



# XX ENANCIB

21 a 25 Outubro/2019 – Florianópolis

A Ciência da Informação e a era da Ciência de Dados

ISSN 2177-3688

**GT-7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação**

**DADOS CIENTÍFICOS ABERTOS: GESTÃO BASEADA NOS PRINCÍPIOS FAIR**

**OPEN SCIENTIFIC DATA: MANAGEMENT BASED ON FAIR PRINCIPLES**

Jane Lecardelli - Universidade do Estado de Santa Catarina

Elaine Rosangela de Oliveira Lucas - Universidade do Estado de Santa Catarina

**Modalidade: Resumo Expandido**

**Resumo:** As políticas das agências de fomento à pesquisa passaram a recomendar e/ou exigir um plano de gestão dos dados aos pesquisadores como requisito à obtenção de recursos para as pesquisas. Este trabalho, apresenta, uma proposta metodológica para elaborar um modelo de plano de gestão de dados científicos, baseado nos princípios FAIR. Considerando as opções metodológicas o estudo se caracteriza como pesquisa exploratória, descritiva e de caráter qualitativo. Utilizamos como modelo base o plano de gestão de dados do *Digital Curation Centre* (DCC), incorporando os princípios FAIR como guias no processo de gestão de dados científicos abertos.

**Palavras-Chave:** Ciência aberta; Dados científicos abertos; Plano de gestão de dados; Princípios FAIR.

**Abstract:** Research funding agency policies now recommend and/or require a data management plan for researchers as a requirement for research funding. This paper presents a methodological proposal to elaborate a model of scientific data management plan, based on the FAIR principles. Regarding the methodological options the study is characterized as exploratory, descriptive and qualitative research. We use the Digital Curation Center (DCC) data management plan as a base model, incorporating FAIR principles as guides in the process of open scientific data management.

**Keywords:** Open science; Open scientific data; Data management plan; FAIR Principles.

## 1 INTRODUÇÃO

O resumo apresenta um recorte de pesquisa que vem sendo desenvolvida no Mestrado Profissional em Gestão de Unidades de Informação (UDESC), com a temática geral que envolve dados científicos abertos e princípios FAIR (do acrônimo: *Findable, Accessible, Interoperable e Reusable*) propostos por Wilkinson, *et al*, (2016), traduzidos como Localizável, Acessível, Interoperável e Reutilizável.

O campo científico passa por transformações e, para que a sociedade de fato se aproprie dos resultados das pesquisas oriundas de investimentos públicos, surge um movimento global de incentivo às ações em prol da ciência aberta, em que o acesso aberto aos dados científicos é priorizado.

As demandas da sociedade por transparência e sustentabilidade, aliadas às novas políticas das agências de fomento para que os dados científicos sejam abertos, passam a recomendar e/ou exigir a apresentação de plano de gestão de dados como requisito para submissão de projetos e obtenção de recursos para pesquisas.

Para que a gestão dos dados seja eficaz e os dados tratados se mantenham íntegros para o reuso com segurança, existem recomendações de boas práticas em gestão, dentre elas, destacam-se os princípios FAIR. Estes princípios recomendam que os dados científicos devam manter uma identificação consistente no decorrer do tempo que permita interpretá-los de forma inequívoca para novas pesquisas. É fundamental ter metadados detalhados de significação e contexto, considerando a complexidade e a diversidade dos dados científicos em diferentes domínios, além de estruturas que permitam a interoperabilidade e as devidas atribuições de licenças para que os dados estejam de fato abertos e prontos para reuso.

Não ter os dados científicos bem documentados pode impactar no desenvolvimento de novas pesquisas, acarretar custos extras para gerar novos dados (por vezes nem sendo possível esta alternativa), desperdiçar tempo e ocasionar retrabalho aos pesquisadores. Para enfrentar tais desafios, delinea-se o objetivo principal do estudo: propor um modelo para elaboração de plano de gestão de dados científicos baseado nos princípios FAIR, a partir de análise das políticas e recomendações de Agências de fomento nacionais e internacionais.

Observa-se que para implementar os princípios FAIR globalmente é importante a adoção de requisitos comuns que facilitem a descoberta, o acesso, a interoperabilidade e o compartilhamento dos dados para o reuso.

Este resumo busca detalhar as opções metodológicas propostas para atingir o objetivo principal da pesquisa e apresenta, também, breve contexto teórico da temática.

## **2 CIÊNCIA ABERTA: Dados científicos, Gestão, Plano de Gestão e Princípios FAIR**

A Ciência Aberta (*open science*) é retratada por um conceito que representa uma abertura para a democratização do conhecimento. Muitos são os benefícios da ciência aberta a todos os envolvidos, tanto aos pesquisadores e às instituições de pesquisa e

fomento, quanto para a sociedade ao promover a qualidade e eficiência da pesquisa e produção de novos conhecimentos.

Santos (2017, p.12) enfatiza como principais benefícios: “reprodutibilidade, transparência científica, velocidade de circulação da informação e reuso de dados, resultando numa ciência de maior qualidade e progressos mais rápidos”.

Dentre os benefícios citados, destaca-se o reuso de dados como atividade fundamental. Segundo Curty (2016, p. 03), o reuso de dados científicos é definido:

como toda nova aplicação de dados por meio de [re-análise] e replicação, ou como a combinação de diferentes conjuntos de dados por meio de integração ou meta-análise, a partir de novas perguntas de pesquisa, novos métodos de análise, com propósitos similares ou distintos daqueles empreendidos no estudo original, com ou sem a participação do reutilizador.

O reuso de dados potencializa novos estudos sob diferentes vieses, minimiza custos, favorece a validação das pesquisas, proporciona maior visibilidade aos pesquisadores originais, o que pode favorecer novas parcerias e recursos para projetos futuros, contribuindo, inclusive, com o desenvolvimento econômico e o bem-estar social.

Os dados científicos são considerados protagonistas no movimento em prol da ciência aberta. A Organização para a Cooperação e Desenvolvimento Econômico (OECD) define dados de pesquisa como

[...] registros factuais usados como fontes primárias na pesquisa científica, e que são geralmente aceitos na comunidade científica como sendo necessários para validar os resultados de pesquisa. Um conjunto de dados de pesquisa constitui uma representação parcial e sistemática do objeto de investigação. (OECD, 2007 *apud* BERTIN; VISOLI; DRUCKER, 2017, p. 38).

O valor dos dados gerados no decorrer das pesquisas é pautado em sua possibilidade de replicação e reuso, pelo contexto histórico, econômico e social. Os dados são heterogêneos e podem variar conforme os domínios científicos a que pertencem. Também diferem pela forma de geração, por exemplo, coletados via sensores ou através de entrevistas, processos laboratoriais ou amostras únicas, por isso a importância em registrar o contexto para facilitar o entendimento por grupos distintos de pesquisadores e interessados ao longo do tempo.

Os autores Rocha, Sales e Sayão (2017) observam que reutilizar dados científicos estruturados evita retrabalho, acelera os avanços da ciência, além de se consolidar como boa prática no meio científico.

Para Silva (2016, p. 391), “a gestão de dados científicos está se tornando um requisito dos financiadores e instituições que pagam para coleta de dados, uma vez que demonstram preocupação com a disponibilidade dos dados no futuro”. As agências de fomento à pesquisa no país já sinalizam para esta tendência, por ser economicamente mais viável e porque os dados podem ser validados e testados por outros pesquisadores em novos cenários e até novas áreas, por meio do reuso de dados.

O plano de gestão de dados (PGD) é um requisito que agências de fomento estão adotando para que os dados de pesquisas financiadas sejam planejados pelos responsáveis, desde o início do ciclo. De forma geral, o plano deve contemplar questões como: quais dados serão gerados e como serão preservados, compartilhados e disponibilizados, considerando questões éticas, legais, de confidencialidade e outras. (FAPESP, 2019).

Como é uma iniciativa recente no Brasil, está no patamar de recomendação para compor os projetos de pesquisa. Porém, já existem agências de fomento que condicionam a aprovação de projetos ao envio do plano de gestão de dados como elemento obrigatório, é o caso da FAPESP, por exemplo, para propostas de pesquisas em determinadas áreas. O Plano de Gestão de Dados (PGD) se constitui de um documento que

descreve que dados serão coletados ou gerados; quais as metodologias e padrões que serão utilizados nesses processos; se, como e sob que condições esses dados serão compartilhados e/ou tornados abertos para a comunidade de pesquisa; e como eles serão curados e preservados. (SIBi/USP, 2019).

Para que o PGD possa ser efetivo, “adotar os princípios FAIR é uma boa maneira de manter os dados disponíveis, acessíveis para reuso por outros interessados, mesmo que não sejam da comunidade científica”. (HENNING, *et al*, 2018, p. 5060). A ideia é que qualquer interessado possa encontrar, acessar e utilizar o material, mesmo que seja de forma diferente do original, desde que respeite as questões éticas, legais e as atribuições de licenças.

No país ainda são recentes as iniciativas das agências em solicitar a elaboração dos PGD como requisito em projetos de pesquisa, portanto, um dos objetivos deste estudo é verificar se as agências de fomento selecionadas recomendam a adoção dos princípios FAIR. Destaca-se ser uma tendência no cenário internacional usar os princípios FAIR como base para que os dados possam estar acessíveis, interoperáveis e sejam citáveis no processo de compartilhamento e reuso.

Estes princípios constam na literatura internacional desde 2014, quando ocorreram discussões acerca dos novos contextos gerados pela *e-Science*, e foram publicados em 2016 por Wilkinson, *et al.* após discussões realizadas por especialistas de diferentes áreas de domínio. Surgem como balizadores para facilitar a descoberta, interoperabilidade e reuso dos dados científicos.

Os princípios FAIR representam “um conjunto mínimo de princípios orientadores e práticas aceitas pela comunidade para que os produtores e os usuários, humanos ou computadores, [possam] usar mais facilmente os dados e citá-los corretamente.” (AVENTURIER, 2017, não paginado, tradução nossa).

Percebe-se, no entanto, que a conscientização e o uso de infraestruturas apropriadas, bem como o investimento na curadoria de dados, são requisitos importantes na implementação dos princípios FAIR, assim como o apoio e financiamento para reunir, de forma abrangente, as melhores práticas em gestão de dados científicos em domínios diferentes.

### **3 OPÇÕES TEÓRICO-METODOLÓGICAS: a proposta**

A pesquisa caracteriza-se como exploratório-descritiva, e para as análises e inferência pretende-se utilizar a técnica da análise de conteúdo. A pesquisa busca uma análise detalhada das informações nos *sites* oficiais das agências de fomento selecionadas. O critério de seleção das Agências no *ranking* se pautou nas iniciativas/políticas de gestão de dados abertos identificadas conforme estudos de Monteiro (2019). O *corpus* da pesquisa foi composto por dez agências de fomento à pesquisa, sendo cinco nacionais (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo à Pesquisa do Estado de São Paulo, Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Empresa Brasileira de Pesquisa Agropecuária, Fundação Oswaldo Cruz), e cinco internacionais (National Institutes of Health, World Health Organization, European Union, Fundação para a Ciência e a Tecnologia de Portugal, National Research Foundation). O recorte das dez agências se fez para possibilitar aprofundar as análises conforme a metodologia proposta e do tempo disponível para o estudo.

Utilizar-se-á da pesquisa bibliográfica em bases de dados nacionais e internacionais da área de Ciência da Informação e interdisciplinares para compor o referencial teórico. Quanto à abordagem, terá caráter qualitativo. Como instrumento de coleta de dados buscar-se-á fazer levantamento das informações por meio de consulta aos *sites* oficiais das Agências

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019  
21 a 25 de outubro de 2019 – Florianópolis – SC**

selecionadas. Para executar as análises se pretende utilizar o *checklist* traduzido e adaptado da *Digital Curation Centre* (Quadro 01) como base para as verificações das práticas das Agências, e os princípios e facetas FAIR (Quadro 02) como elementos que devem ser contemplados na elaboração do Plano de Gestão de Dados, objetivo central do estudo.

Apresentam-se, no quadro 01, as categorias propostas pelo DCC e *checklist* com perguntas e informações de cada item que compõe o PGD. Intenta-se fazer a análise com documentos de cada uma das dez Agências para verificar se atendem ou não aos critérios propostos pelo DCC. No quadro 02 elencam-se os princípios e facetas FAIR para a análise em cada Agência selecionada. Para ambas as análises (quadros 01 e 02) os documentos serão consultados e será feita a verificação se: Atende totalmente (AT), Atende parcialmente (AP), Não atende (NA) aos critérios ou se a informação Não foi identificada (NI). Após a coleta, os dados serão categorizados e analisados por meio da análise de conteúdo e os resultados apresentados através de infografia e texto.

**Quadro 01 – Proposta de análise dos documentos de cada Agência – Plano de Gestão de dados.**

Planos de Gestão de Dados (PGD) - (Síntese - checklist das perguntas a serem consideradas - Modelo DCC).		Agências Nacionais e Internacionais		
		Agência 1 ...	PGD	Edição
<b>Dados administrativos</b>	Um ID pertinente, conforme determinado pelo financiador e / ou instituição. <b>Descrição do Projeto. // Políticas relacionadas.</b> - Existem procedimentos nos quais você baseará sua abordagem? - O seu departamento / grupo tem diretrizes de gerenciamento de dados? - Sua instituição ou financiador possui uma política de Gestão de Dados?			
<b>Coleta de dados</b>	<b>Quais dados você coletará ou criará?</b> - Que tipo, formato e volume de dados? - Os formatos e softwares escolhidos permitem o compartilhamento e o acesso de longo prazo aos dados? - Existem dados que você pode reutilizar? <b>Como os dados serão coletados ou criados?</b> - Quais padrões ou metodologias você usará? - Como você vai lidar com versionamento? - Quais processos de garantia de qualidade você adotará?			
<b>Documentação e Metadados</b>	Que documentação e metadados acompanharão os dados? - Quais informações são necessárias para que os dados sejam lidos e interpretados no futuro? - Como você vai capturar/criar esta documentação e metadados? - Quais padrões de metadados você usará e por quê?			
<b>Ética e conformidade legal</b>	Você obteve permissão para preservação e compartilhamento de dados? - Como você protegerá a identidade dos participantes, se necessário? Por exemplo, via anonimização. - Como os dados confidenciais serão tratados para garantir que eles sejam armazenados e transferidos com segurança? Como você vai gerenciar os direitos autorais e questões de direitos de propriedade intelectual (IPR)? - Como os dados serão licenciados para reutilização? - Existem restrições à reutilização de dados de terceiros? - O compartilhamento de dados será adiado / restrito?			
<b>Armazenamento e Backup</b>	<b>Como os dados serão armazenados durante a pesquisa?</b> - Você tem armazenamento suficiente ou precisa incluir taxas para serviços adicionais? - Quem será responsável pelo backup e recuperação? - Como os dados serão recuperados no caso de um incidente? <b>Como você vai gerenciar o acesso e a segurança?</b> - Quais são os riscos para a segurança de dados e como eles serão gerenciados? - Como você controlará o acesso para manter os dados seguros? - Ao criar ou coletar dados no campo, como você garantirá sua transferência segura para seus principais sistemas protegidos? <b>Obs:</b> Se você optar por usar um serviço de terceiros, certifique-se de que isso não conflite com nenhuma política de financiador, institucional, departamental ou de grupo, por exemplo, em termos de jurisdição legal em que os dados são mantidos ou da proteção de dados confidenciais.			
<b>Seleção e Preservação</b>	Quais dados devem ser retidos/destruídos para fins contratuais, legais ou regulamentares? - Como vai decidir quais outros dados manter? Quais são os usos de pesquisa previsíveis para os dados? Quanto tempo os dados serão retidos e preservados? <b>Qual é o plano de preservação a longo prazo do conjunto de dados?</b> - Onde e em qual repositório ou arquivo os dados serão mantidos? - Quais custos, se houver, serão o repositório de dados ou a taxa de arquivamento selecionada?			
<b>Compartilhamento de dados</b>	Por quanto tempo você precisa de uso exclusivo dos dados e por quê? - Será necessário um contrato de compartilhamento de dados (ou equivalente)?			
<b>Responsabilidades e Recursos</b>	Quem é responsável pela implementação do PGD e por assegurar que seja revisado e revisado? - Quem será responsável por cada atividade? - Como as responsabilidades serão divididas em projetos de pesquisa colaborativa? - <b>Quais recursos você precisará para entregar seu plano?</b> - É necessário especialista adicional (ou treinamento para o pessoal existente)? - Você precisa de hardware ou software que seja adicional ou excepcional para a provisão institucional existente? - As cobranças serão aplicadas pelos repositórios de dados?			
<b>Informações em relação ao PGD para cada agência. Se:</b>				
<b>Atende totalmente (AT)</b>	<b>Atende parcialmente (AP)</b>			
<b>Não atende (NA)</b>	<b>Não identificada (NI)</b>			
Fonte: Adaptado e traduzido de DCC (2013). Disponível em: <a href="http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf">http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf</a> .				

**Quadro 02 – Proposta de análise dos documentos de cada Agência, em atenção aos princípios FAIR**

Princípios e Facetas - FAIR	Agências Nacionais e Internacionais		
	Agência1 ...		
	Doc1	Doc2	Doc3
<b>Para serem encontráveis: <i>Findable</i></b>			
F1. (Metadados) são atribuídos identificadores globais, persistentes e identificáveis			
F2. Os dados são descritos com ricos metadados (definidos no R1 abaixo)			
F3. Os Metadados incluem claramente e explicitamente os identificadores dos dados que os descrevem			
F4. (Metadados) são registrados ou indexados por intermédio de um recurso pesquisável			
<b>Para serem acessíveis: <i>Accessible</i></b>			
A1. (Metadados) são recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado			
A1.1 O protocolo é aberto, gratuito e universalmente implementável			
A1.2 O protocolo permite procedimentos de autenticação e autorização, quando necessário			
A2. (Metadados) são acessíveis, mesmo quando os dados não estão mais disponíveis			
<b>Para serem interoperáveis: <i>Interoperable</i></b>			
I1. (Metadados) usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.			
I2. (Metadados) usam vocabulários que seguem os princípios FAIR			
I3. (Metadados) incluem referências qualificadas para outros (metadados)			
<b>Para serem reutilizáveis: <i>Reusable</i></b>			
R1. (Metadados) são ricamente descritos com uma pluralidade de atributos precisos e relevantes			
R1.1. (Metadados) são liberados com licenças de uso de dados claras e acessíveis			
R1.2. (Metadados) estão associados a precedências detalhadas			
R1.3. (Meta)dados encontram domínios relevantes de padrões comunitários.			
<b>Documentos de cada Agência selecionada/em relação as facetas do FAIR. Avaliar se:</b>			
<b>Atende totalmente (AT)</b>	<b>Não atende (NA)</b>		
<b>Atende parcialmente (AP)</b>	<b>Não identificada (NI)</b>		
Fonte: Adaptação de Wilkinson, <i>et al</i> (2016) com tradução de Santos (2017, p.55-56).			

Os procedimentos metodológicos propostos aliados à análise de conteúdo permitirão ao final do processo, além de atingir os objetivos definidos, gerar um guia (produto) que possa fornecer subsídios para que profissionais se apropriem de informações relevantes no cenário da gestão de dados científicos abertos e, com isso, possam enfrentar os desafios no contexto da ciência aberta.

## REFERÊNCIAS

AVENTURIER, P. **Princípios FAIR**: critérios de qualidade para dados de pesquisa. 2017. Disponível em: <https://publicient.hypotheses.org/1456>. Acesso em: 04 jun. 2019.

BERTIN, P.R.B.; VISOLI, M. C.; DRUCKER, D. P. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de P&D. **Ponto de Acesso**, Salvador, v.11, n.2, p. 34-48, ago. 2017. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1081741/1/Agestaodedadosdepesquisa....pdf>. Acesso em: 27 mar. 2019.

CURTY, R. G. A. As diferentes dimensões do reuso de dados científicos. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 9, n. 2, 2016. Disponível em: <http://www.ufpb.br/evento/index.php/enancib2016/enancib2016/paper/view/4056>. Acesso em: 29 jan. 2019.

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

DCC. Digital Curation Centre. **Checklist for a Data Management Plan**. v.4.0. 2013. Disponível em: <http://www.dcc.ac.uk/resources/data-management-plans>. Acesso em: 29 jun. 2019.

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. **Plano de Gestão de Dados – FAPESP**. Disponível em: <http://www.fapesp.br/gestaodedados/>. Acesso em: 10 mar. 2019.

HENNING, P., *et al.* Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. In: Encontro Nacional de Pesquisa em Ciência da Informação, ENANCIB, 19. **Anais...** Londrina, 2018, UEL, 22 a 26 out.2018. p.5047-5067. Disponível em: <http://www.brapci.inf.br/index.php/res/v/103243>. Acesso em: 10 mar. 2019.

MONTEIRO, G. **Mapeamento e análise das políticas institucionais de financiadores da pesquisa brasileira**: cenário dos dados científicos abertos. 2019. 132 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Gestão da Informação, do Centro de Ciências Humanas e da Educação, Universidade do Estado de Santa Catarina (UDESC), Florianópolis, 2019.

ROCHA; L. L.; SALES, L. F.; SAYÃO, L. F. Uso de cadernos eletrônicos de laboratório para as práticas de ciência aberta e preservação de dados de pesquisa. **Ponto de Acesso**, Salvador, v.11, n.3, p. 2-16, dez. 2017. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/24945>. Acesso em: 25 abr. 2019.

SANTOS, P. X. (coord.). **Livro Verde - Ciência aberta e dados abertos**: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Rio de Janeiro: Fiocruz, 2017. 141 p. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/24117/2/Livro-Verde-07-06-2018.pdf>. Acesso em: 10 mar. 2019.

SIBi/USP – Sistema Integrado de Bibliotecas da USP. **Plano de Gestão de dados**. Disponível em: <https://www.sibi.usp.br/apoio-pesquisador/dados-pesquisa/plano-gestao-dados-2/>. Acesso em: 25 abr. 2019.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n.1, p. 160018, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 25 fev. 2019.