais atendem aos princípios dos dados FAIR. Segundo Hodson et al., (2018, p.55) esta é uma tarefa desafiadora tendo em vista que as métricas devem fornecer uma indicação clara do que está sendo medido e definir um processo reproduzível para atingir essas medidas.

Neste sentido, o autodenominado *FAIR Metrics Group*⁷ criou um *Template*⁸ que contém questões importantes a serem discutidas, relativas a cada uma das letras do FAIR e voltadas para a sua medição. Para a formulação destas questões partiu-se do seguinte pressuposto: o que é considerado FAIR em uma comunidade pode não ser para outra comunidade. Por esta razão, o grupo sugere que, além de um conjunto comum de métricas, outras que sejam de interesse específico devam ser geradas pelos membros de cada comunidade, respondendo, inicialmente, às perguntas de acordo com as características ou recursos de cada área, permitindo, ainda, a criação de novas perguntas ou métricas adicionais que poderão ser compartilhadas pela própria comunidade (FAIR METRICS, 2017).

Outra iniciativa desenvolvida pela OzNome, na Austrália, foi a ferramenta *5 Star Data Rating*⁹. Esta faz uma classificação baseada em cinco estrelas relacionadas aos cinco atributos:

⁶ “[...] The Principles are aspirational, in that they do not strictly define how to achieve a state of “FAIRness”, but rather they describe a continuum of features, attributes, and behaviors that will move a digital resource closer to that goal.” (WILKINSON, et al., 2017, p. 1).

⁷ Disponível em: <www.fairmetrics.org>. Acesso em: 06 ago. 2018.

⁸ Disponível em: <<https://www.biorxiv.org/content/early/2017/12/01/225490>>. Acesso em: 06 ago. 2018.

⁹ Disponível em: <<http://ozname.csiro.au/5star>>. Acesso em: 06 ago. 2018.

localizável, acessível, interoperável, reutilizável e confiável, possibilitando que os gestores realizem uma autoavaliação dos seus dados.

Por intermédio de perguntas específicas os dados são classificados de acordo com as cinco estrelas de qualidade, podendo identificar o grau de *FAIRness* dos seus objetos digitais. Desta forma é possível melhorar a coleta de dados e perceber como eles são acessados por outras pessoas, além de permitir melhorias no processo de coleta e publicação de dados (OZNOME, 2018).

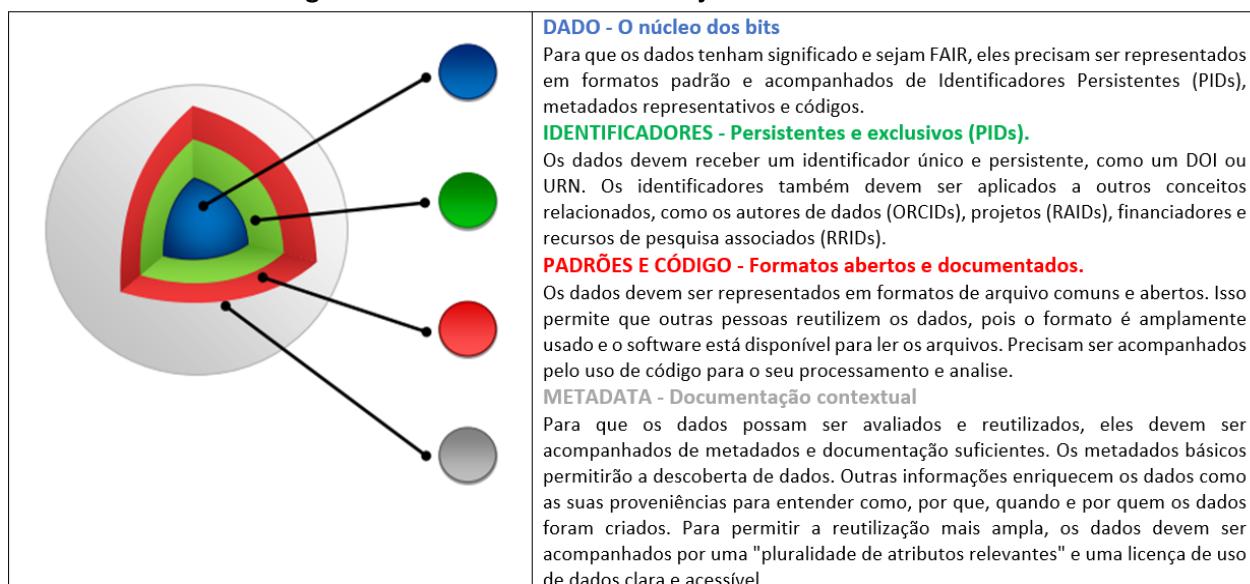
As métricas são recursos poderosos na adequação do comportamento individual e institucional. Nesse sentido, Hodson et al., (2018), sugerem que as escalas de avaliação FAIR sejam desenvolvidas como um certo nível de maturidade, de modo a incentivar os criadores de dados a tornarem seus recursos cada vez mais enriquecidos e reutilizáveis.

3 O ECOSISTEMA DOS DADOS FAIR

Para sustentar a implantação dos princípios FAIR, foi necessário definir os elementos centrais dos objetos de dados envolvidos e, a partir dessa definição, desenvolver um ecossistema que inclua os serviços necessários para criar, gerir e partilhar os objetos de maneira FAIR (HODSON et al., 2018, p.78).

A figura 1 exibe o modelo para objetos de dados que encapsula os elementos básicos necessários para que os dados sejam FAIR.

Figura 1: Modelo estrutural do Objeto de dados FAIR.

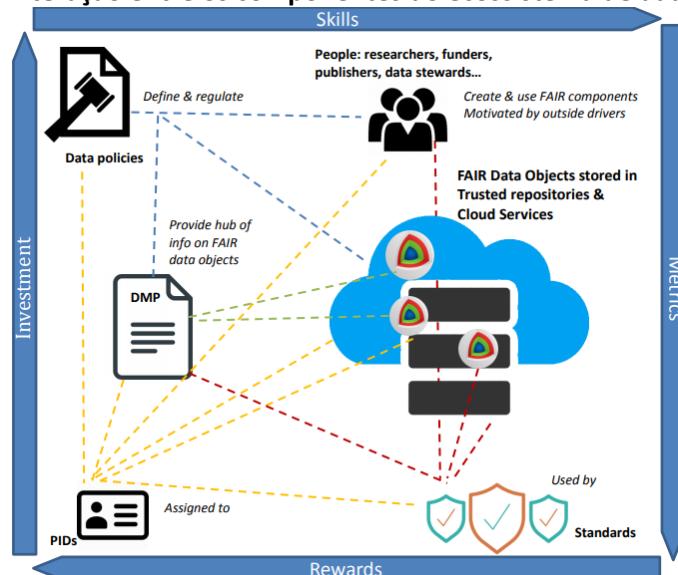


Fonte: HODSON et al., 2018, p.38, adaptado e traduzido pelos autores.

Para Hodson *et al.* (2018, p.38), dados são uma sequência *bitstream* ou binária que, para serem acessados e reutilizados, necessitam ser acrescidos de significados. Por essa razão, precisam ser padronizados e identificados com Identificadores Persistentes (PIDs) de diferentes naturezas e funções; devem, ainda, ser acrescidos de metadados enriquecidos com as suas decodificações. Eles recomendam que sejam documentados e que utilizem formatos abertos e padrões, sendo, ainda, importante a definição clara das licenças de uso que autorizam o nível de reutilização, possibilitando a interoperabilidade entre eles.

A proposta de um Ecossistema de dados FAIR compreende, no mínimo, cinco componentes essenciais: políticas, planos de gestão de dados, identificadores, padrões e repositórios. A figura 2 apresenta os cinco componentes e o fluxo de trabalho entre eles. (HODSON *et al.*, 2018, p.21).

Figura 2: Interação entre os componentes do ecossistema de dados FAIR.



Fonte: HODSON *et al.*, 2018, p.39.

Observa-se que o fluxo de trabalho desse Ecossistema está amparado por habilidades, métricas, reconhecimento e investimentos. No entanto, são as políticas de dados, definidas por diferentes atores (*stakeholders*), que exercem um papel fundamental na regulação, na execução dos serviços e nos investimentos. Por sua vez, o plano de gestão de dados relaciona questões relevantes a serem contempladas nas boas práticas do gerenciamento e curadoria dos dados, além de fazer o link com os componentes FAIR. Devem incluir ainda a gestão do *software* e outros materiais de pesquisa. Já os Identificadores Persistentes, de diferentes tipos e funções, devem ser atribuídos em várias partes do ecossistema, fornecendo, assim, maior contexto aos dados e possibilitando a interligação entre eles. Os padrões de metadados, os

vocabulários controlados, as ontologias e os tesouros servem para enriquecer a descrição de dados, desse modo, permitir a transferência e a troca de protocolos e a maior precisão na recuperação dos dados.

Percebe-se, também, que o ecossistema de dados FAIR é altamente distribuído, sendo os repositórios digitais confiáveis responsáveis pelos armazenamento, gerenciamento e preservação dos dados. (HODSON *et al.*, 2018, p.39).

E que, para o ecossistema de dados FAIR funcionar de forma satisfatória é importante o envolvimento de diferentes atores. HODSON *et al.* (2018, p. 66) destacam a participação da comunidade de pesquisa; dos serviços de dados; dos administradores de dados; dos organismos geradores de normas e padrões; dos fóruns globais, dos formuladores de políticas; das agências financiadoras de pesquisa; das instituições de ensino e pesquisa e dos editores. Todas estas entidades ocupam seu papel no projeto e implantação do ecossistema, conforme recomendações iniciais e ações fundamentais apresentadas no quadro 2.

Quadro 2: Recomendações do Plano de Ação para implantação do Ecossistema

RECOMENDAÇÕES INICIAIS	
Fase 1: Definir e aplicar o FAIR adequadamente	Rec. 1: Adotar as definições do FAIR Rec. 2: Estabelecer mandatos e limites de abertura Rec. 3: Elaborar um modelo para objetos de dados do FAIR
Fase 2: Desenvolver e apoiar um Ecossistema de dados FAIR sustentável	Rec. 4: Identificar os componentes do Ecossistema de dados FAIR Rec. 5: Propor e definir o financiamento sustentável para componentes do FAIR Rec. 6: Estabelecer estratégias e financiamentos baseado em evidências
Fase 3: Garantir que os dados sejam FAIR assim como os serviços certificados de apoio ao FAIR	Rec. 7: Definir o arcabouço da interoperabilidade entre componentes Rec. 8: Estabelecer os níveis FAIRness nas áreas transversais Rec. 9: Desenvolver métricas robustas para os dados FAIR Rec. 10: Utilizar Repositórios Digitais Confiáveis Rec. 11: Desenvolver métricas para avaliar e certificar serviços de dados
Fase 4: Incorporar uma cultura do FAIR na prática de pesquisa	Rec. 12: Estabelecer Gerenciamento de dados via Plano de Gestão de Dados Rec. 13: Profissionalizar a ciência de dados e as funções da administração de dados Rec. 14: Promover o reconhecimento e recompensar a adoção dos dados do FAIR e os processos para a administração dos dados
DESENVOLVIMENTO DE AÇÕES FUNDAMENTAIS NAS:	
Política de dados FAIR	Rec. 15: Promover a harmonização de políticas Rec. 16: Promover a ampla aplicação do FAIR
Cultura de dados FAIR	Rec. 17: Selecionar e priorizar Objetos de Dados do FAIR

	Rec. 18: Promover o uso e depósito em Repositórios Digitais Confiáveis Rec. 19: Incentivar a reutilização de dados Rec. 20: Apoiar na migração de dados existentes para serem FAIR Rec. 21: Usar informações contidas em Planos de Gerenciamento de Dados
Tecnologia para FAIR:	Rec. 22: Desenvolver componentes FAIR para atender às necessidades de pesquisa Rec. 23: Incentivar serviços para apoiar os dados do FAIR Rec. 24: Apoiar as tecnologias semânticas Rec. 25: Facilitar o processamento automatizado
Competências e Papéis para FAIR:	Rec. 26: Evoluir os perfis em Ciência de dados e habilidades de administração e gestão de dados Rec. 27: Promover Planos de capacitação e funções de intermediação Rec. 28: Evoluir constantemente as estruturas curriculares e Treinamento
Métricas FAIR	Rec. 29: Implementar métricas FAIR Rec. 30: Monitorar o FAIR Rec. 31: Acompanhar as citações de dados com novas métricas
Custos e investimentos em FAIR	Rec. 32: Avaliar o Custo do Gerenciamento de dados Rec. 33: Promover o uso de Modelos de negócios sustentáveis Rec. 34: Alavancar os serviços de dados existentes tendo em vista as diretrizes do <i>European Open Science Cloud (EOSC)</i>

Fonte: HODSON *et al.* 2018, p. 65, traduzido e adaptado pelos autores.

Para tornar os princípios do FAIR uma realidade, todas as recomendações e ações devem ser consideradas. São propostas de mudanças em âmbitos político, cultural e técnico, direcionadas essencialmente para as práticas de pesquisa.

Neste sentido, pode-se inferir que o Plano de Ação para os Dados FAIR é um bom exemplo de esforço de implantação. Este Plano está expresso no documento *Turning FAIR data into reality*¹⁰, criado por um grupo de especialistas europeus que reconhecem, acima de tudo, que tais princípios “[...] desempenham um papel essencial nos objetivos da ciência aberta para melhorar e acelerar a pesquisa científica, aumentar o engajamento da sociedade

¹⁰ Disponível em: <<https://doi.org/10.5281/zenodo.1285272>>. Acesso em: 06 ago. 2018.

e contribuir significativamente para o crescimento econômico”¹¹ (HODSON *et al.*, 2018, p.4, tradução dos autores).

Por fim, é importante ressaltar que a proposta deste Ecossistema encontra respaldo na Declaração do *European Open Science Cloud* (EOSC) que afirma: “[...] Somente uma mudança cultural considerável permitirá a reutilização, a longo prazo, para a ciência e para a inovação de dados criados por atividades de pesquisa: nenhuma disciplina, instituição ou país deve ser deixado para trás”¹² (EOSC DECLARATION, 2017, p.1, tradução dos autores).

E, conquanto seja uma iniciativa europeia, busca a troca de experiências com o objetivo de garantir que os dados científicos e os serviços desenvolvidos na Europa sejam interoperáveis internacionalmente.

4 ESTUDOS E NECESSIDADES

Estudo recente realizado com a comunidade acadêmica do Reino Unido, que participa do grupo de trabalho da FORCE 11¹³, expõe o impacto e o potencial dos princípios dos dados FAIR na prática, para tanto, considerando diferentes domínios disciplinares. Ainda que o FAIR esteja bem-estabelecido em termos conceituais, quando colocado em prática, gera dúvidas e incertezas. O relatório do *Joint Information Systems Committee* (JISC) aponta que:

[...] Baixos níveis de compreensão em torno da propriedade de dados pela comunidade de pesquisa; A visão sobre o que significa a adesão ao FAIR na prática e como evidenciá-los são inconsistentes; Forte apoio ao crescimento do conjunto de ferramentas e recursos disponíveis que reduzem a carga do gerenciamento de dados; Falta de consciência e prática na legibilidade de dados por máquina; Existe um volume significativo de dados que não são efetivamente gerenciados, particularmente dados brutos ou não processados e, a infraestrutura de suporte está ausente; Falta de boas ferramentas para suportar a captura de metadados na geração de dados¹⁴(ALLEN; HARTLAND, 2018, p. 6, traduzido pelos autores).

Os autores supramencionados identificaram divergências na percepção sobre os dados FAIR nas áreas disciplinares das Ciências Biológicas, Humanidades Digitais, Química e Ciências

¹¹ “[...] play an essential role in the objectives of Open Science to improve and accelerate scientific research, to increase the engagement of society, and to contribute significantly to economic growth.”

¹² “[...] Only a considerable cultural change will enable long-term reuse for science and for innovation of data created by research activities: no disciplines, institutions or countries must be left behind”.

¹³ Disponível em: <<https://www.force11.org>>. Acesso em: 06 ago. 2018.

¹⁴ “[...] There are low levels of understanding around data ownership in the research community; Views on what adherence to FAIR means in practice and how to evidence it are inconsistent; Strong support for growing the body of tools and resources available that reduced the burden of data management; Lack of awareness and practice in machine readability of data; There is a significant volume of data that is not effectively managed, particularly raw or unprocessed data, and the supporting infrastructure is absent; Lack of good tooling to support metadata capture at data generation”(ALLEN; HARTLAND, 2018, p.6).

Sociais, chegando-se à conclusão de que existe entre elas um forte distanciamento, pois verifica-se:

“[...]diversidade de tipos de dados; variação de ferramentas e sistemas correspondentes para suporte de gerenciamento de dados; opiniões divergentes sobre a “carga” do gerenciamento de dados; atitudes para compartilhar e perceber os benefícios individuais e influência da pressão de publicação e problemas de carreira”¹⁵ (ALLEN; HARTLAND, 2018, p. 6, traduzido pelos autores).

Se por um lado o uso adequado do FAIR pode ser considerado um incremento no âmbito do acesso aberto, oferecendo maior flexibilidade e clareza no nível de acesso aos diferentes dados e sua estratégia de licenciamento; por outro lado, existe uma desigualdade nas adesões ao FAIR, entre pesquisadores, disciplinas e universidades. Isto se dá decorrente de diferentes usos disciplinares das práticas de pesquisa e diferenças culturais relacionados à abertura e ao compartilhamento dos dados. Existe, ainda, uma falta de conformidade em torno do nível de *FAIRness*, além de problemas relacionados às questões de computabilidade do FAIR necessárias para a interoperabilidade, entendido como o elemento mais desafiador das práticas do FAIR (ALLEN; HARTLAND, 2108, p.23, SURF, 2018, p.2).

Percebe-se, no entanto, que a conscientização e o uso de infraestruturas apropriadas, bem como o investimento na curadoria de dados são requisitos importantes na implementação do FAIR, assim como o acesso a financiamentos e habilidades para reunir de forma coerente os processos das melhores práticas da gestão de dados.

Em outra pesquisa recente que avalia o nível de *FAIRness* de alguns repositórios Europeus, Dunning; Smaele; Bohmer (2017, p.187) identificam algumas fragilidades na adoção do FAIR. Estas fragilidades decorrem da existência de múltiplos atributos de avaliação que possuem objetivos próximos, além de serem questões abertas que podem ser interpretadas de maneira vaga. Outro ponto destacado por estes autores está ligado ao nível de profundidade das questões, pois ora parecem ser iminentemente de âmbito técnico enquanto outras são mais orientadas por políticas.

Observou-se ainda, que os repositórios analisados estão pouco alinhados aos princípios FAIR. Deduz-se que isto decorre de políticas básicas que necessitam ser ajustadas, pois estas precisam contemplar aspectos ligados a identificadores, metadados, ontologias,

¹⁵ “[...] *Diversity of data types; Variation in corresponding tools and systems to support data management; Divergent views on the “burden” of data management; Attitudes to sharing and perceived individual benefits; The influence of publication pressure and career issues*” (ALLEN; HARTLAND, 2018, p. 6).

licenciamento e protocolo. Satisfazer os princípios FAIR exige que as habilidades de diferentes atores estejam reunidas (DUNNING; SMAELE; BOHMER, 2017, p.188).

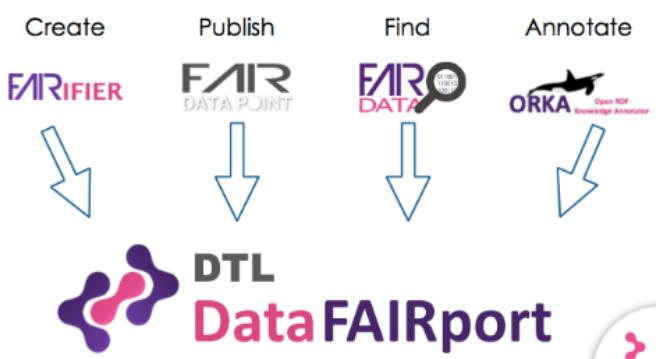
Por fim, em documento recente que apresenta seis práticas do FAIR em diferentes áreas ligadas a iniciativas holandesas, concluiu-se que os princípios FAIR são vistos como parte integrante de uma grande mudança cultural. Aspectos ligados às políticas são fundamentais e não devem estar apenas em conformidade com os princípios FAIR, mas sim, ser um reflexo das abordagens ligadas ao domínio do conhecimento em análise. Afirma-se que adotar o FAIR exige um grande esforço, mas vale a pena (SURF, 2018, p.44).

Pesquisas desta natureza nos levam a perceber a diversidade ainda existentes nas interpretações desses princípios e as dificuldades na sua adoção, o que nos impossibilitou fazer uma comparação dos resultados desses estudos e, o potencial de pesquisa e trabalho existentes para a sua adequada aplicação. Uma das formas de minimizar tais dificuldades está na proposição de ferramentas, que funcionem como artefatos para agilizar e auxiliar o processo de implantação.

5 FERRAMENTAS E EXEMPLO DE APLICAÇÃO

Envolvido desde o início do movimento FAIR, o Dutch Techcenter for Life Sciences (DTL), em conjunto com seus parceiros, desenvolveu um conjunto de soluções tecnológicas para apoiar a criação, publicação, busca e anotação de dados FAIR, conforme demonstra a figura 3.

Figura 3: Soluções FAIR do DTL



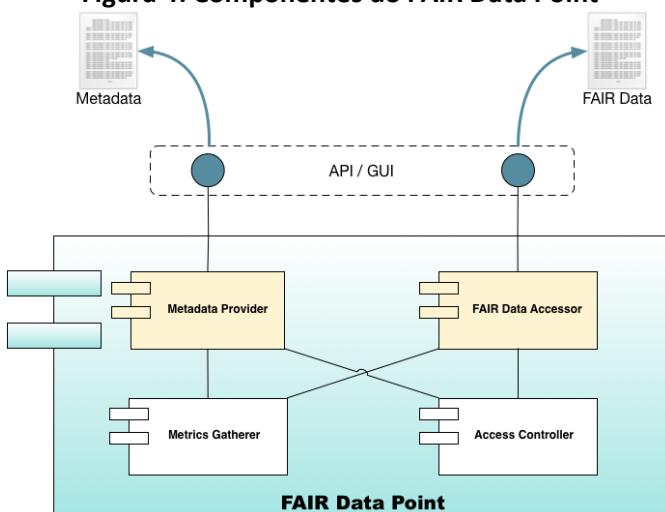
Fonte: SANTOS et.al. - 2016

As ferramentas ilustradas na parte superior da figura 3 foram projetadas e desenvolvidas para serem utilizadas de forma independente, mas também podem ser integradas em uma plataforma de interoperabilidade de dados chamada Data FAIRPort. Ela

utiliza uma camada de *software* para fazer interface com os conjuntos de dados, permitindo expor o conteúdo destes recursos (dados e metadados) em conformidade com os princípios FAIR, podendo ter seus serviços acessíveis tanto por humanos quanto por mecanismos automatizados em computadores. Essa ferramenta possibilita que os gestores de dados exibam seus conjuntos de dados em conformidade com tais princípios, permitindo que os usuários de dados encontrem informações, relativas aos conjuntos de dados, mediante seus metadados, sendo o acesso aos dados liberados conforme as licenças estabelecidas (SANTOS *et al.*, 2016).

O FAIR Data Point (FDP), um dos componentes do Data FAIRPort, é a ferramenta voltada para a publicação de metadados de forma FAIR, desenvolvida com tecnologias da *web semântica*¹⁶, especificamente *Linked Data*¹⁷ com a linguagem *Resource Description Framework* (RDF)¹⁸.

Figura 4: Componentes do FAIR Data Point



Fonte: SANTOS *et al.*, 2016, p.6.

A figura 4 apresenta os quatro componentes do FAIR Data Point:

- ✓ O *Metadata Provider*: responsável por fornecer acesso aos metadados estabelecidos.
- ✓ O *FAIR Data Accessor*: responsável por fornecer acesso ao conteúdo dos dados contidos no conjunto de dados. Na implementação do FDP são adotados metadados vinculados ao W3C.
- ✓ O *Metrics Gatherer*: responsável por monitorar vários aspectos do uso do FDP. Essas métricas podem ser usadas pelo proprietário do FDP para avaliar a carga em seu FDP e, assim, fornecer ao proprietário as informações necessárias para ajustar

¹⁶ Disponível em:<<https://www.w3.org/standards/semanticweb/>>. Acesso em: 06 ago. 2018.

¹⁷ Disponível em: <<https://www.w3.org/standards/semanticweb/data>>. Acesso em 06 ago. 2018.

¹⁸Disponível em:<<https://www.w3.org/TR/rdf-concepts/>>. Acesso em 06 ago. 2018.

a infraestrutura. Além disso, possibilita que os donos dos dados tenham informações sobre o quanto eles são acessados.

- ✓ O Access Controller: atua como um *gatekeeper* protegendo o acesso aos (meta) dados de acordo com as licenças dadas. Ademais, é responsável por garantir o acesso aos dados, assegurando o armazenamento e a transmissão dos dados por meio de criptografia, quando necessário, como no caso de dados sensíveis, que também contam com pseudonimização ou anonimização (SANTOS *et al.*, 2016, p.6).

É importante ressaltar que as especificações do FDP não trazem novas definições de metadados. Pelo contrário, as especificações orientam como utilizar padrões existentes e consolidados para descrever os metadados mínimos que um repositório deve seguir. Tais padrões, como as ontologias W3C DCAT¹⁹ e RE3data²⁰, descrevem a terminologia comum para a descrição de conjunto de dados e de repositórios de dados, respectivamente.

Hoje, o FDP está sob responsabilidade da iniciativa GO-FAIR, que visa implementar uma abordagem *bottom-up* para a *European Open Science Cloud* (EOSC). A iniciativa “GO-FAIR”²¹, apoiada inicialmente pelos governos holandês, alemão e francês, surge como a proposta de apoiar e coordenar as comunidades para a definição de formas comuns de concretizar os princípios FAIR. O objetivo é discutir e definir soluções voltadas para a implantação de dados e serviços FAIR. O GO-FAIR colabora com o *FAIR Metrics Group* no aprimoramento e aplicação das métricas FAIR apresentadas na subseção 2.1.

Um exemplo recente de aplicação do FDP em um repositório digital está presente no relato de Santos; Thompson e Staiger (2018) apresentado na última Conferência do EUDAT, realizada em janeiro de 2018, em Porto - Portugal, intitulada “*Putting the EOSC vision into practice*”. Nesta Conferência foram levantadas questões relacionadas a Dados Sensíveis, à Web Semântica, à Gestão de Dados de Pesquisa, à Interoperabilidade de dados e a Ontologias.

Em específico, tal relato inclui a experiência do piloto de uso do FDP no acesso real aos dados, em formato interoperável, por meio do alinhamento dos metadados do Repositório EUDAT/B2share²² aos metadados do FDP.

Nesse contexto é possível afirmar a necessidade de criação de uma metodologia rígida para alinhamento dos metadados, usados nos repositórios de dados, aos metadados do FDP.

¹⁹Disponível em: <<https://www.w3.org/TR/vocab-dcat/>>. Acesso em: 06 ago. 2018.

²⁰Disponível em: <<https://github.com/re3data/ontology>>. Acesso em: 06 ago. 2018.

²¹ Disponível em: <> Acesso em: 06 ago. 2018.

²² Disponível em:<<https://www.eudat.eu/services/b2share>>. Acesso em: 06 ago. 2018.

Além disso, questões recorrentes da pesquisa de tecnologias semânticas e ontologias formais surgiram. Por exemplo, como encontrar e avaliar ontologias existentes em relação a determinado domínio e como mensurar o uso de uma ontologia. Embora existam muitas pesquisas relacionadas ao tema, confirmou-se estas serem lacunas de pesquisa relacionada a implantação dos princípios de dados FAIR.

6 CONSIDERAÇÕES FINAIS

Apesar de ter seu lançamento datado em 2016, a adoção e implementação dos princípios FAIR, pelas instituições de pesquisa, só começaram a se consolidar efetivamente em 2017, quando a Comissão Europeia passou a exigir um plano de gestão de dados alinhados ao FAIR como condição para liberação de recursos. Desde então, propostas de soluções para implementação de tais princípios vêm sendo criadas na tentativa de amenizar problemas futuros.

O Ecossistema para os dados FAIR sugerido por Hodson *et al.*, (2018) é uma proposta de estrutura de implementação do FAIR na sua totalidade. Esta proposta preconiza a necessidade de adoção de políticas e serviços de informação de forma consistente, voltadas para os diferentes domínios; além de investimentos no desenvolvimento padrões e tecnologias apropriados que possam suportar os dados FAIR de forma harmonizada e coerente. A criação de culturas e práticas comuns que atendam às necessidades específicas de cada área também devem ser consideradas. Muitas soluções técnicas possuem características e objetivos comuns, mas precisam ser implementadas de maneira que façam sentido para a disciplina, de forma a atender às necessidades particulares da pesquisa que está sendo conduzida. Uma maior conscientização de todos envolvidos no âmbito das práticas científicas, desde pesquisadores, comunidades de pesquisa, universidade e agências financeiras, ainda é necessária.

O Data FAIRPort materializa uma série de componentes de *software*, representando tentativas tecnológicas, com amplitude de uso universal, podendo ser adotados em qualquer plataforma que enfrente problemas de gerenciamento de dados e interoperabilidade. Tal ferramenta desempenha um papel importante no processo de padronização da interoperabilidade entre os repositórios digitais sendo a experiência do piloto, usado no

B2share/EUDAT, considerada um caminho para solucionar tais dificuldades que se apresentam à ciência de dados.

As discussões apresentadas neste artigo são consideradas importantes para situar sobre o estado da arte das aplicações do FAIR em contextos e dimensões variados. Preconizam algumas ações em curso voltadas para o uso das práticas FAIR, assim como destacam algumas lacunas ainda a serem trabalhadas, configurando um enorme potencial a ser explorado direcionado à gestão de dados, não apenas no mundo científico e acadêmico, mas, também, no âmbito administrativo, governamental e empresarial.

Para finalizar, é possível supor que os princípios FAIR não sejam mais uma novidade para a comunidade científica, mas que a sua aplicação continua gerando problemas devido às diferentes interpretações a eles associadas. Colocá-los em prática ainda é uma tarefa árdua em função do seu alto grau de subjetividade e complexidade, uma questão que vem sendo tratada pelas métricas FAIR. Essas métricas são consideradas ferramentas úteis para medir o nível de *FAIRness* de um recurso digital e, portanto, garantir a aderência do recurso digital a esses princípios. Nesse contexto, o grande desafio está em respeitar as particularidades de cada área do conhecimento, sendo capaz de acomodar os requisitos de dados específicos de cada domínio, ao invés de utilizar as métricas como um padrão único.

No Brasil, a exigência de um plano de gestão de dados alinhados ao FAIR para liberação de recursos, já é uma realidade dentro da Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP).²³ Por outro lado, por perceber o potencial desses princípios, o Instituto Brasileiro de Informações em Ciência e Tecnológica (IBICT) iniciou as negociações de parceria com a iniciativa GO-FAIR, com o intuito de encaminhar a adoção desses princípios dando os primeiros passos para a inserção brasileira no processo de compartilhamento e reuso dos dados de pesquisa alinhados aos princípios FAIR.

REFERÊNCIAS

- ALLEN, R.; HARTLAND, D. “FAIR in practice Jisc report on the Findable Accessible Interoperable and Reusable Data Principles. 2018. Disponível em: <<https://zenodo.org/record/1245568#.W2BaOdlzZaQ>>. Acesso em: 31 jul. 2018.

²³ Disponível em: <<http://www.fapesp.br/gestaodedados/>>. Acesso em: 06 ago. 2018.

DUNNING, A.; SMAELE, M.; BÖHMER, J. Are the FAIR Data Principles Fair? **International Journal of Digital Curation**, v. 12, p. 177-195, 2017. Disponível em:
<<http://www.ijdc.net/article/view/567/493>>. Acesso em: 31 jul. 2018.

EUROPEAN COMMISSION. (EOSC) Commission high level expert group on the European open science cloud. Realizing the European Open Science Cloud, European Commission, 2016. Disponível em:
<https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf>. Acesso em: 31 jul. 2018.

EOSC DECLARATION, European Open Science Clau. New Research & Innovation Opportunities, 2017. Disponível em:
<https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemode=none> Acesso em: 31 jul. 2018.

FAIR METRICS GROUP. Disponível em: <<http://www.fairmetrics.org>>. Acesso em: 31 jul. 2018.

HODSON, S. *et al.* **Turning FAIR data into reality**. Interim report of the European Commission Expert Group on FAIR data. 2018. Disponível em:
<<https://doi.org/10.5281/zenodo.1285272>>. Acesso em: 31, jul. 2018.

MONS, B. **Data Stewardship For Open Science**. Implementing FAIR Principles. CRC Press Taylor & Francis Group. 2018. Disponível em:
<www.taylorfrancis.com/books/9781498753180>. Acesso em: 02 ago. 2018.

OZNOME. Disponível em: <<https://research.csiro.au/oznome/tools/oznome-5-star-data>>. Acesso em: 31 jul. 2018.

SANTOS, L. O. B. da S. *et al.* **FAIR Data Points Supporting Big Data Interoperability**. Enterprise Interoperability - Proceedings of the Workshops of the Eighth International Conference I-ESA 2016, p. 1. Disponível em:
<https://www.researchgate.net/publication/309468587_FAIR_Data_Points_Supporting_Big_Data_Interoperability>. Acesso em: 31 jul. 2018.

SANTOS, L. O. B. da S.; THOMPSON, M.; STAIGER, C. **An EUDAT-based FAIR Data Approach for Data Interoperability**, 2018. Disponível em: <<https://www.eudat.eu/communities/an-eudat-based-fair-data-approach-for-data-interoperability>>. Acesso em: 31 jul. 2018.

SAYÃO, L. F.; SALES, L. F. **A ciência invisível**: os dados da cauda longa da pesquisa científica. Disponível em: <<http://www.cnen.gov.br/images/CIN/PDFs/CURSO-Say%C3%A3o-DEZ2017.pdf>>. Acesso em: 31 jul. 2018.

SURF. FAIR DATA ADVANCED USE CASES: From Principles to Practice in the Netherlands. 2019. Disponível em: <<https://zenodo.org/record/1250535#.Wyl3RkiFNaQ>>. Acesso em: 31 jul. 2018.

WILKINSON, M. D. *et al.* A design framework and exemplar metrics for *FAIRness*. nov., 2017. Disponível em:
<<https://www.biorxiv.org/content/biorxiv/early/2017/11/29/225490.full.pdf>>. Acesso em: 31 jul. 2018.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, 2016. Disponível em:
<<https://www.nature.com/articles/sdata201618>>. Acesso em: 31 jul. 2018.