



ARTIGO

PRINCÍPIOS FAIR: GESTÃO DE DADOS PARA HUMANOS E MÁQUINAS

POR

Maria Luiza M. Campos, Vânia Borges, João Luiz R. Moreira
m luiza@ppgi.ufrj.br, vjborges30@ufrj.br e j.luizrebelomoreira@utwente.nl

Os princípios FAIR [6], publicados em 2016, constituem um conjunto de 13 boas práticas que visam orientar a publicação de dados, em particular de pesquisas científicas, de modo a serem localizáveis (em inglês, *Findable*), acessíveis (em inglês, *Accessible*), interoperáveis (em inglês, *Interoperable*) e reusáveis (em inglês, *Reusable*). São 4 princípios para o F, 3 para o A, 3 para o I, e 3 para o R. Iniciativas recentes consideram vital garantir que os dados sejam FAIR, no sentido original do acrônimo, e, também, no sentido de "**Federated, AI-Ready**", ou seja, "Dados Federados e Prontos para Inteligência Artificial (IA)", portanto legíveis e acionáveis por máqui-

nas ¹. Dados legíveis significam dados em forma digital que possam ser lidos tanto por humanos quanto por máquinas, ou seja dados formatados com padrões de interoperabilidade sintática como, por exemplo, CSV, XML, e JSON. Dados acionáveis por máquina significam dados em forma digital no qual aplicações de IA possam 'raciocinar', a exemplo de dados na forma de grafos de conhecimento.

A iniciativa GO FAIR² vem complementar outras iniciativas já em andamento no Brasil, voltadas para a abertura de dados de pesquisa, a exemplo da participação brasileira na *Research Data*

1 <https://www.go-fair.org/wp-content/uploads/2020/03/VODAN-IN-Manifesto.pdf>

2 <https://www.go-fair.org/>

Alliance (RDA)³, da rede de repositórios da FAPESP (apelidado de 'metabuscador'⁴), do LattesData⁵, e do suporte da Rede Nacional de Pesquisa (RNP) à implantação piloto de repositórios em universidades e centros de pesquisa⁶. É importante ressaltar que implementações associadas aos princípios FAIR foram impulsionadas em diversas iniciativas globais a partir de 2016, após a publicação original na revista *Nature* [6]. Os princípios FAIR são endossados pelo G7, pelo G20, pelo Fórum Econômico Mundial, e pela Comissão Europeia (entre outros), sendo a base da nuvem europeia de ciência aberta (EOSC)⁷.

O Escritório de Apoio de Coordenação GO FAIR Brasil⁸, liderado pelo IBICT com o suporte da UFRJ, FIOCRUZ, RNP e EMBRAPA, tem a responsabilidade de fomentar a adoção dos princípios FAIR no país, disseminando, apoiando e coordenando as atividades para sua implementação [4]. Além disso, provê suporte às necessidades das redes de implementação (RI) GO FAIR ativas, alinhado com as recomendações do RDA Brasil. Até o momento, 6 redes vêm sendo estruturadas, cada uma liderada por uma instituição chave em seu domínio: Saúde (FIOCRUZ), Agro (EMBRAPA), Biodiversidade (Jardim Botânico do Rio de Janeiro), Humanidades (IBICT), Ciências Nucleares (CNEN) e Ensino, Ciência, Tecnologia e Inovação (UNESP).

3 <https://www.rd-alliance.org/groups/rda-brazil>

4 <https://metabuscador.uspdigital.usp.br>

5 <https://www.gov.br/cgu/pt-br/governo-aberto/noticias/2020/2/cnpq-e-ibict-lancam-lattes-data>

6 <https://www.rnp.br/noticias/rnp-implimentara-federacao-de-repositorios-para-dados-de-pesquisa-ate-2020>

7 <https://eosc-portal.eu/>

8 <https://www.go-fair-brasil.org/>

Embora os princípios FAIR sejam boas práticas de gestão de dados empregadas na Computação, sua grande diferença para outras iniciativas é o 'empacotamento' (de forma complementar) dessas boas práticas, e o direcionamento tanto para o uso humano quanto para o uso por máquinas. Os princípios FAIR refletem a convergência de uma série de linhas de pesquisa da Computação e da Metaciência, assim como de Organização do Conhecimento e da Ciência da Informação. Desse modo, observa-se que princípios relacionados à gestão de metadados são diretrizes típicas da Administração de Dados, enquanto princípios sobre protocolos de comunicação padronizados para acesso e reuso de dados são práticas comuns da Engenharia de Software, e os princípios sobre representação de conhecimento com dados distribuídos são apoiados por técnicas em dados interligados e IA.

Para a implementação dos princípios FAIR, desde 2020 a comunidade GO FAIR passou a trabalhar em um *framework* baseado nas experiências obtidas por grupos de trabalhos das diferentes redes de implementação. Conhecido como *framework* de *FAIRificação*, visa contribuir com os desafios identificados em processos de geração de dados FAIR, sendo constituído por três elementos-chave [5]. O primeiro elemento de análise são os metadados, estruturados de acordo com o domínio específico da pesquisa, legíveis e acionáveis por máquina, sendo por isso denominados de Metadados para Máquinas (em inglês, *Metadata for Machine - M4M*). O segundo elemento refere-se

à seleção da estratégia de implementação FAIR a ser adotada na pesquisa. Para auxiliar essa escolha, tem-se como referência o Perfil de Implementação FAIR (em inglês, *FAIR Implementation Profile* – FIP) que relaciona as estratégias, padrões e ferramentas adotadas por diferentes comunidades. O terceiro elemento do *framework* corresponde à tecnologia para publicar os metadados e serviços de acesso aos dados. O FAIR *Data Point* (FAIR DP) é um repositório de metadados que promove a visibilidade dos dados de pesquisa através de uma hierarquia de descritores padrão, possibilitando acesso automatizado, de acordo com o esquema de autorização definido (aberto ou tão restrito quanto necessário, empregando algum mecanismo de autenticação [2]).

Complementar a esses elementos do *framework* (M4M, FIP e FAIR DP), outros elementos se mostram essenciais quando se trata de enfrentar os desafios de apoiar a interoperabilidade semântica de dados, intra e inter domínios. Conforme pontuado por Guizzardi [1], os princípios FAIR, associados à interoperabilidade, necessitam não apenas de vocabulários, mas sim de ontologias de domínio bem fundamentadas. Essas ontologias devem obedecer práticas de engenharia de ontologias e modelagem conceitual baseadas em categorias básicas consistentes definidas por ontologias de fundamentação, de forma a permitir a correta associação entre elementos de conceitualizações de diferentes sistemas. A pesquisa brasileira nessa área é, hoje, referência mundial, com um histórico de longa data em eventos e grupos de pesquisa que são reconhecidos

internacionalmente, como o Ontobras⁹ e o Núcleo de Estudos em Modelagem Conceitual e Ontologias (NEMO) da UFES¹⁰.

Como uma evolução dos princípios inicialmente voltados para recursos de dados, a comunidade FAIR passou a discutir e considerar um escopo mais amplo para o “R” em FAIR. Assim, além do reuso, reprodutibilidade requer que as operações sobre os dados representadas como fluxos de atividades nos assim chamados *workflows* científicos, sejam também gerenciados e aderentes aos princípios FAIR.

Muitas das soluções tecnológicas para apoiar os princípios FAIR vêm sendo experimentadas no contexto da *RI Virus Outbreak Data Network* (VODAN), criada no início da pandemia da COVID-19 e concebida para desenvolver uma infraestrutura distribuída de apoio à interoperabilidade de dados sobre surtos virais correntes e futuros [3]. No Brasil, o VODAN BR faz parte da rede GO FAIR Brasil Saúde, com colaboração multi-institucional com a UFRJ, a UNIRIO, e hospitais parceiros.

Ao considerar estratégias para a gestão de dados de pesquisa no Brasil, é importante adotar uma abordagem evolutiva, na qual *FAIRness* ainda representa um caminho a ser trilhado, combinando mudança de cultura, de políticas e de serviços de apoio ao pesquisador e, não menos importante, de suporte tecnológico. A iniciativa GO FAIR se estrutura neste sentido, através de seus 3 pilares, GO CHANGE, GO TRAIN e GO BUILD, orga-

9 <https://www.inf.ufrgs.br/ontobras/en/14th-seminar-on-ontology-research-in-brazil/>
10 <https://nemo.inf.ufes.br/>

nização também adotada pelo Escritório GO FAIR Brasil. Como primeiro passo, a cultura de implantação de repositórios vem se disseminando, com ampla participação de bibliotecários e de pesquisadores da Ciência da Informação. Ao mesmo tempo, pesquisadores brasileiros de áreas diversas participam de grupos de trabalho e de interesse associados à RI FAIR, assim como à iniciativa da RDA, transformando as propostas ali concebidas em políticas institucionais e treinamentos para seus pares. De forma complementar, mas já baseado em resultados de pesquisas anteriores no Brasil nas

áreas de gestão de metadados, engenharia de ontologias, tecnologias e padrões da web semântica e dados interligados, segue-se investindo na construção de infraestruturas de apoio aos dados FAIR. Em especial, face aos desafios para se chegar ao amplo reuso e interoperabilidade, há que se desenvolver mecanismos de apoio baseados em IA, particularmente em técnicas de Aprendizado de Máquina e Processamento de Linguagem Natural.

Referências

1. GUIZZARDI, G. Ontology, ontologies and the "I" of FAIR. *Data Intelligence* 2(2020), 181–191.
2. LANDI A et al. The "A" of FAIR – As Open as Possible, as Closed as Necessary. *Data Intelligence*, 2020, 2(1–2) 47–55.
3. MONS, B. The VODAN IN: support of a FAIR-based infrastructure for COVID-19. *European Journal of Human Genetics*. 2020; 28. 1-4.
4. SALES, L. et al. GO FAIR Brazil: a challenge for Brazilian data science. *Data Intelligence*, 2019, 1(1) 238-245.
5. SCHULTES, E. et al. Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. *ER 2020 Workshops, Austria, Proceedings*. Springer Science and Business Media Deutschland GmbH. 2020. p. 138-147.
6. WILKINSON, M. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016; 3:160018.



MARIA LUIZA MACHADO CAMPOS é Professora Associada no Instituto de Computação da Universidade Federal do Rio de Janeiro e atua na coordenação executiva do GO FAIR Brasil como representante internacional e na coordenação tecnológica do Projeto VODAN BR. Seus principais temas de pesquisa estão associados à integração de informações heterogêneas, abordando principalmente gestão de metadados, engenharia de ontologias, modelagem conceitual, dados abertos e web semântica.



JOÃO LUIZ RABELO MOREIRA é Professor Assistente na Universidade de Twente e atua na coordenação executiva do GO FAIR Brasil, como representante na Europa. Sua pesquisa de doutorado seguiu a linha de interoperabilidade semântica, com ênfase na integração de dados de serviços inteligentes de emergência suportados por tecnologias de IoT no contexto de cidades inteligentes. Obteve seu pós-doutorado pela Universidade VU Amsterdam com pesquisa em engenharia de ontologias para workflows científicos com princípios FAIR.



VÂNIA BORGES é Militar da Reserva da Marinha, onde atuou por 9 anos na Diretoria de Saúde da Marinha no acompanhamento e controle das atividades para o desenvolvimento do Sistema Informatizado de Gestão de Saúde. Doutoranda pelo PPGI-UFRJ, atua como coordenadora de desenvolvimento no Projeto VODAN BR. Seus temas de interesse estão associados à interoperabilidade entre bases heterogêneas, gestão de metadados, ontologias, modelagem conceitual e web semântica.