



XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB**ISSN 2177-3688****GT-8 – Informação e Tecnologia****A IMPORTÂNCIA DAS BOAS PRÁTICAS EM REPOSITÓRIOS DE DADOS DE PESQUISA*****THE IMPORTANCE OF BEST PRACTICES IN RESEARCH DATA REPOSITORY*****Letícia Guarany Bonetti. UFSCar.****Ana Carolina Simionato Arakaki. UFSCar.****Modalidade: Resumo Expandido**

Resumo: O compartilhamento de dados de pesquisa traz diversos benefícios para a ciência. Nesse contexto, fala-se da importância dos repositórios de dados, que precisam se atentar às boas práticas internacionalmente adotadas, garantindo sua integridade científica. Por meio de uma pesquisa exploratória, qualitativa e bibliográfica, objetiva-se discutir a importância da adoção de boas práticas e identificar correspondências entre os princípios FAIR, o *COAR Community Framework for Good Practices in Repositories* e as Boas Práticas para dados na Web da W3C. Na comparação foi possível identificar várias facetas correspondentes, expondo aspectos comuns e centrais para boas práticas no compartilhamento de dados de pesquisa. Conclui-se que é importante que o Brasil busque implementá-las, garantindo a credibilidade e a visibilidade dos repositórios de dados de pesquisa nacionais.

Palavras-Chave: Dados científicos. Repositórios digitais. Boas práticas. Acesso Aberto.

Abstract: Sharing research data has many benefits for science. In this context, data repositories are of great importance and need to pay attention to adopted internationally best practices, ensuring their scientific integrity. Through exploratory, qualitative, and bibliographic research, the objective is to discuss the importance of adopting best practices and identify correspondences between the FAIR principles, the COAR Community Framework for Good Practices in Repositories and the W3C's Best Practices for data on the Web. In the comparison it was possible to identify multiple corresponding facets, exposing central aspects for best practices in sharing research data. In conclusion it is important that Brazil seeks to implement them, ensuring the credibility and visibility of national research data repositories.

Keywords: Research data. Digital repositories. Best practices. Open access.

1 INTRODUÇÃO

Os avanços tecnológicos relacionados com a inserção das Tecnologias Digitais de Informação e Comunicação (TDICs) na sociedade atingiram diversos setores, dentre eles o científico (ANJOS; DIAS, 2019). Bertin, Visoli e Drucker (2017, p. 35) afirmam que, nesse contexto, os “[...] pesquisadores têm produzido uma quantidade de dados sem precedentes, muitos dos quais são subutilizados ou pouco explorados em seu potencial para o avanço



científico e tecnológico”. São esses dados de pesquisa que levam aos resultados e fornecem a evidência necessária para o conhecimento científico publicado, ainda mais no contexto do quarto paradigma: a Ciência Aberta (SILVEIRA *et al.*, 2021).

Araújo, Dias e Autran (2021, p. 101) citam a importância do movimento da Ciência Aberta, “[...] cujas contribuições possibilitam que os resultados das investigações científicas sejam mais acessíveis, incluindo o compartilhamento de códigos, dados e artigos científicos”. A Ciência Aberta aumenta a transparência e o reuso dos dados, e esse ponto de vista tem se tornado cada vez mais comum entre financiadores, editores, cientistas e outras partes interessadas na pesquisa como a *National Science Foundation* (NSF), a *National Institutes of Health* (NIH) e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

É restituído aos dados o seu protagonismo histórico (SALES *et al.*, 2020). Mas para que eles possam ser reutilizados e seu benefício máximo possa ser explorado, eles precisam ser compartilhados, e é nesse contexto que entram os repositórios de dados de pesquisa. Responsáveis por dar visibilidade à produção acadêmica, eles trazem diversos benefícios quando se fala em gestão dos dados de pesquisa: aumento do impacto, novos usos potenciais, maximização da transparência e redução de custos e de retrabalho são alguns deles. Mas para que os dados possam ser descobertos e reutilizados, é necessário seguir as boas práticas.

Nessa perspectiva, objetiva-se discutir a importância de sua adoção, que segundo Guandalini, Furnival e Arakaki (2019), ganha cada vez mais atenção das instituições de ensino e das agências financiadoras. Para isso, foi feita a análise de três diferentes estruturas de boas práticas debatidas no cenário internacional, em busca de correlações. Configura-se como uma pesquisa exploratória, que “[...] visa a proporcionar maior familiaridade com o problema, tornando-o explícito” (PRODANOV; FREITAS, 2013, p. 127), qualitativa e bibliográfica.

As buscas foram feitas na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), bem como nos periódicos da área de Ciência da Informação no Portal de Periódicos da CAPES e nas bases *Web of Science*, *Scientific Electronic Library Online* (SciELO) e o *Google Scholar*. As expressões de busca utilizadas foram “boas práticas”, “acesso aberto”, “dados de pesquisa”, “princípios FAIR”, “publicação de dados” e “repositório de dados de pesquisa” e seus equivalentes em inglês, uma vez que os estudos pioneiros na área se concentram nesse idioma. Elaborou-se sentenças de busca utilizando operadores lógicos, permitindo combinações entre os termos. Para a seleção das referências observou-se



a pertinência da abordagem e a compatibilidade com os temas destacados, e foi feito um recorte temporal a partir de 2016, ano em que houve a disseminação mais ampla do FAIR.

2 ACESSO ABERTO E REPOSITÓRIOS DE DADOS DE PESQUISA

O termo "dado", de forma isolada, apresenta significado restrito e pouco informativo. Já quando se fala, especificadamente, dos dados de pesquisa, pode-se defini-los como informações, em particular fatos ou números, coletados para serem examinados e considerados como base para raciocínio, discussão ou cálculo (EUROPEAN COMMISSION, 2017). Eles são heterogêneos e variam de acordo com as diferentes áreas do conhecimento, mas alguns exemplos são: estatísticas, resultados de experimentos, medições, observações resultantes de trabalho de campo, gravações de entrevistas e imagens.

Nos últimos 20 anos foi possível observar várias mudanças na ênfase na área de gestão e compartilhamento de dados de pesquisa (HIGMAN; BANGERT; JONES, 2019). Para que o benefício desse compartilhamento possa ser maximizado é importante citar o movimento de Acesso Aberto (AA), anterior à Ciência Aberta (SILVEIRA *et al.*, 2021). O AA foi formalizado a partir de 2002 com a *Budapest Open Access Initiative* (BOAI). A BOAI foi uma reunião organizada pelo *Open Society Institute* (OSI) que discutiu a ideia de que a produção científica, principalmente a financiada com recursos públicos, deveria estar disponível sem barreiras para a sociedade, inclusive os dados de pesquisa (GUANDALINI; FURNIVAL; ARAKAKI, 2019).

De acordo com Guandalini, Furnival e Arakaki (2019, p. 3), o movimento *Open Data* (dados abertos em português), que surge no contexto do paradigma da Ciência Aberta, “[...] é um termo que propõe a transparência na divulgação dos dados, tendo como propósito estabelecer sua possível reutilização”. O acesso aberto aos dados de pesquisa ajuda a aumentar a visibilidade do pesquisador, monitorar e avaliar as atividades científicas da instituição, aumentar o prestígio e democratizar o acesso e evitar duplicações de esforços e de investimentos (SILVA; SANTAREM SEGUNDO; SILVA, 2018).

Higman, Bangert e Jones (2019) acrescentam 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. A *European Commission* (2017)



defende que não deve haver pagamento por dados financiados com dinheiro público, e que os dados devem ser depositados, preferencialmente, em repositórios de dados de pesquisa.

Rodrigues, Dias e Lourenço (2022, p. 297) definem os repositórios de dados como aqueles que "[...] executam papéis centrais nas infraestruturas do conhecimento como entidades que facilitam o fluxo de dados entre as partes, geralmente ao longo do tempo". Os repositórios gerenciam o acesso a dados e metadados valiosos e oferecem serviços para dar suporte à reutilização, o que muitas vezes está além da capacidade de pesquisadores, pequenas equipes e da maioria dos laboratórios de pesquisa (HODSON *et al.*, 2018). Mas para uma boa gestão, é preciso atentar-se às boas práticas internacionais.

3 BOAS PRÁTICAS PARA COMPARTILHAMENTO DE DADOS DE PESQUISA

De acordo com Guandalini, Furnival e Arakaki (2019, p. 3), o termo boas práticas "[...]" pode ser entendido como condutas adotadas para uma maior divulgação, disseminação e desenvolvimento da ciência, visando a sua abertura à sociedade". Considerando a relevância dos dados de pesquisa e dos repositórios, surgiram vários estudos, guias e estruturas de boas práticas para os diversos aspectos que envolvem os repositórios e os dados depositados.

Segundo Sales *et al.* (2020), a gestão dos dados tem que ir além do armazenamento seguro e do acesso; estando fortemente relacionado à adoção das boas práticas ao longo do ciclo de vida da pesquisa. Essa gestão adequada pode se dar por meio "[...] da estruturação adequada dos dados, da adoção de metadados de qualidade, do potencial de interoperabilidade entre eles, do atendimento a questões legais e éticas" (SALES *et al.*, 2020, p. 239, tradução nossa), dentre outras práticas também associadas aos princípios FAIR.

Os princípios FAIR são diretrizes internacionalmente conhecidas. São um acrônimo para "*findable*", "*accessible*", "*interoperable*" e "*reusable*", que em português significa "encontrável", "acessível", "interoperável" e "reutilizável". Os quatro princípios foram estabelecidos como resultado da conferência internacional '*Jointly designing the data FAIRPORT*', de 2014. A conferência reuniu especialistas de diversos países e áreas para discutir o uso, tratamento e reutilização de dados de pesquisa no âmbito da Ciência Aberta.

A disseminação mais ampla do FAIR começou em março de 2016, com a publicação dos princípios no *Nature's Journal Scientific Data*. Seu objetivo é enfatizar o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente (WILKINSON *et*



al., 2016). Hodson *et al.* (2018, p. 6, tradução nossa) afirmam que a “[...] aplicação de dados FAIR depende, no mínimo, dos seguintes componentes essenciais: políticas, plano de gestão de dados, identificadores, padrões e repositórios”. Apesar de terem se tornado amplamente aceitos e populares, eles carecem de detalhes técnicos e critérios claros e mensuráveis. É nesse contexto que se pode traçar um paralelo com outras estruturas de boas práticas.

O *COAR Community Framework for Good Practices in Repositories*¹ foi elaborado pela *Confederation of Open Access Repositories* (COAR), uma associação internacional que reúne repositórios individuais e redes de repositórios para alinhar práticas. Atualmente há uma série de estruturas de avaliação, mas elas estão espalhadas por diferentes organizações e costumam ser relevantes para apenas uma região ou um tipo de repositório. Por essa razão a COAR revisou as estruturas existentes, avaliou o nível de viabilidade de implementação e criou seu modelo, que guarda relações com outros *frameworks* como os próprios princípios FAIR, o PLOS “*Criteria that Matter*” e o “*Core Trust Seal*”. A principal característica do modelo COAR é, portanto, fornecer uma estrutura global e multidimensional para boas práticas em repositórios, inclusive de dados, ajudando-os a avaliar e melhorar suas operações atuais.

Ainda no contexto dos dados de pesquisa, tem-se as Boas Práticas para dados na *Web*² da *World Wide Web Consortium* (W3C). Elas foram desenvolvidas para prover diretrizes em prol de uma gestão de dados mais consistente, promovendo a padronização, a reutilização e a confiança entre publicadores e consumidores de dados. Além de apoiar um ecossistema autossustentável e a descoberta dos dados tanto por pessoas como por máquinas (LÓSCIO; BURLE; CALEGARI, 2017). Incluem diferentes aspectos ligados à publicação de dados como: os formatos, o acesso e os metadados, e foram sumarizadas em 35 boas práticas (BP).

Percebe-se que as iniciativas acima buscam prover diretrizes que auxiliam na consistência da publicação dos dados de pesquisa. Isso porque para a descoberta de dados relevantes e o processamento automático “[...] para identificar padrões e correlações não visíveis aos olhos humanos, necessita-se de dados bem descritos e acessíveis que estejam em conformidade com os padrões de suas respectivas comunidades” (RODRIGUES; DIAS;

¹ Disponível em: <https://www.coar-repositories.org/coar-community-framework-for-good-practices-in-repositories/>

² Disponível em: <https://www.w3.org/Translations/DWBP-pt-BR/>.



LOURENÇO, 2022, p. 300). A adoção de boas práticas está intimamente ligada à capacidade de reutilização por humanos e máquinas, construindo práticas de dados confiáveis.

Com isso em mente, os quadros abaixo buscam correlacionar os quatro princípios FAIR, as 35 boas práticas (BP) da W3C e as nove características do modelo COAR, com o objetivo de analisar e sintetizar aspectos centrais para boas práticas no compartilhamento de dados de pesquisa. Este trabalho apresenta uma proposta complementar ao estudo similar de Silva, Santarem Segundo e Silva (2018). É importante salientar que os princípios FAIR foram usados como a base de comparação e que algumas diretrizes da COAR e da W3C se enquadram em mais de uma faceta, enquanto outras não possuem uma correspondência exata.

Quadro 1 – Faceta encontrabilidade: FAIR X COAR X W3C

ENCONTRABILIDADE		
FAIR	COAR	W3C
F1. (meta)dados são atribuídos a um identificador globalmente exclusivo e persistente	1.4 O repositório atribui identificadores persistentes que direcionam para a página inicial do recurso	BP 9 - Usar URIs persistentes como identificadores de conjunto de dados
	1.12 No caso de dados de pesquisa, o repositório suporta identificadores de dados em vários níveis de granularidade, quando apropriado	BP 10 - Usar URIs persistentes como identificadores dentro de conjuntos de dados
	4.2 O repositório suporta revisão dos metadados e controle de versão dos recursos pelo depositante ou administrador	BP 11 - Atribuir URIs a versões de conjuntos de dados
F2. os dados precisam ser descritos com metadados ricos	1.1 O repositório permite que os usuários apliquem metadados básicos do Dublin Core aos seus registros, bem como elementos mais granulares	BP 1 – Fornecer metadados para humanos e máquinas
	1.8 Os metadados no repositório estão disponíveis em formatos legíveis por humanos e legíveis por máquina	BP 2 – Fornecer metadados descritivos
F3. os metadados incluem e explicitamente o identificador dos dados que descreve	1.10 O repositório suporta identificadores persistentes para autores, financiadores, instituições, programas de financiamento e subsídios e outras entidades relevantes	BP 3 – Fornecer metadados estruturados
F4. (meta)dados são registrados ou indexados em um recurso pesquisável	1.6 Os metadados no repositório são indexados por serviços e agregadores externos de descoberta acadêmica	*Não foi encontrada uma correspondência
	1.7 O repositório está incluído em um ou mais registros disciplinares ou gerais de repositórios	

Fonte: Elaborado pelas autoras (2022).

Quando se fala em princípios FAIR, o primeiro passo para poder (re)utilizar dados é encontrá-los. Metadados e dados devem ser fáceis de encontrar tanto para humanos quanto para computadores. Percebe-se que a faceta “encontrabilidade” dos dados de pesquisa é contemplada por todas as três estruturas analisadas. Ela é ligada principalmente aos metadados e identificadores persistentes, sendo fundamental para as boas práticas.

Silva, Santarem Segundo e Silva (2018, p. 92) explicam que para que os dados sejam interpretados pelos computadores “[...] os metadados podem ser fornecidos em um formato de serialização Turtle e JSON, ou podem ser incorporados na página HTML (HTML-RDFA ou JSON-LD), e reutilizar padrões existentes e vocabulários populares, como por exemplo, o padrão de metadados *Dublin Core*”. Tanto a W3C quanto a COAR citam o uso do *Dublin Core*.



Sua forte influência na descrição de objetos digitais, como é o caso dos dados de pesquisa, torna-o um recurso valioso para a adoção de boas práticas. Outro item importante, mas que não obteve correspondência direta com os princípios FAIR, é a inclusão clara da citação da publicação original do conjunto de dados na interface do repositório (BP 35). Isso facilita não só a descoberta, mas também a reutilização dos dados de pesquisa.

Quadro 2 – Faceta acessibilidade: FAIR X COAR X W3C

ACESSIBILIDADE		
FAIR	COAR	W3C
A1. (meta)dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado	1.2 O repositório permite a coleta de metadados usando OAI-PMH	BP 23 – Disponibilizar dados por meio de uma API
	3.3 O repositório adota <i>Signposting</i> para dar suporte ao acesso aos recursos pelas máquinas	
A1.1 o protocolo é aberto, gratuito e universalmente implementável	2.1 Os recursos no repositório estão disponíveis sem nenhum custo para o usuário	BP 24 – Usar padrões da <i>Web</i> como base de API
	2.9 O repositório adere à versão mais recente das Diretrizes de Acessibilidade de Conteúdo da <i>Web</i> do W3C	
	8.5 O repositório é construído em software de código aberto bem suportado	
A1.2 o protocolo permite um procedimento de autenticação e autorização, quando necessário	2.4 Nos casos em que o repositório está coletando dados confidenciais de pesquisa, existem mecanismos que permitem que os proprietários de dados limitem o acesso apenas a usuários autorizados	*Não foi encontrada uma correspondência
	2.7 Nos casos em que há acesso restrito a um recurso, o repositório facilita uma maneira indireta de acessar esse recurso	
A2. os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis	1.3 Nos casos em que o recurso foi retirado, o repositório fornece uma <i>tombstone</i> e o registro de metadados permanece disponível publicamente	BP 22 – Fornecer uma explicação para dados que não estão disponíveis
		BP 27 – Preservar identificadores

Fonte: Elaborado pelas autoras (2022).

Uma vez que o usuário encontre os dados necessários, ele precisa saber como pode acessá-los, possivelmente incluindo autenticação e autorização. Na faceta de acessibilidade, só não foi possível encontrar correspondência quanto a um item (Quadro 2). A W3C não especifica sobre procedimentos de autorização/autenticação em suas BPs, mas no documento oficial é dito que “Espera-se que as políticas de compartilhamento de dados avaliem o risco de exposição e estabeleçam as medidas de segurança apropriadas para a proteção de dados sensíveis, tais como a autenticação segura e a autorização” (LÓSCIO; BURLE; CALEGARI, 2017).

Quanto à acessibilidade é possível notar o foco em protocolos padronizados, além do acesso aberto e gratuito, como é defendido pelo movimento do AA. A COAR oferece algumas diretrizes extras que dizem respeito ao acesso nos repositórios em si. Elas não encontraram correspondência direta com o FAIR. O item 1.5, por exemplo, que está dentro da característica de descoberta, se refere ao repositório oferecer um recurso de busca. Já o item 2.5 diz respeito ao repositório possuir uma interface de usuário amigável para dispositivos móveis. São tópicos que não são cobertos pelo FAIR e nem pela W3C.



Outra questão importante que foi mapeada nas três estruturas é o caso dos metadados estarem disponíveis mesmo quando os dados não estão mais. Os conjuntos de dados tendem a desaparecer com o tempo. Armazenar os metadados geralmente é mais fácil e barato, e dessa forma o usuário terá orientação. Enquanto a COAR recomenda manter o registro de metadados disponível publicamente (1.3), a W3C recomenda fornecer uma explicação do porquê os dados não estão disponíveis (BP 22). Quanto ao critério A2 dos princípios FAIR foi feita também uma correspondência com a BP 27, que diz respeito à preservação dos identificadores. Dessa forma a URI irá sempre redirecionar para informações a respeito.

O único outro item de “acesso” da COAR que não obteve correspondência com os princípios FAIR foi a questão da recomendação de ferramentas para anonimizar dados sensíveis para permitir seu compartilhamento (2.8). É importante lembrar que embora muitas políticas exijam dados FAIR e abertos, os dois não significam a mesma coisa, e as instituições precisam estar atentas à Lei Geral de Proteção de Dados Pessoais (LGPD), nº 13.709/2018.

Já quanto às boas práticas da W3C, alguns itens citados na faceta de “acesso” não encontraram correspondência direta com os princípios FAIR como: usar negociação de conteúdo para disponibilizar dados em formatos múltiplos (BP 19), fornecer acesso em tempo real (BP 20), fornecer dados atualizados (BP 21), fornecer documentação completa para as APIs (BP 25) e evitar alterações que afetem o funcionamento de sua API (BP 26). Uma API oferece aos consumidores dos dados maior flexibilidade e facilidade de processamento. Logo, percebe-se que a facilidade e a rapidez de acesso, sem barreiras como custos e com o uso de identificadores persistentes, são boas práticas fundamentais.

Quadro 3 – Faceta interoperabilidade: FAIR X COAR X W3C

INTEROPERABILIDADE		
FAIR	COAR	W3C
I1. (meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento	1.8 Os metadados no repositório estão disponíveis em formatos legíveis por humanos e legíveis por máquina	BP 15 - Reutilizar vocabulários, dando preferência aos padronizados
I2. (meta)dados utilizam vocabulários e/ou ontologias que seguem os princípios FAIR	1.1 O repositório permite que os usuários apliquem metadados básicos do Dublin Core aos seus registros, bem como elementos mais granulares	BP 16 – Escolher o nível de formalização correto
	1.13 O repositório facilita o uso de vocabulários controlados em seus registros de metadados	
	8.2 O repositório suporta submissão mediada usando protocolos padronizados como SWORD	
I3. (meta)dados incluem referências qualificadas para outros (meta)dados	1.9 O repositório facilita a vinculação no registro de metadados entre conteúdos relacionados, como pré-prints, artigos publicados, dados e software (incluindo PIDs para recursos relacionados mantidos em outros lugares)	BP 22 – Fornecer uma explicação para dados que não estão disponíveis

Fonte: Elaborado pelas autoras (2022).



Os dados geralmente precisam ser integrados com outros dados. Além disso, precisam interoperar com aplicativos ou fluxos de trabalho para análise, armazenamento e processamento. Conforme mostra o Quadro 3, foram encontradas correspondências entre todas as três estruturas para todos os aspectos da faceta “interoperabilidade”. Nela é possível observar o foco no uso de vocabulários e metadados (LÓSCIO; BURLE; CALEGARI, 2017). De acordo com Hodson *et al.* (2018), os dados devem ser representados em formatos comuns, e idealmente abertos, e ser ricamente documentados usando padrões de metadados e vocabulários adotados pela comunidade de pesquisa para permitir a interoperabilidade e a reutilização, uma vez que eles dependem de um alto grau de contextualização.

Indo um pouco além, a COAR apresenta uma categoria de “garantia de qualidade” dos metadados, com o repositório realizando uma revisão dos metadados básicos após o envio do recurso (5.1) e fornecendo documentação/política que descreva quais processos de curadoria são aplicados aos recursos e metadados (5.2). Apesar desses itens não terem encontrado uma correspondência direta junto aos princípios FAIR, percebe-se que é uma tarefa importante ligada diretamente aos repositórios. Quando se fala sobre o auto arquivamento, comum nesses sistemas, é preciso ter controle de qualidade para garantir que os dados sejam padronizados e permitam a interoperabilidade.

Por fim tem-se a faceta da reutilização. O objetivo final dos princípios FAIR é assegurar que os dados possam ser reutilizados tanto por humanos quanto por máquinas, acelerando a interdisciplinaridade e a inovação. Para atingir essa meta, os metadados e dados devem ser bem descritos para que possam ser replicados e/ou combinados em diferentes configurações. (WILKINSON *et al.*, 2016). É possível notar, pelo Quadro 4, que apenas um aspecto de “reutilização” não foi coberto pelas boas práticas da W3C.



Quadro 4 – Faceta reutilização: FAIR X COAR X W3C

REUTILIZAÇÃO		
FAIR	COAR	W3C
R1. (meta)dados são ricamente descritos com uma pluralidade de atributos precisos e relevantes	3.2 As páginas de destino no repositório incluem metadados necessários para a citação dos recursos e estão em formato legível por máquina	BP 1 – Fornecer metadados para humanos e máquinas
		BP 2 – Fornecer metadados descritivos
		BP 3 – Fornecer metadados estruturados
		BP 31 – Enriquecer dados por meio da geração de novos dados
	4.2 O repositório suporta revisão dos metadados e controle de versão dos recursos pelo depositante ou administrador	BP 7 – Fornecer indicador de versão
		BP 8 – Fornecer o histórico de versão
R1.1. (meta)dados são publicados com uma licença de uso de dados clara e acessível	3.1 O repositório inclui informações de licenciamento no registro de metadados que estipulam condições de reutilização para o recurso	BP 4 – Fornecer informações sobre a licença dos dados
	1.11 Os metadados no repositório são disponibilizados sob uma licença de domínio público Creative Commons (CC0)	
R1.2. (meta)dados estão associados com a proveniência detalhada	6.3 O repositório coleta metadados básicos de preservação, incluindo proveniência, data de upload e formato de arquivo	BP 5 – Fornecer informações de procedência dos dados
R1.3. (meta)dados cumprem os padrões da comunidade relevantes para o domínio	1.1 O repositório permite que os usuários apliquem metadados básicos do Dublin Core aos seus registros, bem como elementos mais granulares	* Não foi encontrada uma correspondência

Fonte: Elaborado pelas autoras (2022).

A descrição rica dos conjuntos de dados (R1) é fundamental para a citação adequada dos recursos, e nesse aspecto é preciso ter um controle claro de, por exemplo, versões do mesmo conjunto de dados. Fornecer indicadores e histórico de versão auxilia e gera maior confiança para o pesquisador que está reutilizando os dados. A licença de uso (R1.1) é outro aspecto de suma importância e, segundo Silva, Santarem Segundo e Silva (2018, p. 98), “O termo pode ser explicitado pelo elemento *dc:terms:license* do padrão DCTERMS, *cc:license* do *Creative Commons*, *schema:license* do *Schema.org* ou *xhtml:license* do XHTML”.

A proveniência é outro elemento ligado à reutilização dos dados que traz confiança para os consumidores. Como já visto, eles demandam uma contextualização rica para sua interpretação, e sem as informações de proveniência como *dc:creator*, *dc:publisher* e *dc:issued* a reutilização fica comprometida. Os princípios FAIR, a COAR e a W3C apresentam diretrizes neste sentido, expondo mais um aspecto em comum para adoção de boas práticas.

Com relação à reutilização, a W3C apresenta alguns itens adicionais que não obtiveram correspondência direta com os princípios FAIR como: fornecer dados em formatos múltiplos (BP 14) e avaliar a cobertura do conjunto de dados (BP 28), que diz respeito à preservação dos conjuntos de dados. Os princípios FAIR concentram-se no acesso e não abordam de forma explícita a preservação a longo prazo (HODSON *et al.*, 2018), ao contrário da W3C e da COAR que possuem categorias inteiras destinadas às práticas de preservação.



Por fim, é importante destacar que as boas práticas da W3C, assim como as da COAR, possuem algumas diretrizes mais administrativas. Recomendações como coletar e compartilhar *feedbacks* dos consumidores de dados (BP 29/30); interface amigável (2.5 COAR); e possuir informações de contato para suporte aos usuários (7.2/7.3 COAR) são alguns exemplos que não encontram correspondência direta com os princípios FAIR.

As boas práticas da COAR e da W3C oferecem recomendações mais verificáveis, diferente dos princípios FAIR, que não trazem muitos esclarecimentos sobre implementação. A W3C oferece um “como testar” para cada uma das 35 BPs, além de oferecer exemplos e abordagens para implementação. Já o modelo COAR foi elaborado para os repositórios poderem avaliar suas operações baseados em diretrizes aplicáveis. As três propostas convergem em vários aspectos, mas também se mostram complementares em outros.

4 CONSIDERAÇÕES FINAIS

No paradigma da Ciência Aberta, quanto mais dados abertos são compartilhados, maiores são os benefícios para a ciência. Cada vez mais partes interessadas em pesquisa ao redor do mundo estão adotando e incentivando a disponibilização dos dados de pesquisa em Acesso Aberto. Para que os dados sejam compartilhados de forma a serem encontráveis, acessíveis, interoperáveis e reutilizáveis, tem-se boas práticas: diretrizes e guias que buscam garantir a integridade científica. Alguns exemplos são os princípios FAIR, o *COAR Community Framework for Good Practices in Repositories* e as Boas Práticas para Dados na Web.

Na comparação entre as três estruturas foi possível identificar vários aspectos correspondentes, expondo diretrizes que são centrais para boas práticas como: metadados ricos, licenças bem declaradas, proveniência dos dados, uso de identificadores persistentes e vocabulários e protocolos padronizados. Entretanto, outros itens do modelo COAR e da W3C não encontraram correspondência, indo além do escopo mais conceitual dos princípios FAIR, como os ligados à preservação e administração dos dados de pesquisa.

Sendo assim, é possível concluir que as boas práticas recomendadas convergem em diretrizes comuns, e é importante que o Brasil busque implementá-las, garantindo a credibilidade e a visibilidade dos repositórios de dados nacionais. Em pesquisas futuras o objetivo é avaliar o nível de conformidade dos repositórios de dados de pesquisa nacionais quanto às boas práticas citadas neste trabalho.



REFERÊNCIAS

- ANJOS, R. L.; DIAS, G. A. Atuação dos profissionais da informação no Ciclo de Vida dos Dados – DATAONE: um estudo comparado. **Informação & Informação**, Londrina, v. 24, n. 1, p. 80-101, mar. 2019. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/31085>. Acesso em: 28 jan. 2022.
- ARAÚJO, D. G.; DIAS, G. A.; AUTRAN, M. M. M. Compartilhamento de dados no contexto da ciência brasileira um estudo integrativo. **Informação & Informação**, Londrina, v. 26, n. 3, p. 100-124, 2021. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/41035>. Acesso em: 11 abr. 2022.
- BERTIN, P.; VISOLI, M.; DRUCKER, D. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de p&d. **PontodeAcesso**, Salvador, v. 11, n. 2, p. 34-48, 2017. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/21449>. Acesso em: 04 jan. 2022.
- EUROPEAN COMMISSION. **Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020**. 2017. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Acesso em: 15 abr. 2022.
- 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**, Campinas, v. 17, e019034, 2019. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8655895>. Acesso em: 21 jan. 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. Disponível em: <http://insights.uksg.org/articles/10.1629/uksg.468/>. Acesso em: 22 ago. 2022.
- HODSON *et al.* **Turning FAIR data into reality**. Interim report of the European Commission Expert Group on FAIR data. 2018. Disponível em: https://zenodo.org/record/1285272#.YuL_sC35RhA. Acesso em: 30 abr. 2022.
- LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the Web best practices**. W3C, 2017. Texto digital. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 02 jun. 2018.
- PRODANOV, C. C.; FREITAS, E. C. **Metodologia do Trabalho Científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. Novo Hamburgo: Feevale, 2013.
- RODRIGUES, M. M.; DIAS, G. A.; LOURENÇO, C. A. Repositórios de dados científicos na américa do sul: uma análise da conformidade com os princípios fair. **Em Questão**, v. 28, n. 2,



p. 113057, 2022. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/113057>. Acesso em: 28 jul. 2022.

SALES, L. *et al.* GO FAIR Brazil: A Challenge for Brazilian Data Science. **Data Intelligence**, v. 2, n. 1–2, p. 238–245, 2020. Disponível em: <https://direct.mit.edu/dint/article/2/1-2/238-245/10004>. Acesso em: 27 jul. 2022.

SILVA, L. C.; SANTAREM 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. DOI: 10.22478/ufpb.2358-3908.2018v5n2.44812 Acesso em: 12 ago. 2022.

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. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/79646>. Acesso em: 10 abr. 2022.

WILKINSON, M. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 160018, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 20 jan. 2021.