



# Desafios para a implementação de dados FAIR

Challenges for implementation of FAIR data

Letícia Guarany Bonetti, Universidade Federal de São Carlos – leticiagbonetti@gmail.com

Ana Carolina Simionato Arakaki, Universidade Federal de São Carlos – acsimionato@ufscar.br

## Eixo 6 - O mundo digital: apropriação e desafios

### 1 INTRODUÇÃO

A forma como a ciência se modificou durante os últimos anos foi uma constante, ligada às transformações cada vez mais rápidas ocorridas no cenário atual. As tecnologias disponíveis remodelam os meios de comunicação, como também o entendimento das boas práticas para o compartilhamento e disponibilização de recursos na *Web*, muitas vezes pautados pela abertura do acesso e pela democratização do conhecimento trazidas pela filosofia da Ciência Aberta, reconhecida como o quarto paradigma da ciência.

Nesse cenário, tem-se o movimento de dados abertos, ou *Open Data*, ligado à Ciência Aberta. Higman, Bangert e Jones (2019) explicam que os dados abertos são a prática de tornar os dados de pesquisa disponíveis publicamente, acessíveis e reutilizáveis com restrições mínimas. No cenário da Ciência Aberta, os dados abertos tornaram-se cada vez mais uma expectativa de financiadores e formuladores de políticas. Não há dúvidas de que as transformações relacionadas à revolução digital estão mudando radicalmente a vida cotidiana e a pesquisa (SILVEIRA *et al.*, 2021; COSTA; LEITE, 2018; HODSON *et al.*, 2018; CURTY; AVENTURIER, 2017).

Percebe-se que a ciência contemporânea restituiu aos dados de pesquisa seu protagonismo histórico. Para que então seja extraído o máximo benefício deles, evitando duplicações de esforços, agilizando o ciclo da ciência, gerando economia de recursos e garantindo a transparência (SHINKAI, 2011), é importante que os dados sejam compartilhados em conformidade com boas práticas internacionais.

Segundo Guandalini, Furnival e Arakaki (2019), as instituições de ensino e as agências financiadoras se atentam cada vez mais a essas práticas. Alguns exemplos são a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), a *National*



Science Foundation (NSF), a National Institutes of Health (NIH), e a Comissão Europeia, que exigem a gestão dos dados de pesquisa e o estabelecimento de políticas de acesso aberto.

Além disso, o número de órgãos responsáveis pelo financiamento das pesquisas que exigem que os dados sejam localizáveis, acessíveis e abertos também continua a crescer (BEZJAK *et al.*, 2018), indo ao encontro do que é defendido pelos princípios FAIR. Eles já são reconhecidos mundialmente como elementos-chave para boas práticas em todos os processos de gestão de dados, demonstrando sua relevância na ciência contemporânea (SALES *et al.*, 2020). Apesar de FAIR e aberto não serem sinônimos e não representarem a mesma coisa, muitas políticas internacionais já exigem dados FAIR e abertos.

Logo, a realidade em questão requer dos pesquisadores uma maior familiaridade com a temática. Mas, segundo Dias, Anjos e Rodrigues (2019), a adoção de práticas que permitam aos pesquisadores brasileiros acessar e reutilizar os dados de pesquisa ainda acontece de maneira tímida, em estágios iniciais. Assim justifica-se a importância do tema abordado neste trabalho, que tem o objetivo de apresentar alguns desafios para a implementação dos princípios FAIR em dados de pesquisa no cenário da Ciência Aberta, preenchendo essa lacuna.

## 2 FUNDAMENTAÇÃO TEÓRICA

Os princípios FAIR, um acrônimo para "*Findable*", "*Accessible*", "*Interoperable*" e "*Reusable*", em português: "Encontrável", "Acessível", "Interoperável" e "Reutilizável", são diretrizes conhecidas internacionalmente. Os 4 princípios foram estabelecidos como resultado da conferência internacional '*Jointly designing the data FAIRPORT*' realizada em janeiro de 2014 no *Lorentz Center* em Leiden, Holanda (HENNING *et al.*, 2021). A conferência reuniu especialistas de diversos países e de diferentes áreas de pesquisa para discutir o uso, tratamento e reutilização de dados.

A disseminação mais ampla dos princípios FAIR começou em março de 2016 com a publicação dos princípios no *Nature's Journal Scientific Data*, 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" (HENNING *et al.*, 2018, p.175).

Segundo Wilkinson *et al.* (2016), os princípios FAIR enfatizam o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente, além de apoiar sua reutilização por indivíduos. Isso porque o ecossistema digital existente em torno da publicação de dados impede a extração do benefício máximo dos investimentos. A razão não é a falta de tecnologia apropriada, mas sim a falta de um preparo desses dados, para que eles possam ser acionáveis por máquina. Mons *et al.* (2020) explicam que a principal contribuição que os princípios FAIR devem trazer para a pesquisa do século 21 é apoiar máquinas e humanos em seu trabalho colaborativo.

Sendo assim, os princípios "desempenham um papel essencial nos objetivos da Ciência Aberta para melhorar e acelerar a pesquisa científica, aumentar o engajamento da sociedade e contribuir significativamente para o crescimento econômico" (HODSON *et al.*, 2018, p.3, tradução nossa). A agenda em prol da Ciência Aberta tem a ambição de tornar o compartilhamento de dados FAIR o padrão para pesquisa científica, o que exige acordos globais para garantir uma interoperabilidade o mais ampla possível dos dados.

Mas a implementação dos princípios FAIR não é uma tarefa simples. Eles não são e nem exigem padrões ou ferramentas específicas, deixando certas escolhas para a comunidade que irá aplicá-los aos seus conjuntos de dados. Além disso, "os elementos dos princípios FAIR estão relacionados, mas são independentes e separáveis e podem ser implementados em qualquer combinação" (SILVA; SANTARÉM SEGUNDO; SILVA, 2018, p.5201). Não existe, portanto, um passo a passo de em que ordem ou de que forma os implementar para garantir níveis mais elevados de aderência, uma vez que o FAIR não é binário (FAIR/não FAIR), mas sim um espectro ao longo do qual vários graus de *FAIRness* são possíveis (HIGMAN; BANGERT; JONES, 2019).

Mesmo com essas lacunas, os princípios ganham um papel de destaque nas discussões internacionais. Vários eventos e publicações indicam a atualidade do tema dentro da rede internacional de pesquisa: o *FAIR Festival 2021* realizado pela GO FAIR, iniciativa apoiada pela Europa, Estados Unidos, Austrália e África; o *Open Repositories 2021*, que tem membros de instituições como *Indiana University*, *British*





*Library*, *DataCite*, *University of California* e do Instituto Brasileiro de Informação em Ciência e Tecnologia no comitê permanente; e o *The Dataverse Community Meeting*, organizado pelo *Dataverse Project* filiado à Universidade de Harvard. Os eventos citados demonstram a importância dos estudos e aplicações dos princípios FAIR mundialmente. Mas para que isso seja possível, é importante estar ciente de alguns desafios para alcançar maiores níveis de *FAIRness* (WILKINSON *et al.*, 2016).

### 3 MÉTODO DA PESQUISA

A pesquisa configura-se como exploratória, de abordagem qualitativa, com uma revisão de literatura para a fundamentação teórica e análise do tema. Para tanto, foi realizado levantamento bibliográfico nas bases: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), *Scientific Electronic Library Online* (SciELO), e no motor de busca Google Acadêmico, no período de janeiro a junho de 2022.

As expressões de busca utilizadas foram “dados de pesquisa”, “princípios FAIR” e “Ciência Aberta”, com os termos equivalentes em inglês. Elaborou-se sentenças de busca utilizando operadores lógicos, permitindo combinações entre os termos. As buscas retornaram um total de sete (BRAPCI), um (SciELO) e 13.500 (Google Acadêmico) resultados para a combinação dos dois primeiros termos em português. As mesmas buscas com os termos em inglês resultaram em dois (BRAPCI), quatro (SciELO) e 629.000 (Google Acadêmico) trabalhos. Para a seleção dos textos relevantes observou-se a pertinência da abordagem, a compatibilidade com o assunto tratado e o número de citações, buscando trabalhos que, além de contextualizarem o tema, citassem desafios quanto à adoção dos princípios FAIR para dados de pesquisa. Foi feito um recorte temporal a partir de 2016, ano em que houve a disseminação mais ampla dos princípios FAIR.

### 4 RESULTADOS E DISCUSSÕES

A implementação dos princípios FAIR pode trazer inúmeros benefícios para a ciência, permitindo que máquinas acessem e processem grandes quantidades de dados de forma automática. Mas sua implementação traz vários desafios para a



comunidade acadêmica, uma vez que eles não foram devidamente detalhados na sua publicação original (WILKINSON *et al.*, 2016), sendo mais bem estabelecidos enquanto conceitos. Isso pode estar diretamente ligado à dificuldade de pesquisadores e instituições aplicarem os princípios FAIR aos seus dados de pesquisa, uma vez que a própria publicação original abre margem para interpretações ambíguas. Sem contar as políticas de autoarquivamento dos repositórios de dados que podem gerar discrepâncias entre as descrições e os níveis de conformidade dos dados de pesquisa aos princípios FAIR.

Na prática, nota-se que os princípios FAIR estão sujeitos a diferentes interpretações, levando também a diferentes implementações e, conseqüentemente, reduzindo sua eficácia quando se pensa na importância da padronização e interoperabilidade. Para tornar o compartilhamento de dados FAIR o padrão para pesquisa científica, se faz necessário acordos globais. Mas, ao mesmo tempo, tem-se que aquilo "que é considerado FAIR em uma comunidade pode não ser para outra comunidade" (HENNING *et al.*, 2018, p.181). Há, portanto, um conflito.

Nesse contexto pode-se citar o trabalho *Turning FAIR into reality*<sup>1</sup>, escrito por um grupo de especialistas europeus que expuseram recomendações e ações para a aplicação dos princípios FAIR. Hodson *et al.* (2018) afirmam que surgem dúvidas sobre o significado exato de certos princípios e como eles se aplicam a diferentes disciplinas de pesquisa. Por isso é importante definir passos para implementá-los adequadamente, seguindo padrões internacionais, o que foi abordado no documento.

As recomendações primárias são agrupadas em quatro etapas no *Turning FAIR into reality*, para tornar os dados FAIR uma realidade: 1- definir e aplicar FAIR adequadamente; 2- desenvolver e apoiar um ecossistema de dados FAIR sustentável; 3- garantir dados FAIR e serviços certificados; 4- incorporar uma cultura FAIR na prática de pesquisa. As instruções podem ser vistas em todos os seus detalhes no documento oficial, mas percebe-se uma tentativa de definir etapas e instruções para a implementação dos princípios com certa padronização, permitindo que certificações internacionais avaliem o nível de conformidade com o que é dito FAIR.

Em uma das etapas, Hodson *et al.* (2018, p.6, tradução nossa) afirmam que a

---

<sup>1</sup> Disponível em: <https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en>



“aplicação de dados FAIR depende, no mínimo, dos seguintes componentes essenciais: políticas, plano de gestão de dados, identificadores, padrões e repositórios”. Cada um desses componentes pode apresentar desafios para a aderência aos princípios. Os identificadores únicos e persistentes, por exemplo, são citados também por outros autores como um elemento essencial (MONS *et al.*, 2020).

De acordo com Dias, Anjos e Rodrigues (2019, p.182) para “maximizar a encontrabilidade dos dados, é muito importante que sejam atribuídos aos conjuntos de dados identificadores persistentes, únicos e globais”. O identificador persistente mais utilizado para identificar objetos digitais, com aproximadamente 275 milhões de atribuições por meio de agências de registro em todo o mundo, é o *Digital Object Identifier* (DOI). Ele não é gratuito e o valor a ser desembolsado pela instituição pode representar uma barreira para a implementação de um elemento essencial para dados FAIR. Sabe-se que, no contexto nacional, recursos financeiros são um desafio comum quando se trata da implementação de novas tecnologias.

Quanto aos repositórios de dados, há uma carência de tempo, dinheiro e habilidades para implementar as políticas necessárias para serem compatíveis aos princípios FAIR. Hodson *et al.* (2018, p.58, tradução nossa) propõem que todos os repositórios de dados sejam certificados de acordo com os critérios *CoreTrustSeal*, uma vez que “mecanismos de certificação existentes ajudarão os repositórios a implementar práticas que os ajudem a garantir que seus acervos de dados sejam FAIR”. Mas a certificação citada também não é gratuita, exigindo um investimento de mil euros (conforme consta no site oficial), outra barreira financeira.

Henning *et al.* (2021) também apontam desafios que foram encontrados na implementação dos princípios FAIR: falta de apoio à mudança de cultura, a necessidade de políticas, regulamentos, licenciamento, protocolos e conhecimento de metadados. Para alcançar níveis maiores de conformidade com os princípios FAIR, os metadados são fundamentais, mas requerem habilidades técnicas que muitos bibliotecários não possuem, o que traz a necessidade de equipes multidisciplinares.

Somado a isso, há os desafios relacionados com a tensão entre as necessidades específicas de cada domínio e a interoperabilidade entre domínios, uma vez que as idiosincrasias de cada área disciplinar são inerentes (SALES *et al.*, 2020). Entretanto, os benefícios máximos do FAIR só serão alcançados quando os dados puderem ser





compartilhados e reutilizados de forma confiável em todos os contextos. Logo, é fundamental que as estruturas disciplinares sejam articuladas de maneira comum e adotem padrões globais para permitir a reutilização interdisciplinar sempre que possível (HODSON *et al.*, 2018, p.7). Tem-se, portanto, esse conflito na tomada de decisões quando se diz respeito ao uso de padrões para diferentes áreas com diferentes necessidades.

A interoperabilidade, que diz respeito ao terceiro princípio, é, inclusive, o mais complexo de ser implementado (HENNING *et al.*, 2021). A necessidade de habilidades técnicas e de entendimento quanto aos padrões que devem ser adotados pode ser um dos fatores que contribui para a dificuldade de se alcançar a interoperabilidade desejada em repositórios de dados de pesquisa, por exemplo, sendo um desafio que exige a capacitação dos funcionários ou a contratação de pessoal qualificado. Dias, Anjos e Rodrigues (2019, p.178) afirmam que:

a maioria dos pesquisadores da área da CI alega conhecer o processo de gestão de dados de pesquisa, mas têm dúvidas sobre a devida operacionalização do processo e mencionam a falta de tempo e de recursos para prosseguir com a prática.

Com isso em mente, Henning *et al.* (2021, p.734, tradução nossa) concluem que:

o principal obstáculo à implementação dos princípios FAIR não é a sua interpretação subjetiva ou mesmo a sua aceitação pela comunidade científica, mas sim o custo inicial de operacionalização total, que pode ser totalmente recuperado posteriormente através dos benefícios de alcançar níveis mais elevados de conformidade com os princípios FAIR.

Percebe-se que apesar dos desafios quanto à mudança cultural das práticas científicas, e da interpretação ambígua dos princípios no documento oficial, a questão de custo ainda se mostra o maior obstáculo a ser superado. Quanto ao processo inicial de implementação dos princípios FAIR, tem-se também o desafio da capacitação. De acordo com Mons *et al.* (2020, p.7, tradução nossa), “exigirá muitas habilidades diferentes que não são tradicionalmente cobertas pelos currículos de pesquisa de estudantes e pesquisadores contemporâneos”.

Nesse sentido, a implementação dos princípios FAIR envolve uma ampla gama



de habilidades que precisarão ser empregadas por indivíduos e que exigem, anteriormente, treinamentos, com a elaboração de materiais de capacitação. As habilidades necessárias “devem ser claramente definidas dentro de uma estrutura educacional, particularmente uma que vincule habilidades com competências” (SHANAHAN; HOEBELHEINRICH; WHYTE, 2021, p.2, tradução nossa).

A elaboração de estruturas educacionais é essencial porque fornece orientação para aqueles que constroem currículos e cursos sobre quais objetivos de aprendizagem, resultados e critérios de avaliação de desempenho devem ser considerados e podem então ser incluídos nas estruturas curriculares correspondentes. Entretanto, sabe-se que a mudança curricular em universidades brasileiras conta com barreiras burocráticas, o que pode levar a longos períodos para atualização. Não é uma tarefa simples e, por isso, treinamentos em padrões e habilidades FAIR ainda estão praticamente em desenvolvimento (SHANAHAN; HOEBELHEINRICH; WHYTE, 2021), podendo ser mais bem trabalhados em cursos de extensão e fora do currículo oficial das instituições de ensino superior.

Em âmbito nacional, tem-se a iniciativa GO FAIR Brasil, onde um dos pilares é o GO-Train (Treinamento), que:

busca orientar e treinar o maior número de gestores, pesquisadores, alunos, professores, analistas de sistemas, curadores e administradores de dados na geração, armazenamento, acesso e preservação de dados FAIR, incluindo questões informacionais, políticas e de infraestrutura. Concentra-se na conscientização da aplicação dos princípios FAIR e no treinamento de desenvolvimento de habilidades para comunidades que desejem tornar seus dados FAIR (GO FAIR BRASIL, 2021).

Apesar das iniciativas e do progresso significativo desde que o termo foi cunhado em 2016, autores ainda afirmam que o treinamento em habilidades relacionadas aos princípios FAIR precisará ser ampliado muito em breve (SHANAHAN; HOEBELHEINRICH; WHYTE, 2021). É preciso uma harmonização das competências para implementação dos princípios, além de acordos internacionais sobre boas práticas, sendo essa a espinha dorsal da futura inovação social (MONS *et al.*, 2020). É fundamental, portanto, os trabalhos em prol da superação dos obstáculos, buscando adotar soluções padronizadas, garantindo que o benefício máximo possa ser extraído dos dados, tanto pelas máquinas quanto pelos humanos.





Sendo assim, os desafios para a implementação dos princípios FAIR no cenário da pesquisa são vários: questões financeiras, barreiras técnicas, criação de políticas, necessidade de capacitação e de acordos globais, mudança de paradigma, interpretações ambíguas geradas pelo documento oficial, dentre outros. Mas é essencial que a comunidade científica se atente para os estudos e aplicações dos princípios FAIR, uma vez que são um meio de melhorar a qualidade da pesquisa e ajudar a obter maiores benefícios para a ciência e inovação (HODSON *et al.*, 2018), principalmente no contexto do grande volume de dados.

## 5 CONSIDERAÇÕES FINAIS

O compartilhamento e gestão dos dados de pesquisa traz inúmeros benefícios para a ciência: transparência, economia de recursos, reprodutibilidade, agilização do ciclo científico, dentre outros. Mas para alcançar o benefício máximo, não basta compartilhar os dados na *Web*, sem nenhum preparo adequado. É preciso levar em conta boas práticas que possibilitem que os dados sejam encontráveis, acessíveis, interoperáveis e reutilizáveis, o que é proposto pelos princípios FAIR.

Entretanto, a implementação dos princípios traz também inúmeros desafios que precisam ser superados, como os altos investimentos, a necessidade de capacitação, a interpretação subjetiva, a falta de guias para a aplicação, a aceitação pela comunidade científica, dentre outros. Este trabalho buscou apresentar algum deles, trazendo um pouco do contexto internacional de aplicação e desafios que regiões como os Estados Unidos e a Europa têm enfrentado. Assim, é possível visualizar questões que o Brasil enfrenta e ainda precisará enfrentar nos próximos anos para que os benefícios dos dados de pesquisa FAIR possam ser maximizados.

Em estudos futuros pretende-se analisar as iniciativas internacionais para compreender as soluções que vêm sendo empregadas para os avanços em prol de dados FAIR. Assim será possível visualizar um panorama de ações que o Brasil pode tomar rumo a uma ciência mais colaborativa, eficiente, aberta e FAIR, seguindo as tendências mundialmente adotadas.

## REFERÊNCIAS



BEZJAK, S. *et al.* **The Open Science Training Handbook**. 2018. Zenodo.

BORGMAN, C. L. *et al.* Knowledge infrastructures in science: data, diversity, and digital libraries. **International Journal on Digital Libraries**, v. 16, p. 207–227, 2015.

COSTA, M. P. da; LEITE, F. C. L. Fatores que exercem influência na comunicação dos dados de pesquisa: uma revisão sistematizada da literatura no campo da Ciência da Informação. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais [...]** Londrina: UEL, 2018.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017. **Anais [...]** Marília: UNESP, 2017.

DIAS, G. A.; ANJOS, R. L.; RODRIGUES, A. A. Os Princípios FAIR: viabilizando o reuso de dados científicos. *In*: DIAS, G. A.; OLIVEIRA, B. M. F. (org.). **Dados científicos: perspectivas e desafios**. Paraíba: Editora UFPB, 2019. p. 177-187.

GUANDALINI, C. A.; FURNIVAL, A. C. M.; ARAKAKI, A. C. S. Boas práticas científicas na elaboração de planos de gestão de dados. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 17, e019034, 2019.

GO FAIR BRASIL. **Pilares**. 2021. Disponível em: <https://www.go-fair-brasil.org>. Acesso em: 10 mai. 2022.

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, L. R.; SANTOS, L. O. B. da S. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, João Pessoa, v. 14, n. 3, p. 175-192, 2019.

HENNING, P. C.; SILVA, L. O. B da; PIRES, L. F.; SINDEREN, M. V.; MOREIRA, J. L. R. The FAIRness of data management plans: an assessment of some European DMPs. **RECIIS: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, 2021.

HIGMAN, R.; BANGERT, D.; JONES, S. Three camps, one destination: the intersections of research data management, FAIR and Open. **Insights**, v. 32, n. 1, 2019.

HODSON, *et al.* **Turning FAIR data into reality**: interim report of the European Commission Expert Group on FAIR data. 2018.

MONS, B.; SCHULTES, E.; LIU, F.; JACOBSEN, A. The FAIR Principles: First Generation Implementation Choices and Challenges. **Data Intelligence**, v. 2, n. 1, p. 1–9, 2020.



SALES, L.; HENNING, P.; VEIGA, V.; COSTA, M. M.; SAYÃO, L. F.; SANTOS, L. O. B. da S.; PIRES, L. F. GO FAIR Brazil: A Challenge for Brazilian Data Science. **Data Intelligence**, n. 2, v. 1-2, p. 234–245, 2020.

SHANAHAN, H.; HOEBELHEINRICH, N.; WHYTE, A. Progress toward a comprehensive teaching approach to the FAIR data principles. **Patterns** 2, n. 10, v. 2, p. 1-8, 2021.

SHINKAI, R. S. Integridade na pesquisa e ética na publicação. **Scientia Medica**, Porto Alegre, v. 21, n. 1, p. 2-3, 2011.

SILVA, L. C.; SANTARÉM SEGUNDO, J. E.; SILVA, M. F. Princípios fair e melhores práticas do linked data na publicação de dados de pesquisa. **Informação & Tecnologia**, v. 5, n. 2, p. 81-103, 2018.

SILVEIRA, L. *et al.* Ciência aberta na perspectiva de especialistas brasileiros: proposta de taxonomia. **Encontros Bibli: Revista eletrônica de Biblioteconomia e Ciência da informação**, v. 26, p. 1-27, 2021.

WILKINSON, M. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 160018, 2016.

### **Agências financiadoras**

Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).