

GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta

Patrícia Corrêa Henning

Doutora; Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brasil;
henningpatricia@gmail.com

Cláudio José Silva Ribeiro

Doutor; Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brasil;
claudio.ribeiro@unirio.br

Luiz Olavo Bonino Santos

Doutor; GO FAIR International Support and Coordination Office, Leiden, Holanda;
luiz.bonino@go-fair.org

Paula Xavier dos Santos

Doutora; Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brasil;
paula.xavier@fiocruz.br

Resumo: Este artigo tem o objetivo de apresentar os princípios FAIR e a iniciativa Global Open FAIR que busca disseminar esses princípios em todos os países interessados na aplicação dos dados FAIR (*Findable, Accessible, Interoperable, Reusable*) em seus serviços de informação. Propõe ainda a divulgação e capacitação de instituições de ensino e pesquisa nesses princípios, com o intuito de promover a normalização no tratamento da gestão dos dados garantindo a interoperabilidade entre eles. Como procedimento metodológico, utiliza a revisão bibliográfica e documental para o embasamento teórico sobre ciência aberta, acesso aberto à informação científica e aos dados de pesquisa, visando fundamentar os princípios FAIR em aplicações e serviços de gestão de dados de pesquisa. Ressalta a importância desse tipo de iniciativa para a expansão mundial de abertura dos dados de pesquisa no âmbito da ciência aberta. Ao final, aponta para a necessidade de uma mudança nos processos de pesquisa em ciência e tecnologia na direção da adoção desses princípios.

Palavras-chave: Ciência aberta. Dados abertos. Gestão de dados. Princípios FAIR. GO FAIR.

1 Introdução

Vivenciamos um cenário de mudança de paradigma no mundo científico, onde as práticas de produção, acesso e disseminação do conhecimento estão em franca revisão. Percebe-se que nunca existiram tantos pesquisadores produzindo ciência no mundo como nos últimos anos. Markram (2017) relata em *TEDxBrussels* que, anualmente, cerca de 8 milhões de pesquisadores em todo o

mundo gastam mais de 2,3 trilhões de dólares com pesquisas realizadas. O resultado deste investimento gera, por ano, aproximadamente dois milhões de artigos científicos. No entanto, apenas 10% deles são abertos, ou seja, de acesso livre. É mais frustrante ainda se pensarmos que a maior parte dessas pesquisas foi financiada com verbas públicas e que dificilmente teremos acesso ao seu resultado. Portanto, algumas ações precisavam ser conduzidas para mudar esse cenário e é neste sentido que o acesso aberto às publicações científicas e, mais recentemente, a ciência aberta se consolidaram como novas modalidades do fazer científico. A adoção de diferentes dinâmicas, voltadas para o trabalho colaborativo, que visam os interesses sociais e coletivos, além do compartilhamento e reutilização dos dados e da informação, são premissas que fazem parte desta nova maneira de conduzir a ciência.

Além disso, pode-se afirmar que a ciência aberta expressa um novo modelo do processo da produção e comunicação do conhecimento refletida nas relações entre ciência, tecnologia, informação e inovação. Descreve, acima de tudo, um movimento em curso, um processo e um *status* de acesso às práticas de pesquisa e de produção de conhecimento. Nesse novo cenário o cidadão é livre para usar, reutilizar e distribuir abertamente a informação assim como os dados científicos, sem restrições tecnológicas e sociais, em um ciclo de pesquisa transparente e aberto, voltado para a colaboração, onde o acesso livre é a prática comum e a restrição legal de acesso é a exceção.

É nesse contexto que o conhecimento visto como um bem comum auxilia e embasa as reflexões sobre as práticas da ciência contemporânea, como pode ser observado no livro *Reinventing Discovery*, que apresenta a ciência aberta com a ideia de que o conhecimento científico, de todos os tipos, deve ser compartilhado abertamente, tão cedo quanto possível, desde o início do seu processo de descoberta (NIELSEN, 2012).

É também considerada um conceito amplo que abarca diversas práticas livres e abertas da ciência. Tem no seu centro a percepção de que o conhecimento científico é patrimônio da humanidade e, portanto, deve ser disponibilizado livremente para que todos os cidadãos possam acessá-lo livremente, sem quaisquer restrições econômicas ou legais.

Entre tais práticas destacamos as relacionadas por Albagli, Clinio e Raychtock (2014, p. 438), que apresentam uma amostra dos principais tipos de ações em curso nos quais o conhecimento comum perpassa todas elas, tais como: acesso aberto à informação científica, educação aberta e recursos educacionais abertos, dados científicos abertos, ferramentas e materiais científicos abertos, ciência cidadã e cadernos de pesquisa abertos. O acesso aberto à informação científica é a prática mais consolidada devido aos 16 anos de discussão e aplicação em todo mundo. É resultado da *Budapest Open Access Initiative* (BOAI), que pode ser considerada a primeira expressão, da sociedade científica, em prol do acesso aberto e que culminou na declaração que o acesso à literatura científica deveria ser sem barreiras de acesso, livre de custos, imediato e disponível de forma *on-line*. Parte do pressuposto segundo o qual o acesso aberto não tenha empecilhos financeiros, legais ou técnicos, além daqueles próprios do acesso à Internet. A única restrição à reprodução e distribuição e a única função do *copyright*, nesse contexto, devem ser o controle dos autores sobre a integridade de sua obra e o direito de serem adequadamente reconhecidos e citados (BUDAPEST OPEN ACCESS INITIATIVE, 2002).

Porém, foi em 2003, na Declaração de Berlim sobre o Acesso Aberto ao Conhecimento em Ciências e Humanidades que o conceito de acesso aberto se ampliou dando espaço ao tratamento dos “[...] dados brutos e metadados, fontes originais, representações digitais de materiais pictóricos e gráficos, além de material multimídia acadêmico”, pois estes devem ser abertamente acessíveis e utilizáveis (PAMPEL; DALLMEIER-TIESSSEN, 2014, p. 214, tradução nossa).

Em suma, pode-se afirmar que estes movimentos pavimentaram o percurso na direção da ciência aberta, buscando contemplar os interesses da sociedade. Esta percepção é convalidada por Monino e Sedkaoui que afirmam “[...] o livre acesso aos dados contribui para o aprimoramento das instituições democráticas do ponto de vista do cidadão.” (MONINO; SEDKAOUi, 2016, p. 23, tradução nossa).

Este artigo apresenta uma breve exposição sobre gestão de dados de pesquisa, no âmbito da ciência aberta e, sua correlação com a iniciativa GO FAIR e os princípios FAIR que visam trazer soluções para a falta de

comunicação entre os dados, as dificuldades de serem encontrados, interligados e reproduzíveis, o que leva à sua invisibilidade, inibindo o aumento da capacidade das descobertas e geração de novos conhecimentos.

2 Procedimentos metodológicos

Este relato é de natureza conceitual e descritiva com uma abordagem exploratória, apoiado em análise do conteúdo (MENDONÇA, 2007). Adotou-se como método a pesquisa de documentos institucionais e bibliográficos na base de dados *Web of Science*, além de buscas utilizando o mecanismo *Google Scholar*, todos relacionados ao tema ciência aberta, dados abertos, gestão de dados, princípios FAIR e GO FAIR. Para análise e interpretação dos textos obtidos, foram selecionados os conteúdos de natureza histórica e aqueles que pudessem contextualizar o tema gestão de dados de pesquisa no âmbito da ciência aberta. A intenção foi obter na literatura científica existente e em documentos de natureza institucional, registros que descrevessem o desenvolvimento da ciência aberta e das suas práticas, desde a gênese deste movimento.

3 Os dados abertos de pesquisa

As questões relativas aos dados de pesquisa são bem mais complexas do que podemos imaginar, uma vez que a sua definição ainda não é um consenso geral. De maneira mais objetiva, podem ser definidos como os dados gerados e/ou utilizados durante o processo de desenvolvimento da pesquisa (COSTA; LEITE, 2017, p. 89). No contexto deste artigo, esta afirmação pode ser complementada por registro feito pela *Organisation for Economic Co-operation and Development* (OECD), que consideram dados aqueles utilizados como fonte primária de uma pesquisa (ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007).

O *National Institutes of Health* (NIH), instituto de pesquisa em saúde dos Estados Unidos, pode ser considerado um exemplo que demonstra a

preocupação com os dados abertos já há alguns anos. Esta instituição foi a pioneira na criação de uma política de compartilhamento de dados de pesquisa. O documento gerado em outubro de 2003 determinou que: “Os dados devem estar disponíveis livremente o mais rápido possível, salvaguardando a privacidade dos participantes e protegendo dados confidenciais e proprietários” (NATIONAL INSTITUTES OF HEALTH, 2003¹ *apud* COSTA; LEITE, 2017, p. 92).

O tema dados abertos de pesquisa vêm sendo estudado desde janeiro de 2004, quando os ministros de Ciência e Tecnologia dos países membros da OECD se reuniram em Paris, para discutir a necessidade de um guia internacional voltado para o acesso aos dados de pesquisa. Esta reunião culminou na *Declaration on access to research data from public funding*, que estabelece objetivos e princípios relacionados com as ações de abertura dos dados, tais como: transparência, conformidade legal, responsabilidade formal, profissionalismo, proteção da propriedade intelectual, interoperabilidade, qualidade e segurança, além da eficiência e prestação de contas para a sociedade. Em 2007 esta mesma instituição lança uma nova versão dessa declaração, incluindo quatro novos princípios: qualidade, segurança, flexibilidade e sustentabilidade (ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007).

Nesta mesma direção, o relatório da *Royal Society*, lançado em julho de 2012, analisa os desafios e as oportunidades trazidas pelas novas formas de reunir, armazenar, manipular e transmitir os dados e as informações sobre pesquisas científicas. Este é um importante documento para a prática de dados científicos abertos, pois aborda a maneira como os cientistas precisam adaptar-se às mudanças nos cenários tecnológico, social e político, reunindo recomendações a pesquisadores, universidades, institutos de pesquisa, agências de fomento, governo, editoras de revistas científicas, associações e organismos profissionais (JONES, 2013).

Os dados de pesquisa podem ser considerados um dos pilares para as investigações e apresentam enormes benefícios tanto para ciência quanto para a sociedade. Eles podem ser de tipos e origens diversas tais como: dados públicos,

dados privados ou com algum grau de sensibilidade, dados gerados por instituições ligadas ao setor público e financiados com recursos do estado, dados gerados por empresas do setor privado e dados ligados ao uso restrito com patentes. Dentro de cada contexto institucional, todos estes dados podem estar disponíveis para serem acessados, desde que sejam adotadas licenças apropriadas para sua utilização. É possível afirmar que quanto mais abertos estiverem, mais serão usados, reusados e combinados com outros dados, promovendo o crescimento econômico, a inovação e o desenvolvimento.

Giglia e Swan (2012, p.741-715) comprovam esta percepção quando relacionam as vantagens do compartilhamento que foram sintetizadas no quadro 1.

Quadro 1 - Vantagens dos dados abertos para a ciência e para a sociedade

Para a ciência é possível:	Para a sociedade é possível:
Acessar os dados para compará-los, contrastá-los e debatê-los. Podemos ter uma ciência mais sólida baseada em dados.	Reutilização para inovação e aumento da competitividade. A abertura dos dados pode ajudar a gerar novos produtos, serviços e empregos.
Adicionar novos dados para criar novo conhecimento. Os dados abertos são mais visíveis, se propagam mais rapidamente, aceleram a criação do conhecimento e promovem o progresso da ciência.	Reutilização para a criação de valores por intermédio de mineração de dados. A exploração desse vasto recurso de dados e informação pode gerar benefícios para a economia, aprimorando a produtividade e o valor adicional para o consumidor.
Reutilizar os dados para novas pesquisas sem custos para coletá-los novamente. Uma vez coletados, analisados e depositados, poderão ser usados em novas pesquisas em diferentes perspectivas.	Reutilização para as novas mixagens. A abertura dos dados pode ser utilizada tanto em aplicações para computador e dispositivos móveis.
Reutilizar os dados para verificar e detectar a fraude e/ou a falsificação. Mais transparência poderá evitar fraude.	Reutilização para preservação. A preservação será mais fácil se os dados estiverem abertos e interoperáveis via padrões comuns.
Criar novos dados com técnicas de mineração de dados. Aplicável a dados abertos e licenciados para reutilização, proporciona aumento de valor e economia de tempo.	Reuso para formulação de tomada de decisão. Os formuladores de políticas podem ser capazes de tomar decisões baseadas em evidências e monitorar o impacto dessas decisões.
	Reutilização para o conhecimento da sociedade. A abertura de dados permite que os cidadãos sejam mais informados nos níveis político, cultural e científico.

Fonte: Quadro elaborado pelos autores, adaptado de Giglia e Swan (2012, p.741-715, tradução nossa).

Embora essas vantagens sejam claras e cientificamente aceitáveis, existem outras barreiras a serem vencidas para alcançarmos uma melhor compreensão pela comunidade científica.

Conforme aponta o relatório da FORCE11 (2017), existem diferentes desafios quando se pretende trabalhar com dados de pesquisa. Estes desafios decorrem da complexidade desta atividade e dos múltiplos interesses envolvidos: pesquisadores que desejam analisar e compartilhar seus dados; especialistas em dados que oferecem serviços, *software* e ferramentas para auxiliar a análise e processamento; agências de financiamento (privadas e públicas), que estão cada vez mais preocupadas com a adequada gestão dos dados; comunidade da ciência de mineração de dados, que se preocupa com a integração e análise do resultado para o avanço das descobertas; e a análise computacional com vistas a descobrir padrões para a interligação das atividades de pesquisa.

Relatos dessa natureza nos levam a perceber que um novo paradigma vem se configurando no campo da pesquisa científica, ampliando as práticas coletivas, o compartilhamento da utilização e reutilização dos dados em outras pesquisas, além de possibilitar a geração dos chamados *spillovers* do conhecimento. Todos estes servem de catalisadores para o aproveitamento de novas ideias em diferentes contextos e aplicações.

Diante de tais constatações, vislumbra-se claramente a necessidade do envolvimento de diferentes atores e aparatos tecnológicos nestes espaços que se configuram. Infere-se ainda a importância de repensar diretrizes e políticas, que proponham novos caminhos para atender às novas demandas na construção da ciência.

4 O gerenciamento dos dados de pesquisa

No âmbito da ciência aberta, o gerenciamento de dados de pesquisa é uma prática que vem sendo adotada como uma exigência das agências de fomento às pesquisas internacionais. Em janeiro de 2017, a Comissão Europeia possibilitou um avanço na direção da abertura dos dados de pesquisa no âmbito das

pesquisas do continente, quando passou a exigir dos beneficiários por ela financiados, dentro do Programa Horizonte 2020 um plano de gestão de dados abertos alinhados com os princípios para dados FAIR, como condição para o recebimento da verba de financiamento de suas pesquisas. Para apoiar tal decisão, a Comissão Europeia oferece o Open Research Data Pilot in Horizon 2020 (OPENAIRE, 2015), que auxilia na elaboração dos planos de gestão de dados de pesquisa. Esta exigência contribui para a reutilização destes dados, permitindo também um maior equilíbrio entre a abertura e a proteção da informação científica. Evidencia outras formas de uso e compartilhamento dos direitos de propriedade intelectual, indo ao encontro da premissa “[...] tão aberto quanto possível, tão fechado quanto necessário.” (EUROPEAN COMMISSION, 2016, p. 4, tradução nossa).

No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) vem promovendo ações para o avanço da ciência nesta temática, por meio do Manifesto de Acesso Aberto a dados da pesquisa brasileira. Este documento demonstra o valor estratégico e informacional dos dados de pesquisa, estabelecendo responsabilidades para o desenvolvimento de ações de apoio, além do incentivo à adesão e às iniciativas para a Ciência Aberta e a Ciência Cidadã. Como desdobramento do seu manifesto, o IBICT criou o Plano de Dados Abertos 2017-2018, no qual estabelece como objetivo a promoção, a ampliação e o aprimoramento da abertura de dados da instituição, garantindo a sua publicidade, transparência e eficiência na administração pública, além de facilitar o intercâmbio de dados com outros órgãos para a melhoria da gestão pública (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2017).

Recentemente, a Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP), anunciou que em determinadas modalidades e chamadas será necessário anexar às propostas de submissão um “Plano de Gestão de Dados”. Este plano deve responder a duas perguntas básicas: (1) quais dados serão gerados pelo projeto; e (2) como estes serão preservados e disponibilizados, considerando questões éticas, legais, de confidencialidade e outras

(FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2017).

Diante deste cenário é possível inferir que o gerenciamento dos dados de pesquisa é fator preponderante para o sucesso das investigações, levantando questões complexas com ênfase nas responsabilidades e habilidades diversas no processo de pesquisa. Dentre tais questões, as mais preocupantes referem-se aos problemas legais, ao uso indevido de dados e a falta de padronização. Estes aspectos demonstram a necessidade de estabelecer medidas organizacionais, técnicas e políticas, que visem uma infraestrutura confiável para possibilitar a reutilização de dados e sua integridade permanente. A falta de padronização para o tratamento, armazenamento, preservação e recuperação ainda é uma dificuldade encontrada com frequência. Se houvesse homogeneização e convergência na descrição e licenciamento dos dados, o processo de interoperabilidade entre as instituições poderia acontecer de forma transparente, evitando esforços de integração adicionais, o que certamente promoveria o avanço da ciência.

Neste sentido, diversos guias e cursos de gestão de dados de pesquisa adotados no Brasil e no mundo advertem sobre a importância dos metadados como elemento primordial para a gestão e curadoria dos dados. O Guia de Gestão de Dados de Pesquisa para bibliotecários e pesquisadores, por exemplo, apresenta todas as etapas necessárias para a elaboração de um plano de gestão de dados, ressaltando: ciclo de vida dos dados, documentação, proteção, preservação, compartilhamento, formato, garantia da qualidade, ética, consentimento e qualidade. Descreve a importância dos metadados como:

[...] sem uma descrição minuciosa do contexto tecnológico dos arquivos de dados, do contexto no qual os dados foram criados ou coletados, das medidas que foram feitas, dos detalhes espaciais e temporais, dos instrumentos usados, dos parâmetros e unidades e da qualidade dos dados e da sua proveniência, é improvável que os dados possam ser descobertos, interpretados, gerenciados e efetivamente usados e reusados. (SAYÃO; SALES, 2015, p. 20).

Outro exemplo pode ser encontrado no curso *online Research Data Netherlands* (2019). Desenvolvido por três instituições holandesas o *Centre for*

Research Data - 4TU (2019), a *Data Archiving and Networked Services* (2019) e a *SURF* (2019), o curso tem por objetivo promover orientações para o arquivamento a longo prazo e a reutilização de dados de pesquisa, ressaltando as etapas de documentação e uso de metadados, sejam eles descritivos, estruturais, técnicos ou administrativos.

Percebe-se que a gestão de dados de pesquisa está criando um novo perfil de atuação profissional na gestão de investigações em Ciência. Segundo Giglia e Swan (2012, p. 715), o profissional desta área precisará ser um especialista em metadados, conhecer os protocolos de interoperabilidade e ter noções de recursos orientados para a arquitetura de soluções que façam uso da Web. Estas atividades poderão ser assumidas pelo profissional de Ciência da Informação.

Portanto, em busca de ações para fomentar a padronização e melhorar o gerenciamento dos dados de pesquisa é que os princípios FAIR, que serão abordados na próxima seção, se apresentam como um conjunto mínimo de requisitos orientadores. Estes requisitos devem ser aceitos pela comunidade de produtores e consumidores de dados de pesquisa, com a finalidade de incorporar boas práticas para a publicação e compartilhamento de dados científicos.

5 Os princípios FAIR

As primeiras manifestações referentes aos princípios FAIR surgiram no início de 2014, quando o nó holandês do grupo *Elixir*, em cooperação com as instituições: *Netherlands eScience Center*, o *Dutch Techcentre for Life Sciences (DTL)* e o *Centro Lorentz*, organizaram uma conferência intitulada *Jointly designing a data FAIRPORT*. Nesta conferência foram discutidos os obstáculos relativos à utilização e reutilização de dados e as propostas para a solução destes problemas, o que seria feito por intermédio da criação de uma infraestrutura global para dados no contexto do *eScience*. Estiveram presentes especialistas e pesquisadores, membros de institutos de pesquisa, editores, especialistas em web semântica, inovadores e cientistas da computação (DATA FAIRPORT, 2014).

O resultado da reunião foi um consenso geral quanto à necessidade de criar uma infraestrutura global que suporte a abertura dos dados de pesquisa voltados para a publicação, o compartilhamento e a reutilização de dados. Além destes aspectos, houve também um conjunto de discussões sobre os requisitos para gerenciamento dos dados de forma aberta que foi promovido pelas agências de fomento à pesquisa. Outro resultado de destaque foi a proposta de construção de um *backbone* para possibilitar a interoperabilidade global dos dados, viabilizando que os computadores possam interagir e descobrir de forma automática todos os conjuntos de dados disponíveis para uma pesquisa específica (DATA FAIRPORT, 2014).

Estas propostas suscitaram a criação de princípios e práticas que pudessem orientar a descoberta, o acesso, a integração e a reutilização da vasta quantidade de dados e informação gerada pela ciência. Esses princípios são regulados por intermédio de protocolos que definem a interoperabilidade semântica básica dos conjuntos de dados, quando possível, usando-se padrões e protocolos comunitários já aprovados ou emergentes (WILKINSON *et al.*, 2016).

O acrônimo FAIR significa (*Findable Accessible, Interoperable, Reusable*) visa implementar um conjunto de metadados definidos tanto para uso por mecanismos computacionais automatizados, quanto para uso por pessoas. Estes, se forem devidamente adotados viabilizam a interoperabilidade entre diferentes ambientes de dados. O resultado foi o *FAIR Guiding Principles* apresentados por Wilkinson *et al.* (2016) e detalhados nas subseções a seguir:

5.1 (F) – Findable – localizáveis

Dados e metadados devem ser encontrados por pessoas e mecanismos automatizados. Os metadados devem ser legíveis por estes mecanismos e são essenciais para a descoberta automática dos conjuntos de dados e serviços:

- a) **(F1)** para metadados devem ser atribuídos identificadores globais, persistentes e identificáveis. É considerado o atributo mais fundamental

de todos, pois sem identificadores globais únicos e persistentes será difícil conseguir outros elementos de dados FAIR;

- b) **(F2)** os dados são descritos fazendo uso de metadados enriquecidos. Os metadados enriquecidos devem incorporar várias propriedades sobre os dados, viabilizando a execução de diversas tarefas rotineiras (por exemplo: verificação de autoria e licenciamento, limpeza, atualidade e manutenção de seu conteúdo) de forma automática com o desenvolvimento de tarefas que exigem muita atenção dos pesquisadores. Portanto, não se deve presumir que determinados metadados não são úteis. Seja generoso e ofereça-os sempre;
- c) **(F3)** os metadados incluem clara e explicitamente os identificadores dos dados que são descritos: os metadados e o conjunto de dados são geralmente apresentados em arquivos separados. A associação entre um arquivo de metadados e o conjunto de dados deve ser explicitada. O identificador persistente de um conjunto de dados deve estar presente nos metadados;
- d) **(F4)** os metadados são registrados ou indexados por intermédio de um recurso pesquisável: os identificadores e as descrições de metadados não garantirão a “capacidade de encontrar” na internet. Os princípios F1 e F3 fornecerão os elementos principais para a indexação detalhada por alguns repositórios atuais e para serviços futuros.

5.2 (A) – Accessible –acessíveis

As limitações sobre o uso de dados, os protocolos para consulta e reuso de dados são explicitados tanto para pessoas quanto para mecanismos automatizados.

- a) **(A1)** Os metadados são recuperáveis pelos seus identificadores usando se um protocolo de comunicação padronizado: A recuperação de dados FAIR deve ser mediada sem ferramentas especializadas e sem utilizar métodos de comunicação específicos. A autorizações de uso devem estar

claramente definidas, por exemplo estabelecendo quem pode acessar os dados reais e de que forma isto pode ser feito,

- **(A1.1)** O protocolo é livre, aberto e universalmente implementável.

Para maximizar a reutilização de dados, o protocolo deve ser livre, aberto (disponível para ajustes) e, portanto, implementável globalmente para facilitar a recuperação de dados. Este critério poderá afetar a escolha do repositório onde os dados serão compartilhados,

- **(A1.2)** O protocolo permite procedimentos de autenticação e autorização, quando necessário. O protocolo também deve atender às condições sobre perfis específicos para atender à privacidade, pois o 'A' em FAIR não significa necessariamente estar disponível para o público em geral. Portanto, mesmo os dados privados e altamente protegidos podem atender aos princípios FAIR;

b) **(A2)** Os metadados são acessíveis mesmo quando os dados não estão mais disponíveis. Os conjuntos de dados tendem a degradar-se ou a desaparecer completamente. Este princípio diz que os metadados devem persistir mesmo quando os dados não são mais sustentados, o que está relacionado com os problemas de registro e indexação descritos no (F4).

5.3 (I) – Interoperable – interoperáveis

Deve ser possível que os mecanismos automatizados reconheçam os dados para que eles possam ser combinados automaticamente com outros. A interoperabilidade dos dados pode ser vista como um problema de longo prazo e não trivial, que exigirá um esforço mais criativo na criação dos dados FAIR.

a) **(I1)** os metadados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento. Com o apoio de mecanismos automatizados deve ser possível que os usuários troquem e interpretem os dados. Para garantir a capacidade de busca automática e a interoperabilidade

dos conjuntos de dados é fundamental utilizar: (1) vocabulários controlados, ontologias e tesouros, fazendo uso de identificadores globais únicos, persistentes e identificáveis, conforme definido em 4.1; e (2) um bom modelo de dados e uma estrutura bem definida para descrever e estruturar os dados e metadados;

- b) **(I2)** Os metadados usam vocabulários que seguem os princípios FAIR. O vocabulário controlado usado para descrever conjuntos de dados precisa ser documentado, usando identificadores globais únicos e persistentes;
- c) **(I3)** Os metadados incluem referências qualificadas para outros metadados; Uma “referência qualificada” descreve a sua intenção. De forma mais precisa, se um conjunto de dados se basear em outro conjunto, isto precisa ser especificado. Os links entre os conjuntos de dados precisam ser descritos semanticamente por meio de rótulos. Além disso, a proveniência de todos os conjuntos de dados precisa ser devidamente citada (incluindo seus identificadores globais únicos e persistentes).

5.4 (R) – Reusable –reutilizáveis

Dados e metadados devem ser suficientemente descritos, para possibilitar o reuso por pessoas e/ou mecanismos automatizados, a fim de que possam ser replicados ou combinados em futuras pesquisas.

- a) **(R1)** os metadados são descritos com uma pluralidade de atributos precisos e relevantes. Este objetivo difere de F2 (ver 4.1) na medida em que nos concentramos na capacidade do usuário (pessoa ou mecanismo automatizado) decidir se os dados que acabou de encontrar são realmente úteis em seu contexto particular. Para tomar essa decisão, o responsável pela liberação dos dados deve fornecer não apenas metadados que permitem a descoberta, mas também outros que descrevam o contexto no qual esses dados foram gerados;

- b) **(R1.1)** os metadados são liberados com licenças de uso de dados claras e acessíveis. Este atributo se refere à noção de interoperabilidade legal. Quais os direitos de uso ligados aos dados? Os direitos poderão limitar severamente a reutilização dos dados, especialmente em função da necessidade de cumprimento das restrições de licenciamento;
- c) **(R1.2)** os metadados estão associados à proveniência detalhada. Deve-se saber de onde os dados vieram, esclarecer a origem, mas também é preciso conhecer se estes foram reutilizados e/ou como o autor deseja ser reconhecido. É preciso incluir a descrição sobre a geração, coleta os dados e o histórico de disponibilização e reuso, além de indicar se houve algum processo de transformação para o caso de serem derivados de alguma operação. Este fluxo de informação deve ser descrito em um formato legível por mecanismos automatizados;
- d) **(R1.3)** os metadados fazem parte de domínios com uso de padrões compartilhados em comunidades. A reutilização dos conjuntos de dados deve utilizar os mesmos tipos de origem. Organizá-los de forma padronizada, com formatos de arquivo bem definidos e de ampla utilização, além de possuir a documentação fazendo uso de modelo e vocabulário comum (DUTCH TECHCENTRE FOR LIFE SCIENCES, 2017).

É importante lembrar que foi na perspectiva de uso dos princípios FAIR, que a Comissão Europeia lançou em 2016 o relatório *Realising the European Open Science Cloud* (EUROPEAN COMMISSION, 2016), da Commission High Level Expert Group on the European Open Science Cloud (EOSC). Este relatório apresenta um conjunto de recomendações que apoiam e aceleram a atual transição mundial para uma ciência com inovação. É a ciência de forma mais aberta, eficaz e atendendo de maneira ágil a demandas específicas, decorrentes das definições políticas, governamentais e de outras implementações. A EOSC é uma iniciativa europeia, mas que não se limita ao continente europeu, pois pode ser aplicada globalmente. No entanto, esta iniciativa ilumina o caminho para a interoperabilidade de dados e serviços com

todas as instituições que seguirem os princípios FAIR (EUROPEAN COMMISSION, 2016, p. 8).

Como pode ser observado em SCIENTIFIC ELECTRONIC LIBRARY ONLINE (2016), além da EOSC, outras adesões aos princípios FAIR vêm acontecendo no mundo, como a iniciativa *Big Data to Knowledge* (BD2K), do *American National Institutes for Health* (NIH). Este esforço apresenta orientações claras para garantir que todos os dados e serviços associados à emergente Internet de Dados sejam compatíveis com a iniciativa FAIR.

Durante conferência *Jointly designing a Data FAIRPORT* a Scientific Electronic Library Online (Scielo) declarou que promove os princípios FAIR e que pretende adotá-los para a gestão dos seus dados científicos, ao mesmo tempo que afirma: “[...] parece certo que estes princípios rapidamente se tornarão uma base crucial para inovação no movimento global em direção a ambientes de Ciência Aberta.” (SCIENTIFIC ELECTRONIC LIBRARY ONLINE, 2016).

A abertura dos dados está presente no movimento da ciência aberta e os princípios FAIR tratam estes dados como inteligentemente abertos. Isto significa que estarão livremente disponíveis, desde que a sua privacidade e reutilização sejam preservadas e estejam condicionadas às licenças específicas, além dos créditos reconhecidos e das citações devidamente adotadas (SCIENTIFIC ELECTRONIC LIBRARY ONLINE, 2016).

No entanto, é importante destacar que tais princípios ainda que recentes no mundo científico nacional e internacional, quando colocados em prática, podem gerar equívocos devido às dificuldades no seu entendimento e na falta de experiências na sua aplicação e uso. Em estudo recente, realizado na *University of Delft*, na Holanda, Dunning; Smaele; Böhmer, (2017) avaliaram o alinhamento a esses princípios de 38 repositórios digitais utilizados naquele país. Chegou-se à conclusão que todos eles estavam com baixo grau de compatibilidade aos princípios FAIR.

É dentro desse contexto de dúvidas e incertezas, que surge a iniciativa GO FAIR com a missão de promover a implantação de práticas e serviços que adotem os princípios FAIR na sua totalidade (GO FAIR, 2018).

6 A iniciativa GO FAIR

Diante das dificuldades percebidas no processo de implantação dos princípios FAIR, o *Dutch Techcentre for Life Sciences* (DTL), apoiado pelos governos da Holanda, da Alemanha e, mais recentemente o da França, criou a iniciativa “GO FAIR”. Esta iniciativa busca oferecer um modelo de implantação desses princípios a ser utilizado no desenvolvimento de uma infraestrutura global de interoperabilidade de dados, como ilustrado na figura 1.

Figura 1 - Iniciativas de implantação do GO FAIR



Fonte: Dutch Techcentre for Life Sciences, DTL (2017).

Em maio de 2017 os representantes da Holanda e Alemanha assinaram o documento *Joint Position Paper on the European Open Science Cloud* com o objetivo de unir esforços para participar do EOSC por intermédio da iniciativa GO FAIR. Estes países participam também da construção da *Open Science Policy Platform* (OSPP) onde os grupos de trabalho visavam garantir a sustentabilidade e transparência do acesso aberto, assim como criar incentivos,

mecanismos de avaliação da pesquisa e aumento do envolvimento da sociedade. (JOINT POSITION PAPER, 2017).

Esta iniciativa também ensejou o plano nacional dos Países Baixos denominado *Nationaal Plan Open Science* (WEZENBEEK *et al.*, 2017), elaborado com o apoio do DTL, onde se estabelecem as ações que serão desenvolvidas na Holanda para ciência aberta até o ano de 2020. Este plano é composto por propostas que concentram ações em três direções: (1) promoção do acesso às publicações científicas (*open access*); (2) promoção da utilização e reutilização dos dados de pesquisa (*open data*); e (3) adaptação aos sistemas de avaliação e de propriedade intelectual, todos eles voltados para a ciência aberta (WEZENBEEK *et al.*, 2017).

As atividades planejadas pela iniciativa GO FAIR atendem à segunda ação do *Nationaal Plan Open Science* holandês, uma vez que visam à promoção da utilização e reutilização dos dados e à elaboração de um modelo de gerenciamento global de dados de pesquisa. Estas definições encontram-se detalhadas em diferentes níveis, podendo contemplar níveis mais abrangentes (macro) ou menos abrangentes (micro) sendo sustentadas por três pilares apresentados na figura 2 e detalhados posteriormente (DUTCH TECHCENTRE FOR LIFE SCIENCES, 2017).

Figura 2 - Atividades do GO FAIR



Fonte: Dutch Techcentre for Life Sciences, DTL (2017).

GO-CHANGE – Promove mudanças culturais visando tornar os princípios FAIR um padrão de trabalho na ciência, além de reformar os sistemas de recompensa de modo que a avaliação das carreiras de pesquisa reconheça as atividades de ciência aberta.

GO-TRAIN – Promove treinamento dentro e fora da Europa, de modo a localizar, criar, manter e sustentar o conhecimento sobre o gerenciamento de dados. A meta desse treinamento é criar especialistas certificados em dados em cada estado-membro para cada área.

GO-BUILD – Trata da necessidade de criar infraestruturas para dados interoperáveis, criando padrões, protocolos e serviços compatíveis e possibilitando que os pesquisadores depositem, acessem e analisem dados científicos de todas as áreas.

A partir destas ações verifica-se que a iniciativa GO FAIR vem contribuindo de forma efetiva para a criação de infraestruturas voltadas para interoperabilidade de dados, de forma alinhada aos princípios FAIR.

No Brasil, o movimento começa a se manifestar por meio de algumas iniciativas, como por exemplo a apresentação do Prof. Barend Mons, da Universidade de Leiden, na Holanda, intitulada *The Internet of FAIR data and services*, apresentada na Conferência Luso-Brasileira de acesso aberto (ConfOA) realizada em outubro de 2017, no Rio de Janeiro.

Mais recentemente, a adesão brasileira à iniciativa para adoção dos princípios FAIR vem se consolidando por intermédio do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). O delineamento GO FAIR/Brasil, que deve ser estruturado com representantes segundo uma visão de domínios ou áreas do conhecimento, possibilitará a inserção dos dados de pesquisa brasileiros nas nuvens da ciência aberta global.

7 Considerações finais

Como observado na introdução deste relato, apesar de existir um bom quantitativo de pesquisadores produzindo ciência no mundo, a abertura dos

dados de pesquisa ainda vem ocorrendo de maneira tímida. A adoção de planos de gestão de dados para gerenciar tal abertura pode ser entendida como uma prática recente no meio científico e acadêmico de maneira global.

Neste sentido é possível considerar que estamos vivenciando uma mudança de paradigma do fazer científico na qual novas práticas, regras e políticas estão sendo reconfiguradas. Esta evolução vai ao encontro de diversos fatores, tais como: eficiência, visibilidade, transparência, confiabilidade e controle das pesquisas científicas. O compartilhamento de dados, o incentivo ao seu reuso, o envolvimento do cidadão no processo de desenvolvimento das pesquisas, dentre outras práticas, são características que possibilitam a criação do conhecimento, o progresso da ciência e impulsionam a criação de novas oportunidades de trabalho.

Muitas vezes, tais regras estão sendo impostas por meio de políticas mandatórias. Um exemplo foi a decisão tomada pela Comissão Europeia, em janeiro de 2017, que estabelece que os seus beneficiários só receberão verbas para pesquisas se entregarem os planos de gestão de dados e disponibilizarem os dados e o resultado das pesquisas em acesso aberto (EUROPEAN COMMISSION, 2014)

Por outro lado, existem exemplos de disponibilização sem qualquer imposição institucional. É possível citar os projetos: (1) *The Human Genome Project* que promoveu a abertura dos seus dados e possibilitou o avanço dos estudos; (2) o acidente nuclear de Fukushima, ocorrido em março de 2011, em que cientistas da área de radiação nuclear e poluição abriram seus dados no dia seguinte ao acidente; c) o depósito dos dados da primeira sequência do vírus da gripe H5N1 no *GenBank*, mesmo contra as orientações da *World Health Organization* (WHO). Desde então, várias organizações da área da saúde pública, que atuam com pesquisa sobre gripes, vêm adotado essa prática de abertura dos dados (GIGLIA; SWAN, 2012).

Os esforços aqui apresentados indicam que a execução de práticas da abertura dos dados de pesquisa e do acesso à informação científica começam a se comportar como um caminho sem volta. Nesta jornada, pesquisadores, professores, alunos e outros atores da sociedade precisarão se adaptar a esta

realidade. A iniciativa GO FAIR não é a única a tratar deste assunto, no entanto, os relatos presentes neste estudo indicam que o seu processo de implantação está bem estruturado e baseado em experiências internacionais para a implementação dos princípios FAIR na gestão dos dados de pesquisa. A partir destas constatações, percebe-se o surgimento da construção de um arcabouço para o desenvolvimento de uma infraestrutura global para compartilhamento e interoperabilidade de ciência de dados.

Referências

ALBAGLI, Sarita; CLINIO, Anne; RAYCHTOCK, Sabryna. Ciência Aberta: correntes interpretativas e tipos de ação. **LIINC em revista**, Rio de Janeiro, v. 10, n. 2, 2014.

BUDAPEST OPEN ACCESS INITIATIVE. 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/>. Acesso em: 8 abr. 2019.

CENTRE FOR RESEARCH DATA - 4TU. 2019. Disponível em: <https://www.4tu.nl/en/>. Acesso em: 8 abr. 2019.

COSTA, Michelli; LEITE, Fernando César Lima. Princípios e recomendações basilares para a comunicação dos dados de pesquisa. **Em Questão**, Porto Alegre, v. 23, n. 1, p. 87-112, jan/abr. 2017.

DATA ARCHIVING AND NETWORKED SERVICES. 2019. Disponível em: <https://dans.knaw.nl/en>. Acesso em: 8 abr. 2019.

DATA FAIRPORT. **Jointly designing a data FAIRPORT**. 2014. Disponível em: <https://www.lorentzcenter.nl/lc/web/2014/602/info.php3?wsid=602>. Acesso em: 8 abr. 2019.

DUNNING, Alastair; SMAELE, Madeleine de; BÖHMER, Jasmim. Are the FAIR Data Principles fair? *In*: INTERNATIONAL DIGITAL CURATION CONFERENCE, 12., 2017, Edinburgh. **Anais [...]**. 2017.

DUTCH TECHCENTRE FOR LIFE SCIENCES (DTL). **GO FAIR initiative**. 2017. Disponível em: <https://www.dtls.nl/fair-data/go-fair>. Acesso em: 8 abr. 2019.

EUROPEAN COMMISSION. **Guidelines on open access to scientific publications and research data in Horizon 2020**. 2014. Disponível em: <https://www.openaire.eu/guidelines-on-open-access-to-scientific-publications-and-research-data-in-horizon-2020>. Acesso em: 8 abr. 2019.

EUROPEN COMMISSION. Commission High Level Expert Group on the European Open Science Cloud. **Realising the European open science cloud**. Brussels: European Commission, 2016. Disponível em: https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf. Acesso em: 8 abr. 2019.

EUROPEN COMMISSION. Guidelines on FAIR data management in horizon 2020. 2016. Disponível em: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Acesso em: 8 abr. 2019.

FORCE11. **Guiding principles for findable, accessible, interoperable and reusable data publishing version b1.0**. 2017. Disponível em: <https://www.force11.org/node/6062#annex3>. Acesso em: 8 abr. 2019.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Plano de gestão de dados**. São Paulo, 2017. Disponível em: <http://www.fapesp.br/gestaodedados>. Acesso em: 8 abr. 2019.

GIGLIA, Elena; SWAN, Alma. Open Access to data for a new open science. **European Journal of Physical and Rehabilitation Medicine**, [S. l.], v. 48, n. 4, p. 713-716, dez. 2012.

GO FAIR International Support and Coordination Office. **GO FAIR initiative**. Leiden, 2018. Disponível em: <https://www.go-fair.org/go-fair-initiative/>. Acesso em: 8 abr. 2019.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Plano de dados abertos, 2017-2018**. Brasília: IBICT, 2017.

JOINT POSITION PAPER: on the European Open Science Cloud Germany and the Netherlands. 2017. Disponível em: <https://www.dtls.nl/wp-content/uploads/2017/05/DE-NL-Joint-Paper-FINAL.pdf>. Acesso em: 08 abr. 2019.

JONES, Frances. Editor-chefe da Nature fala sobre a abertura da ciência. **Agência FAPESP**, São Paulo, 6 mar. 2013.

MARKRAM, Kamila. **A ciência aberta pode salvar o planeta**. Publicado pelo canal TEDxBussels. 2017. 1 vídeo (13 min.) Disponível em: <https://youtu.be/uPtP6-nAjJ0>. Acesso em: 8 abr. 2019.

MENDONÇA, Ana Valéria Machado. O uso da análise do discurso do sujeito coletivo em Ciência da Informação. In: MUELLER, Suzana Pinheiro Machado (org.). **Métodos para a pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007.

MONINO, Jean-Louis; SEDKAOUI, Soraya. Open Data: a new challenge. In: MONINO, Jean-Louis; SEDKAOUI, Soraya. **Big data, open data and data development**. London: Wiley, 2016.

NIELSEN, Michael. **Reinventing discovery: the new era of networked science**. Princeton, 2012.

OPENAIRE. **Open research data pilot in horizon 2020**. 2015. Disponível em: <https://www.openaire.eu/item/open-research-data-pilot-in-h2020>. Acesso em: 8 abr. 2019.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles and guidelines for access to research data from public funding**. Paris: OECD, 2007.

PAMPEL, Heinz; DALLMEIER-TIESSEN, Sunjè. Open research data: from vision to practice. In: BARTLING, Sönke; FRIESIKE, Sascha (ed.) **Open science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing**. Heidelberg: Sönke Bartling German Cancer Research Center, 2014. p. 213-224.

RESEARCH DATA NETHERLANDS. 2019. Disponível em: <https://researchdata.nl/en>. Acesso em: 8 abr. 2019.

SAYÃO, Luís Fernando; SALES Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE. Princípios reitores FAIR publicados em periódico do Nature Publishing Group. **SciELO em Perspectiva**, [S.l.], 2016.

SURF. **About SURF**. 2019. Disponível em: <https://www.surf.nl/en/about-surf>. Acesso em: 8 abr. 2019.

WEZENBEEK, Wilma van *et al.* **Nationaal plan open science**. Haia: Ministerie van Onderwijs, Cultuur en Wetenschap, 2017. Disponível em: <https://doi.org/10.4233/uuid:9e9fa82e-06c1-4d0d-9e20-5620259a6c65>. Acesso em: 17 abr. 2019.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, [s. l.], v. 3, article n. 60018, mar. 2016.

GO FAIR AND FAIR PRINCIPLES: what do they represent for the expansion of data in open Science?

Abstract: This paper aims to present the FAIR principles and the Global Open FAIR (GO FAIR) initiative, which seeks to disseminate these principles in all countries interested in the application of FAIR (Findable Accessible, Interoperable, Reusable) data in their information services. It also proposes the dissemination and training of these principles at education and research institutions with the aim of promoting normalization of the data management, ensuring interoperability between them. As a methodological procedure, it uses bibliographic and documentary revision for the theoretical foundation on open science, open access to scientific information and research data, aiming to base the FAIR principles on applications and services of research data management. Highlight the importance of this type of initiative for the worldwide expansion of open research data in open science. In the end, it points to changes necessity in the processes of research in science and technology towards the adoption of these principles.

Keywords: Open science. Open data. Data management. FAIR principles. GO FAIR.

Recebido: 11/07/2018
Aceito: 13/11/2018

¹ NATIONAL INSTITUTES OF HEALTH. NIH data sharing policy and implementation guidance. 2003. Apud Costa e Leite (2017, p. 92).