



PLANO DE GESTÃO DE DADOS NO APOIO À GESTÃO DE DADOS CIENTÍFICOS FAIR

**Elizabete Cristina de Souza de Aguiar Monteiro¹, Fabiano Couto Corrêa da Silva²,
Marcello Mundim Rodrigues³, Miguel Ángel Márdero Arellano⁴ e
Alexandre Faria de Oliveira⁴**

¹*Universidade Estadual Paulista (Unesp), Brasil*

²*Universidade Federal do Rio Grande do Sul (UFRGS), Brasil*

³*Universidade Federal de Minas Gerais (UFMG), Brasil*

⁴*Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasil*

Resumo: Relata os avanços no desenvolvimento da plataforma que servirá como Cadastro Nacional de Planos de Gestão de Dados, sediada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), em busca de uma ciência brasileira mais aberta e acessível. Este roteiro é desenhado de acordo com tendências globais, padrões-ouro como os Princípios FAIR e as 5 Estrelas para Dados Abertos de Tim Berners-Lee, os quais enquadram as decisões que foram tomadas para se alinhar com a situação das organizações financiadoras de projetos de pesquisa e entidades acadêmicas. Para estruturar o PGD IBICT, analisamos práticas utilizadas no desenvolvimento de ferramentas para PGDs em nível internacional, e descrevemos os principais recursos que foram adaptados ao modelo em construção. Descrevemos, além disso, a experiência e os resultados do projeto piloto, incluindo a template e principais funcionalidades. Finalmente, é apresentado o workflow de funcionamento do PGD IBICT.

Palavras-chave: Dados científicos, gestão de dados, princípios FAIR, plano de gestão de dados, 5 estrelas para dados abertos.

DATA MANAGEMENT PLAN TO SUPPORT FAIR SCIENTIFIC MANAGEMENT DATA

Abstract: It reports on progress in the development of the platform that will serve as the National Registry of Data Management Plans, hosted by the Brazilian Institute for Information in Science and Technology (known as IBICT), in pursuit of a more open and accessible Brazilian science. This roadmap is designed according to global trends, gold standards such as the FAIR Principles and Tim Berners-Lee's 5-Star Open Data, which frame the decisions that have been made to align with the situation of research funding organizations and academic entities. To structure the DMP by IBICT, we analyze practices used in the development of tools for DMPs at the international level, and describe the main features that were adapted to the model under construction. We also describe the experience and results of the pilot project, including the template and main features. Finally, the workflow for the operation of the DMP by IBICT is presented.

Keywords: Scientific data, data management, FAIR principles, data management plan, 5-star open data.

Correspondência para: (correspondence to:) elizabete.monteiro@marilia.unesp.br

INTRODUÇÃO

É consenso entre a comunidade científica que os dados devem ser preservados para garantir o seu acesso no futuro. Há uma grande diversidade de situações sobre gestão

de dados em função das práticas de cada disciplina ou grupo de pesquisadores. Um estudo demonstrou que mais de 60% dos dados científicos estavam presentes unicamente no computador dos

pesquisadores e o risco de perder este conhecimento é muito grande (FERGUSON, 2014). Ademais, gradativamente, governos, financiadores, agências de fomento e universidades estão começando a recomendar e/ou exigir o desenvolvimento de Planos de Gestão de Dados (PGDs) no momento da solicitação de financiamento e, bem como junto aos relatórios parciais e finais, nos casos em que o financiamento é alocado. Um PGD é um documento formal, normalmente elaborado antes do início do projeto, descrevendo o tratamento que os dados de pesquisa científica coletados ou gerados durante o andamento do projeto receberão.

No Brasil, o uso de PGD como requisito obrigatório no recebimento de recursos públicos para realização de uma pesquisa tem se tornado uma realidade. O caso mais expressivo é o da Fundação de Amparo à Pesquisa de São Paulo (FAPESP), que passou a exigir um Plano de Gestão de Dados como condição para financiar pesquisas em todas as áreas e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) solicitando em algumas modalidades. Embora seja um passo importante, considerando que umas das origens dessa exigência é da maior fundação de apoio em âmbito nacional, ainda falta uma estrutura coordenada para reunir e compartilhar esses PGDs.

O Plano de Gestão de Dados FAIR do IBICT é um modelo destinado a qualquer tipo de conjunto de dados, disciplina, projeto de pesquisa e repositório. Baseia-se numa síntese dos requisitos das agências de financiamento, diretrizes institucionais e melhores práticas. A ferramenta tem como principal característica atender aos Princípios FAIR, que definem um conjunto mínimo de princípios e práticas (relacionados, porém independentes e separáveis), os quais permitem que máquinas e humanos encontrem, acessem,

interoperem e reutilizem dados e metadados científicos (RESEARCH DATA ALLIANCE, 2020, p. 4, tradução nossa).

Por meio de pesquisa descritiva, este artigo tem como objetivo apresentar o projeto Plano de Gestão de Dados FAIR do IBICT como ferramenta para elaboração de PGDs e seu *workflow* de funcionamento.

MATERIAIS E MÉTODOS

A pesquisa é de natureza qualitativa do tipo descritiva, com descrição da implementação do PGD IBICT. Foi feito um levantamento dos requisitos técnicos e características para que a ferramenta seja acionável por máquina (maDMP)¹ e atenda os Princípios FAIR. A implantação e testes estão sendo feitos por etapas, alinhando com as necessidades dos fluxos de trabalho e gestão de dados e informações das instituições, iniciando com as agências de fomento.

RESULTADOS

O PGD IBICT é composto por perguntas abertas e fechadas a fim de atender os requisitos para um PGD acionável por máquina (maDMP), com dados estruturados, aos Princípios FAIR, diretrizes a serem seguidas na busca por sistemas e dados encontráveis, acessíveis, interoperáveis e reutilizáveis, e às 5 Estrelas para Dados Abertos de Berners-Lee que tem o objetivo de disponibilizar dados em formatos abertos, assim como mantê-los organizados, acessíveis e interconectados. Buscou-se contemplar todas as etapas necessárias à preservação digital de dados científicos. Para as etapas do preenchimento do PGD, são indicadas as orientações de como atender aos princípios FAIR e às 5 Estrelas para Dados Abertos. Ademais, são apresentados exemplos de preenchimento para as perguntas abertas.

Com o depósito e a descrição dos conjuntos de dados científicos feitos pelo próprio pesquisador, em um repositório e/ou base

que se utilize de auto depósito, ou quando o solicitante ao financiamento preencher os dados na plataforma da agência de fomento, algumas informações dos campos de metadados serão aproveitadas promovendo a interoperabilidade, com a troca de metadados entre as plataformas. Portanto, para que possa se alinhar aos Princípios FAIR, um conjunto de dados precisa ser depositado em um repositório FAIR, como o Dataverse; precisa ser bem descrito; precisa ser indexado; entre outros.

Está previsto que o depósito do PGD IBICT seja em um repositório FAIR, que atribua identificadores persistentes a seus objetos digitais, que publique e disponibilize seus dados e metadados, que restrinja o acesso quando pertinente, que ligue dados e metadados a outros ou a documentos inter-relacionados com o uso de hiperlinks persistentes, entre outros. Os PGDs ficarão armazenados no repositório Dataverse do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Os repositórios implementados por meio do Dataverse estão em maior conformidade com os Princípios FAIR (RODRIGUES; DIAS e LOURENÇO, 2022).

DISCUSSÃO

Para um país com dimensões continentais como o Brasil, um cadastro nacional de PGDs pode ser considerado um recurso estratégico, tendo em vista a necessidade permanente de oferecer visibilidade e transparência sobre as pesquisas realizadas, na mesma medida. Trata-se de um ciclo virtuoso, em que todos são beneficiados, tanto produtores quanto consumidores de dados científicos.

AGRADECIMENTOS

Ao Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) pelo financiamento da pesquisa.

REFERÊNCIAS

FERGUSON, A.; NIELSON, J.; CRAGIN, M.; BANDROWSKI, A. E; MARTONE, M. E. Big data from small data: data-sharing in the 'long tail' of neuroscience. **Nature Neuroscience**, 17, 1442–47, 2014.

Disponível em:

<https://doi.org/10.1038/nn.3838>. Acesso em: 20 abr. 2022.

RESEARCH DATA ALLIANCE. **FAIR data maturity model**: specification and guidelines. RDA FAIR data maturity model Working Group, 2020. Disponível em: DOI: 10.15497/rda00045. Acesso em: 20 abr. 2022.

NOTAS

¹ Machine-actionable Data Management Plan