



UNIVERSIDADE DO ESTADO DE SANTA CATARINA - UDESC
CENTRO DE CIÊNCIAS HUMANAS E DA EDUCAÇÃO – FAED
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO DA INFORMAÇÃO – PPGINFO

DISSERTAÇÃO DE MESTRADO
DADOS CIENTÍFICOS ABERTOS EM AGÊNCIAS DE FOMENTO À PESQUISA: CENÁRIO DOS PLANOS DE GESTÃO DE DADOS (PGD) E PRINCÍPIOS FAIR

JANE LECARDELLI

Florianópolis, SC

2020

JANE LECARDELLI

**DADOS CIENTÍFICOS ABERTOS EM AGÊNCIAS DE FOMENTO À PESQUISA:
CENÁRIO DOS PLANOS DE GESTÃO DE DADOS (PGD) E PRINCÍPIOS FAIR**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação - PPGInfo, do Centro de Ciências Humanas e da Educação – FAED, da Universidade do Estado de Santa Catarina – UDESC, como requisito para obtenção ao grau de mestra em Gestão de Unidades de Informação.

Linha de pesquisa: Gestão da Informação

Orientadora Prof^ª: Dra. Elaine Rosangela de Oliveira Lucas

Florianópolis, SC

2020

L456d Lecardelli, Jane

Dados Científicos Abertos em Agências de Fomento à Pesquisa:
Cenário dos Planos de Gestão de Dados (PGD) e Princípios FAIR /
Jane Lecardelli. – Florianópolis, SC, 2020.

119 f.; 30 cm.

Orientadora: Profa. Dra. Elaine Rosangela de Oliveira Lucas.

Dissertação (Mestrado) -- Universidade do Estado de Santa Catarina,
Centro de Ciências Humanas e da Educação, Programa de Mestrado
Profissional em Gestão de Unidades de Informação, Florianópolis, SC,
2020.

Inclui ilustrações

1. Dados científicos abertos. 2. Gestão de dados. 3. Princípios FAIR. I.
Lucas, Elaine Rosangela de Oliveira. II. Universidade do Estado de
Santa Catarina, Centro de Ciências Humanas e da Educação, Programa
de Pós-Graduação Profissional em Gestão de Unidades de Informação.
IV. Título.

CDD 020

Ficha catalográfica elaborada pela bibliotecária Gilvane Maria Lecardelli (CRB 14/1226)

JANE LECARDELLI

**DADOS CIENTÍFICOS ABERTOS EM AGÊNCIAS DE FOMENTO À PESQUISA:
CENÁRIO DOS PLANOS DE GESTÃO DE DADOS (PGD) E PRINCÍPIOS FAIR**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação, do Centro de Ciências Humanas e da Educação da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do título de Mestra em Gestão de Unidades de Informação.

Banca Examinadora:

Orientadora:

Prof^a. Dra. Elaine Rosângela de Oliveira Lucas
Universidade do Estado de Santa Catarina (UDESC)

Membros:

Prof. Dr. Fabiano Couto Corrêa da Silva
Universidade Federal do Rio Grande do Sul (UFRGS)

Prof^a. Dra. Paula Carina Araújo
Universidade do Estado de Santa Catarina (UDESC)

Florianópolis, 22 de outubro de 2020.

“Observar pelo canto do olho é, em ciência, começar a elaborar a hipótese. O que é observado pelo centro do olho é o evidente, o óbvio, aquilo que é partilhado pela multidão.

Na Ciência, como no mundo das invenções, observar pelo canto do olho é ver o pormenor diferente, aquele que é o começo de qualquer coisa de significativo.

Observar a realidade pelo canto do olho, isto é: pensar ligeiramente ao lado. A isto chama-se criatividade. Daqui saíram todas as teorias científicas importantes.”

(Gonçalo M. Tavares. Notas sobre Ciência).

Aos recomeços da vida...

AGRADECIMENTOS

Não poderia iniciar agradecendo a outra pessoa, por ser tão singular e importante nesta etapa da minha vida, minha orientadora, Professora Lani Lucas, por ter me conduzido neste percurso de estudos de forma a motivar e amar cada vez mais a pesquisa. Pelas longas conversas, orientações, apoio e suporte para o estágio em Madrid. Além, de todas as palavras de carinho e incentivo em momentos pessoais conturbados durante o processo de produção/escrita desta dissertação. Ela foi além de ser a orientadora de mestrado, foi uma incentivadora, me motivou a seguir! Essa aventura não teria sido tão maravilhosa sem sua presença e as suas dicas fantásticas! Espero que essa parceria continue! Gratidão!

Agradecer aos amigos “Desorientados” com os quais fui privilegiada em poder compartilhar momentos durante o mestrado, pela parceria, convivência, estudos, discussões, aprendizados, risadas, aventuras... Condições *sine qua non* para alegrar meus dias!

Ao Toni Picalho por propiciar minha “inclusão no mundo digital”, pelas dicas tecnológicas e em especial pelas conversas! Era sempre um prazer te encontrar!

À Gabriela Monteiro pela parceira de pesquisa, de conversas aleatórias e de estudos, com quem compartilhei minhas inquietações e momentos de *insights*... Ela desbravou aspectos de estudo na temática e me deu a possibilidade de analisar outros vieses a partir de suas descobertas. Tornou-se uma amiga especial!

À Helouise Viola que foi uma parceria desde meu início no mestrado, na monitoria, nos estudos, nos papos adoráveis sobre tudo, desbravando e aprendendo juntas nas trocas de experiências sobre o mestrado. Todos vocês fizeram meus dias de aprendizado mais felizes, com toda certeza!

Agradecer ao amigo Werton pelo auxílio na parte gráfica e programação da ferramenta em Python, além das boas conversas, atenção e bom humor sempre presentes!

À amiga Ana Dantas pela parceria de longa data, incentivo e por toda ajuda com as imagens, de forma rápida e tão atenta a todos os detalhes! Muito obrigada!

À Adriana Cativelli pelo tempo em Madrid, por literalmente me conduzir pelos “caminhos dentro da UC3M” e que também foi uma parceria nos descobrimentos da cultura espanhola, com quem tive momentos muito divertidos e excelentes recordações.

Agradecer a oportunidade de poder conhecer e aprender na Universidad Carlos III de Madrid, em especial a professora Virgínia Ortiz-Repiso, minha tutora pelas contribuições. Ao Professor Carlos García-Zorita (UC3M) pelas conversas sobre cultura, a universidade e pela

extrema gentileza de nos apresentar a culinária típica, juntamente com o Sérgio Marugán Lázaro, sempre disposto a auxiliar com a maior atenção, e apresentar novas ferramentas e possibilidades. Ao pesquisador Ivan, por me auxiliar a desvendar um pouco mais os princípios FAIR, e também, por sua extrema gentileza em nos situar nos costumes e festividades locais no final de 2019... Enfim, a todos da UC3M que tive o prazer de conviver, sempre vou lembrar as nossas conversas e almoços com muito carinho, vocês, fizeram com que minha estadia se tornasse inesquecível, me encantei pela Espanha, sua cultura, costumes e histórias! Hasta Luego!

Ao PPGInfo representado pela coordenação e pelos professores sempre empenhados em nos proporcionar aulas de excelência. Aos servidores técnicos, em especial ao Christian e Mateus por toda atenção, apoio e disponibilidade em auxiliar durante estes dois anos, muito obrigada!

Agradecer aos Colegas da ‘Turma 2018’, pela troca de experiências, conversas, estudos colaborativos, discussões, aprendizados que permitiram ampliar o repertório de experiências acadêmicas.

De forma especial, agradecer a minha família pelo amor incondicional mesmo à distância. Minha mãe por ser tão maravilhosa e amorosa, por me apoiar e saber estar sempre presente com suas palavras de carinho, ânimo e alegria. Minha irmã Vane pelo carinho e amor, pelas intermináveis conversas e risadas das muitas histórias! Aos meus irmãos Adilson e Gilberto e ao meu Pai, vocês são indispensáveis na minha vida!

A todos que acompanharam este processo de perto ou longe, entenderam ausências, apoiaram, questionaram, sugeriram, alegraram, divertiram, auxiliaram com a pesquisa de alguma forma, minha gratidão!

Esta dissertação é a concretude de um sonho e de um tempo de trabalho, que apesar de todos os desafios já deixa saudades.

“O motor da ciência não é o conhecimento, e sim a falta dele”.
(Stuart Firestein, 2019).

RESUMO

As 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 suas pesquisas. Este estudo buscou mapear o cenário dos Planos de Gestão de Dados científicos, e dos Princípios FAIR a partir de análises aos documentos em Agências que atuam com fomento em áreas de domínio multidisciplinares. Quanto às opções e procedimentos metodológicos o estudo se caracteriza como pesquisa exploratória, descritiva, de caráter qualitativo. Utiliza como base de análise os critérios do Plano de Gestão de Dados (PGD) do Digital Curation Centre (DCC) e investigou se os modelos abordam os Princípios FAIR como norteadores no processo de gestão de dados científicos abertos. Para análise do PGD os oito critérios analisados foram: dados administrativos, coleta de dados, documentação e metadados, ética e conformidade legal, armazenamento e backup, seleção e preservação, compartilhamento dos dados e responsabilidades e recursos. Dentre as seis Agências selecionadas, três são nacionais e três internacionais. Nos documentos analisados das Agências nacionais foram identificadas exigências/recomendações que atendem aos critérios do PGD, sendo a FAPESP com maior incidência atendendo total/majoritariamente aos oito critérios, seguida do CNPq com atendimento parcial a todos os critérios e a FAPEMIG com atendimento parcial a três dos oito critérios. Nas Agências internacionais tem-se a EU com atendimento total/majoritário a todos os critérios analisados, a FCT com atendimento total/majoritário a seis e parcial a dois critérios, e a NRF atendendo de forma parcial a sete critérios e um não foi identificada informação. Em relação aos Princípios FAIR nas três Agências nacionais, não foram identificadas informações sobre sua menção junto ao PGD. Nas internacionais os Princípios FAIR têm atendimento total/majoritário na EU com política específica para esta finalidade e nas Agências FCT e NRF não foram identificadas informações específicas sobre os subprincípios que compõem o FAIR. A análise buscou identificar menção/adesão a cada um dos quinze subprincípios. Os resultados demonstram que o maior percentual de cobertura dos critérios ao PGD está entre as Agências EU e FAPESP com atendimento total/majoritário a todos os critérios, seguidos da FCT, CNPq, NRF e FAPEMIG, respectivamente. Quanto aos princípios FAIR a única Agência que apresenta atendimento total/majoritário a todos os princípios é a EU. Em todas as demais Agências não foram identificadas informações específicas aos princípios e/ou subprincípios FAIR.

Palavras-chave: Ciência aberta. Dados Científicos Abertos. Plano de Gestão de Dados. Princípios FAIR.

ABSTRACT

The Research Promotion Agencies started to recommend and / or require a Data Management Plan to researchers as a requirement to obtain resources for their researches. This study sought to map the scenario of scientific Data Management Plans, and the FAIR Principles based on analysis of documents in Agencies that work with promotion in multidisciplinary areas. As for the methodological options and procedures, the study is characterized as exploratory, descriptive, qualitative research. It uses as a basis of analysis the criteria of the Data Management Plan (PGD) of the Digital Curation Center (DCC) and investigated whether the models address the FAIR Principles as guidelines in the process of open scientific data management. For the PGD analysis, the eight criteria analyzed were: administrative data, data collection, documentation and metadata, ethics and legal compliance, storage and backup, selection and preservation, data sharing and responsibilities and resources. Among the six selected Agencies, three are national and three international. In the documents analyzed by the national agencies, requirements / recommendations that meet the PGD criteria were identified, with FAPESP having the highest incidence meeting all / mostly of the eight criteria, followed by CNPq with partial service to all criteria and FAPEMIG with partial service to three of the eight criteria. The international agencies have the EU with total / majority service to all criteria analyzed, the FCT with total / majority service to six and partial to two criteria, and the NRF partially meeting seven criteria and one was not identified. Regarding the FAIR Principles in the three national Agencies, information about their mention with the PGD was not identified. In the international ones, the FAIR Principles have full / majority assistance in the EU with a specific policy for this purpose and in the FCT and NRF Agencies, no specific information was identified on the subprinciples that make up the FAIR. The analysis sought to identify mention / adherence to each of the fifteen subprinciples. The results show that the highest percentage of coverage of the criteria for the PGD is between the EU and FAPESP agencies with full / majority compliance with all criteria, followed by FCT, CNPq, NRF and FAPEMIG, respectively. As for the FAIR principles, the only Agency that has full / majority compliance with all the principles is the EU. In all the other Agencies, no specific information was identified regarding the FAIR principles and / or subprinciples.

Keywords: Open science. Open scientific data. Data Management Plan. FAIR Principles.

LISTA DE FIGURAS

Figura 01 - Estrutura da dissertação	23
Figura 02 - Pilares da ciência aberta.....	24
Figura 03 – Espectro dos dados.....	31
Figura 04 – Aspectos para dados de qualidade.....	33
Figura 05 - Ciclo de vida dos dados abertos.....	36
Figura 06 – Princípios para bons metadados	45
Figura 07 – Você conhece os Princípios FAIR?.....	50
Figura 08 – Etapas das Opções e Procedimentos Metodológicos.	62
Figura 09 – Posicionamento geográfico das Agências estudadas	88
Figura 10 – Representação dos resultados nas três Agências Nacionais – PGD.....	92
Figura 11 – Representação dos resultados nas três Agências Nacionais - Princípios FAIR.....	93
Figura 12 – Representação dos resultados das três Agências Internacionais – PGD.....	97
Figura 13 – Representação dos resultados nas três Agências Internacionais – Princípios FAIR.....	98
Figura 14 - Infografia para Projetos de Pesquisa - PGD com abordagem FAIR.....	101

LISTA DE QUADROS

Quadro 01 - Repositórios brasileiros cadastrados no Re3data	41
Quadro 02 - Princípios FAIR.....	48
Quadro 03 – Comparativo - Dados <i>Big Data</i> e Cauda Longa	51
Quadro 04 - Indicação para adoção dos Princípios FAIR na gestão de dados	52
Quadro 05 - Termos de busca para levantamento bibliográfico.....	55
Quadro 06 - <i>Ranking</i> das Agências Nacionais que mais fomentam pesquisas brasileiras	56
Quadro 07 - <i>Ranking</i> das Agências Internacionais que mais fomentam pesquisas no Brasil	57
Quadro 08 - Composição do <i>Corpus</i> de análise da Pesquisa: Agências de Fomento	57
Quadro 09 – Ferramenta Python – Quantitativo de documentos PDFs analisados.....	58
Quadro 10 – Plano de Gestão de Dados – Síntese de questões orientadoras do <i>Checklist</i> DCC	60
Quadro 11 - Modelo de análise - Princípios FAIR e subprincípios para os documentos de cada Agência.....	61
Quadro 12 – Análise dos critérios em atenção ao Plano de Gestão de Dados - CNPq	64
Quadro 13 – Análise à menção dos Princípios FAIR – CNPq	65
Quadro 14 – Análise dos critérios em atenção ao Plano de Gestão de dados – FAPESP	68
Quadro 15 – Análise à menção dos Princípios FAIR – FAPESP.....	69
Quadro 16 – Análise dos critérios em atenção ao Plano de Gestão de dados – FAPEMIG.....	71
Quadro 17 – Análise à menção dos Princípios FAIR – FAPEMIG	73
Quadro 18 – Análise dos critérios em atenção ao Plano de Gestão de dados - (EU).....	75
Quadro 19 – Análise à menção dos Princípios FAIR – EU	77
Quadro 20 – Análise dos critérios em atenção ao Plano de Gestão de Dados - FCT.....	80
Quadro 21 – Análise à menção dos Princípios FAIR – FCT	82
Quadro 22 – Análise dos critérios em atenção ao Plano de Gestão de Dados - NRF	85
Quadro 23 – Análise à menção dos Princípios FAIR - NRF.....	86
Quadro 24 – Resultado da análise aos critérios do PGD – CNPq.....	89
Quadro 25 – Resultado da análise aos critérios do PGD – FAPESP.....	90
Quadro 26 – Resultado da análise aos critérios do PGD - FAPEMIG	91
Quadro 27 – Resultado da análise aos critérios do PGD – EU.....	94
Quadro 28 – Resultado da análise aos critérios do PGD – FCT.....	95
Quadro 29 – Resultado da análise aos critérios do PGD – NRF	96

LISTA DE SIGLAS E TERMOS

AA	Acesso Aberto
BDTD	Base de Teses e Dissertações
BENANCIB	Base dos Encontros Nacionais de Pesquisa e Pós-Graduação em Ciência da Informação
BOAI	Budapest Open Access Initiative
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação da Universidade do Paraná
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COAR	Confederation of Open Access Repositories
ConfOA	Conferência Luso-Brasileira sobre Acesso Aberto
CONFAP	Conselho Nacional das Agências de Financiamento do Estado
DCC	Digital Curation Centre
DCMI	Dublin Core Metadata Initiative
DMP	Data Management Plan
EC	European Commission
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
EOSC	European Open Science Cloud.
EU	European Union
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FAIR	Findable, Accessible, Interoperable, Reusable.
FCT	Fundação para a Ciência e a Tecnologia de Portugal
Finep	Financiadora de Estudos e Projetos
FIOCRUZ	Fundação Oswaldo Cruz
FORCE11	Uma comunidade de acadêmicos, bibliotecários, arquivistas, editores e financiadores de pesquisa, que surgiu organicamente para ajudar a facilitar a mudança para uma melhor criação e compartilhamento de conhecimento, iniciada em 2011.
FOSTER	Facilitate Open Science Training for European Research
GT-PNCA	Grupo de Trabalho para a Política Nacional de Ciência Aberta (Portugal)
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
LISA	Library and Information Science Abstracts

LISTA	Library Information Science & Technology Abstracts
MCTI	Ministério da Ciência, Tecnologia e Inovação
MCTES	Ministério da Ciência, Tecnologia e Ensino Superior da República Portuguesa
MOOCs	Massive Open Online Courses
NIH	National Institutes of Health
NRF	National Research Foundation
OAD	Open Access Directory
OAI	Open Archives Initiative
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
ODI	Open Data Institute
OECD	Organization for Economic Cooperation and Development
OPIDoR	Otimização do Compartilhamento e Interoperabilidade de Dados de Pesquisa
ORCID	Open Researcher and Contributor Identifier
OSI	Open Society Institute
PGD	Plano de Gestão de Dados
PPGInfo	Programa de Pós-Graduação em Gestão da Informação
PNCA	Política Nacional de Ciência Aberta (Portugal)
RCAAP	Repositório Científico de Acesso Aberto de Portugal
RDA	Research Data Alliance
RDF	Resource Description Framework
Re3data	Registry of Research Data Repositories
SIExp	Sistema de Gestão de Dados Experimentais
SNCTI	Sistema Nacional de Ciência, Tecnologia e Inovação
UDESC	Universidade do Estado de Santa Catarina
UNESCO	United Nations Educational, Scientific and Cultural Organization
UE	União Europeia
WHO	World Health Organization

SUMÁRIO

1	INTRODUÇÃO	16
1.1	OBJETIVO GERAL.....	20
1.1.1	Objetivos Específicos	20
1.2	JUSTIFICATIVA	21
1.3	ESTRUTURA DA DISSERTAÇÃO	22
2	CIÊNCIA ABERTA	24
2.1	DADOS CIENTÍFICOS ABERTOS.....	28
2.1.1	Curadoria de dados científicos	32
2.2	GESTÃO DE DADOS CIENTÍFICOS	35
2.2.1	Plano de Gestão de Dados	37
2.2.2	Repositórios de dados abertos	40
2.3	PRINCÍPIOS FAIR	46
3	OPÇÕES E PROCEDIMENTOS METODOLÓGICOS	54
3.1	CARACTERIZAÇÃO DA PESQUISA	54
3.2	FORMAÇÃO DO CORPUS DA PESQUISA	56
3.2	INSTRUMENTO DE COLETA DE DADOS E OPÇÕES DE ANÁLISE	58
4	ANÁLISE DOS DADOS POR AGÊNCIA DE FOMENTO	63
4.1	AGÊNCIAS NACIONAIS	63
4.2	AGÊNCIAS INTERNACIONAIS	73
5	RESULTADOS CONSOLIDADOS	88
5.1	AGÊNCIAS NACIONAIS	89
5.2	AGÊNCIAS INTERNACIONAIS	94
5.3	INFOGRAFIA PARA UM PGD COM ABORDAGEM FAIR	99
6	CONSIDERAÇÕES FINAIS	102
	REFERÊNCIAS	105
	APÊNDICES	116
	APÊNDICE A - Resultado gerado pela Ferramenta Python – Incidência de termos	117
	APÊNDICE B - Modelo da planilha de análise dos documentos de cada Agência - PGD	118
	APÊNDICE C - Modelo de análise aos documentos de cada Agência - Princípios FAIR	119

1 INTRODUÇÃO

O campo científico passa por transformações. Observa-se o movimento recente da sociedade pelo acesso aberto às inovações e aos processos de fazer ciência. Esta nova abordagem envolve a colaboração entre os agentes de pesquisa, as agências de fomento, instituições de ensino e a própria sociedade.

O momento que vivemos, com a pandemia (COVID-19), intensificou o olhar sobre a importância de cooperação entre as áreas científicas, seus atores, pesquisadores e instituições sejam públicas ou privadas. Também a população começou a ver a relevância do compromisso social por parte das instituições e a responsabilidade de cada cidadão para o bem estar coletivo. A ciência acelerou inovações para atender as novas demandas geradas pelo contexto social vigente e a comunicação científica também passou por alterações com este novo cenário. Os periódicos científicos ofereceram acesso aberto a seus conteúdos num esforço conjunto de cooperação entre cientistas/especialistas para favorecer o avanço mais rápido das pesquisas, especialmente no campo da saúde com enfoque em combater o avanço da contaminação pelo coronavírus (COVID-19).

As tecnologias digitais se apresentam como grandes aliadas neste processo em prol do conhecimento aberto. 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 defesa da ciência aberta, em que o acesso aberto aos dados é priorizado e onde são devidamente respeitadas as questões éticas e legais para as eventuais restrições.

Este contexto informacional dinâmico aproxima especialistas de seu público e propicia maior troca informacional entre os indivíduos, sejam especialistas ou não. Toda essa produção e interação gera uma necessidade de análise e tratamento do que deve ser preservado e disponibilizado para novos estudos e pesquisas, a fim de evitar custos desnecessários e retrabalho, e possibilitando gerar novos conhecimentos através da reanálise de dados científicos sob novas perspectivas, por meio do reuso de dados tratados e já disponíveis.

O reuso de dados é considerado tendência para acelerar o avanço das ciências de forma mais equitativa. Beneficia a sociedade com o uso dos recursos e investimentos de forma inteligente e economicamente viável, preservando os dados como ativos úteis para novas gerações.

A ciência aberta faz uso intensivo das tecnologias que possibilitam a colaboração *online* para atividades científicas em todos os campos do conhecimento, de forma aberta e transparente, permitindo maior envolvimento social e favorecendo a cidadania.

Compartilhar informações científicas é primordial para potencializar os avanços da ciência, poupar de retrabalho, possibilitar reproduzir pesquisas para validação e permitir o reuso de informações por novos interessados para novas descobertas.

Segundo Henning *et al* (2018, p. 11), a “abertura dos dados científicos é uma das principais propostas da ciência aberta - tema que tem sido abordado como prioridade no âmbito internacional, com intenso debate sobre princípios e padrões para o acesso aos dados de pesquisas”.

A ciência aberta abrange mais do que somente o acesso aberto às publicações e dados científicos, é uma espécie de ‘guarda-chuva’ para outros termos, tais como: dados abertos, ciência cidadã, revisão aberta por pares, código aberto, caderno aberto de laboratório, recursos educacionais abertos e redes sociais científicas. “Nesse contexto, produção e comunicação científicas constituem processos indissociáveis, a comunicação sendo diretamente produtiva”. (ALBAGLI; CLINIO; RAYCHTOCK, 2014, p. 436).

Dessa forma, todas as etapas do processo científico passam a ser abertas, sendo uma iniciativa positiva para os atores envolvidos, tanto aos pesquisadores quanto às instituições de pesquisa, agências de fomento e, por fim, à sociedade.

A abertura de todo o processo de produção do conhecimento, incluindo os dados coletados ou gerados durante a pesquisa, é vista como oportunidade para fomentar a colaboração entre pesquisadores, o avanço no conhecimento e a criação de soluções que atendam às demandas da sociedade. (FIOCRUZ, 2019d).

Esta emergente realidade informacional requer dos envolvidos maior competência em informação para lidar com dados científicos, envolve capacidade crítica, questionamento e observação apurada para análises de contextos além das aparências, a fim de selecionar, preservar e disponibilizar os dados de fato relevantes. A curadoria de dados científicos deve propiciar aos pesquisadores condições para efetivar a gestão desses dados, desde o início do seu ciclo de vida para que possam ser reutilizados em novas pesquisas em favor do avanço científico.

Mas o que são dados científicos? De acordo com Silva (2019, p. 21),

são as informações registradas ou produzidas através de qualquer forma ou meio durante o decurso de uma pesquisa. Podem ser de tipos numérico, descritivo ou visual e reproduzir-se em formato de papel (incluindo notas de pesquisa em cadernos, fotografias, etc.) ou digital. O conceito de dados científicos também faz referência às distintas ferramentas como protocolos, códigos numéricos, gráficos e tabelas que são necessárias para recolher e organizar os dados, tanto em trabalhos de campo quanto em laboratório.

Essa variedade de dados primários provenientes de pesquisas nas distintas áreas do conhecimento humano necessita de “publicização” para que se tornem dados científicos

abertos. Essa ação é importante para permitir a “reprodutibilidade e reutilização em pesquisas derivadas ou não, além de permitir o amplo escrutínio - o que pode contribuir para expor inconsistências, baixa qualidade, plágio ou fraude” das pesquisas. (ALBAGLI; CLINIO, RAYCHTOC, 2014, p.440).

No entanto, para que eles estejam realmente abertos, devem manter uma identificação consistente no decorrer do tempo que permita interpretá-los de forma coerente para novas pesquisas. Metadados detalhados de significação e contexto são fundamentais, considerando a complexidade e a heterogeneidade dos dados científicos em diferentes domínios, além de estruturas que permitam interoperabilidade entre sistemas e as devidas atribuições de licenças.

Portanto, para que a gestão de dados científicos 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.

Os princípios FAIR – do acrônimo: *Findable, Accessible, Interoperable e Reusable*, traduzidos como: Encontrável, Acessível, Interoperável e Reutilizável – 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 cita-los corretamente.” (AVENTURIER, 2017, tradução nossa).

As demandas da sociedade por transparência e sustentabilidade, aliadas às novas políticas das Agências de Fomento para que os dados de pesquisa sejam abertos, passam a recomendar e/ou exigir a apresentação de Planos de Gestão de Dados como requisito para submissão de projetos e obtenção de recursos para pesquisas. As instituições inseridas neste cenário desempenham importante função no movimento de acesso aberto, e reforçam “o papel da ciência e sua efetiva contribuição para o desenvolvimento de sociedades mais equitativas e sustentáveis” (SANTOS, 2017, p. 11).

A quantidade de dados científicos gerados é bastante expressiva, porém, toda essa gama informacional necessita de pessoas qualificadas para realizarem sua análise dentro do ciclo informacional e posterior preservação, armazenagem ou descarte e disseminação/disponibilização. Diante disso, é possível dimensionar a necessidade de estudos e pesquisas neste campo da Ciência da Informação para aprimorar serviços de curadoria e gestão de dados, especialmente os dados científicos abertos.

O que ocorre é que esse grande volume de dados e informações ainda não estão sendo adequadamente tratados por todos os segmentos produtores. Esta curadoria de dados exige atenção e cautela durante todo o processo, desde a criação até a sua disponibilização final para o reuso.

Isto posto, avalia-se que a intensa produção de dados científicos necessita de ações de curadoria e gestão desde o início do ciclo de vida destes dados.

Essas atividades ganham destaque e as pesquisas sobre a temática estão em fase de implementação. Na Europa e Estados Unidos da América (EUA) já existem estudos mais consolidados; no Brasil ainda se está buscando estabelecer diretrizes para a gestão de dados científicos. Tais iniciativas estão sendo impulsionadas principalmente pelas Agências de Fomento à pesquisa. E estas recentes exigências/recomendações acabam por refletir novas perspectivas de atuação dos profissionais de informação, tanto dos que atuam em bibliotecas universitárias como dos pesquisadores (laboratórios de pesquisa, etc.).

Frequentemente, e mesmo nas organizações de pesquisa mais modernas, os dados de pesquisa estão dispersos, mal documentados ou inacessíveis, o que impossibilita sua reutilização em novas análises e a obtenção de respostas a novas perguntas. Dentro de um mesmo grupo de pesquisa, por exemplo, não é incomum que uma nova questão científica provoque um movimento de ‘garimpo’ de dados, por meio de contatos com ex-alunos, buscas em computadores antigos de membros do projeto – em geral, máquinas encostadas nos cantos dos laboratórios, aguardando por doação – ou, até mesmo, em mídias diversas contendo *backups*, por vezes obsoletas ou inadequadas para a preservação de dados em longo prazo. (DRUCKER, 2012 *apud* BERTIN; VISOLI; DRUCKER, 2017, p. 36).

Estes problemas derivados da falta de gestão dos dados podem impactar no desenvolvimento de novas pesquisas científicas, acarretando custos extras para gerar novos dados (por vezes, nem sendo possível essa alternativa), desperdiçar tempo dos pesquisadores, perder prazos em projetos e, até mesmo, gerar retrabalho aos pesquisadores. Tais situações geram ônus para todos os envolvidos, desde os financiadores públicos e/ou privados da pesquisa científica até os próprios pesquisadores que se frustram com o retrabalho e com atrasos em cronogramas/prazos.

Quando reutilizados, os dados poderão converter-se em embriões de novos projetos, assim como conduzir à experimentação e à verificação de hipóteses não consideradas quando da investigação original. A reutilização de dados científicos poderá, também, dar lugar a novas áreas de trabalho, potencializadoras de empregabilidade, de incremento do investimento público e de produção de riqueza. E, embora o retorno do financiamento varie em função dos diferentes domínios de investigação, não restam dúvidas de que a não duplicação de recolha de dados constitui, por si só, um benefício. (RODRIGUES; SARAIVA, 2010, p. 31).

A curadoria atua neste contexto; o Digital Curation Centre (DCC, 2019, tradução nossa) cita que “além de reduzir a duplicação de esforços na criação de dados de pesquisa, a curadoria aumenta o valor em longo prazo dos dados existentes, tornando-os disponíveis para pesquisas de alta qualidade”, a novos interessados. Além disso, possibilita validar métodos e análises já efetuados como forma de verificar possíveis problemas e evitar fraudes, permitindo maior transparência às pesquisas realizadas.

Diante do exposto, esta pesquisa propõe estudar a gestão de dados científicos abertos e os Princípios FAIR. Assunto para o qual ainda existe eminente necessidade de informações para fundamentar e impulsionar a nova realidade da gestão de dados científicos.

Sendo assim, indaga-se: As Agências de Fomento à pesquisa científica, dentro da realidade brasileira, apresentam iniciativas e/ou políticas para gestão dados científicos abertos? E quais as exigências e/ou recomendações destas Agências para elaboração de um Plano de Gestão de Dados (PGD), abordando os Princípios FAIR dentro do contexto brasileiro?

Para concretizar esta pesquisa foram delineados os seguintes objetivos geral e específicos.

1.1 OBJETIVO GERAL

Mapear o cenário dos dados científicos abertos, no que diz respeito ao Plano de Gestão de Dados e aos princípios FAIR, a partir das exigências e/ou recomendações das Agências de Fomento que mais financiam a pesquisa brasileira.

1.1.1 Objetivos Específicos

- a) Identificar entre Agências que fomentam a pesquisa científica no Brasil, quais delas têm iniciativas e/ou políticas que contemplam a gestão de dados científicos;
- b) Analisar os critérios relativos às exigências/recomendações para a elaboração de Planos de Gestão de Dados (PGD), das Agências de Fomento selecionadas no estudo;
- c) Verificar se as Agências de Fomento selecionadas citam/adotam os Princípios FAIR;
- d) Propor uma infografia, a partir dos critérios analisados, para o Plano de Gestão de Dados com abordagem nos Princípios FAIR.

No contexto da ciência aberta, a gestão eficaz de dados científicos abertos se apresenta como um importante elemento propulsor de novos estudos, além de favorecer a transparência e a divulgação dos resultados das pesquisas para a sociedade.

1.2 JUSTIFICATIVA

O interesse pela temática surgiu por meio de um levantamento inicial em que se observou existir um campo de investigação com lacunas passíveis de maior detalhamento no tema relativo aos dados científicos abertos. Dessa forma, considerou-se relevante buscar familiaridade e apropriação de conhecimentos no assunto para a aplicação prática em serviços e na gestão de unidades de informação, em diferentes domínios do saber. Analisar as práticas das Agências de Fomento neste momento se mostra importante a fim de verificar de que forma estão gerenciando suas demandas, num cenário de intensa busca por soluções frente aos desafios decorrentes da Pandemia gerada pela COVID-19, ao passo que reforçam a necessidade e importância da ciência, do compartilhamento e reuso de dados científicos.

Tendo em vista vivenciar a era da sociedade de informação, que traz consigo um ‘tsunami de dados’” diário como diria Santos (2017, p.56), intensifica-se o interesse em pesquisar esta temática para atuar de forma técnica, ética e efetiva no tratamento de dados científicos. Com isso, buscou-se apresentar uma infografia como proposta de roteiro de PGD que possa auxiliar os profissionais e os pesquisadores que atuam com atividades relacionadas à gestão de dados científicos, na elaboração do Plano de Gestão de Dados baseado em princípios construídos no movimento do acesso aberto e tendo os Princípios FAIR como orientação inicial.

Acredita-se que essa pesquisa possa fornecer subsídios para que profissionais da informação e/ou pesquisadores de diferentes áreas se adequem a mercados carentes de *expertise* no cenário da gestão de dados científicos abertos.

Para Rodrigues e Saraiva (2010), ainda existem barreiras para que os dados sejam compartilhados e reutilizados, dentre elas, destacam que poucos pesquisadores têm experiência e conhecimentos técnicos para o tratamento. Além disso, apontam carência de critérios estabelecidos para esta gestão, em especial, na “[...] criação de metadados capazes de garantir o acesso aos dados durante o seu ciclo de vida, bem como de assegurar a sua preservação a longo termo”. (RODRIGUES; SARAIVA, 2010, p. 33). Analisando as barreiras mencionadas, avaliou-se ser importante pesquisar o tema para fomentar maior conhecimento pessoal e para a área da Ciência da Informação, pois a demanda pelos serviços já existe e requer profissionais preparados para tratar eficazmente essa gama de dados científicos.

Segundo Santos (2014, p. 41),

a curadoria é apontada como novo campo de trabalho para o profissional da informação. Já existe demanda do mercado por ‘curadores’ [...] contudo ainda não fica claro qual o papel desses profissionais, uma vez que há poucos exemplos de

serviços dentro de instituições de informação. O momento atual ainda parece ser o de saber quais seriam os serviços de curadoria dentro desses espaços, que formação é necessária para fazer ‘curadoria’.

Essa pesquisa propõe ampliar o leque de informações sobre as novas demandas das Agências de Fomento à pesquisa, e auxiliar os pesquisadores e profissionais na gestão dos ativos de pesquisa. Objetiva-se, além de contribuir para a construção de novos conhecimentos e estudos no campo da Ciência da Informação, descortinar perspectivas para a atuação dos profissionais da informação neste dinâmico cenário de produção, análise, descoberta, uso e reuso de dados científicos em diferentes formatos/suportes, especialmente o digital.

1.3 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está estruturada em seis seções. **A introdução** que contempla a problemática de pesquisa, objetivo geral, específicos, justificativa e estrutura da dissertação.

A fundamentação teórica faz parte da segunda seção, se divide em subseções e abrange os temas: ciência aberta, dados científicos abertos, curadoria de dados científicos, gestão de dados científicos, Plano de Gestão de Dados (PGD), repositórios de dados e os Princípios FAIR.

Na terceira seção são delineadas as **opções e procedimentos metodológicos**, com as propostas de análise baseadas nos modelos do Plano de Gestão de Dados do DCC (Digital Curation Centre) e dos Princípios FAIR.

Na sequência, como quarta seção, apresenta-se a **Análise dos dados por Agência de Fomento** que expõe as informações coletadas. Cada Agência de Fomento será representada numa subseção, com rol de documentos e os critérios de análise baseados nas opções e procedimentos metodológicos descritos anteriormente.

A quinta seção apresenta os **Resultados Consolidados** obtidos durante a fase anterior. Expõe a representação dos resultados consolidados das análises e infografia com algumas ferramentas e padrões relevantes para gerar um Plano de Gestão de Dados alinhado aos Princípios FAIR.

A partir disso e alinhadas a cada objetivo proposto, elaboram-se as **Considerações finais** na seção seis.

Para melhor visualização a estrutura da dissertação é apresentada na forma de figura 01, abaixo.

Figura 01 - Estrutura da dissertação

Estrutura da Dissertação



Fonte: Dados da Pesquisa (2020).

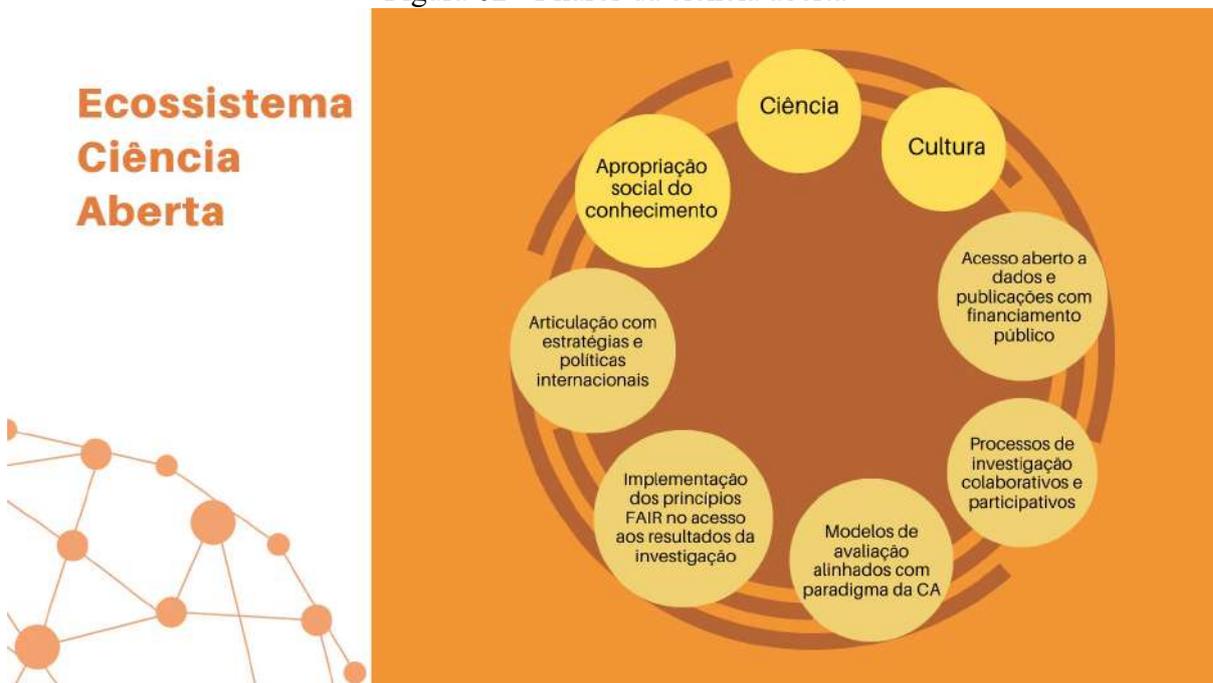
2 CIÊNCIA ABERTA

A ciência aberta envolve mais do que o acesso aberto às publicações e dados científicos. Para além da abertura dos resultados de pesquisa, a prática da ciência aberta implica que o processo de pesquisa seja ele próprio aberto usando métodos, ferramentas e *workflows* que facilitem a partilha, reutilização e colaboração. (FIOCRUZ, 2019a).

O movimento da ciência aberta representa uma mudança benéfica à sociedade, em que o acesso é aberto, a construção é coletiva, o compartilhamento é estimulado, a transparência e sustentabilidade são valorizadas e os novos conhecimentos construídos favorecem ao bem-estar coletivo.

A figura 02 ilustra o ecossistema da ciência aberta. Envolvidos neste ecossistema tem-se o acesso aberto aos dados e publicações derivadas de investimento público, aos processos de pesquisa colaborativos, novos modelos de avaliação, a implementação dos Princípios FAIR e articulação com políticas internacionais, uma vez que pretende ter abrangência global.

Figura 02 - Pilares da ciência aberta



Fonte: Adaptado de (CIÊNCIA ABERTA, 2019).

Retrata um ecossistema dinâmico e plural que envolve diferentes atores, culturas, processos, políticas, marcos legais, infraestruturas distintas e saberes interdisciplinares para que, em conjunto, possam atuar em benefício da democratização do conhecimento e do bem-estar social.

Dentre os marcos para o movimento global de acesso aberto ao conhecimento científico, destaca-se a “Declaração de Budapeste (*Budapest Open Access Initiative* - BOAI), publicada em 2002, na qual foram propostos o conceito e as estratégias para o acesso aberto por intermédio da Via Dourada¹ e da Via Verde²” (SANTOS, 2017, p. 11). Na sequência, em 2003 publica-se a Declaração de Berlim³ sobre Acesso Livre ao Conhecimento nas Ciências e Humanidades, em defesa do movimento de acesso aberto para incentivar os pesquisadores a publicar seus trabalhos em pelo menos um repositório.

Os autores Rios, Lucas e Amorim (2019, p. 166) fazem um apanhado das Declarações em prol do movimento de acesso aberto, e sintetizam o que propõe cada uma:

Convenção de Santa Fé (1999)

- Marcou o lançamento do Open Archives Initiative (OAI)
- Arquivos *e-prints*
- Proposta de um novo formato para avaliação dos artigos científicos, através do auto arquivamento.

Declaração de Budapeste (2002)

- Usa o termo *Open Access* pela primeira vez para definir o acesso livre a informação
- Apresenta o conceito de Acesso Aberto
- Traz duas estratégias para alcançar o Acesso Aberto. O auto arquivamento e as revistas de Acesso Aberto.
- Oferece apoio para as instituições que tem interesse em aderir ao movimento.

Declaração de Bethesda (2003)

- Apresenta recomendações para os envolvidos com a informação científica: Instituições e agências de financiamento; bibliotecários e editores; sociedade científica

Declaração de Berlim (2003)

- Promover a Internet como o instrumento funcional ao serviço de uma base de conhecimento científico global e do pensamento humano
- Busca estabelecer o paradigma do Acesso Aberto eletrônico.

Declaração de Haia (2014)

- Fala sobre a mineração de dados e *Big Data*
- Estabelecer as estratégias práticas para a informação alcançar características de Acesso Aberto. Recomenda o uso de ORCID, XML, *Creative Commons*.

Na Declaração de Haia de 2014, os dados já são mencionados. Ocorre no mesmo ano em que surgem as discussões para a formulação dos Princípios FAIR, estes devido ao interesse em publicar e reutilizar os dados científicos.

No Brasil não existe uma ‘política pública nacional mandatária para a ciência aberta’, o que cria uma lacuna, se compararmos o país com as iniciativas do cenário internacional.

Por outro lado, registra-se o crescimento de iniciativas nesta direção, oriundas de agentes de fomento, universidades e institutos, destacando-se a FAPESP e a sua

¹Via Dourada: Acesso Aberto por meio da publicação de artigos em periódicos com Acesso Aberto (AA).

²Via Verde: Acesso Aberto por meio do depósito/autoarquivamento de artigos publicados em periódicos, anais e apresentações de conferências, revisados por pares, além de teses e dissertações, em um repositório de Acesso Aberto.

³https://openaccess.mpg.de/67605/berlin_declaration_engl.pdf.

recente exigência do Plano de Gestão de Dados como requisito para submissão de propostas a seus editais. (FIOCRUZ, 2019b).

No cenário internacional, o Projeto FOSTER é uma iniciativa da comunidade europeia com parcerias em seis países e intuito de promover e implantar a prática da ciência aberta, estimulando uma “mudança real e duradoura no comportamento dos pesquisadores europeus para garantir que a *Open Science* (OS) seja a norma”. (FOSTER, 2019).

Segundo o projeto FOSTER (2019), ciência aberta pode ser definida como:

a prática da ciência de tal forma que outros podem colaborar e contribuir, na qual os dados de pesquisa, as notas de laboratório e outros processos de pesquisa estão disponíveis livremente, em condições que permitem a reutilização, redistribuição e reprodução da pesquisa e dos dados e métodos subjacentes.

A ciência aberta (*open science*) é retratada por um conceito que representa uma abertura para a democratização do conhecimento.

Abarca diversas práticas do fazer científico, desde o acesso aberto à literatura científica e aos dados abertos de pesquisa, passando pela ciência cidadã, recursos educacionais abertos, *software* e *hardware* abertos, entre outras. Faz referência a um modelo metodológico alinhado à filosofia da cultura digital, da colaboração e do compartilhamento das práticas e dos conteúdos científicos. (SANTOS, 2017, p. 122).

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”.

As iniciativas em ciência aberta favorecem maior “eficiência” nas pesquisas e apropriação no processo do ‘fazer científico’, possibilitam estudos com mais qualidade (permitem testes/verificações), despertam e agilizam explorar novos temas, ampliam a visibilidade dos autores originais (com as citações) e promovem “o retorno científico para as instituições” (CIÊNCIA ABERTA, 2019).

Segundo Tenopir *et al.* (2015 *apud* SILVA; SANTAREM SEGUNDO; SILVA, 2018, p.5199).

Do ponto de vista das entidades financiadoras de pesquisa, o depósito em acesso aberto das investigações em projetos justifica o investimento, cria transparência e evita duplicação de financiamento. Para a pesquisa, portanto, poder avançar para beneficiar a comunidade científica global.

Dentre os benefícios citados, destaca-se a reutilização como atividade fundamental. O

reuso de dados científicos segundo Curty (2016, p. 03) 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 – o que pode favorecer novas parcerias e recursos para projetos futuros, contribuindo, inclusive, com o desenvolvimento econômico e o bem-estar social. É possível afirmar que facilitar a descoberta e o reuso de dados é o objetivo, também, do plano de gestão de dados.

Os autores Curty e Aventurier (2017, p. 03) corroboram ao afirmar que investir na ciência aberta se justifica pelo efetivo reuso dos dados. “Neste sentido, a sustentabilidade do ciclo da ciência aberta depende da busca por formas eficientes de maximizar o reuso de dados científicos, ao invés de meramente estocá-los como volumes ociosos em repositórios”.

Por meio da afirmação acima é possível observar que na ciência aberta são necessários instrumentos de apoio à gestão dos dados. O uso de metadados na implementação dos Princípios FAIR já faz parte das ações adotadas por instituições de pesquisa e Agências de Fomento Internacionais⁴ para a gestão dos seus ativos científicos. A questão central não está somente no armazenamento, engloba as premissas de como os dados podem ser descobertos, acessados, interoperáveis para a ressignificação em novos contextos. Na ciência aberta o conhecimento é considerado um bem coletivo que deve estar acessível a todos, para o exercício da cidadania.

Nessa vertente a Fiocruz (2020), elenca alguns benefícios do **acesso aberto** ao conhecimento, conforme segue:

- Disponibiliza para toda a sociedade o acesso público e gratuito do conhecimento científico produzido por uma instituição pública;
- Aumenta a capacidade de recuperação e o uso da literatura científica relevante;
- Amplia a visibilidade, a leitura e o impacto da produção científica;
- Facilita a pesquisa interdisciplinar;
- Agiliza a pesquisa, a descoberta e a inovação;
- Garante o controle sobre a autoria dos artigos e constitui uma ferramenta contra plágios, uma vez que ao depositarem seus trabalhos, os autores passam a ter uma cópia registrada no respectivo repositório;
- Oferece um ponto de referência único para os trabalhos - acessíveis, sem barreira de tempo ou espaço;
- Contribui para a formulação de indicadores confiáveis de produção intelectual;
- Supre as demandas das agências de fomento para a disseminação de sua produção científica;

⁴Por exemplo: “National Institutes of Health (NIH), Data Commons, G20 Hangzhao Consensus, Amsterdam Call for Action on Open Science, European Open Science Cloud,” dentre outros. (CURTY; AVENTURIER, 2017, p. 3).

- Preserva a memória institucional;
- Contribui para a gestão de investimentos em pesquisa e ensino. (FIOCRUZ, 2020c).

Nesta rede da ciência aberta muitos são os atores que fazem parte, citam-se os pesquisadores, as Agências de fomento público, editores, bibliotecários, gestores de dados, os centros de pesquisa, organizações de apoio à inovação e investigação, universidades, órgãos governamentais e a sociedade de forma geral.

Por ser composta de heterogeneidades e permear em áreas interdisciplinares é importante que as discussões acerca de políticas e diretrizes contêm com a cooperação do maior número de envolvidos, de forma a ser representativas e corresponder às necessidades e particularidades tecnológicas, sociais e culturais dos membros desta rede.

2.1 DADOS CIENTÍFICOS ABERTOS

A nomenclatura quanto ao termo ainda não está consolidada, de forma que alguns autores denominam de ‘dados de pesquisa abertos’. Porém, nesta pesquisa, o termo adotado será ‘dados científicos abertos’. O enfoque será especificamente nos dados oriundos de pesquisas científicas, não abarcando dados abertos de cunho governamental.

Os dados científicos são considerados protagonistas no movimento em favor da ciência aberta. Passam a fazer parte de estudos e pesquisas que visam analisar todo o ciclo de vida dos dados, por isso a importância em torná-los acessíveis, prontos para descoberta e significativos para o reuso.

Segundo a definição da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), dados de pesquisa são

registros factuais (pontuações numéricas, registros textuais, imagens e sons) utilizados como fontes primárias para pesquisa científica e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação sistemática e parcial do assunto que está sendo investigado. (OECD, 2007, p. 13, tradução nossa).

Conforme os autores Fernandes e Oliveira (2018, p. 257), dados de pesquisa no cenário atual se caracterizam:

não apenas alicerces de pesquisas realizadas, mas como subsídios essenciais para outros pesquisadores, instituições acadêmicas e agências de fomento. Esses dados eram descartados ou até mesmo armazenados de forma inapropriada e sem nenhuma gestão, sendo quase sempre guardados com os próprios geradores da informação, perdendo-se no tempo e no espaço.

Outro conceito de dados científicos, de acordo com Medeiros (2016, p. 299), “[...] os dados científicos são insumos para novas pesquisas e também o produto delas, consideramos que eles aumentam o poder comunicacional e informacional dos pesquisadores”.

O valor dos dados gerados no decorrer das pesquisas é pautado em sua possibilidade de replicabilidade científica e reuso, ou 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, obtidos por sensores ou coletados através de entrevistas, experimentos laboratoriais ou amostras únicas, etc.; por isso a importância em registrar o contexto para facilitar o entendimento por grupos distintos de pesquisadores e interessados ao longo do tempo. Todo dado é dependente de interpretação contextual.

Segundo o Guia de Gestão de Dados de Pesquisa para bibliotecários e pesquisadores, esses dados são categorizados:

[...] segundo sua **natureza**: **(a) observacionais**: gerados por meio de observações diretas que podem ser associadas a lugares e/ou tempo específico, por sua natureza instantânea qualificam-se como registros históricos, pois não podem ser coletados uma segunda vez e, portanto, devem ser submetidos a processos de curadoria que os preservem para sempre; **(b) computacionais**, são o resultado da execução de modelos computacionais ou de simulações; **(c) experimentais**, são provenientes de situações controladas em laboratórios. [...] podem ser números, imagens, vídeos, áudio, software, algoritmos, equações, animações, modelos e simulações. Por fim, conforme o **fluxo da pesquisa**, **(a) dados brutos**, coletados diretamente de instrumentos científicos de pesquisa; **(b) derivados**, resultados do processamento ou combinação de dados brutos ou de outros dados; **(c) canônicos ou referenciais**, coleções de dados consolidados e arquivados geralmente em grandes centros de dados. (SAYÃO; SALES, 2015, p. 7-9).

Todas estas categorias de dados são passíveis de tratamento, porém é importante ter os objetivos de pesquisa claros para planejar a gestão destes dados em conjunto com pesquisadores e equipes multidisciplinares. Essa gama de diferentes dados impacta no processo de gestão e no agir dos gestores da informação, que precisam estar atentos às novas demandas e preparados para cenários cada vez mais dinâmicos.

No entendimento de Medeiros e Caregnato (2012, p. 320):

a *e-Science* promove algo que está ligado à constituição da ciência, isto é, maior possibilidade de cooperação, colaboração e interdisciplinaridade. Necessita-se, entre outros quesitos, de esforços e estudos na forma como os dados são liberados, podendo ser dados fortemente estruturados ou ainda que passaram por um processo de curadoria.

Percebe-se que ainda cabem muitos estudos nesta temática da *e-science*/dados brutos, e as pesquisas que já estiverem nos padrões da ciência aberta são muito importantes para acelerar este processo científico. Estas ações estimulam a colaboração entre os cientistas e a comunidade, promovem parcerias, permitem o compartilhamento e reuso dos dados. Destaca-

se a necessidade de atenção aos princípios éticos e legais de menção às fontes, aos direitos autorais, as licenças previstas, a legislação vigente sobre proteção de dados sensíveis (pessoais, privacidade).

Importante mencionar que nem todos os dados científicos podem ser disponibilizados em acesso aberto. Pode ocorrer de por questões contratuais exigir que os dados permaneçam “restritos/fechados” por um período de embargo, ou, por interesse comercial para solicitação de patentes, ou publicações, etc. Noutros casos, pode requerer processo de anonimização dos dados. Para alguns estudos em saúde, por exemplo, onde os pacientes devem ter sua privacidade/identidade preservada, deve ser utilizada a técnica de anonimização. Este processo deve ser minucioso, e o ideal é que esteja previsto desde o início do ciclo de vida dos dados para a pesquisa. Conforme define o glossário da FIOCRUZ (2020), a anonimização consiste na “utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo”.

Conforme o Open Data Institute ⁵ (ODI, 2020) os dados têm diferentes perspectivas e podem ser fechados, compartilhados ou abertos.

Fechados podem ser dados de cunho pessoal, protegidos por lei, contratos ou internos de uma empresa (por exemplo, relatórios de vendas).

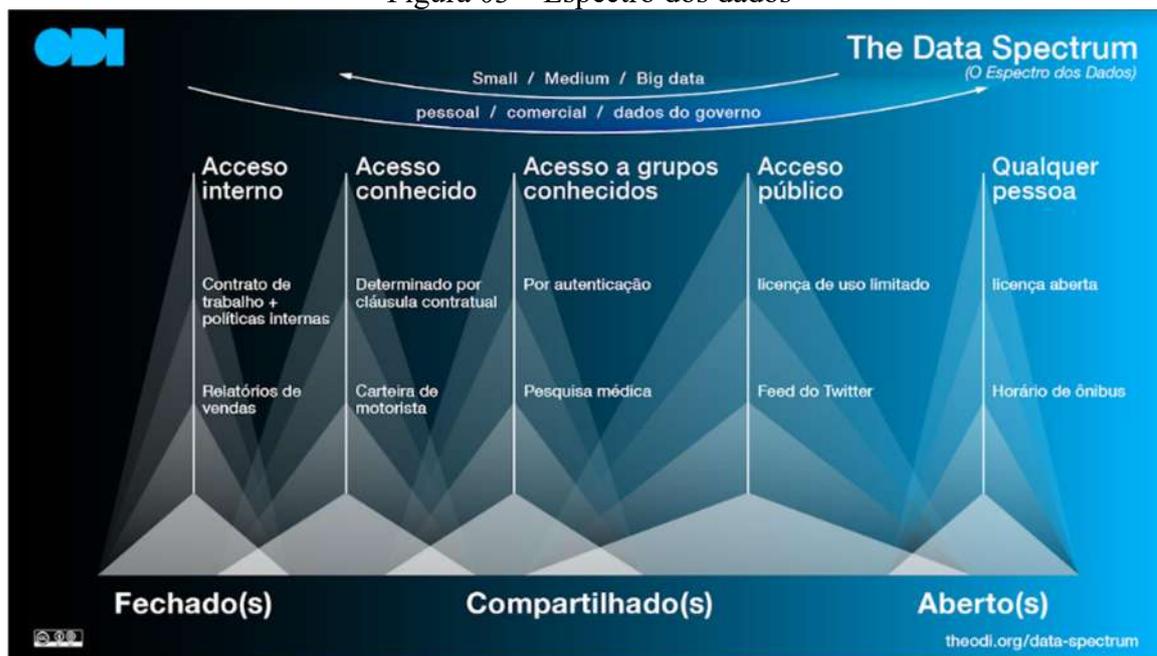
Dados compartilhados são aqueles cujo acesso, reuso e redistribuição estão limitados a determinadas pessoas ou grupos. Seja através do estabelecimento de convênios entre instituições, seja por conta de uma relação de confiança entre pesquisadores que colaboram informalmente. Tais acordos costumam determinar as condições de uso, especialmente em relação à possibilidade de exploração comercial e a atribuição de autoria. (FIOCRUZ, 2020).

E dados abertos que podem ser livremente distribuídos e reutilizados sujeitos a no máximo menção à fonte e uso da mesma licença que o original.

A figura 03 ilustra melhor esta dinâmica.

⁵ O Open Data Institute (ODI), foi co-fundado em 2012 pelo inventor da web, Sir Tim Berners-Lee e o especialista em Inteligência Artificial Sir Nigel Shadbolt para defender o uso inovador de dados abertos e fomentar mudanças positivas em todo o mundo.

Figura 03 – Espectro dos dados



Fonte: ODI (2020).

Quanto aos dados científicos abertos, destaca-se a iniciativa dos Princípios FAIR (encontrável, acessível, interoperável e reutilizável) em fornecer orientações para boas práticas na gestão dos dados e, por conseguinte, favorecer a apropriação social do conhecimento. Oliveira e Silva (2016, p. 10) afirmam que “a *e-Science* se propõe a ser aberta, gratuita e de livre acesso”. Mas para que isso se concretize é fundamental a cooperação entre os pesquisadores de diferentes domínios, e neste contexto, as orientações para boas práticas do FAIR, incentivadas em nível global, irão favorecer este intercâmbio. Um dado alinhado aos Princípios FAIR, não significa obrigatoriamente, dado aberto.

De acordo com Cavalcanti e Sales (2017, p. 83), o acesso aberto aos dados de pesquisa tem como principal objetivo,

[...] colaborar para o desenvolvimento da pesquisa científica em escala mundial, pois o acesso a eles contribuiria para dinamizar novas pesquisas a partir da eliminação de etapas que já tenham sido realizadas por outros pesquisadores e que se tornaram conhecidas com a publicação dos dados coletados. O impacto também ocorre na economia de recursos com a diminuição de pesquisas duplicadas e com o direcionamento destes recursos para novas pesquisas.

Os autores Rocha, Sales e Sayão (2017, p. 11) coadunam com os benefícios apresentados acima ao observar que reutilizar dados científicos estruturados evita retrabalho e acelera os avanços da ciência. “O acesso livre às informações não só publicadas através de teses e dissertações, mas também de dados utilizados ao longo das pesquisas, vêm se consolidando cada vez mais nas comunidades científicas como uma prática bem vista”.

No Brasil, a governança pública da ciência aberta para dados científicos tem como marco inicial o “Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã”, lançado em setembro de 2016, pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), com iniciativas e ações em prol do movimento de acesso aberto (SANTOS, 2017).

Conforme Santos (2017, p. 30), o ‘Manifesto’ menciona que

[...] os dados de pesquisa transitam em diferentes instâncias, portanto apresentam recomendações específicas às universidades e aos institutos de pesquisas, às sociedades científicas e academias de ciência, aos órgãos de fomento à pesquisa, aos editores de revistas ou periódicos científicos, aos cursos de pós-graduação e graduação nas áreas de informação, aos gestores e executores de programas e projetos de dados de pesquisa e aos pesquisadores.

Dentre estas diferentes instâncias, este estudo busca analisar os dados científicos sob o aspecto da curadoria e gestão voltada para a elaboração do Plano de Gestão de Dados alinhado aos Princípios FAIR, conforme as recomendações das Agências de Fomento à pesquisa.

2.1.1 Curadoria de dados científicos

O termo curadoria está etimologicamente relacionado ao “ato de curar, zelar, vigiar por algo”. Deriva do campo do direito, mas posteriormente o termo passou

a relacionar-se com o campo das artes, dos museus e de seus respectivos acervos. As representações de curadoria vigentes vinculam-se à ação humana e, ampliadas para qualquer contexto social referem-se sobremaneira a atividades de seleção, organização e apresentação de algo a partir de algum critério inerente ao indivíduo curador. (CORRÊA; BERTOCCHI, 2012, p. 28-29).

Segundo Sayão e Sales (2013, p. 06), a curadoria de dados pode ser definida “pelo lado mais pragmático e operacional” como o

conjunto de atividades gerenciais, técnicas e informacionais fortemente padronizadas chamado coletivamente de curadoria de dados de pesquisa, permite que os dados possam ser tratados, arquivados em ambientes digitais confiáveis, preservados e reconfigurados de forma que possam ser aplicados em novos contextos científicos; sirvam de base para novas pesquisas; sejam aproveitados para fins educacionais; e, sobretudo, colaborem para minimizar a duplicação de esforços nas estratégias de criação de dados.

A curadoria está relacionada às atividades de manutenção de dados científicos em formato digital e demais materiais ao longo do seu ciclo de vida, com o intuito de disponibilizá-los às futuras gerações. Para além dos processos de arquivamento e preservação é importante incluir a “criação de dados de qualidade e gestão, e a capacidade de acrescentar

valor aos dados para produção de novas fontes de informação e conhecimento”. (SANTOS, 2014, p. 111).

Alguns pontos para dados de qualidade que podem ser tratados durante o processo de curadoria e gestão, seguem representados e contextualizados na figura 04.

Figura 04 – Aspectos para dados de qualidade



Fonte: Adaptado de Fiocruz (2020d).

Segundo os autores portugueses Rodrigues e Saraiva (2010, p. 48, grifo nosso), a curadoria de dados pode ser entendida como um grupo de ações que:

garantem que um conjunto de dados é genuíno, permitindo o seu uso por outros que não os seus produtores. A curadoria pode envolver ações de descrição dos dados, de ligação destes a outros que os tornem inteligíveis, de registro dos usos que tenham e dos resultados a que tenham dado origem. A curadoria envolve também ações de preservação, em que a representação dos dados e os seus metadados tenham de ser modificados. **As ações de curadoria e de gestão de dados têm alguma intersecção, sendo as de gestão mais independentes do conteúdo e do uso.**

A curadoria seleciona, cuida, contextualiza os dados para que se possa preservar e manter sua autenticidade no decorrer do tempo.

Para Machado e Vianna (2016, p. 2), “curadoria digital é aquela que se articula com todas as ações de uso e preservação em relação ao meio digital que visam à reutilização de dados digitais autênticos [...]”.

Araújo (2017, p. 15) afirma que a curadoria “está ligada à percepção da importância da certificação de confiabilidade, da obsolescência e da evolução dos formatos (com o risco de perda pelo desenvolvimento tecnológico e pela fragilidade das mídias digitais)”.

Hoje em dia, os suportes tecnológicos se atualizam com muita rapidez, então, é imprescindível que os profissionais estejam preparados e atentos a estas mudanças para manter os ativos adequadamente armazenados e preservados ao longo do tempo. É relevante destacar que estas atividades de curadoria e apoio aos pesquisadores são contínuas, exigem estudos, capacitações, investimentos e recursos adequados. (CAVALCANTI; SALES; SAYÃO, 2016).

Os conceitos denotam que a curadoria abrange todo o ciclo de vida dos dados, desde o processo de planejamento e geração até a etapa de preservação para seu possível reuso.

São gerados muitos dados nas rotinas de pesquisa das comunidades científicas, e estes dados passaram a ser ativos preciosos para além dos pesquisadores e cientistas envolvidos. Porém, para que possam ser descobertos e reutilizados devem estar devidamente identificados, ter metadados associados, estar acessíveis e oferecer possibilidade de interoperabilidade.

Para Blanco (2014, p. 84), o profissional curador é como um “consultor que presta os seus serviços gerindo as informações mais relevantes sobre determinado setor de atividade”.

Considerando a extrema variação dos dados, os ambientes mais efetivos de gestão de curadoria são aqueles que permitem uma troca dinâmica de *expertise*, práticas e conhecimentos entre membros da equipe. O compartilhamento de *expertise* desempenha um papel central nas operações em curso e no desenvolvimento de qualquer solução em curadoria de dados. (MAYERNIK, *et al.*, 2012, p. 12 *apud* SAYÃO; SALES, 2013, não paginado).

Com a vasta produção de dados científicos o papel de curador se faz cada dia mais relevante e aponta para um novo perfil profissional, mais dinâmico, especializado e inserido em equipes interdisciplinares no universo da pesquisa científica. Este profissional deve ser competente em informação e ter uma cultura do apreender contínuo, para que possa construir novos saberes junto a diferentes categorias de atores e áreas científicas.

Deve ter domínio da área específica de pesquisa para atuação conjunta com pesquisadores e novas categorias profissionais, construindo um ciclo informacional diferenciado e de qualidade. “A curadoria não se limita a um único nicho de profissionais, encontramos curadores entre profissionais da área de Comunicação Social, Biblioteconomia, Ciências da Informação, Ciências Humanas e outras diversas áreas do conhecimento”. (SILVA GRAÇA *et al.*, 2016, p. 71). Notadamente, equipes de trabalho com competências complementares são importantes durante o processo, que deve ser baseado em critérios preestabelecidos e no uso de metadados adequados.

Esse trabalho cooperativo é enfatizado por Fernandes e Oliveira (2018, p.258), ao afirmar que “a gestão e o compartilhamento dos dados de pesquisa exigem um trabalho em parceria entre pesquisadores e prestadores de serviços de curadoria, muitas vezes responsáveis por um repositório de dados de pesquisa”.

A curadoria busca adicionar valor aos dados por meio de metadados pertinentes, atividades de descrição, identificação, proveniência e contexto para apoio a gestão de forma ampla.

2.2 GESTÃO DE DADOS CIENTÍFICOS

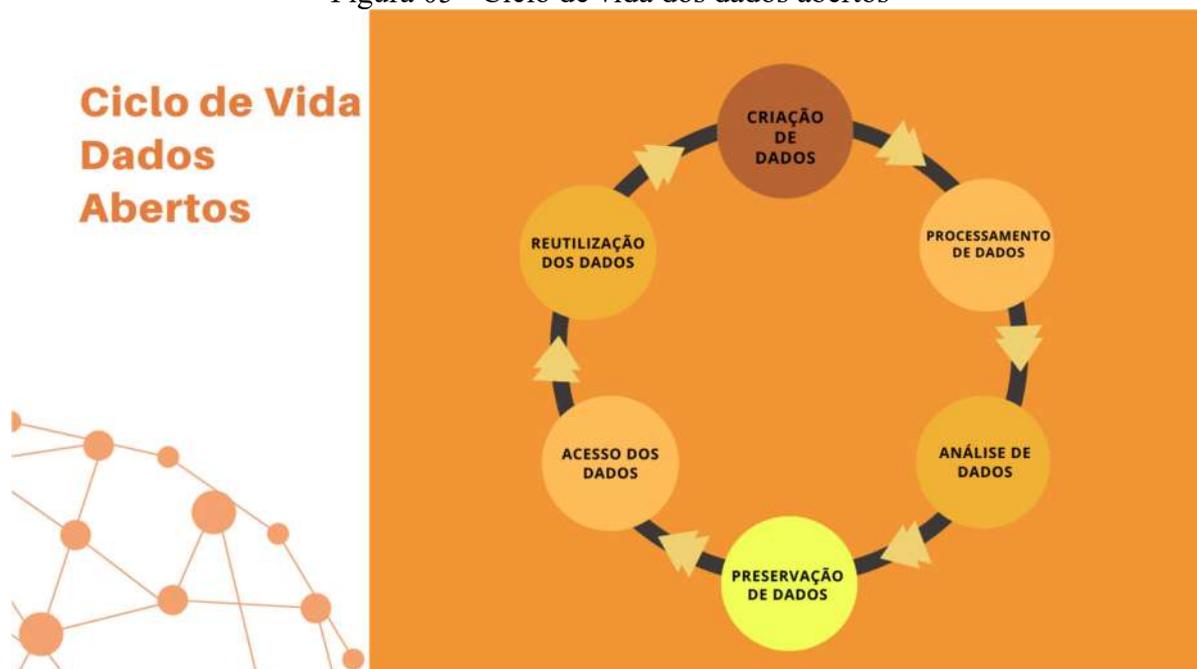
A gestão de dados científicos envolve planejar a administração dos dados da pesquisa durante todo o seu ciclo de vida, desde o processo de criação, coleta e documentação até a preservação. Segundo material sobre ciência aberta, disponibilizado pela Fiocruz em parceria com a Universidade do Minho (Portugal), uma boa gestão de dados, além de poupar tempo, facilita a toda equipe:

encontrar, entender e analisar os dados. A longo prazo, seguir boas práticas de gestão de dados significa que pesquisadores não envolvidos com o projeto podem encontrar, entender os dados e usá-los no futuro. Ao documentar os dados providenciando a forma adequada de citá-los, será mais fácil obter crédito pelo reuso dos seus dados. (FIOCRUZ, 2019a).

Documentar e adotar boas práticas na gestão dos dados permite que estes sejam descobertos e citados em novos contextos de estudos, o que confere maior visibilidade ao autor original. Contudo, o processo de gestão deve ser contínuo em todas as etapas do ciclo de vida dos dados.

A figura 05 ilustra o ciclo básico de vida dos dados abertos, que envolve desde sua criação, processamento, análise, preservação, acesso e, por fim, o reuso.

Figura 05 - Ciclo de vida dos dados abertos



Fonte: Adaptado de (FIOCRUZ, 2019a)⁶

No ambiente científico e acadêmico muitas pesquisas são realizadas e os dados, que poderiam ser disponibilizados em acesso aberto, acabam por ficar restritos somente com o pesquisador, e com o passar do tempo se perdem devido à falta de atualizações de versões dos *softwares*, *hardwares*, equipamentos obsoletos, novas mídias não compatíveis, ou até por ausência de executar *backups* pelos envolvidos nas pesquisas.

Diante deste cenário, observa-se a necessidade de ações de planejamento e envolvimento interdisciplinar dos profissionais e pesquisadores que atuam com curadoria e gestão já na fase inicial das pesquisas, para facilitar e garantir o uso e acesso aos dados no decorrer de todo o processo e ao longo do tempo.

Do ponto de vista do acesso aos dados da investigação, as bibliotecas estão desenvolvendo serviços de apoio durante as fases do ciclo de vida dos dados científicos, ou seja, quando os pesquisadores estão gerando e utilizando os dados em seu plano de trabalho. Muitas vezes, estes serviços devem ser prestados em estreita colaboração com pesquisadores e podem incluir o desenvolvimento **de planos de gestão** para documentar e organizar os dados através do desenvolvimento de ferramentas ou recursos para armazenar dados de forma segura. (SILVA, 2016, p. 388, grifo nosso).

Os estudos relatam que, neste contexto, o bibliotecário deve interagir com equipes interdisciplinares para compartilhar *expertises* distintas e, com isso, obter melhores resultados no processo de curadoria.

⁶Curso: O que é Ciência aberta? Link: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie1/curso1/aula3.html>.

Curadores de dados provenientes das bibliotecas especializadas, dos arquivos e de setores da tecnologia da informação são capazes de gerir, inserir nos sistemas e preservar coleções de dados de pesquisa, entretanto os especialistas em assunto é que serão capazes de fazer as análises necessárias à reinterpretação e reuso dessas coleções (SAYÃO; SALES, 2013, não paginado).

É válido destacar a relação com a característica interdisciplinar da Ciência da Informação e seu contexto de estudos baseados nos fluxos informacionais. Conforme Costa (2017, p. 25), “na Ciência da Informação, a *e-science* traz implicações relevantes sobre a comunicação científica, afinal os dados oriundos da *e-science* são de fato dados científicos primários”. A gestão de dados científicos está nesta conjuntura de mudança e permeada pelas tendências de pesquisas impulsionadas pelos contextos sociais vigentes.

Para Curty e Aventurier (2017, p. 3-4), a gestão de dados científicos “sinaliza para uma iminente movimentação em torno de políticas e iniciativas que valorizem a publicação de dados e seu potencial de reuso” Este movimento suscita mudanças na área da publicação científica e impulsiona a estabelecer diretrizes e ações para que seja possível atender estas novas demandas de forma estruturada, transparente e sustentável para a sociedade.

“Uma eficiente gestão de dados reduz a quantidade de trabalho necessária para a interpretação e compilação de informações obtidas no final de um projeto” (SILVA, 2016, p. 388). Se a geração dos dados oriundos das pesquisas científicas é planejada desde o início, com o instrumento denominado Plano de Gestão de Dados, ao efetivar o término dos projetos toda essa gama documental já estará com a curadoria necessária para ser disponibilizada conforme as previsões legais e contratuais de especificidades das áreas de domínio, sem necessitar de esforço extra para a compilação e curadoria final.

2.2.1 Plano de Gestão de Dados

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

É uma iniciativa ainda recente no Brasil, está no patamar de recomendação para compor os projetos de pesquisa. Porém, já existem Agências de fomento à pesquisa 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).

O Sistema integrado de bibliotecas da USP documenta que é na elaboração do plano de gestão de dados que se define:

quais dados coletar, manter, o que descartar e em que formato de arquivo salvar estabelece a decisão sobre onde e como armazenar seus dados, e estima os custos de preservá-los, a curto e longo prazo. [...] Planejar com antecedência significa que os dados recolhidos serão facilmente identificados, encontrados, acessados e reutilizados por você e por outros pesquisadores (quando apropriado). (SIBi, 2019).

Observa-se que a utilização/adoção dos Princípios FAIR na etapa de planejamento fará com que os dados sejam descobertos e possam ser reutilizados com qualidade. Além disso, favorece ter a clareza do que será de fato necessário na pesquisa, proporciona economia e maior assertividade na própria coleta e geração dos dados, evitando material excedente e desnecessário.

Existem ferramentas que auxiliam na criação do plano de gestão de dados. Uma delas é desenvolvida pelo Digital Curation Centre (DCC), o *DMPonline*⁷; outra é o *DMPTool*⁸, fornecido pelo Centro de Curadoria da Universidade da Califórnia. Uma terceira ferramenta é a *EasyDMP*⁹ da UNINETT Sigma2AS (Sigma2), organização que “gerencia a e-infraestrutura nacional para dados em grande escala e ciência da computação da Noruega”. Esta ferramenta atende com *templates* do PGD direcionados para projetos no Horizonte 2020, Ciência Europeia, mas também possibilita adaptar para outras instituições específicas que atuam com financiamento público. (UNINETT SIGMA2, 2020, tradução nossa). Deve-se atentar aos modelos disponíveis que ainda são versões beta, e dessa forma não permitem o uso em todas as etapas pré/durante/pós-projeto.

Todos os modelos citados oferecem possibilidade de consultar planos de gestão de dados genéricos e públicos, divulgados pelos usuários da ferramenta *online*, porém estes documentos não são avaliados por qualidade, integridade ou aderência a órgãos financiadores.

⁷ <https://dmponline.dcc.ac.uk/>. Nos idiomas alemão, espanhol, francês e inglês (GB, US).

⁸ https://dmptool.org/general_guidance. Nos idiomas inglês e português (BR).

⁹ <https://easydmp.sigma2.no/>. No idioma inglês.

O Digital Curation Centre (DCC) oferece também um *checklist*¹⁰ para a elaboração de um plano de gestão de dados, atualmente na versão 4.0. São perguntas e orientações de como fazer, com um documento bastante prático para quem deseja ou necessita elaborar um PGD. Aborda detalhadamente as seguintes categorias de informações: dados administrativos, coleta de dados, documentação e metadados, questões éticas e conformidades legais, armazenamento e *backup*, seleção e preservação, compartilhamento de dados e responsabilidades, e recursos.

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 e legais das atribuições de licenças.

No Brasil, as Agências buscam se adequar ao contexto internacional e solicitar a elaboração dos PGD como requisito em projetos de pesquisa. Portanto, um dos objetivos do estudo é verificar se as agências de fomento selecionadas recomendam a adoção dos Princípios FAIR na elaboração do PGD. Destaca-se que esta é uma tendência no cenário internacional: usar o FAIR como base para que os dados possam estar acessíveis, interoperáveis e sejam citáveis no processo de compartilhamento e reuso.

No contexto internacional, o Horizonte 2020¹¹ é o maior programa de pesquisa e inovação da União Europeia, e tem o

apoio político dos dirigentes europeus e dos membros do Parlamento Europeu. Estes decidiram que o investimento em investigação e inovação é crucial para o futuro da Europa e, assim, colocaram-no no centro da estratégia Europa 2020 para um crescimento inteligente, sustentável e inclusivo. (COMISSÃO EUROPEIA, 2019, p. 05).

Por primar pela inovação e excelência científica em pesquisa, a Europa determinou que, a partir de 2017,

todos os projetos gerados no âmbito do H2020 participam automaticamente do piloto para abertura de dados *Open Research Data Pilot* [ORD] que tem os seguintes requisitos: 1) desenvolver e manter atualizado um Plano de Gerenciamento de Dados; 2) depositar os dados em um repositório de dados de pesquisa; 3) assegurar que terceiros possam livremente acessá-los, explorá-los, reproduzi-los e disseminá-los; 4) esclarecer quais ferramentas serão necessárias para o uso dos dados brutos utilizados na validação dos resultados da pesquisa (ou fornecer as próprias ferramentas). (SANTOS, 2017, p. 139, grifo nosso).

Ao observar os investimentos desses países em pesquisas para inovação e projetos voltados à ciência aberta, avalia-se que as iniciativas possam servir como norteadoras na

¹⁰ http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf.

¹¹ Programa H2020 “com quase 80 bilhões de euros de financiamento disponível ao longo de sete anos (2014 a 2020)”. (SANTOS, 2017, p.139).

construção de uma ciência mais democrática para promover justiça e cidadania de forma global.

Outra iniciativa robusta da Comunidade Europeia é o denominado *Plan S*. Desenvolvido por uma aliança/coalizão entre 11 Agências de Fomento (iniciado em Set. 2018 - cOAlition S¹²), incluindo a European Commission and the European Research Council (ERC), e que pretende acelerar a transição para o amplo acesso aberto.

O *Plan S* tem como princípio que

todas as publicações acadêmicas sobre os resultados de pesquisas financiadas por doações públicas ou privadas fornecidas por conselhos e organismos de financiamento nacionais, regionais e internacionais, devem ser publicadas em periódicos de acesso aberto, em plataformas de acesso aberto ou feitas imediatamente disponíveis através de repositórios de acesso aberto sem embargo. (EUROPEAN SCIENCE FOUNDATION, 2020).

O *Plan S* vigorará a partir de 1º de janeiro de 2021, mas existirá um prazo a partir da adesão para que os envolvidos (pesquisadores, instituições, editores e repositórios) façam ajustes/alterações e para que as políticas dos financiadores se desenvolvam, adequem e se consolidem. Estes vários projetos de colaboração entre importantes organizações de pesquisa, deixam claro o interesse da Comunidade Europeia em acelerar o processo em prol de uma ciência aberta e global.

2.2.2 Repositórios de dados abertos

Para que os dados científicos fiquem preservados ao longo do tempo são necessárias infraestruturas que os mantenham armazenados e acessíveis, e os repositórios de dados fazem parte desta infraestrutura tecnológica e gerencial durante o ciclo de vida dos dados.

No centro desse arcabouço estão os **repositórios digitais de dados de pesquisa** que, por muitas razões e demandas, rapidamente se tornam parte essencial da infraestrutura mundial de pesquisa. Duas dessas demandas são determinantes para a ampliação das ações em torno da gestão de dados e para o seu ordenamento: as políticas mandatórias das agências financiadoras de pesquisa e a incorporação pelas instituições de pesquisa e pela sociedade em geral dos valores e princípios preconizados pela Ciência Aberta. (SAYÃO; SALES, 2016, p. 92).

No contexto internacional o Re3data, criado em 2012 e financiado pela German Research Foundation (DFG), “promove uma cultura de compartilhamento, maior acesso e visibilidade dos dados da pesquisa.” O Re3data “é um registro global de repositórios de dados

¹²Mais informações e diretrizes para a implementação do *Plan S*, podem ser encontradas em: <https://www.coalition-s.org/>. A partir de janeiro de 2020, todos os financiadores que concordarem em adotar o Plan S em suas políticas, terão um ano após contrato para se adequar.

de pesquisa que abrange diferentes áreas do conhecimento”. (PAVÃO, *et al.*, 2018, p. 06). Ele é uma ferramenta que pode auxiliar os pesquisadores a identificar repositórios por áreas de domínio e particularidades de suas pesquisas.

Através de consulta aos registros no Re3data, aplicando o filtro por país, foram identificados dez repositórios brasileiros em domínios diferentes, seis em parcerias com outras instituições internacionais e quatro como exclusivamente nacionais. Em 2019 eram oito, sendo destes três exclusivamente nacionais.

O quadro a seguir apresenta informações relativas à identificação dos repositórios, as áreas temáticas, tipo de repositório, o padrão de metadados utilizado e o país(es) de origem.

Quadro 01 - Repositórios brasileiros cadastrados no Re3data

Repositórios Brasileiros Cadastrados no Re3data

Nome do Repositório	Áreas / Assuntos	Tipo de Repositório	Padrão de Metadados	País(es)
Base de Dados Científicos da Universidade Federal do Paraná.	Ciências da Engenharia; Ciências Naturais; Ciências da Vida; Humanidades e Ciências Sociais.	Institucional	Dublin Core	Brasil
Repositório de dados de pesquisa CEDAP - dados de pesquisa. Universidade Federal do Rio Grande do Sul.	Ciências da Engenharia; Ciências Naturais; Ciências da Vida; Humanidades e Ciências Sociais.	Institucional	Dublin Core	Brasil
Dados Abertos De Pesquisas - Repositório de Dados de Pesquisa do Instituto Federal de Goiânia.	Ciências da Engenharia; Ciências Naturais; Ciência da Computação, Engenharia Elétrica e de Sistemas; Química; Ciências da Vida; Agricultura, Silvicultura, Horticultura e Medicina Veterinária; Biologia.	Institucional	OAI-ORE - Open Archives Initiative Object Reuse and Exchange; Dublin Core; Esquema de Metadados DataCite; DDI - Iniciativa de Documentação de Dados.	EUA Brasil Internacional
Banco de Dados de Exploração e Produção - BDEP. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis.	Ciências Naturais; Geoquímica, Mineralogia e Cristalografia; Geofísica e Geodésia; Geociências (incluindo geografia).	Não informado	Não informado	Brasil
GLOBE (Global Collaboration Engine) - Universidade de Maryland, Condado de Baltimore.	Biologia; Ciências de Plantas; Geociências (incluindo geografia); Ciência Atmosférica e Oceanografia; Ciências da Vida; e Ciências Naturais.	Disciplinar	Não informado	EUA Brasil

Continuação

Nome do Repositório	Áreas / Assuntos	Tipo de Repositório	Padrão de Metadados	País(es)
Rede IBICT Cariniana Dataverse - Instituto Brasileiro de Informação em Ciência e Tecnologia Rede Cariniana Dataverse.	Humanidades; História; História da Ciência; Ciências Sociais e Comportamentais; Ciências da Vida; Ciências da Engenharia; Ciências Naturais; Humanidades e Ciências Sociais.	Institucional	DDI - Iniciativa de Documentação de Dados	Brasil Internacional
Programa Internacional de Descoberta do Oceano - IODP.	Geociências (incluindo geografia); Oceanografia; Geofísica e Geodésia; Geologia e Paleontologia; Geoquímica, Mineralogia e Cristalografia; Ciências Naturais; Ciência Atmosférica e Oceanografia.	Disciplinar	ISO 19115: 2003. Informações geográficas - Metadados	EUA - Japão Alemanha - China União Europeia República da Coreia Austrália - Índia Brasil
Open Research Data @ PUC-Rio. [1] Coleção Digital - Dados de Pesquisa @ PUC-Rio	Engenharia de sistemas; Ciência da Computação, Engenharia Elétrica e de Sistemas; Ciências da Engenharia.	Institucional	Não informado	Brasil
Repositório de Dados PPBio - Programa de Pesquisa em Biodiversidade. (Rede de Pesquisa Ecológica de Longo Prazo).	Ciências da Vida; Biologia; Ecologia Vegetal e Análise de Ecossistemas; Ecologia Animal, Biodiversidade e Pesquisa de Ecossistemas; Ciências de Plantas; Zoologia.	Disciplinar	EML - Linguagem de Metadados Ecológicos	Brasil Internacional EUA
WorldClim - Global Climate Data. Centro de Referência em Informação Ambiental - CRIA.	Ciência Atmosférica e Oceanografia; Geociências (incluindo geografia); Ciências Naturais.	Disciplinar	Não informado	EUA - Brasil Austrália - Índia Países Baixos

Fonte: Adaptado de Re3data (2020, tradução nossa).

Em agosto de 2020, o Re3data tinha em seus registros 2563 repositórios cadastrados, porém, este dado é constantemente atualizado, o que denota iniciativas crescentes para a gestão dos dados científicos em longo prazo, favorecendo as possibilidades de reuso.

Segundo Rodrigues (2018, p. 37), os repositórios devem ser mais ativos no processo de comunicação científica, deixar de ser “recipientes passivos e isolados das versões finais dos resultados de pesquisa publicados”. Menciona a visão da nova geração de repositórios, propagada pela Confederation of Open Access Repositories (COAR), focada em:

Posicionar os repositórios como a base para uma infraestrutura de comunicação acadêmica e científica distribuída e globalmente interligada, sobre a qual serviços de valor acrescentado podem ser disponibilizados, transformando o atual sistema e tornando-o mais centrado nas necessidades da pesquisa, aberto e favorável à

inovação, e coletivamente gerido pela comunidade acadêmica. (RODRIGUES, 2018, p. 39).

Conforme Uzwyshyn (2016, p. 01), um repositório de dados “permite exame, prova, revisão, transparência e validação dos resultados de um pesquisador por outros especialistas além do artigo acadêmico publicado”. Pode publicar os dados negativos de pesquisa para que outros pesquisadores evitem trilhar os caminhos infrutíferos das experiências que não deram certo. Isso acelera e melhora o processo da produção científica de forma global.

De acordo com a Fiocruz (2020), “inovações orientadas a dados vem surgindo, tais como a sua disponibilização como material suplementar em artigos científicos, os repositórios de dados, os artigos de dados (*data papers*¹³) entre outros”. Essas publicações em geral, se destinam a publicar a metodologia de coleta de dados, detalhando softwares utilizados, forma de processamento, estrutura, potencial de reuso e link para acesso ao repositório dos dados (FIOCRUZ, 2020). Essas inovações compõem novas infraestruturas e derivam da relevância informacional dos dados para todas as áreas da pesquisa e ensino.

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) lançou em outubro de 2020, um conjunto de publicações nas quais constam “diretrizes voltadas para repositórios de dados e de publicações científicas”, com abordagem em padrões de interoperabilidade e boas práticas em repositórios. Estas publicações podem servir como norteadoras para os gestores e as instituições no desenvolvimento de políticas e na implementação de repositórios de dados, favorecendo a gestão de dados científicos no âmbito da Ciência Aberta no Brasil. (IBICT, 2020).

Segundo o autor Michener (2015), armazenar dados por um longo período requer estruturas mais robustas e repositórios de dados adequados ao domínio da pesquisa.

Muitas disciplinas mantêm repositórios específicos, como o GenBank, para dados de sequências de nucleotídeos e o banco de dados de proteínas para sequências de proteínas. Da mesma forma, muitas universidades e organizações também hospedam repositórios institucionais, e existem vários repositórios gerais de dados científicos, como o Dryad (<http://datadryad.org/>), o Figshare (<http://figshare.com/>) e o Zenodo (<http://zenodo.org/>). Como alternativa, pode-se procurar facilmente repositórios específicos de disciplina e de uso geral por meio de catálogos on-line, como <http://www.re3data.org/> (ou seja, REgistry of REsearch data REpositories) e <http://www.biosharing.org> (isto é, BioSharing). Muitas vezes, é considerado uma boa prática depositar código em um repositório host como o GitHub, que é especializado em gerenciamento de código-fonte, bem como em alguns tipos de dados, como arquivos grandes e dados tabulares (<https://github.com/>). (MICHENER, 2015, não paginado, tradução nossa).

¹³Os *data journals* (periódico de dados) são publicações regulares, com sistema de revisão por pares, concebidas para compartilhar dados com potencial de reuso por terceiros. Eles publicam *data papers* (artigos de dados) nos quais os autores compartilham dados brutos de maneira estruturada e legível para humanos. (FIOCRUZ, 2020).

Outro exemplo de repositório multidisciplinar de dados é o Harvard Dataverse¹⁴. Ele “permite o compartilhamento, preservação, citação, exploração e análise dados de pesquisa. Está aberto a todos os pesquisadores, dentro e fora da comunidade de Harvard”. (FIOCRUZ, 2020e). É considerada uma infraestrutura robusta que pode ser avaliada no processo de escolha de um repositório de dados multidisciplinar.

O relatório ‘Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores’, publicado em 2018, enfatiza a necessidade de divulgar esta infraestrutura e menciona a necessidade de “especialistas treinados para a gestão de dados de pesquisa, de modo a garantir a usabilidade destes dados em projetos futuros”. (VANZ, *et al*, 2018, p. 36).

O relatório aponta ainda, que precisam ser esclarecidos conceitos acerca de dados científicos, dos próprios repositórios de dados e de “práticas de disponibilização e reuso”. O que reforça a importância de elucidar o PGD em atender aos critérios que permitam o compartilhamento e reuso com segurança e confiabilidade.

A estreita relação entre os repositórios institucionais e os metadados é indiscutível. Os metadados são usados para definir permissões, direitos de acesso, compartilhamento, reutilização, redistribuição e políticas, bem como os requisitos técnicos para visualização, acesso ou preservação de objetos digitalizados ou concebidos originalmente em formato digital. (PAVÃO, *et al*, 2015, p.104).

“Durante a seleção e desenvolvimento do esquema de metadados que será utilizado no repositório deve-se fazer o confronto contínuo com os seis princípios dos ‘bons metadados’ publicados pela NISO (2007)”. (PAVÃO, *et al*, 2015, p.109).

A figura 06 aponta estes princípios.

¹⁴ Endereço: <https://dataverse.harvard.edu/>

Figura 06 – Princípios para bons metadados



Fonte: Adaptado de (PAVÃO, *et al*, 2015, p. 109).

Na escolha do padrão de metadados é importante observar se estão condizentes e atendem aos padrões da comunidade, tanto para usuários atuais como futuros. Metadados que atendam aos princípios da figura 06 favorecem a descoberta de conjuntos de dados mesmo quando estes ainda não estejam em acesso aberto, pois detalham as restrições ou motivos do embargo.

Alguns exemplos de esquemas de metadados são citados, tais como: Dublin Core (DCMS), b) Learning Object Metadata (LOM), c) Visual Resources Association (VRA), d) Encoded Archival Description (EAD), e) Conceptual Reference Model (CIDOC/CRM), f) Metadata for Images in XML (NISO/MIX), g) MACHine-Readable Cataloging (MARC), h) Metadata Object Description Standard (MODS), i) Metadata Encoding & Transmission Standard (METS), j) Preservation Metadata. (Adaptado de PAVÃO, *et al*, 2015, p.105-106).¹⁵

A DCMI é provavelmente a iniciativa mundial mais conhecida no que diz respeito a esquemas de metadados para descrição de objetos digitais, principalmente textuais, em repositórios digitais, e proporciona uma base para o desenvolvimento dos mesmos. O Dublin Core (DC) adota a sintaxe do Resource Description Framework (RDF) e surgiu como uma alternativa para simplificar e, conseqüentemente, tornar a descrição e recuperação de objetos digitais na web mais rápida e econômica sem, contudo, deixar de seguir um padrão mínimo, indispensável para a interoperabilidade entre os sistemas. (PAVÃO, *et al*, 2015, p. 106).

¹⁵Para detalhes de cada esquema de metadados pode ser consultado o artigo: Metadados e Repositórios Institucionais: uma relação indissociável para a qualidade da recuperação e visibilidade da informação - (PAVÃO, *et al*, 2015, p.115-116). DOI: 10.9771/1981-6766rpa.v9i3.15163.

Os metadados compõem elementos primordiais a várias facetas ou subprincípios dos denominados Princípios FAIR, além, de corresponder a fatores preponderantes para a sua implementação.

2.3 PRINCÍPIOS FAIR

Estes princípios constam da literatura internacional desde 2014¹⁶, quando ocorreram discussões acerca dos novos contextos gerados pela *e-Science*, e foram publicados em 2016¹⁷. Surgindo, inicialmente, como recomendações de boas práticas para a ciência aberta, porém, observam-se iniciativas na comunidade europeia para que estes princípios sejam estabelecidos como requisitos à gestão de dados.

Os Princípios FAIR foram formulados após o Lorentz Center workshop, em janeiro de 2014, quando um grupo diversos atores, compartilhando interesse na publicação e reutilização de dados científicos, reuniu-se para discutir as características requeridas dos ambientes contemporâneos de publicação de dados científicos. [...] O primeiro projeto dos Princípios FAIR foi publicado no *site* Force11 para avaliação e comentários por parte de uma comunidade mais ampla – um processo que durou quase dois anos. (SCIENTIFIC ELECTRONIC LIBRARY ONLINE, 2016).

Ressalta-se que estes princípios não são ‘fechados’, segundo os próprios autores mencionam, eles devem ser “inteligentemente abertos” para que possam respeitar as particularidades das diferentes áreas do saber científico. Eles orientam para que as escolhas na gestão de dados possam ser assertivas e permitam o compartilhamento e reuso no contexto da ciência aberta, respeitando as questões éticas e legais. A Comissão Europeia preconiza que os dados científicos devam ser “tão abertos quanto possível, tão fechados quanto necessário”. Concerne destacar que nem todos os dados FAIR, estão em acesso aberto, por isso, a grande importância de metadados ricos e detalhados para contexto e significado.

“Os princípios do FAIR fornecem diretrizes, não requerem tecnologias específicas e permitem que as comunidades de partes interessadas definam padrões FAIR específicos e desenvolvam métricas para quantificá-los”. (DUMONTIER; WESLEY, 2018, p. 39, tradução nossa).

Constituem-se em Princípios FAIR – aqueles que “têm por objetivo tornar os dados de

¹⁶ Integram o relatório da conferência "Projetando Conjuntamente um DATA FAIRPORT", realizada de 13 a 16 de janeiro de 2014, em Leiden, Holanda. O principal desafio abordado foi a rápida mudança para dados impulsionados pela *e-Science*, desencadeada pela capacidade de criar conjuntos de dados cada vez maiores e mais complexos para a descoberta do conhecimento. (Resumo e tradução nossa).

¹⁷ Publicados em 2016, após discussões realizadas por especialistas de diferentes áreas de domínio. Surgem como balizadores para facilitar a descoberta, interoperabilidade e reuso dos dados científicos. Publicação de WILKINSON, M., et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*, n. 3, 2016.

pesquisa (por intermédio de metadados) encontráveis, acessíveis, interoperáveis e reutilizáveis (*findable, accessible, interoperable and reusable*), assegurando, dessa forma, uma boa gestão dos dados”. (SANTOS, 2017, p. 124). Esta definição é também corroborada através do glossário de termos da Fiocruz (2020a), quanto a tornar os dados de pesquisas qualificados a novas descobertas, através do uso de metadados.

Nesse sentido, é válido evidenciar a relevância da padronização dos metadados. Conforme os autores Viana, Márdero Arellano; Shintaku (2005, p.15), “os metadados dão significado semântico a dados aparentemente isolados, provendo um contexto ao processo de busca”. Um dos formatos desta regulamentação é o padrão Dublin Core. Estes padrões visam promover a interoperabilidade dos dados entre quem os cria e os usuários posteriores.

Este contexto com significação é fundamental para o processo de reutilização dos dados científicos. Em situações em que os dados não possam estar à disposição em acesso aberto, os metadados podem servir como elementos primordiais à descoberta de pesquisas, que porventura tenham os dados em período de embargo, ou indisponíveis por motivos como privacidade, confidencialidade, questões contratuais e legais, processos de patentes, finalização de estudos, etc.

Vale destacar que o sigilo é temporal e os prazos devem ser previstos nas políticas para dados científicos abertos, a exemplo das classificações prevista na Lei Acesso a Informação (Lei nº 12.527, de 18 de novembro de 2011). Nos casos de restrição de acesso, os metadados podem ofertar informações sobre os responsáveis e como conseguir os dados, em casos especiais. Ex: atualmente é fundamental o acesso, compartilhamento e reuso de dados na área da saúde, no menor tempo possível para acelerar as pesquisas no intuito de desenvolver uma vacina contra o vírus SARS-CoV-2, a fim de mitigar o avanço da Pandemia (COVID-19). Nesse caso, o tempo é fator decisivo na preservação de milhares de vidas e a ciência está sendo pressionada a avançar de forma mais rápida.

Porém, somente compartilhar não basta, pois se os dados não estiverem estruturados de forma a permitir acesso e interoperabilidade, eles perdem seu valor para reuso.

Apontados como orientadores para as boas práticas na gestão de dados científicos apresentam-se, no quadro 02, os Princípios FAIR e seus subprincípios.

Quadro 02 - Princípios FAIR



Fonte: Adaptado de WILKINSON, *et al* (2016). Tradução: Santos (2017, p. 55-56).

O termo '(Meta)dados' significa que o Princípio deve ser aplicado aos metadados e aos dados. (WILKINSON, *et al.*, 2016).

Para atender ao princípio de ser **encontrável**, o objeto de dados deve ter um identificador único e persistente, ter metadados¹⁸ claros que o distingam unicamente, e isso deve ser possível de ser feito também por máquinas. Os dados serão **acessíveis** “na medida em que podem ser sempre obtidos por máquinas e seres humanos”. (AVENTURIER, 2017, não paginado).

Atenderão ao princípio de ser **interoperáveis**, se os metadados forem passíveis de serem processados por máquinas, os formatos de dados utilizarem vocabulários compartilhados e/ou ontologias e os dados dentro do objeto forem tanto sintática quanto semanticamente processáveis por máquinas. Serão **reutilizáveis**, se cumprirem aos três princípios anteriores, além de ter metadados ricos e bem descritos que permitam indicar sua proveniência, facilitar a indicação de fonte e possam ser comparados a outras fontes de dados. (AVENTURIER, 2017, não paginado).

Em uma ação recente, novembro de 2018, foi publicado o livro ‘Transformando FAIR em realidade: relatório final e plano de ação do grupo de peritos da Comissão Europeia sobre

¹⁸Metadado – informação que descreve características de determinado dado, explicando-o em certo contexto de uso. (FIOCRUZ, 2019).

dados FAIR¹⁹, este documento visa incentivar o avanço do movimento para a ciência aberta Europeia e global. Através da divulgação de um conjunto de recomendações para facilitar a criação e implementação dos Princípios FAIR, não somente na comunidade europeia, mas também em quaisquer outras instituições que tenham interesse em gerir adequadamente seus dados científicos, para “acelerar a pesquisa, envolver o poder da análise de máquinas em escala, garantindo a transparência, a reprodutibilidade e a utilidade social, [dos] dados e outros objetos digitais criados e utilizados para pesquisa”. (EUROPEAN COMMISSION, 2018).

Importante destacar que existe a necessidade de adotar

políticas e serviços de informação de forma consistente, voltadas para os diferentes domínios; além de investimentos no desenvolvimento padrões e tecnologias apropriados que possam suportar os dados FAIR de forma harmonizada e coerente. [...] Muitas soluções técnicas possuem características e objetivos comuns, mas precisam ser implementadas de maneira que façam sentido para a disciplina, de forma a atender às necessidades particulares da pesquisa que está sendo conduzida. (HENNING *et al.*, 2018, p. 5064).

A autora afirma ainda, que é necessária uma “maior conscientização de todos envolvidos no âmbito das práticas científicas, desde pesquisadores, comunidades de pesquisa, universidade e agências financiadoras” para que a adoção dos princípios se consolide como uma boa prática. (HENNING *et al.*, 2018, p. 5064).

Atualmente, os Princípios FAIR para a gestão de dados estão sendo recomendados e podem ter adesão voluntária, porém a tendência é que se tornem mandatórios, inicialmente pelas Instituições de Fomento Públicas, até como uma resposta dos investimentos em pesquisa para a sociedade.

Um relatório publicado em 2018 (*State of Open Data* de 2018, publicado pela Digital Science²⁰) realizado com 748 pesquisadores (independente do campo), aponta que “somente 144 dos pesquisadores sabem, ou estão familiarizados sobre os Princípios FAIR”, o que corresponde a 19,25% dos pesquisados. Este estudo faz também uma apresentação por áreas do conhecimento humano. Em Ciências Sociais, de 134 respondentes, tem-se: 26 que afirmaram estar familiarizados com os Princípios FAIR, 35 já ouviram falar dos Princípios FAIR, mas não os conhecem, e, 73 afirmaram nunca ter ouvido falar dos Princípios FAIR. (BROCK, 2019).

¹⁹ Tradução de “Turning FAIR into reality: Final Report and Action Plan on FAIR Data”.

²⁰ Este relatório foi publicado pela Digital Science, que faz parte do Holtzbrinck Publishing Group, uma empresa de mídia global dedicada à ciência e educação. (Tradução nossa).

Figura 07 – Você conhece os Princípios FAIR?



Fonte: Elaborado a partir de Fane e Treadway (2018, p.10).

Conforme a figura 07, referente à pergunta ‘se conhece os Princípios FAIR’ Fane e Treadway (2018, p.11), afirmam no relatório que “a porcentagem de entrevistados que relataram estar familiarizados com os Princípios FAIR foi de apenas 15%, sendo que 25% já ouviram falar sobre FAIR e 60% nunca ouviram falar deles.” (Tradução nossa).

Os autores ainda observam que “essa falta de conscientização é preocupante, pois os Princípios FAIR estão sendo rapidamente adotados pelos editores, financiadores e instituições em todo o mundo, mas há uma lacuna crucial na educação dos pesquisadores sobre o que se espera deles”. (FANE; TREADWAY, 2018, p.11, tradução nossa).

Também afirmam que o princípio que mais gera dúvidas quanto à definição é o que se refere à interoperabilidade (apontado por 41% dos participantes da pesquisa). O que indica para a necessidade de iniciativas de conscientização e esclarecimento junto à comunidade científica (pesquisadores, editores, bibliotecários, gestores de dados...) no intuito de que tenham a clareza dos Princípios FAIR e compreendam a importância da gestão durante todo o ciclo dos dados, como um diferencial para a ciência na atualidade.

Uma vez que a produção de dados é exponencial, não estruturada e heterogênea, especialmente os gerados pela cauda longa da ciência, se reforça a importância de diretrizes que possam facilitar a gestão de todo o ciclo desse conjunto de dados.

A ciência da cauda longa não é um sinônimo de questões científicas menos relevantes ou de uma ciência menor, tampouco os dados que ela gera podem ser desconsiderados em um contexto cuja diversidade e integração de dados são a chave para respostas para novas indagações da ciência. Ao contrário, é um território com altos coeficientes de autonomia, que favorece a inovação e a geração de conhecimentos multi e interdisciplinares. (SAYÃO; SALES, 2019a, p.41-42).

Na sequência apresenta-se quadro com algumas características dos dados gerados pela ciência nos contextos da *big data* e cauda longa.

Quadro 03 – Comparativo - Dados *Big Data* e Cauda Longa

Quadro comparativo das principais características		
	Big Data	Cauda Longa
Volume de Dados	Grande	Pequeno
Uniformidade	Homogêneos	Heterogêneos
Coleta de Dados	Automatizada	Manual
Documentação	Estruturada	Menos estruturada
Curadoria	Centralizada	Individual
Diretrizes e normas	Uniformes e estabelecidas	Únicas ou inexistentes
Repositório de preferência	Disciplinar ou temático	Institucional, quando depositado
Visibilidade para terceiros	Visível	Invisível
Acesso	Aberto	Fechado
Reutilização	Frequente	Rara
Desafio	Atuar sobre grandes volumes	Melhorar a curadoria

Fonte: Elaborado pela Fiocruz (2020) baseado em Heidorn (2008).

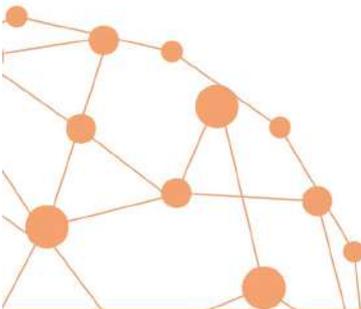
A partir desse quadro é possível dimensionar os desafios que a gestão de dados científicos necessita enfrentar/gerenciar. Segundo Sayão (2020), “existem duas a três vezes mais dados sendo produzidos na cauda longa da ciência do que no *big data*. E esses dados são gerados por pequenos laboratórios, grupos de pesquisa, Agências menores e sua disponibilização ainda é mais restrita”. Nacionalmente, já existem iniciativas para buscar gerenciar eficazmente e tornar essa gama informacional disponível por meio de boas práticas de pesquisa.

Exemplo brasileiro de ação para implementação dos Princípios FAIR na gestão de dados pode ser observado no quadro 03, se refere ao documento “Termo de Referência - Gestão e Abertura de Dados para Pesquisa na Fiocruz”. Iniciativa recente, mas que aponta para o interesse das instituições de fomento à pesquisa em aderir às boas práticas na gestão de seus dados científicos, acompanhando o movimento de acesso aberto ao conhecimento.

Além de prever boas práticas por meio da adoção dos Princípios FAIR, sempre que possível, também busca instrumentos para avaliar e monitorar o grau ou nível de atendimento

aos Princípios, num processo denominado FAIRness.

Quadro 04 - Indicação para adoção dos Princípios FAIR na gestão de dados

<p>D 2.9</p> <p>Adotar os princípios FAIR na gestão de dados e utilizar indicadores e métricas para avaliar o grau de aplicação dos princípios FAIR (fairness):</p>	<p>Tornar os dados encontráveis, sempre que possível, por intermédio de identificadores globais, persistentes e únicos, com metadados detalhados.</p>
	<p>Tornar os dados acessíveis, sempre que possível, possibilitando a sua recuperação por protocolos de comunicação padronizáveis, abertos, gratuitos e universalmente implementáveis. Os metadados devem estar acessíveis, mesmo quando os dados não estiverem disponíveis.</p> <p>Tornar os dados interoperáveis, sempre que possível, sendo processáveis por máquina, usando uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento, incluindo vocabulários controlados disponíveis dentro dos mesmos princípios FAIR e referências qualificadas para outros metadados.</p> <p>Tornar os dados reutilizáveis, sempre que possível, utilizando-se de metadados ricamente descritos, com uma pluralidade de atributos precisos, relevantes, liberados com licenças apropriadas de uso, associadas às suas procedências.</p>

Fonte: Adaptado de FIOCRUZ (2019d, p.08).

Percebe-se, no entanto, que a conscientização e o uso de infraestruturas apropriadas, bem como o investimento em serviços de 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.

A iniciativa internacional GO FAIR busca auxiliar na implementação dos Princípios FAIR de forma global. Atua em todas as áreas de domínio e no Brasil tem um Escritório de Apoio, denominado ‘GO FAIR Brasil’, sediado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT. Tem foco em “difundir, apoiar e coordenar as atividades relacionadas à adoção da estratégia de implementação dos princípios FAIR definida pela iniciativa GO FAIR em todo o território brasileiro”. (GO FAIR, 2020).

A iniciativa GO FAIR apresenta opções de recursos e ferramentas passíveis de uso para cada um dos subprincípios FAIR. Mas é importante esclarecer que ela não limita ou indica ferramentas tecnológicas específicas para o processo de implementação, mas sim serve de apoio.

Enfatiza que os Princípios de dados

FAIR se aplicam a metadados, dados e infraestrutura de suporte (por exemplo, mecanismos de pesquisa). A maioria dos requisitos de localização e acessibilidade

pode ser alcançada no nível de metadados. Interoperabilidade e reutilização exigem mais esforços no nível de dados. (GO FAIR, 2019, tradução nossa).

A iniciativa GO FAIR se apoia em três pilares. De acordo com Henning, *et al* (2019, p. 407), estes pilares envolvem:

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.

De maneira geral, os aspectos mencionados - mudanças culturais, treinamentos e infraestruturas para dados - estão interligados. É necessário conhecer, ter especialistas capacitados em gestão de dados, em diferentes locais e áreas de domínio para estimular as mudanças entre os pesquisadores e profissionais, além, é claro, de ter infraestruturas adequadas. Estas condições requerem investimentos e divulgação para a concretização, porém, as transformações no campo científico decorrentes do movimento de acesso aberto já sinalizam mudanças.

Importante destacar que o processo ideal e menos oneroso, seria efetuar a curadoria dos dados desde o início para que sejam gerados seguindo as orientações FAIR. Na impossibilidade disso, devem ser avaliados quais dados devem passar pelo processo, qual sua relevância para novos contextos, as possibilidades de reuso, valor histórico, cultural, social, etc.

Adotar as recomendações FAIR estimula novas práticas nos ambientes de pesquisa e adiciona valor aos dados científicos no cenário da ciência aberta.

3 OPÇÕES E PROCEDIMENTOS METODOLÓGICOS

Esta seção detalha a execução da pesquisa. A metodologia que determina como os dados foram coletados e analisados para alcançarem os objetivos. Também é responsável por garantir o desenvolvimento da pesquisa atestando o caráter científico, a validação da pesquisa em si e do conhecimento resultante deste processo investigativo.

Destaca-se que em ciência não interessam somente os resultados, e sim todos os métodos e procedimentos adotados pelo pesquisador no decorrer do processo de construção do estudo. Este caminho percorrido e documentado corretamente pode ser avaliado por pares e membros da comunidade científica, o que confere maior legitimidade ao conhecimento produzido (BRAGA, 2007).

Ao final desta seção se apresenta a figura 08 no intuito de ilustrar as opções e procedimentos metodológicos adotados na pesquisa.

3.1 CARACTERIZAÇÃO DA PESQUISA

Esta pesquisa caracteriza-se como exploratório-descritiva. Exploratória ao buscar maior familiaridade com a temática, devido ao tema ser emergente e a escassez de produção bibliográfica a respeito. É descritiva, pois realizou uma análise detalhada das informações a partir do acesso aos *sites* oficiais das Agências de Fomento selecionadas (mediante o *ranking* das agências nacionais e internacionais que mais financiam pesquisas no Brasil)²¹.

A construção da fundamentação teórica que subsidiou o estudo utilizou-se de levantamento bibliográfico para identificar os principais autores, analisar as publicações mais relevantes e os modelos conceituais abordados dentro do princípio de acesso aberto e atinentes à temática de investigação. Complementarmente, a pesquisa documental consistiu em acessar aos documentos específicos sobre Plano de Gestão de Dados e/ou políticas de gestão, de cada uma das instituições pesquisadas.

O levantamento ocorreu por meio das seguintes bases de dados nacionais e internacionais: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); Library, Information Science & Technology Abstracts (LISTA);

²¹ As agências que mais financiam pesquisas no Brasil (nacionais e internacionais) foram mapeadas por meio da análise de projetos de pesquisas nacionais envolvendo pesquisadores PQs (com bolsa de produtividade pelo CNPq). Os dados abertos da pesquisa de Monteiro (2019b) constam no PPGInfo - Programa de Pós Graduação em Gestão da Informação (UDESC), via repositório ZENODO - DOI: 10.5281/zenodo.3372826.

Library and Information Science Abstracts (LISA); Web of Science e Scopus.

Os termos de busca para recuperação das publicações nas bases de dados foram definidos conforme quadro 05, nos idiomas português e inglês, a fim de que a recuperação seja a mais adequada possível.

Quadro 05 - Termos de busca para levantamento bibliográfico

TERMOS DE BUSCA	
Português <ul style="list-style-type: none"> • Ciência aberta • Acesso aberto • Princípios FAIR • Gestão de dados científicos • Gestão de dados de pesquisa • Plano de gestão de dados • Reuso de dados • Reutilização de dados de pesquisa • Dados abertos • Dados científicos abertos • Curadoria de dados 	Inglês <ul style="list-style-type: none"> • Open Science • Open Access • FAIR principles • Research data management • Data management plan • Data reuse • Open data • Open research data • Data curation

Fonte: Dados da pesquisa (2020).

Foram consultadas as bases mencionadas, no período entre janeiro a março de 2019, com uma nova busca sistematizada em junho de 2020, para identificação de novas publicações e cotejamento dos resultados anteriores. Ao final das buscas foram recuperados 57 documentos e realizadas leituras dos resumos para familiaridade e identificação de pertinência aos objetivos da pesquisa. Dentre eles, o artigo de Wilkinson, *et al* (2016), serviu como base na análise quanto aos Princípios FAIR. Ele elenca os quatro pilares que representam o acrônimo FAIR e detalha subprincípios desejáveis para a adoção desta boa prática.

Além destes documentos bibliográficos, outros foram utilizados para compor o referencial teórico. Estes encontrados através de buscas exploratórias (em acesso aberto) ou por indicação da Prof^a Dra Virginia Ortíz-Repiso Jiménez, da Universidad Carlos III de Madrid, tutora na Espanha, onde tive a oportunidade de realizar estágio de pesquisa durante o mestrado e trocar informações com outros pesquisadores.

Quanto à abordagem, tem caráter qualitativo, pois resultou de uma análise aprofundada aos documentos quanto aos critérios relativos às exigências/recomendações para

elaboração dos Planos de Gestão de Dados e com abordagem aos Princípios FAIR.

A pesquisa qualitativa responde a questões muito particulares. Ela se ocupa, nas Ciências Sociais, com um nível de realidade que não pode ou não deveria ser quantificado. Ou seja, ela trabalha com o universo dos significados, dos motivos, das aspirações, das crenças, dos valores e das atitudes. (MINAYO, 2011, p.21).

Segundo Lakatos e Marconi (2007, p. 271), “na pesquisa qualitativa há um mínimo de estruturação prévia. Não se admitem regras precisas, como problemas, hipóteses e variáveis antecipadas, e as teorias aplicáveis deverão ser empregadas no decorrer da investigação”.

3.2 FORMAÇÃO DO CORPUS DA PESQUISA

O *corpus* da pesquisa foi delineado a partir da pesquisa de Monteiro (2019a), que identificou as Agências (Instituições) que mais fomentam pesquisas brasileiras.

Para essa pesquisa, inicialmente foram selecionadas as cinco Agências Nacionais com políticas e/ou instruções para dados científicos abertos, identificadas (quadro 06). A partir desse número, foram selecionadas também as cinco primeiras instituições internacionais do *ranking* das Agências que mais fomentam pesquisas no Brasil nesta categoria (quadro 07).

Quadro 06 - *Ranking* das Agências Nacionais que mais fomentam pesquisas brasileiras

Ranking Nacional	Financiadores	Sigla	Qtde de Projetos	Políticas / Instruções de Dados Abertos
1º	Conselho Nacional de Desenvolvimento Científico e Tecnológico	CNPq	63.636	Identificadas
2º	Fundação de Amparo à Pesquisa do Estado de São Paulo	FAPESP	26.058	Identificadas
3º	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior	CAPES	14.870	Não identificadas
4º	Fundação de Amparo à Pesquisa do Estado de Minas Gerais	FAPEMIG	8.116	Identificadas
5º	Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro	FAPERJ	6.955	Não identificadas
6º	Financiadora de Estudos e Projetos	FINEP	4.098	Não identificadas
7º	Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul	FAPERGS	2.988	Não identificadas
8º	Empresa Brasileira de Pesquisa Agropecuária	EMBRAPA	2.337	Identificadas
9º	Petróleo Brasileiro S.A.	PETROBRAS	1.998	Não identificadas
10º	Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Estado do Paraná	FA	1.877	Não identificadas
11º	Fundação de Amparo à Pesquisa do Estado da Bahia	FAPESB	1.240	Não identificadas
12º	Fundação de Amparo à Ciência e Tecnologia de Pernambuco	FACEPE	1.151	Não identificadas
13º	Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico	FUNCAP	832	Não identificadas
14º	Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina	FAPESC	650	Não identificadas
15º	Fundação Oswaldo Cruz	FIOCRUZ	607	Identificadas

Fonte: Monteiro (2019a).

Quadro 07 - *Ranking* das Agências Internacionais que mais fomentam pesquisas no Brasil

Ranking Internac.	Financiadores	Sigla	Qtde. Projetos	País/Sede	Políticas / Instruções de Dados Abertos
1º	National Institutes of Health	NIH	392	EUA	Identificadas
2º	World Health Organization	WHO	376	Suíça	Identificadas
3º	European Union	EU	369	Bélgica	Identificadas
4º	Ford Foundation	FF	242	EUA	Não identificadas
5º	Comité Français d'Evaluation de la Coopération Universitaire et Scientifique avec le Brésil	COFECUB	183	França	Não identificadas
6º	Deutscher Akademischer Austauschdienst	DAAD	183	Alemanha	Não identificadas
7º	Fundação para a Ciência e a Tecnologia de Portugal	FCT	170	Portugal	Identificadas
8º	National Research Foundation	NRF	167	África do Sul	Identificadas
9º	Centre National de la Recherche Scientifique	CNRS	141	França	Identificadas
10º	United Nations Educational, Scientific and Cultural Organization	UNESCO	137	França	Identificadas
11º	International Atomic Energy Agency	IAEA	123	Áustria	Não identificadas
12º	International Foundation for Science	IFS	117	Suécia	Não identificadas
13º	United Nations Development Programme	UNDP	93	EUA	Não identificadas
14º	National Aeronautics and Space Administration	NASA	92	EUA	Identificadas
15º	National Science Foundation	NSF	89	EUA	Identificadas

Fonte: Monteiro (2019a).

Posteriormente, por entender que cada área de domínio tem suas especificidades, foi feito um recorte com as Agências que financiam pesquisas em áreas multidisciplinares, por serem, dentre as selecionadas, as com maior percentual de incidência. Além disso, as Agências que atuam em áreas multidisciplinares, geralmente apresentam *templates* mais variados e adaptáveis aos diferentes campos da pesquisa científica.

Sendo assim, o *corpus* de análise desta pesquisa é composto por seis Agências de Fomento, sendo três nacionais e três internacionais selecionadas a partir do *ranking* das Agências que mais financiam a pesquisa no Brasil. Como sintetizado no quadro 08.

Quadro 08 - Composição do *Corpus* de análise da Pesquisa: Agências de Fomento

Ordem	Financiadores	Sigla/Local
1	Conselho Nacional de Desenvolvimento Científico e Tecnológico	(CNPq) - Brasília/DF
2	Fundação de Amparo à Pesquisa do Estado de São Paulo	(FAPESP) - São Paulo/SP
3	Fundação de Amparo à Pesquisa do Estado de Minas Gerais	(FAPEMIG) Belo Horizonte/MG
4	European Union	(EU) – Bélgica
5	Fundação para a Ciência e a Tecnologia de Portugal	(FCT) – Portugal
6	National Research Foundation	(NRF) - África do Sul

Fonte: Dados da pesquisa (2020).

Para cada uma das Agências selecionadas investigou-se as exigências/recomendações específicas para a elaboração do Plano de Gestão de Dados e, na sequência, a verificação se os Princípios FAIR são mencionados e/ou contemplados nestes documentos.

3.2 INSTRUMENTO DE COLETA DE DADOS E OPÇÕES DE ANÁLISE

Como local de coleta de dados, efetuou-se levantamento das informações de interesse da pesquisa, por meio de consulta aos *sites* oficiais das Agências que compuseram o *Corpus* de análise da pesquisa.

Os dados e informações de interesse da pesquisa aqui proposta se relacionam especialmente aos dados científicos abertos, em especial a exigência e/ou recomendação para os Planos de Gestão de Dados, se estas atendem ou mencionam os Princípios FAIR para garantir que os dados sejam de fato encontráveis, acessíveis, interoperáveis e reutilizáveis. Em levantamento exploratório ao site oficial das Agências os documentos localizados pertinentes a temática que não estivessem em formato pdf., foram convertidos para este formato.

Para filtrar os documentos encontrados foi desenvolvida, junto com um profissional desenvolvedor de software, uma ferramenta em Python, que resultava na incidência de termos em cada documento PDF, para identificar sua relevância (por Agência). A ferramenta retorna o quantitativo de ocorrências dos termos-chave conforme uma lista definida previamente pela pesquisadora (espécie de vocabulário de termos) para facilitar e focar as análises. O apêndice A ilustra o retorno aos termos a partir da ferramenta Python.

Termos usados para essa seleção: [armazenamento, coleta de dados, curadoria, dados abertos, data management plan, data reuse, disponibilização de dados, DMP, FAIR data management, FAIR data, FAIR principles, gerenciamento de dados, gestão de dados, identificador persistente, interoperabilidade, metadados, metadata, open data, plano de gerenciamento de dados, plano de gestão de dados, plano de gestão, Princípios FAIR, repositório, repositórios de dados, repository, research data management, research management plan, responsabilidade legal, reuso, reutilização].

O quadro 09 demonstra a quantidade de documentos preliminares recuperados e a quantidade selecionada para análise após uso da ferramenta. Aqui foram excluídos também os que tratavam de dados abertos governamentais, por não ser foco desta pesquisa.

Quadro 09 – Ferramenta Python – Quantitativo de documentos PDFs analisados

Agências de Fomento analisadas							
	CNPq	FAPESP	FAPEMIG	EU	FCT	NRF	Total
Documentos preliminares	15	19	13	45	24	11	127
Selecionados para pesquisa (Python)	04	04	06	05	05	05	29

Fonte: Dados da pesquisa (2020).

A ferramenta desenvolvida possibilitou a verificação sintática dos termos selecionados, porém, a análise semântica dos documentos atendendo aos objetivos propostos na pesquisa foi executada pela pesquisadora através da leitura individual a cada um dos documentos selecionados.

Ao final classificaram-se como relevantes para o estudo um total de 29 documentos. As análises do conteúdo e inferências são descritas na seção 4 Análise dos Dados por Agência de Fomento. A seção mencionada seguirá um roteiro estrutural básico de análise para cada Agência com:

- a) Identificação da agência;
- b) Quadro de análises com critérios para Plano de Gestão de Dados (e o rol de documentos selecionados);
- c) Quadro de análise menção aos Princípios FAIR e subprincípios.

Para executar as análises ao Plano de Gestão de Dados utilizou-se como elemento norteador uma síntese de questões do *checklist* proposto pela Digital Curation Centre²² (DCC) e os Princípios FAIR com seus subprincípios - como elementos a serem mencionados/adotados pelas Agências na elaboração do PGD.

No quadro 10, apresentam-se os critérios propostos pelo DCC e *síntese* com perguntas e informações de cada item que compõe o PGD.

²² http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf.

Quadro 10 – Plano de Gestão de Dados – Síntese de questões orientadoras do *Checklist* DCC

Critérios	Questões orientadoras para análise
Dados administrativos	Um ID pertinente, conforme determinado pelo financiador e / ou instituição. Descrição do Projeto. // Políticas relacionadas. - 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?
Coleta de dados	Quais dados coletará ou criará? - 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? Como os dados serão coletados ou criados? - Quais padrões ou metodologias você usará? - Como vai lidar com versionamento? - Quais processos de garantia de qualidade adotará?
Documentação e Metadados	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?
Ética e conformidade legal	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? - Como os dados serão licenciados para reutilização? - Existem restrições à reutilização de dados de terceiros? - O compartilhamento de dados será adiado/restrito?
Armazenamento e Backup	Como os dados serão armazenados durante a pesquisa? - 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? Como você vai gerenciar o acesso e a segurança? - 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? Obs: 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 da jurisdição legal em que os dados são mantidos ou da proteção de dados confidenciais.
Seleção e Preservação	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? Qual é o plano de preservação a longo prazo do conjunto de dados? - Onde e em qual repositório ou arquivo os dados serão mantidos? - Quais custos, se houver, serão do repositório de dados ou a taxa de arquivamento selecionada?
Compartilhamento de dados	Por quanto tempo você precisa de uso exclusivo dos dados e por quê? - Será necessário um contrato de compartilhamento de dados (ou equivalente)?
Responsabilidades e Recursos	Quem é responsável pela implementação do PGD e por assegurar que seja revisado? - Quem será responsável por cada atividade? - Como as responsabilidades serão divididas em projetos de pesquisa colaborativa? - Quais recursos você precisará para entregar seu plano? - 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?

Fonte: Adaptado de DCC (2019).

No quadro 11 elencam-se os Princípios FAIR e seus subprincípios para a análise em cada agência selecionada. Para efetivar a pesquisa (quadros 10 e 11), os documentos selecionados foram analisados individualmente, efetuando a verificação a cada critério, princípio e subprincípio se: **Atende totalmente/majoritariamente (AT)**; **Atende parcialmente (AP)**; ou se a informação **Não foi identificada (NI)**. Esta legenda será representada graficamente conforme exposto no quadro 11.

Mediante os critérios de análise definidos se observou a necessidade de adotar o termo Atende totalmente ou ‘majoritariamente’, pois os critérios são essencialmente qualitativos e amplos. A análise semântica dos documentos tem grau de subjetividade uma vez que requer/depende de interpretação do pesquisador que as realiza.

As informações relativas a cada um dos critérios para PGD identificadas nos documentos foram organizadas em uma planilha de Excel, conforme modelo do Apêndice B.

O quadro 11 apresenta o modelo de análise aos Princípios FAIR e seus subprincípios e os dados coletados foram organizados conforme a planilha do Apêndice C.

Quadro 11 - Modelo de análise - Princípios FAIR e subprincípios para os documentos de cada Agência

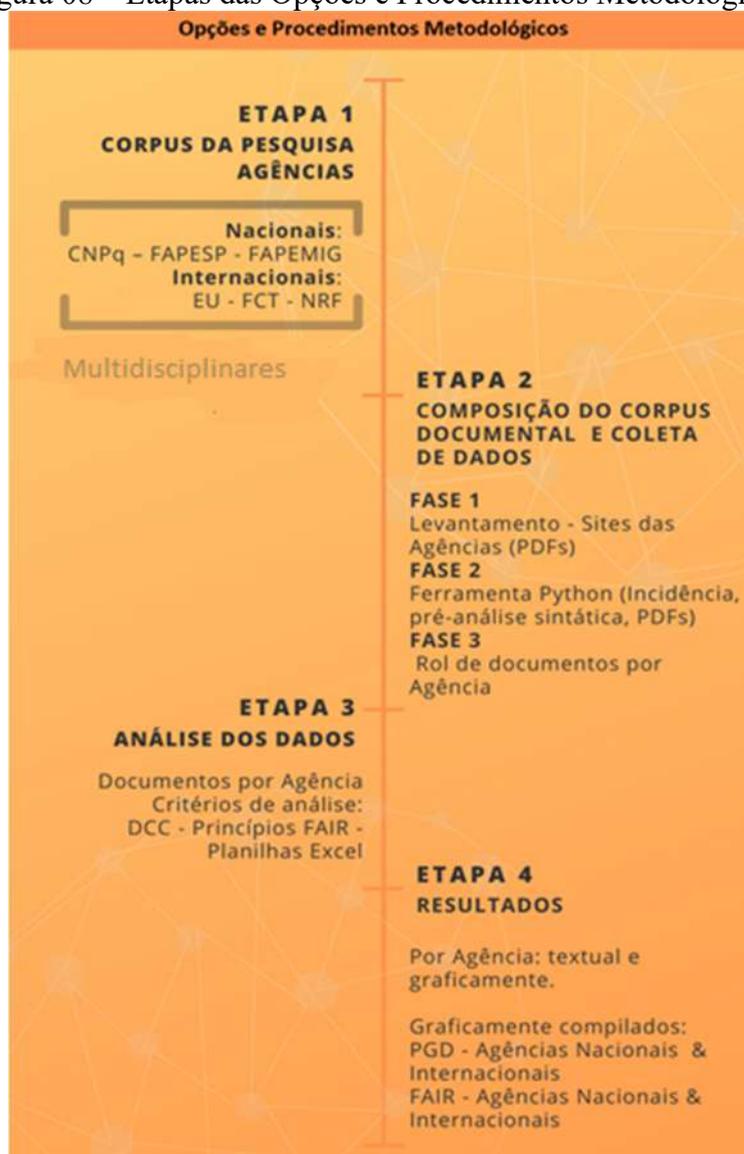
Princípios FAIR		Agência
Princípios	Subprincípios	
Encontráveis 	F1. (Metadados) são atribuídos identificadores globais, persistentes e identificáveis	
	F2. Os dados são descritos com ricos metadados (definidos no R1 abaixo)	
	F3. Os Metadados incluem claramente e explicitamente os identificadores dos dados que os descrevem	
	F4. (Metadados) são registrados ou indexados por intermédio de um recurso pesquisável	
Acessíveis 	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	
Interoperáveis 	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)	
Reutilizáveis 	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.	
LEGENDA	AT - Atende Totalmente / Majoritariamente	
	AP - Atende parcialmente	
	NI - Não identificada (informação)	

Fonte: Dados da pesquisa (2020).

A organização dos dados primários foi executada com uso de *software* planilha eletrônica (Excel) para cada Agência de Fomento com o rol de documentos analisados.

A figura 08 representa sucintamente as opções metodológicas utilizadas na pesquisa.

Figura 08 – Etapas das Opções e Procedimentos Metodológicos.



Fonte: Dados da pesquisa (2020).

As opções e procedimentos metodológicos mencionados aliados à análise documental permitiram ao final do processo, ‘mapear o cenário dos Planos de Gestão de Dados Científicos e dos Princípios FAIR, em Agências de Fomento, objetivo central da pesquisa.

Espera-se fornecer subsídios informacionais para que profissionais desenvolvam PDG e atuem no âmbito da gestão de dados científicos abertos, contribuindo assim com o novo cenário da ciência aberta.

4 ANÁLISE DOS DADOS POR AGÊNCIA DE FOMENTO

Os dados coletados e as análises efetuadas para atender aos objetivos propostos são descritas nesta seção.

A seção aborda as três Agências Nacionais, seus dados e análises e, na sequência as três Agências Internacionais, conforme roteiro estrutural já mencionado.

Ressalta-se que os dados apresentados nas análises resultam do recorte documental proposto nas opções e procedimentos metodológicos e representam o cenário das Agências que atuam com Fomento em áreas multidisciplinares selecionadas.

4.1 AGÊNCIAS NACIONAIS

Esta subseção apresenta os dados coletados e análises das Agências Nacionais, na seguinte ordem: a) Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq; b) Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP; e, c) Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG.

a) Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq

É uma Fundação Pública de âmbito nacional, criada em 1951, com sede em Brasília (DF) subordinada ao Ministério da Ciência, Tecnologia, Inovações e Comunicações, “tem como principais atribuições fomentar a pesquisa científica, tecnológica e de inovação e promover a formação de recursos humanos qualificados para a pesquisa, em todas as áreas do conhecimento”. (CNPq, 2020a).

O CNPq “desempenha papel primordial na formulação e condução das políticas públicas de ciência, tecnologia e inovação”. Analisando o *ranking* das Agências de fomento à pesquisa brasileira, por quantidade de projetos financiados, o CNPq está em primeiro lugar, conforme aponta Monteiro (2019a). “Sua atuação é essencial para o desenvolvimento nacional e o reconhecimento das instituições de pesquisa e dos pesquisadores brasileiros pela comunidade científica internacional”. (CNPq, 2020a).

Tem grande capilaridade por meio de parcerias internacionais com instituições em mais de 20 países, além de acordos e convênios com mais de uma centena de instituições de pesquisas em todo território nacional. Disponibiliza bolsas e auxílios para a pesquisa científica, em todas as áreas do conhecimento, tanto nacional quanto internacionalmente.

“Investe, ainda, em ações de divulgação científica e tecnológica [...], promoção de eventos científicos e à participação de estudantes e pesquisadores [...] em eventos nacionais e internacionais na área de ciência e tecnologia”. (CNPq, 2020b).

A partir da análise aos documentos constatou-se que o CNPq apesar de toda sua abrangência, não apresenta um modelo de PGD específico da Agência, ou então indica modelos de PGD de instituições parceiras para gestão de dados científicos.

O quadro 12 apresenta o rol de documentos e a análise dos critérios com os respectivos dados resultantes.

Quadro 12 – Análise dos critérios em atenção ao Plano de Gestão de Dados – CNPq

PLANO DE GESTÃO DE DADOS (PGD)		CNPq			
		Chamada CNPq/MCTI/CONFAP-FAPs/PELD Nº 21/2020 - Programa de Pesquisa Ecológica de Longa Duração - PELD	PELD-CNPq - Dez Anos do Programa de Pesquisas Ecológicas de Longa Duração no Brasil: Achados, Lições e Perspectivas.	Política de Dados do Programa de Pesquisa em Biodiversidade – PPBio.	RN-009/2016 - Institui a Política de Dados do Programa de Pesquisa Ecológica de Longa Duração (PELD)
DCC	Dados administrativos	●	●	●	●
	Coleta de dados	●	●	●	●
	Documentação e Metadados	●	●	●	●
	Ética e conformidade legal	●	●	●	●
	Armazenamento e Backup	●	●	●	●
	Seleção e Preservação	●	●	●	●
	Compartilhamento de dados	●	●	●	●
	Responsabilidades e Recursos	●	●	●	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●			
	AP - Atende parcialmente	●			
	NI - Não identificada (informação)	●			

Fonte: Dados da pesquisa (2020).

Apesar de não ter modelo de PGD instituído, no documento com a ‘Política de Dados do Programa de Pesquisa Ecológica de Longa Duração (PELD)’ menciona o valor inestimável dos dados coletados para a gestão ambiental. Este documento atende parcialmente a sete critérios mencionados no quadro 12 (exceto responsabilidades e recursos), por ser membro de organizações internacionais para Pesquisa Ecológica de Longa Duração, tais como

International Long Term Ecological Research - ILTER e Global Biodiversity Information Facility - GBIF, que buscam facilitar o acesso, descoberta, uso de dados sobre espécies pelo mundo, com isso, também tem o dever preservá-los para a ampla colaboração científica. (CNPq, 2016).

Na ‘Chamada CNPq/MCTI/CONFAP-FAPs/PELD N° 21/2020’, não se observa a presença de recomendação/exigência quanto ao Plano de Gestão de Dados ou Princípios FAIR pelo CNPq. Somente menciona que deve existir um “Plano de manejo de dados visando disponibilizar em repositórios e acesso público”. A coletânea dos documentos analisados atende parcialmente a todos os critérios do PGD. (CNPq, 2020c).

Na Agência CNPq não foram identificadas ações específicas em atenção aos Princípios FAIR nos documentos analisados.

Quadro 13 – Análise à menção dos Princípios FAIR – CNPq

Princípios FAIR		CNPq
Princípios	Subprincípios	
Encontráveis 	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	●
Acessíveis 	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	●
Interoperáveis 	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)	●
Reutilizáveis 	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.	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

Apesar de não mencionar os Princípios FAIR, existem ações que visam propiciar acesso, descoberta e interoperabilidade dos dados por meio da disponibilização em repositórios temáticos, conforme as áreas de pesquisa. Por exemplo, os projetos regidos pelo documento: Política de Dados do Programa de Pesquisa em Biodiversidade – PPBio, tem o cuidado de recomendar no Repositório de Dados de Estudos Ecológicos o uso de metadados conforme padrão internacional (Ecological Metadata Language - EML), desenvolvido pela Knowledge Network for Biocomplexity²³. “Uma rede internacional que pretende a integração dos dados de vários sítios de coletas, laboratórios e pesquisadores”. Orientam que os metadados sejam organizados contendo as seguintes informações:

título e resumo, palavras-chave, proprietário, contato, partes associadas, projeto de pesquisa, direitos de uso, cobertura geográfica, cobertura temporal, cobertura taxonômica, métodos, informações sobre acesso, arquivos de dados e informações sobre a tabela de atributos. (PPBio, 2020).

No site do repositório constam as orientações para elaboração dos metadados, para tal, recomendam o uso do *software* gratuito (Morpho²⁴), com respectivo link de acesso, manual de utilização e tutoriais para a atividade. Além disso, disponibilizam “conjunto de tabelas de dados PPBio e as instruções de preenchimento” (PPBio, 2020). Essas informações atreladas aos metadados propiciam entendimento e contexto do conjunto de dados, para análises futuras em diferentes vieses ou até mesmo para a replicabilidade científica, caso seja de interesse.

Oferecer orientações claras para curadoria e gestão desde a coleta, constituem boas práticas que facilitam o acesso, a descoberta, e interpretação adequada por qualquer pessoa interessada nos dados, ampliando o potencial de reuso, conforme preconiza a literatura.

b) Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) foi criada em 1960 e começou a funcionar efetivamente em 1962. A FAPESP “apoia a pesquisa e financia a investigação, o intercâmbio e a divulgação da ciência e da tecnologia produzida em São Paulo”, em todas as áreas do conhecimento (FAPESP, 2020a).

A FAPESP passa a apoiar a Inovação Tecnológica em 1995, com o programa Parceria para a Inovação Tecnológica (PITE). Em 1997, criou o programa Inovação Tecnológica em Pequenas Empresas (PIPE). A partir de 1999, essa modalidade de apoio se multiplicou nos

²³ Disponível em: KNB - <http://knb.ecoinformatics.org/index.jsp>.

²⁴ Disponível em: <https://knb.ecoinformatics.org/tools/morpho>.

programas de Políticas Públicas, Biota, Centros de Pesquisa, Inovação e Difusão (CEPIDS), entre outros. Apesar de ser uma Agência de abrangência estadual, tem parcerias com outras instituições nacionais e internacionais, é considerada uma das principais agências de fomento à pesquisa no país. A partir de 1989, conta com repasses de 1% das receitas ordinárias do Estado de SP. Em 2019 apresentou como receita anual o montante de (R\$ 1.453.195.782) para se ter uma noção de seus investimentos no fomento à pesquisa (FAPESP, 2020b).

Conforme divulgado em suas estatísticas, recebeu no ano de 2019 a quantidade de 19.075 propostas para despachos, somadas em todas as categorias de auxílio e bolsas. As propostas recebidas pela Agência são avaliadas por pares – assessores voluntários, entre respeitados pesquisadores ativos no Estado de SP, do país ou do Exterior – de acordo com a natureza e a área do conhecimento em que se insere cada projeto. (FAPESP, 2020a).

Na FAPESP o Plano de Gestão de Dados (PGD) vigora desde 31 de outubro de 2017 e tem caráter mandatório para determinadas modalidades e chamadas. No caso de propostas para projetos temáticos tornou-se obrigatória a apresentação do documento, como anexo.

A modalidade Auxílio à Pesquisa Projeto Temático destina-se a apoiar projetos de pesquisa com objetivos suficientemente ousados, que justifiquem **a duração de até cinco anos** e as condições especialmente favorecidas de apoio e financiamento, incluindo a possibilidade de concessão de bolsas como item orçamentário e solicitações complementares vinculadas ao projeto (FAPESP, 2020c, grifo nosso).

O PGD atende a maioria dos critérios, não especifica os detalhes de como fazer, porém, disponibiliza o *template* do DCC para a elaboração.

O ‘Código de boas práticas científicas’ prevê que os registros de pesquisas devem estar disponíveis a outros pesquisadores, a fim de possibilitar corrigir, replicar ou dar continuidade aos estudos, exceto por restrições éticas ou legais. Mas não explicita como deve ser feita esta disponibilização, isto pode ser planejado por meio do PGD em consonância aos Princípios FAIR.

De acordo com o documento ‘Normas para Apresentação e Seleção de Propostas de Auxílio à Pesquisa Projeto Temático’, a inadequação e/ou insuficiência do PGD já consta como umas das principais deficiências nos pedidos de auxílio à Pesquisa. Ou seja, a elaboração de um bom PGD já é uma condição essencial aos interessados em obter recursos por intermédio desta Agência de Fomento (FAPESP, 2020c). Cumpre salientar que a obrigatoriedade não se aplica a todas as modalidades e chamadas, porém, como é considerada uma boa prática a tendência é que o escopo se amplie gradativamente.

Seguindo, apresentamos o quadro 14 com os dados da análise aos critérios em atenção ao PGD nos documentos da Agência analisados.

Quadro 14 – Análise dos critérios em atenção ao Plano de Gestão de dados – FAPESP

PLANO DE GESTÃO DE DADOS (PGD)		FAPESP			
		Plano de Gestão de Dados – FAPESP	Código de Boas Práticas Científicas	Portaria CTA nº 01/2019 - Política para Acesso Aberto às Publicações Resultantes de Auxílios e Bolsas FAPESP e Anexo I	Normas para Apresentação e Seleção de Propostas de Auxílio à Pesquisa - Projeto Temático
DCC	Dados administrativos	●	◐	◐	●
	Coleta de dados	●	●	●	●
	Documentação e Metadados	●	●	●	●
	Ética e conformidade legal	●	●	●	●
	Armazenamento e Backup	●	●	●	●
	Seleção e Preservação	●	●	●	●
	Compartilhamento de dados	●	◐	●	●
	Responsabilidades e Recursos	●	●	●	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●			
	AP - Atende parcialmente	◐			
	NI - Não identificada (informação)	○			

Fonte: Dados da pesquisa (2020).

A página eletrônica da FAPESP²⁵ elenca *templates* de Planos de Gestão de Dados, sempre ressaltando que estes documentos apresentam diferenças conforme as áreas de domínio. Ao citar o modelo Europeu adotado no Horizonte 2020 informa que este já contempla os Princípios FAIR. O documento está acessível a qualquer pesquisador que necessite ou tenha interesse em seguir as orientações FAIR na gestão de dados científicos.

Porém, na análise efetuada o modelo de PGD adotado na FAPESP ainda não faz menção aos Princípios FAIR, conforme ilustra o quadro 15.

²⁵ <http://www.fapesp.br/gestaodedados/>.

Quadro 15 – Análise à menção dos Princípios FAIR – FAPESP

Princípios FAIR		FAPESP
Princípios	Subprincípios	
Encontráveis 	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	●
Acessíveis 	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	●
Interoperáveis 	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)	●
Reutilizáveis 	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.	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

Dentre os modelos para elaboração do PGD destacados tem-se o DMPTool (<https://dmptool.org/>) e o DMPOnline que é um dos mais utilizados (<https://dmponline.dcc.ac.uk/>). Este último produto do DCC disponibiliza modelos públicos e permite adaptações, tem um rol de perguntas que podem auxiliar e guiar os pesquisadores na construção do PGD específico por domínio. Por vezes, ao iniciar o projeto ainda nem todas as questões são passíveis de respostas, porém à medida que o projeto avança podem ser dimensionadas e respondidas. O PGD é um documento dinâmico que deve ser ajustado sempre que mudanças ocorrerem no projeto, de forma a registrar contexto. Cada nova versão pode conter metadados contextuais para facilitar o entendimento dos dados no futuro.

A ‘Política para Acesso Aberto às Publicações Resultantes de Auxílios e Bolsas FAPESP e seu Anexo I’ de (2019), se relaciona às publicações “de qualquer artigo ou outro

tipo de comunicação científica, que contenha resultados originados de pesquisas apoiadas, parcial ou totalmente, pela FAPESP, em qualquer modalidade de apoio” (FAPESP, 2020d). Não menciona o PGD, mas enfatiza a importância de disponibilizar os resultados das pesquisas em repositórios institucionais de acesso aberto, além de citar que o trabalho de curadoria, gestão e orientação aos pesquisadores deve ser apoiado pelas bibliotecas institucionais. O que aponta para novas oportunidades de atuação e ao mesmo tempo desafia os profissionais pertencentes a esse cenário para a qualificação na temática da ciência aberta.

Denota pioneirismo com ações em Ciência Aberta e apresenta informações sobre a temática no link “Open Science@FAPESP”. As ações são ancoradas

no princípio de que resultados da pesquisa financiada pela FAPESP devem ser tornados públicos dentro do menor prazo possível, respeitadas as tradições de cada domínio científico, resguardados os princípios da ética científica, privacidade e segurança, bem como a proteção da propriedade intelectual (FAPESP, 2020e).

Na área de dados abertos a FAPESP compôs grupo de trabalho com seis universidades públicas paulistas e o ITA para criação de uma rede estadual de repositórios abertos de dados de pesquisa. Como resultado desse trabalho, em dezembro de 2019 foi disponibilizado o ‘Metabuscador de dados de pesquisa’²⁶, que dá acesso a dados científicos da pesquisa produzida pelos membros do grupo, em todos os domínios do conhecimento. Este repositório está atrelado à exigência do Plano de Gestão de Dados da FAPESP e acolhe os dados resultantes das pesquisas realizadas também pelas instituições parceiras (FAPESP, 2020f).

Outra iniciativa importante é a *Scientific Electronic Library Online* (SciELO), uma infraestrutura que apoia a comunicação científica de acesso aberto e sua divulgação de forma global (FAPESP, 2020e).

c) Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG

A Fundação de Amparo à Pesquisa do Estado de Minas Gerais foi criada em agosto de 1985, é a Agência de incentivo “e fomento à pesquisa e à inovação científica e tecnológica do Estado de Minas Gerais”. Atua nas diversas áreas do conhecimento científico, com apoio a “projetos de natureza científica, tecnológica e de inovação, de instituições ou de pesquisadores individuais, que sejam considerados relevantes para o desenvolvimento científico, tecnológico, econômico e social do Estado” (FAPEMIG, 2020a).

Opera nas seguintes linhas de fomento: Pesquisa, Inovação Tecnológica, Capacitação

²⁶ <https://metabuscador.uspdigital.usp.br/>

de Pessoas, Divulgação Científica e Ações Transversais. Para submissão de pospostas é necessário o cadastro no sistema denominado EVEREST²⁷, que concentra as informações dos pesquisadores interessados, e por meio dele é feita a apresentação de proposta às chamadas para projetos de pesquisa e eventos.

Na análise aos documentos não se identificou a solicitação de PGD para envio de propostas. Foram analisados o Manual da FAPEMIG e as últimas chamadas encerradas, conforme quadro 16.

Quadro 16 – Análise dos critérios em atenção ao Plano de Gestão de dados – FAPEMIG

PLANO DE GESTÃO DE DADOS (PGD)		FAPEMIG						
		Manual da FAPEMIG	Chamada FAPEMIG 08/2019 - Programa PIBITI - Empresa FAPEMIG/CNPq/MCTIC/SEMPI	Chamada FAPEMIG 06/2019 - Programa Santos Dumont	Chamada FAPEMIG 01/2020 - Programa Emergencial de Apoio a ações de Enfrentamento da Pandemia causada pelo novo Coronavírus	Chamada Mobility CONFAP Italy 2019	Chamada Mobility CONFAP Italy 2019 - Diretrizes FAPEMIG	
DCC	Dados administrativos	●	●	●	●	●	●	
	Coleta de dados	●	●	●	●	●	●	
	Documentação e Metadados	●	●	●	●	●	●	
	Ética e conformidade legal	●	●	●	●	●	●	
	Armazenamento e Backup	●	●	●	●	●	●	
	Seleção e Preservação	●	●	●	●	●	●	
	Compartilhamento de dados	●	●	●	●	●	●	
	Responsabilidades e Recursos	●	●	●	●	●	●	
LEGENDA	AT - Atende Totalmente / Majoritariamente	●						
	AP - Atende parcialmente	●						
	NI - Não identificada (informação)	●						

Fonte: Dados da pesquisa (2020).

A partir da documentação analisada se observa que a FAPEMIG atende parcialmente aos critérios de: dados administrativos; seleção e preservação e compartilhamento de dados, para os demais critérios não foi identificada informação, conforme quadro 16.

²⁷<http://everest.fapemig.br/>

Dentre os documentos analisados, somente na ‘Chamada FAPEMIG 08/2019 - Programa PIBITI – Empresa FAPEMIG/CNPq/ MCTIC/SEMPI’, aborda de forma genérica que os dados resultantes de pesquisa, devem ser mantidos à disposição de outros pesquisadores, sempre que possível. Recomenda a publicação dos resultados e material suplementar em repositório de acesso aberto, garantido por instituição acadêmica ou outra reconhecida, que proporcione “distribuição irrestrita, interoperabilidade e arquivamento de longa duração”, como forma de preservação (FAPEMIG, 2020b).

Esses fatores podem ser considerados como boas práticas no contexto da ciência aberta.

Outra iniciativa observada é a do projeto denominado “Minas faz ciência”, que objetiva democratizar o conhecimento para a sociedade e também dar visibilidade as pesquisas realizadas em Minas Gerais e no Brasil. Essa divulgação científica ocorre com o cuidado de utilizar uma linguagem acessível ao público não especializado, por meio de vídeos, palestras, *podcasts*, página na internet (inclusive com versão para crianças), além de perfis nas redes Facebook, Twitter e Instagram e da revista ‘Minas Faz Ciência’.

Segundo a FAPEMIG (2020c), “ao disponibilizar informações, espera-se estimular o envolvimento das pessoas nas decisões sobre as políticas e os rumos da produção de CT&I, além de contribuir para o fortalecimento de uma cultura científica no País”. Identifica-se também aqui convergência aos princípios da ciência aberta.

Com relação aos Princípios orientadores FAIR não se identificou nenhuma informação sobre eles ou sua adoção no rol de documentos analisados, conforme exposto no quadro 17.

Quadro 17 – Análise à menção dos Princípios FAIR – FAPEMIG

Princípios FAIR		FAPEMIG
Princípios	Subprincípios	
Encontráveis 	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	●
Acessíveis 	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	●
Interoperáveis 	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)	●
Reutilizáveis 	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.	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

Dentre as boas práticas a Agência apesar de não recomendar um repositório específico para dados, destaca que devem oferecer acesso aberto e ter cadastro no Re3data.

4.2 AGÊNCIAS INTERNACIONAIS

Na sequência apresentam-se os quadros de dados e as análises referentes às Agências Internacionais: a) European Union (EU), b) Fundação para a Ciência e a Tecnologia de Portugal (FCT) e c) National Research Foundation (NRF).

a) European Union - EU

A União Europeia (UE) atualmente é composta “por 27 países europeus” numa coalizão “econômica e política de características únicas”. Foi criada em 1958, e iniciou com o nome de Comunidade Econômica Europeia (CEE), á época, com o objetivo de “reforçar a cooperação econômica entre seis países: Alemanha, Bélgica, França, Itália, Luxemburgo e Países Baixos”. Com o passar do tempo, outros países aderiram e ela passou a atuar para além da área meramente econômica, em outras áreas de interesse do grupo, como: “o clima, o ambiente e a saúde até às relações externas e à segurança, passando pela justiça e pela migração”. Menciona em seus documentos a atuação em cerca de 35 domínios políticos diferentes. No ano de 1993 passou a denominar-se União Europeia (UE). (COMISSÃO EUROPEIA, 2020a, p.7, tradução nossa).

A União Europeia por meio de investimentos e programas busca ampliar a mobilidade entre os cidadãos dos países integrantes, permitindo a estes, residir ou trabalhar em qualquer um dos países membros, além de prezar pelo avanço científico em colaboração e também respeitar as culturas regionais. São respeitados todos os idiomas e a EU atua com 24 línguas oficiais, dentre elas, o português. “Todos os países da UE devem tratar os cidadãos da UE exatamente como os seus próprios cidadãos no que toca ao emprego, à segurança social e à fiscoalidade”. (UNIÃO EUROPEIA, 2020a, p.7).

Dentre os projetos da União Europeia, destaca-se o Horizonte 2020, como o maior programa em pesquisa e inovação, com investimentos de 77 milhões de euros financiados no decorrer de 2014-2020, com objetivos primordiais de:

- a) incentivar a excelência científica, através do Conselho Europeu de Investigação, e a formação e progressão na carreira no âmbito das ações Marie Skłodowska-Curie;
- b) promover a liderança industrial em áreas como as nanotecnologias, a biotecnologia e as tecnologias da informação e da comunicação, e através do apoio às empresas e aos empreendedores; e
- c) superar os nossos maiores desafios sociais, incluindo a saúde, os transportes, a energia, a ação climática e a proteção da liberdade e da segurança. (UNIÃO EUROPEIA, 2020a, p.23).

“A Comissão Europeia visa igualmente desenvolver políticas que incentivem uma investigação de excelência e impulsionem a inovação. **As novas políticas e ações** podem ser agrupadas em **três temas principais: inovação aberta, ciência aberta e abertura ao mundo**”. (UNIÃO EUROPEIA, 2020a, p.23, grifo nosso). Os investimentos, a definição de políticas estratégicas e iniciativas que a UE está realizando demonstram o grau de importância

que atribui à ciência, e na contemporaneidade à ciência aberta. Essa gama de estruturas científicas alçou a UE a produtora central de conhecimento, com “cerca de um terço da produção científica e tecnológica” mundial. (UNIÃO EUROPEIA, 2020a, p.23).

No intuito de investigar na Agência de Fomento EU quais as exigências e/ou recomendações para elaboração do plano de gestão de dados e se menciona os princípios FAIR, foram examinados os documentos conforme quadro 18.

Quadro 18 – Análise dos critérios em atenção ao Plano de Gestão de dados - (EU)

PLANO DE GESTÃO DE DADOS (PGD)		EU				
		Guidelines on FAIR Data Management in Horizon 2020	Guidelines to the Rules on Open Access to Scientific Publications And Open Access to Research Data in Horizon 2020	Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020	EOSC Declaration	Open Research Data and Data Management Plans: Information for ERC Grantees
DCC	Dados administrativos	●	●	●	●	●
	Coleta de dados	●	○	●	○	●
	Documentação e Metadados	●	●	●	●	●
	Ética e conformidade legal	●	●	●	●	●
	Armazenamento e Backup	●	○	●	○	●
	Seleção e Preservação	●	●	●	●	●
	Compartilhamento de dados	●	●	●	●	●
	Responsabilidades e Recursos	●	●	●	●	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●				
	AP - Atende parcialmente	●				
	NI - Não identificada (informação)	○				

Fonte: Dados da pesquisa (2020).

Pode-se aferir por meio da coletânea de documentos analisados, que a EU atende Totalmente/Majoritariamente a todos os critérios do PGD analisados.

Ao tratar sobre modelos para PGD esclarece que não existe um modelo único, que é um documento dinâmico e varia conforme as disciplinas. Porém, recomendam usar o modelo do European Research Council (ERC) disponível no Portal do Participante do Horizonte

2020²⁸ e também o modelo fornecido pelo Digital Curation Center (DCC), (que serviu de base para as análises nesta pesquisa), via ferramenta DMPonline²⁹, ressaltando que ambos atendem aos critérios do ERC e de várias outras organizações de fomento, com a estrutura adequada para atender aos Princípios FAIR.

O programa Horizonte 2020 comporta um projeto denominado Open Research Data Pilot (ORD), que iniciou com foco em algumas áreas temáticas e agora se aplica a todas que visam melhorar e maximizar acesso e reutilização dos dados de pesquisa gerados no Horizonte 2020. Todos os que participam do Projeto ORD devem enviar um PGD em até seis meses após início da concessão, atualmente podem sair a qualquer tempo, mas são incentivados a manter essa boa prática.

A tendência é que se torne um requisito obrigatório conforme avança a disponibilização de plataformas e infraestruturas tecnológicas que facilitem essa gestão integrada. Outro aspecto a destacar é que para algumas Agências em determinadas áreas já exige o PGD como requisito obrigatório, como condicional a aprovação de projetos.

O documento ‘Guidelines on FAIR Data Management in Horizon 2020’ detalha como adotar os Princípios FAIR na gestão de dados, inclusive apresenta um rol de perguntas norteadoras, para que junto ao modelo de PGD se potencialize o valor dos dados. Sempre é bom esclarecer que os princípios FAIR são elementos agregadores de qualidade para a disponibilização e descoberta dos dados, com a finalidade última de possibilitar o reuso. (EUROPEAN COMMISSION, 2016).

Contudo, não exigem ferramentas específicas ou tecnologias exclusivas, ao contrário incentivam o uso de códigos abertos para que as práticas se alarguem e possam melhorar o contexto e significação dos dados para usos futuros. Este documento tem caráter recomendatório para as pesquisas atreladas ao programa Horizonte 2020, mas incentiva a participar do projeto Open Research Data Pilot (ORD), regido pelo princípio “o mais aberto possível, tão fechado quanto necessário” como diretriz de ação para efetivar a gestão segura de dados e atender as melhores práticas de pesquisa. (EUROPEAN COMMISSION, 2016, tradução nossa).

O documento com as ‘Guidelines to the Rules on Open Access to Scientific Publications And Open Access to Research Data in Horizon 2020’, detalha o projeto (ORD), menciona que nem todos os dados podem ser abertos e elenca uma série de motivos nos quais

²⁸ http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-erctpl-oa-data-mgt-plan_en.odt

²⁹ <https://dmponline.dcc.ac.uk>

os participantes podem encerrar a participação no projeto piloto (ou não necessitam disponibilizar os dados). (EUROPEAN COMMISSION, 2020a).

O quadro 19 aponta a menção aos Princípios FAIR e subprincípios a partir da análise aos documentos da Agência - União Europeia.

Quadro 19 – Análise à menção dos Princípios FAIR – EU

Princípios FAIR		EU
Princípios	Subprincípios	
Encontráveis 	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	●
Acessíveis 	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	●
Interoperáveis 	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)	●
Reutilizáveis 	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.	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

A ‘EOSC Declaration’ (2017) esclarece que a implementação da EOSC³⁰ se caracteriza “como um processo e não um projeto, devido a sua natureza ser iterativo e baseado em aprendizado constante e alinhamento mútuo”. (EUROPEAN COMMISSION, 2017, tradução nossa). Para as definições de políticas e diretrizes as instituições

³⁰ European Open Science Cloud.

especializadas em dados (Research Data Alliance, CODATA, DDI Alliance e outras ativas nas comunidades de pesquisa) devem ser consultadas e discutir sobre a implementação prática dos princípios de dados FAIR a nível europeu e global. Os modelos de governança oriundos no âmbito da EU devem “basear-se na representatividade, proporcionalidade, responsabilidade, inclusão e transparência.” (EUROPEAN COMMISSION, 2017, tradução nossa).

“Os beneficiários devem garantir acesso aberto à versão depositada de suas publicações através do repositório escolhido. O acesso aberto deve ser fornecido o mais rápido possível e, em qualquer caso, não mais de seis meses após a data de publicação oficial” No caso de ciências sociais e humanas se permite um tempo de até doze meses. (EUROPEAN RESEARCH COUNCIL, 2017).

O Conselho Europeu de Investigação incentiva os pesquisadores financiados pela organização a usar repositórios específicos da disciplina para suas publicações. Ex: Ciências da Vida o Europa PubMed Central; Ciências Físicas e Engenharia o repositório arXiv. Para Ciências Sociais e Humanas não fazem indicação no momento. Caso não exista um repositório específico na disciplina, as publicações devem ser disponibilizadas em repositórios institucionais ou centralizados, como o Zenodo, por exemplo. (EUROPEAN RESEARCH COUNCIL, 2017, tradução nossa).

Há um documento de Proposta de Parceria Europeia com European Open Science Cloud (EOSC)³¹, datada de maio de 2020, que prevê a partir de sua vigência o PGD tornar-se-á obrigatório para todos os projetos. Já quanto aos princípios FAIR a adesão será incentivada/recomendada, bem como, o uso de repositórios e infraestruturas conectadas ao EOSC. A documentação sobre a temática na EU está em constante atualização e revisão para aprimorar e acompanhar as inovações necessárias às plataformas, as novas legislações ou diretrizes na busca por avançar a cultura da ciência aberta. Neste caso especificamente, os dados científicos serão abertos por padrão, devendo sua restrição de acesso ser justificada.

A União Europeia sintonizada com o cenário mundial atual lançou em 24 de abril de 2020, um documento com ‘Diretrizes para acesso aberto a publicações, dados e outros resultados de pesquisa’ relacionados ao Projeto Horizonte 2020 que trabalham com a doença coronavírus 2019 (COVID-19) e tópicos relacionados (diagnósticos, vacinas e terapêuticas para este novo vírus). São detalhadas informações sobre PGD e os Princípios FAIR com

31

Link: https://ec.europa.eu/info/sites/info/files/research_and_innovation/funding/documents/ec_rtd_he_partnership-open-science-cloud-eosc.pdf.

vistas a agilizar a cooperação e interoperabilidade de dados científicos para respostas a emergência de saúde pública internacional. Ressalta que os dados devem ter saída FAIR mesmo quando o acesso é restrito. (EUROPEAN COMMISSION, 2020b).

b) Fundação para a Ciência e a Tecnologia de Portugal - FCT

A Fundação para a Ciência e a Tecnologia de Portugal (FCT) é a agência pública do país, foi fundada em 1997, com sede em Lisboa, e tem por objetivo a investigação em ciência, tecnologia e inovação.

Sua missão é promover o avanço do conhecimento científico e tecnológico em Portugal para “atingir os mais altos padrões internacionais de qualidade e competitividade, em todos os domínios científicos e tecnológicos, e incentivar a sua divulgação contribuindo para a sociedade e seu avanço econômico” (FCT, 2020a, tradução nossa).

A FCT financia diferentes tipos de atividades por meio de diversos instrumentos: destacam-se: projetos de investigação, formação avançada, emprego científico, unidades de investigação e cooperação internacional.

No ano de 2019 investiu aproximadamente 83M€ somente na atividade de projetos de investigação.

As demais modalidades receberam investimentos nos seguintes montantes: as bolsas de formação avançada investimentos de 98M€; os contratos de emprego científico no valor total executado de 103M€. “O financiamento a Unidades de I&D e Laboratórios Associados foi de 122M€”, [...] e “programas de investigação bilaterais e multilaterais, e as contribuições para organizações científicas internacionais”, foram investidos 60M€ em 2019 (FCT, 2020b).

Em decorrência das análises efetuadas compõem-se as informações do quadro 20.

Quadro 20 – Análise dos critérios em atenção ao Plano de Gestão de Dados – FCT

PLANO DE GESTÃO DE DADOS (PGD)		FCT					
		Política sobre Acesso Aberto a Publicações Científicas resultantes de Projetos de I&D Financiados pela FCT	Política sobre a Disponibilização de Dados e outros Resultados de Projetos de I&D Financiados Pela FCT	Acesso Aberto - Modelos, Políticas e Custos de Acesso	Política Nacional de Ciência Aberta	Segundo Relatório da Comissão Executiva do Grupo de Trabalho para a Política Nacional de Ciência Aberta	
DCC	Dados administrativos	●	●	●	●	●	
	Coleta de dados	●	●	●	●	●	
	Documentação e Metadados	●	●	●	●	●	
	Ética e conformidade legal	●	●	●	●	●	
	Armazenamento e Backup	●	●	●	●	●	
	Seleção e Preservação	●	●	●	●	●	
	Compartilhamento de dados	●	●	●	●	●	
	Responsabilidades e Recursos	●	●	●	●	●	
LEGENDA	AT - Atende Totalmente / Majoritariamente	●					
	AP - Atende parcialmente	●					
	NI - Não identificada (informação)	●					

Fonte: Dados da pesquisa (2020).

Ao analisar a Agência Portuguesa observam-se iniciativas robustas para o acesso aberto, articuladas (grupos de trabalho) e documentadas através de recomendações para os diferentes atores envolvidos. No caso de dados abertos, “os destinatários alvo são: a) Agentes políticos; b) Financiadores públicos; c) Instituições que realizam investigação” (MCTES, 2016, p.10-11).

Políticas coordenadas dessa forma podem se tornar mais efetivas, pois criam um ecossistema de maior visibilidade e conseqüentemente, norteiam projetos e pesquisas futuras, incentivam o reuso de dados e a colaboração, resultando em uma ciência mais ágil e congruente com os respectivos contextos sociais vigentes. Em relação ao PGD no ‘Segundo Relatório da Comissão Executiva do Grupo de Trabalho para a Política Nacional de Ciência Aberta’, não foram identificadas informações no documento aos critérios sobre coleta de dados, ética e conformidade legal e armazenamento e backup; aos demais critérios atende parcialmente. (MCTES, 2016).

Na ‘Política sobre Acesso Aberto a Publicações Científicas resultantes de Projetos de I&D’, enfatiza a obrigatoriedade das publicações em fazer referência aos dados subjacentes, tem caráter mandatório e atende parcialmente somente ao critério de dados administrativos para o PGD, aos demais critérios não foram identificadas informações.

A Política sobre a Disponibilização de Dados e outros Resultados de Projetos de I&D preconiza que “a FCT assegurará a evolução futura desta política, tentando convergir com as melhores práticas internacionais, designadamente com as iniciativas deste domínio que vierem a ser estabelecidas no âmbito da União Europeia.” (FCT, 2014, p. 2). Esta política sobre Dados tem caráter recomendatório e foi adotada em 05 de maio de 2014. Atende de forma completa a seis critérios do PGD, somente não foram identificadas informações sobre dois: seleção e preservação e responsabilidade e recursos.

Já no documento ‘Acesso Aberto - Modelos, Políticas e Custos de Acesso’ não foram identificadas informações a nenhum dos critérios do PGD. (FCT, 2017).

A Política Nacional de Ciência Aberta (PNCA) é um portal com informações sobre a temática, conceitos, glossário, cronologia, legislações, recursos, mapas de repositórios científicos de Acesso Aberto (AA), iniciativas em ciência aberta, documentos relacionados, etc. Tem abrangência nacional o que fortalece o movimento da ciência aberta no país, pois pode amplificar a visibilidade, servir de referência/marco a várias organizações, além de permitir alinhar amplas cooperações, seja entre pesquisadores ou instituições públicas ou privadas de investigação. Com relação ao PGD a política contempla de forma parcial os critérios: dados administrativos e coleta de dados. (CIÊNCIA ABERTA, 2020).

O ‘Segundo Relatório da Comissão Executiva do Grupo de Trabalho para a Política Nacional de Ciência Aberta’ (MCTES, 2016, p.16), pondera que devem se adequar “regulamentos de Mestrados e Doutoramentos das Instituições de Ensino Superior [...] encorajando o AA e a disponibilização aberta dos dados de investigação eventualmente resultantes, observando os Princípios FAIR (*Findable, Accessible, Interoperable, Reusable*).”

A partir da análise a este documento, não se pode aferir sobre o atendimento a cada subprincípio FAIR, pois somente são mencionados os quatro princípios FAIR de forma ampla e incentivada sua adoção.

Quadro 21 – Análise à menção dos Princípios FAIR – FCT

Princípios FAIR		FCT
Princípios	Subprincípios	
Encontráveis 	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	●
Acessíveis 	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	●
Interoperáveis 	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)	●
Reutilizáveis 	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.	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

Para além dos documentos mencionados o Brasil firmou no ano de 2018, o memorando de entendimento entre Ministério da Ciência, Tecnologia e Ensino Superior da República Portuguesa (MCTES) e o Ministério da Ciência, Inovações e Comunicações da República Federativa do Brasil (MCTIC) para a cooperação no âmbito de ciência aberta.

Este memorando em sua cláusula 1^a, respectivamente as alíneas: d) e e) referem-se especificamente aos dados científicos.

d) explicitação e inclusão da temática da gestão e partilha dos dados científicos, definindo eventuais ações piloto (no domínio da partilha de infraestruturas, boas práticas, políticas e procedimentos, formação, etc.) envolvendo comunidades e/ou instituições científicas de Portugal e do Brasil;

e) Promover atividades de sensibilização, disseminação e formação de investigadores e de pessoal de suporte nas instituições de investigação de ambos os países, promovendo e apoiando adoção das melhores práticas no domínio da ciência aberta e do acesso aberto aos resultados da investigação (publicações e dados). (FCT, 2018, p. 3).

Esta é uma iniciativa governamental que pode ser considerada boa prática, pois estimula a cooperação entre pesquisadores no âmbito internacional, envolve Agências de Fomento e grupos trabalho na busca e desenvolvimento de infraestruturas para a gestão de dados científicos. Além disso, promove ações para sensibilizar, elucidar conceitos, ampliar conhecimentos, disseminar, inovar e adotar as melhores práticas em acesso aberto para a divulgação dos resultados de pesquisas científicas (publicações e dados).

O Brasil como signatário pode permutar experiências para definir ou harmonizar orientações e avançar em políticas e estratégias para a ciência aberta juntamente com Portugal. As equipes portuguesas de trabalho têm seus representantes designados pela FCT e no Brasil por meio do IBICT.

Outra ação a destacar em termos de cooperação entre os dois países, é o “portal português RCAAP e o portal brasileiro ou OasisBR que estão agregados desde 2010 e contém mais de 2,2 milhões de registros”. (FCT, 2018, p.2).

No Memorando de entendimento os países reconhecem ainda:

que o acesso à ciência e ao conhecimento, que a ciência aberta visa promover, é indispensável a uma sociedade mais informada e mais consciente do mundo que habita, contribuindo para tornar mais humana, mais justa e mais democrática e onde o bem-estar seja partilhado por todos. (FCT, 2018, p. 2).

A FCT também apoia e reconhece a relevância dos objetivos do Plan S, fomentado pela Science Europe, especialmente no quesito sobre disponibilização imediata dos resultados das pesquisas provenientes de financiamento público.

Portugal se destaca como um dos países pioneiros em adotar iniciativas consoantes ao movimento da ciência aberta e dados abertos, atualmente com políticas e estruturas nacionais sólidas e alinhadas a grandes projetos internacionais (ex. PNCA, RCAAP, Projeto H2020).

Empenha esforços nesse contexto por entender que “não há [Ciência Aberta] sem compreensão pública da ciência” (MCTES, 2016, p.19).

c) National Research Foundation - NRF

A National Research Foundation (NRF) foi fundada em 1998, é a Agência governamental da África do Sul, com sede em Pretoria. Tem como foco “apoiar e promover a pesquisa por meio de financiamento, desenvolvimento de recursos humanos e fornecimento das instalações de pesquisa necessárias para facilitar a criação de conhecimento, inovação e desenvolvimento em todos os campos da ciência e tecnologia, incluindo os saberes

indígenas.” (NRF, 2020, tradução nossa).

A NRF participa do projeto piloto para a Plataforma Africana de Ciência Aberta, lançado em dezembro de 2018 com prospecção de ações anuais em períodos de curto e médio prazo (1 a 5 anos). Esta plataforma foi financiada, na fase inicial pelo governo com apoio de organismos internacionais interessados no desenvolvimento de capacidades na África (NRF, 2018). A Plataforma tem como missão “colocar os cientistas africanos na vanguarda da ciência contemporânea com uso intensivo de dados como um recurso fundamental para uma sociedade moderna”. (NRF, 2018, p. 4).

São identificadas como temáticas relevantes para a comunidade de pesquisa africana: a) Prevenção e controle de doenças, b) Resiliência climática (risco de desastre); c) Proteção ambiental (biosfera, hidrosfera), d) Segurança alimentar e nutricional, e) Cidades resilientes inteligentes, f) Metas de sustentabilidade, g) Produção de conhecimento aprimorada, estas são questões prioritárias e que fomentam programas e desenvolvimento de pesquisas pela Agência Africana.

Conforme o documento: ‘The African Open Science Platform: The Future Of Science and Science For The Future’

devidos passar de um mundo em que manter o acesso exclusivo aos dados que criamos foi uma chave para a descoberta científica e sua aplicação, para um onde o acesso aos dados é a prioridade. Pode-se acumular pouco ou acessar muito. O primeiro é uma ilha - o último é um oceano. (NRF, 2018, p. 6, tradução nossa):

Apesar de não mencionar especificamente o PGD, os documentos analisados fornecem atendimento parcial aos critérios que o compõem, exceto ao critério armazenamento e backup, conforme quadro 22.

Quadro 22 – Análise dos critérios em atenção ao Plano de Gestão de Dados – NRF

PLANO DE GESTÃO DE DADOS (PGD)		NRF					
		The African Open Science Platform: The Future Of Science and Science For The Future (2016)	National Research Foundation Strategy - 2020 (2016)	General Application Guide 2021- Grants Management and Systems Administration (GMSA). (Version 5, de 21 April 2020)	Statement on Open Access to Research Publications from the National Research Foundation (NRF) - Funded Research (2015)	Annual Performance Plan 2019/20 - 2021/22 (2019)	
DCC	Dados administrativos	●	●	●	●	●	
	Coleta de dados	●	●	●	●	●	
	Documentação e Metadados	●	●	●	●	●	
	Ética e conformidade legal	●	●	●	●	●	
	Armazenamento e Backup	●	●	●	●	●	
	Seleção e Preservação	●	●	●	●	●	
	Compartilhamento de dados	●	●	●	●	●	
	Responsabilidades e Recursos	●	●	●	●	●	
LEGENDA	AT - Atende Totalmente / Majoritariamente	●					
	AP - Atende parcialmente	●					
	NI - Não identificada (informação)	●					

Fonte: Dados da pesquisa (2020).

O projeto ‘The African Open Science Platform’ está em fase de operacionalização (com prognóstico de lançamento da plataforma completa em 2020), mas prevê e enfatiza fortemente a gestão de dados, por meio de plataformas de pesquisa temáticas e colaborativas, somente não se identificou informação sobre armazenamento e *backup* dos dados no processo de pesquisa, aos demais critérios atende parcialmente.

A ‘Declaração sobre Acesso Aberto a Publicações de Pesquisa da NRF’, preconiza que os artigos de pesquisa gerados e os dados que complementam essa publicação, financiados total ou parcialmente pela NRF, devem ser depositados em repositório institucional de acesso aberto, com período de embargo não superior a um ano. Esta prática é mandatória desde 2015, inclusive com a atribuição de DOI o que é uma recomendação dos Princípios FAIR. (NRF, 2015).

Estes princípios também são mencionados no ‘Plano de Desempenho Anual 20/2021/22’, ao tratar sobre a infraestrutura para pesquisas em grande escala, com grande volume de dados. Afirma ser “essencial gerenciar os protocolos de hospedagem, segurança,

acesso e preservação de tais dados para fornecer aos pesquisadores dados encontráveis, acessíveis, interoperáveis e reutilizáveis (FAIR), idealmente em apoio à ciência aberta”, plano este, ainda em processo de construção de infraestruturas/implementação (NRF, 2019, p. 69, tradução nossa). Trata também sobre o uso das TICs avançadas (que denomina *e-Research*) para auxiliar os processos de pesquisas, tornando-os mais colaborativos, eficientes e eficazes, nas atividades de coleta, compartilhamento e preservação, dentre outras.

Na análise da documentação, somente o ‘Plano de Desempenho Anual 20/2021/22’ mencionou os princípios orientadores FAIR, de forma ampla, não indica os subprincípios ou maiores informações sobre sua adesão. Não permite aferir especificamente quais subprincípios seriam atendidos, pois afirma que os dados devem ser FAIR e ainda se apresenta como prospecção para execução entre 2020 e 2022.

Quadro 23 – Análise à menção dos Princípios FAIR - NRF

Princípios FAIR		NRF
Princípios	Subprincípios	
Encontráveis 	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	●
Acessíveis 	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	●
Interoperáveis 	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)	●
Reutilizáveis 	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.	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

Concluída a seção Análise dos Dados por Agência de Fomento, serão apresentados, na seção 5, os Resultados Consolidados, relativos aos quadros e análises descritos nesta seção.

5 RESULTADOS CONSOLIDADOS

Das seis Agências analisadas três são brasileiras, sendo duas de abrangência estadual (FAPESP, FAPEMIG) e uma nacional (CNPq). E das três internacionais, duas com abrangência nacional (FCT, NRF) e uma representa um bloco de países que adotaram padrões para acelerar os avanços científicos e melhorar a mobilidade científica entre seus membros, a União Europeia (EU). Esta última realiza investimentos fortes em projetos e iniciativas conjuntas entre seus membros e conta com importante apoio político em prol da ciência aberta (por exemplo: programa Horizonte 2020, EOSC, *Plan S*).

A título de conhecer onde estão situadas pelo mundo as seis Agências analisadas, segue um mapa com o posicionamento geográfico de cada uma, sendo o ponto marcado correspondente ao local de sua cidade sede.

Figura 09 – Posicionamento geográfico das Agências estudadas



Fonte: Dados da pesquisa (2020).

As Agências Nacionais pertencem às regiões Sudeste (FAPESP, FAPEMIG) e Centro-Oeste (CNPq) e as internacionais aos continentes Africano (NRF) e Europeu (EU, FCT).

5.1 AGÊNCIAS NACIONAIS

Nesta subseção são apresentados os resultados consolidados dos dados e informações tratados na seção 4 Análise dos Dados por Agência de Fomento.

a) Quadros de resultados das análises - Agências Nacionais

Os quadros representativos iniciam pelas Agências Nacionais e posteriormente, as Internacionais. Ressalta-se que estes resultados correspondem ao recorte metodológico explicitado e ao rol de documentos analisados na perspectiva dessa pesquisa.

O quadro 24 retrata o resultado da análise correspondente a Agência de Fomento – CNPq apresentada de forma pormenorizada no quadro 12.

Quadro 24 – Resultado da análise aos critérios do PGD – CNPq

PLANO DE GESTÃO DE DADOS (PGD)		
	Critérios	CNPq
DCC	Dados administrativos	
	Coleta de dados	
	Documentação e Metadados	
	Ética e conformidade legal	
	Armazenamento e Backup	
	Seleção e Preservação	
	Compartilhamento de dados	
	Responsabilidades e Recursos	
	LEGENDA	AT - Atende Totalmente / Majoritariamente
AP - Atende parcialmente		
NI - Não identificada (informação)		

Fonte: Dados da pesquisa (2020).

Foram analisados quatro (04) documentos do CNPq. A Agência não apresenta um modelo de PGD instituído, porém, dos oito critérios estabelecidos na análise a Agência Atende parcialmente, a todos. Isso aponta que já existem ações para a gestão de dados científicos em andamento.

Não foram identificadas informações referentes à menção dos Princípios FAIR.

O quadro 25 representa a consolidação da análise referente à Agência de Fomento – FAPESP resultante das informações pormenorizadas que constam no quadro 14.

Quadro 25 – Resultado da análise aos critérios do PGD – FAPESP

PLANO DE GESTÃO DE DADOS (PGD)		
	Critérios	FAPESP
DCC	Dados administrativos	●
	Coleta de dados	●
	Documentação e Metadados	●
	Ética e conformidade legal	●
	Armazenamento e Backup	●
	Seleção e Preservação	●
	Compartilhamento de dados	●
	Responsabilidades e Recursos	●
LEGENDA	AT - Atende Totalmente / Majoritariamente	●
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

No Brasil, a Agência FAPESP adota um modelo de PGD desde 2017 e pode ser considerada pioneira no país na adesão a esta boa prática de pesquisa.

Foram analisados quatro (04) documentos, dentre eles o PGD. A partir do resultado se observa que a FAPESP Atende totalmente/majoritariamente a todos os critérios analisados. Das Agências Nacionais estudadas é a única que tem o PGD como elemento obrigatório para determinados projetos de pesquisa (no caso as pesquisas de longa duração). Também sinaliza para o trabalho colaborativo com as bibliotecas das instituições para apoio nas atividades de curadoria e gestão de dados científicos.

Em relação aos Princípios FAIR não foram identificadas informações sobre a adesão pela FAPESP, porém, menciona *templates* do projeto Horizonte 2020 com modelos de PGD baseados em FAIR. Apesar de ser uma Agência de abrangência estadual tem importantes iniciativas no contexto da ciência aberta e dados abertos.

O quadro 26 corresponde ao resultado das análises pormenorizadas realizadas aos documentos da FAPEMIG explicitados no quadro 16 em atenção os critérios do PGD.

Quadro 26 – Resultado da análise aos critérios do PGD - FAPEMIG

PLANO DE GESTÃO DE DADOS (PGD)		
	Critérios	FAPEMIG
DCC	Dados administrativos	
	Coleta de dados	
	Documentação e Metadados	
	Ética e conformidade legal	
	Armazenamento e Backup	
	Seleção e Preservação	
	Compartilhamento de dados	
	Responsabilidades e Recursos	
	LEGENDA	AT - Atende Totalmente / Majoritariamente
AP - Atende parcialmente		
NI - Não identificada (informação)		

Fonte: Dados da pesquisa (2020).

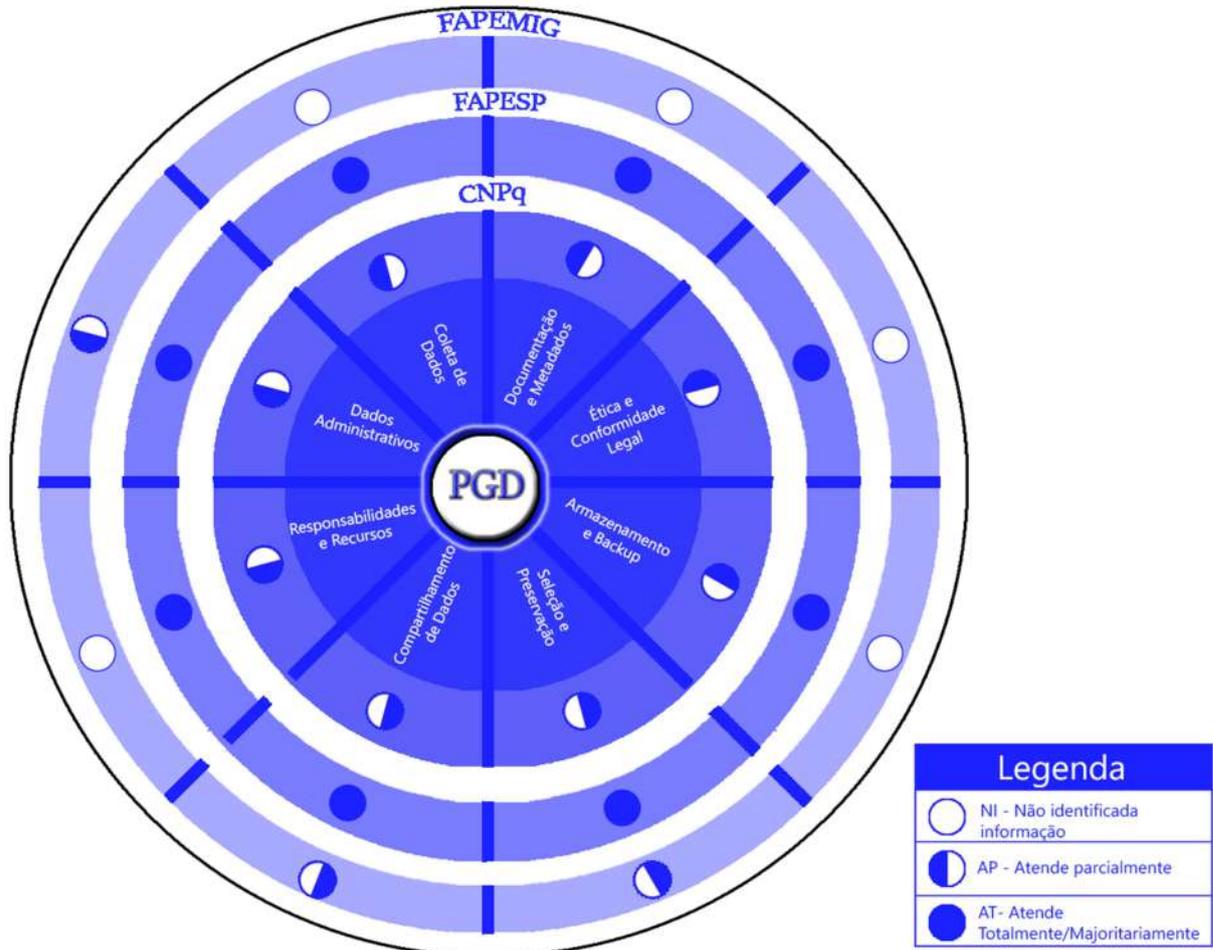
Dentre as Agências Nacionais estudadas a FAPEMIG corresponde a que tem de menor incidência de atenção aos critérios do PGD, com atendimento parcial a somente três, de acordo com o rol de documentos analisados.

Em relação aos Princípios FAIR não foram identificadas informações referentes à sua menção em nenhum dos 06 documentos analisados.

b) Representação gráfica consolidada das três Agências Nacionais - Plano de Gestão de Dados

A figura 10 sintetiza o cenário em atenção aos critérios relativos ao PGD nas três Agências (CNPq, FAPESP, FAPEMIG) a partir da análise aos documentos conforme os quadros 24, 25, 26.

Figura 10 – Representação dos resultados nas três Agências Nacionais – PGD



Fonte: Dados da pesquisa (2020).

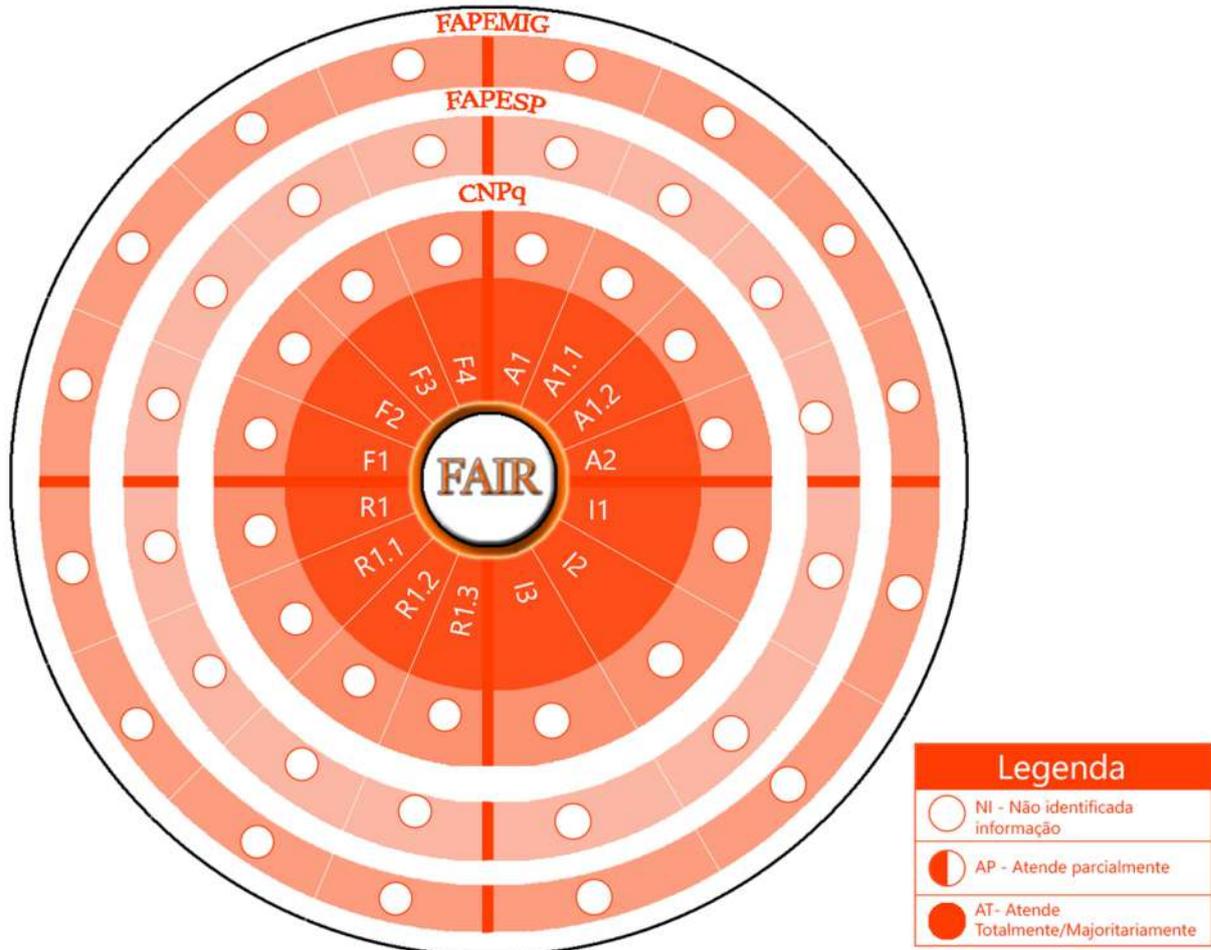
Dentre as três Agências Nacionais investigadas a única que tem um PGD instituído é a FAPESP. As demais Agências tratam da gestão de dados em documentos ou diretrizes relacionados às publicações de resultados de pesquisas, ou chamadas para projetos, em que os dados científicos constam como parte integrante.

Na sequência, a figura 11 condensa a representação dos resultados baseados nas análises aos documentos citados em relação à menção dos Princípios FAIR em cada uma das três Agências Nacionais pesquisadas (CNPq, FAPESP, FAPEMIG). Observa-se que nenhuma ainda faz menção a adoção dos Princípios FAIR. Somente a FAPESP disponibiliza um link para que os interessados possam acessar o material sobre FAIR do programa Horizonte 2020, vinculado à União Europeia.

A figura 11 se divide em quatro quadrantes principais, onde cada um deles representa um dos princípios conforme a letra inicial do código exposto e seus respectivos subprincípios.

Esta figura consolida os dados dos quadros 13, 15 e 17 apresentados na seção 4 Análise dos Dados por Agência de Fomento. Os referidos quadros detalham as informações de cada código FAIR citado na raia central.

Figura 11 – Representação dos resultados nas três Agências Nacionais - Princípios FAIR



Fonte: Dados da pesquisa (2020).

Observando a imagem partindo do centro para o exterior, temos o CNPq, Agência de abrangência nacional, seguido pela FAPESP que apesar de ser Fundação Estadual atua com parcerias em diferentes regiões, seguida da FAPEMIG também de âmbito estadual e seus respectivos resultados a cada subprincípio.

No Brasil, apesar de não ser escopo desta pesquisa, destaca-se a Agência de Fomento Fiocruz com ações e estudos para implementação dos Princípios FAIR na gestão de seus dados científicos abertos.

5.2 AGÊNCIAS INTERNACIONAIS

Esta subseção apresenta os quadros consolidados referentes aos resultados correspondentes às análises nas Agências Internacionais: European Union (EU), Fundação para a Ciência e a Tecnologia de Portugal (FCT) e National Research Foundation (NRF).

a) Quadros de resultados consolidados das análises nas Agências Internacionais

O quadro 27 corresponde ao resultado consolidado da análise pormenorizada apresentada no quadro 18, subseção 4.2 Agências Internacionais. Deriva da análise ao rol de 06 documentos da Agência e atende totalmente/majoritariamente aos oito critérios investigados em atenção ao PGD.

Quadro 27 – Resultado da análise aos critérios do PGD – EU

PLANO DE GESTÃO DE DADOS (PGD)		
	Critérios	EU
DCC	Dados administrativos	●
	Coleta de dados	●
	Documentação e Metadados	●
	Ética e conformidade legal	●
	Armazenamento e Backup	●
	Seleção e Preservação	●
	Compartilhamento de dados	●
	Responsabilidades e Recursos	●
	LEGENDA	AT - Atende Totalmente / Majoritariamente
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

Dentre as Agências analisadas na pesquisa a EU é a única que apresenta documento específico para PGD baseado nos Princípios FAIR, pautado em modelos com políticas/diretrizes abrangentes para a gestão de dados científicos abertos.

Soma-se a isso, contempla no Programa Horizonte 2020 um projeto piloto (Open Research Data Pilot) que objetiva potencializar o acesso e reuso de dados científicos gerados

no Programa H2020. Atualmente, este projeto ainda é de caráter recomendatório, mas é incentivado como boa prática a ser adotada.

O quadro 28 aponta o resultado referente aos dados detalhados no quadro 20 referente aos critérios para o PGD.

Quadro 28 – Resultado da análise aos critérios do PGD – FCT

PLANO DE GESTÃO DE DADOS (PGD)		
	Critérios	FCT
DCC	Dados administrativos	●
	Coleta de dados	●
	Documentação e Metadados	●
	Ética e conformidade legal	●
	Armazenamento e Backup	●
	Seleção e Preservação	◐
	Compartilhamento de dados	●
	Responsabilidades e Recursos	◐
	LEGENDA	
	AT - Atende Totalmente / Majoritariamente	●
	AP - Atende parcialmente	◐
	NI - Não identificada (informação)	○

Fonte: Dados da pesquisa (2020).

Na FCT, foram analisados cinco documentos e tem-se o cenário para o PGD: dos oito critérios analisados atende totalmente/majoritariamente a seis, são atendidos parcialmente dois.

A FCT alude de forma genérica quanto aos Princípios FAIR, apenas em um relatório, de modo que não foi possível analisar aos subprincípios propostos, não sendo identificada informação.

O quadro 29 retrata o resultado dos dados compilados no quadro 22 a partir das análises aos 05 documentos da Agência NRF.

Quadro 29 – Resultado da análise aos critérios do PGD – NRF

PLANO DE GESTÃO DE DADOS (PGD)		
	Critérios	NRF
DCC	Dados administrativos	
	Coleta de dados	
	Documentação e Metadados	
	Ética e conformidade legal	
	Armazenamento e Backup	
	Seleção e Preservação	
	Compartilhamento de dados	
	Responsabilidades e Recursos	
LEGENDA	AT - Atende Totalmente / Majoritariamente	
	AP - Atende parcialmente	
	NI - Não identificada (informação)	

Fonte: Dados da pesquisa (2020).

Comparando-se as Agências Internacionais analisadas atinente a atenção ao PGD, a NRF representa a Agência com menor incidência de atendimento aos critérios, atendendo parcialmente a sete e não sendo identificada informação sobre armazenamento e backup.

Não foi identificada informação detalhada sobre os Princípios FAIR. Somente como prospecção para implementação junto a infraestruturas para pesquisas com grande volume de dados.

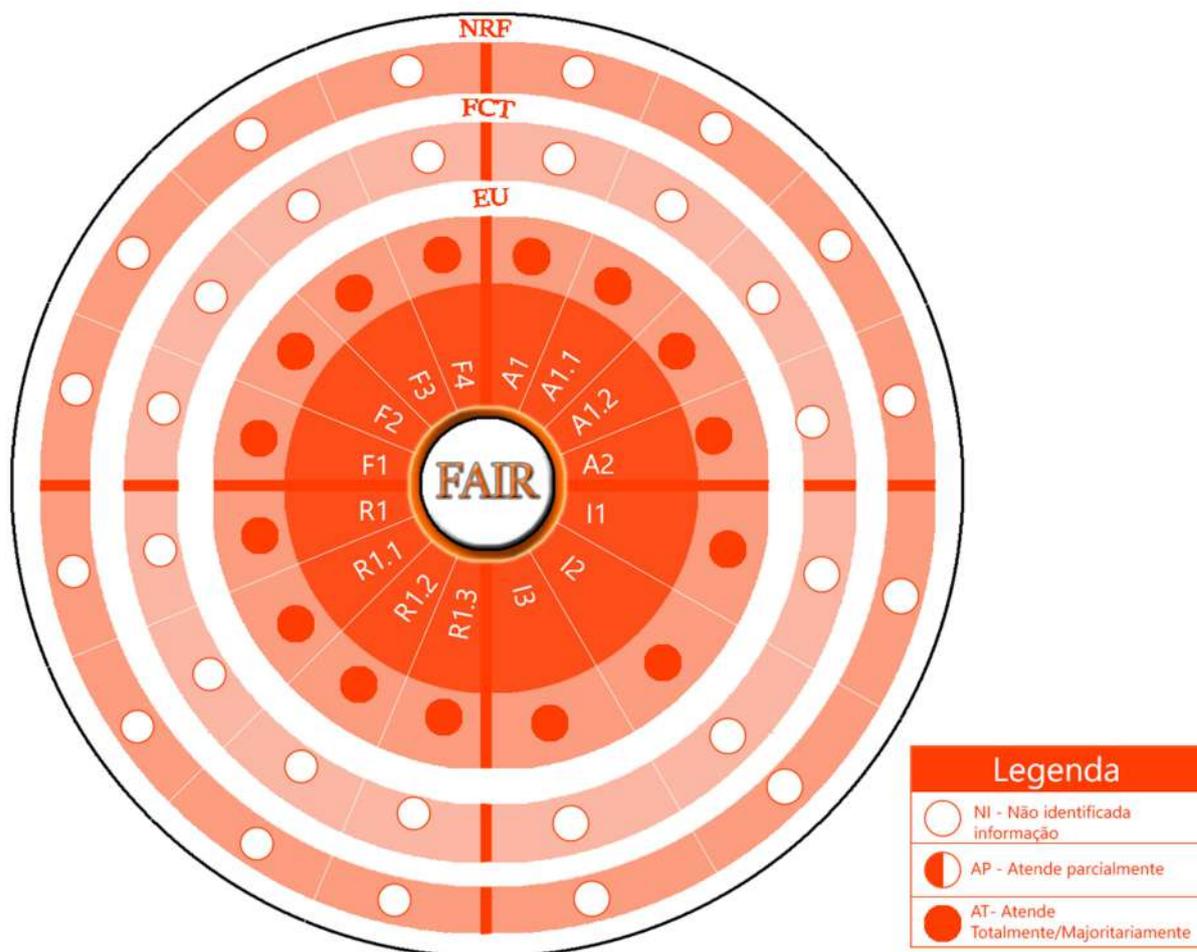
b) Representação consolidada das três Agências Internacionais – PGD

Apresenta de forma agrupada os resultados derivados das análises nas três Agências Internacionais, na figura 12.

b) Representação consolidada das três Agências Internacionais – Princípios FAIR

A figura 13 condensa os quadros 19, 21 e 23 apresentados na seção 4 Análise dos Dados por Agência de Fomento, subseção 4.2 Agências Internacionais.

Figura 13 – Representação dos resultados nas três Agências Internacionais – Princípios FAIR



Fonte: Dados da pesquisa (2020).

Os resultados das análises quanto aos princípios FAIR nas Agências Internacionais denotam que a União Europeia está com iniciativas mais robustas, sendo a única a atender totalmente/majoritariamente a todos os subprincípios. Inclusive com diretrizes específicas para a execução do PGD com amparo dos Princípios FAIR.

A FCT menciona os Princípios FAIR, mas não explana claramente quais seriam os subprincípios, portanto, para evitar imprecisões nos resultados não se considerou a menção nas análises.

Na NRF observou-se que consta no planejamento ainda a ser executado que devem

atender aos Princípios FAIR, mas no momento não foram identificadas informações sobre sua adoção.

No geral, foi possível observar que as políticas/diretrizes para as publicações dos resultados das pesquisas estão mais estruturadas, já quanto à divulgação dos dados científicos apesar de ter iniciativas, ainda fazem parte de projetos pilotos ou então, constam como recomendações. Exceto nas Agências FAPESP e EU, que já tem exigência de caráter mandatório para PGD e divulgação dos dados científicos. O Plano de Gestão de Dados é elemento importante e aliado aos Princípios FAIR, potencializa o valor dos dados ao longo do tempo e seu potencial de reuso.

Destaca-se a importância das bibliotecas e dos profissionais da informação para diminuir esse *gap* entre as exigências/recomendações das Agências e as formulações dos documentos para atendimento a esta nova cultura do fazer científico. Esse conhecimento pode ser construído por meio de grupos de trabalho com equipes multidisciplinares, juntamente com bibliotecários que tem familiaridade com questões importantes de documentação e metadados que são essenciais nesse processo. A gestão de dados científicos oferece uma excelente oportunidade de atuação para estes profissionais no âmbito da pesquisa científica.

5.3 INFOGRAFIA PARA UM PGD COM ABORDAGEM FAIR

A partir das informações observadas por meio das análises dos dados coletados nas Agências, se propõe uma infografia com questões e algumas ferramentas consideradas relevantes para gerar um PGD alinhado aos Princípios FAIR.

A proposta não tem a pretensão de elencar um modelo estático e único, apenas intenta nortear e auxiliar os pesquisadores e outros atores envolvidos que necessitem elaborar um PGD e procuram instrumentos de apoio nesse processo/etapa. Algumas Agências já indicam *templates* próprios ou de organizações parceiras no intuito de facilitar aos pesquisadores e suas equipes a geração deste documento, porém, nem todos fornecem detalhes e exemplos com possíveis padrões, informações reunidas e objetivas do processo.

Nessa perspectiva se delineou uma infografia básica (espécie de roteiro) para elaborar um PGD pautada na literatura adicionada de ferramentais (exemplos de padrões) para auxiliar na gestão dos dados científicos.

Sempre ressaltando que o PGD é um documento dinâmico, e como tal, deve ser atualizado à medida que a pesquisa se desenvolve, pois, algumas questões sobre os dados e usos podem estar nebulosas no início, mas durante o processo de investigação se delineiam e

apresentam maior clareza. As anotações de versionamento devem ser claras e acessíveis para manter o PGD o mais atualizado possível, assim como a indicação dos responsáveis.

O PGD deve abordar quais infraestruturas vai utilizar, padrões de dados e metadados alinhados às áreas de domínio, ou caso necessite criar que estejam documentados e permitam interoperabilidade, prever licenças claras de uso dos dados, repositório, como será a curadoria, o armazenamento, segurança e a preservação ao longo do tempo para que os dados estejam íntegros para o reuso.

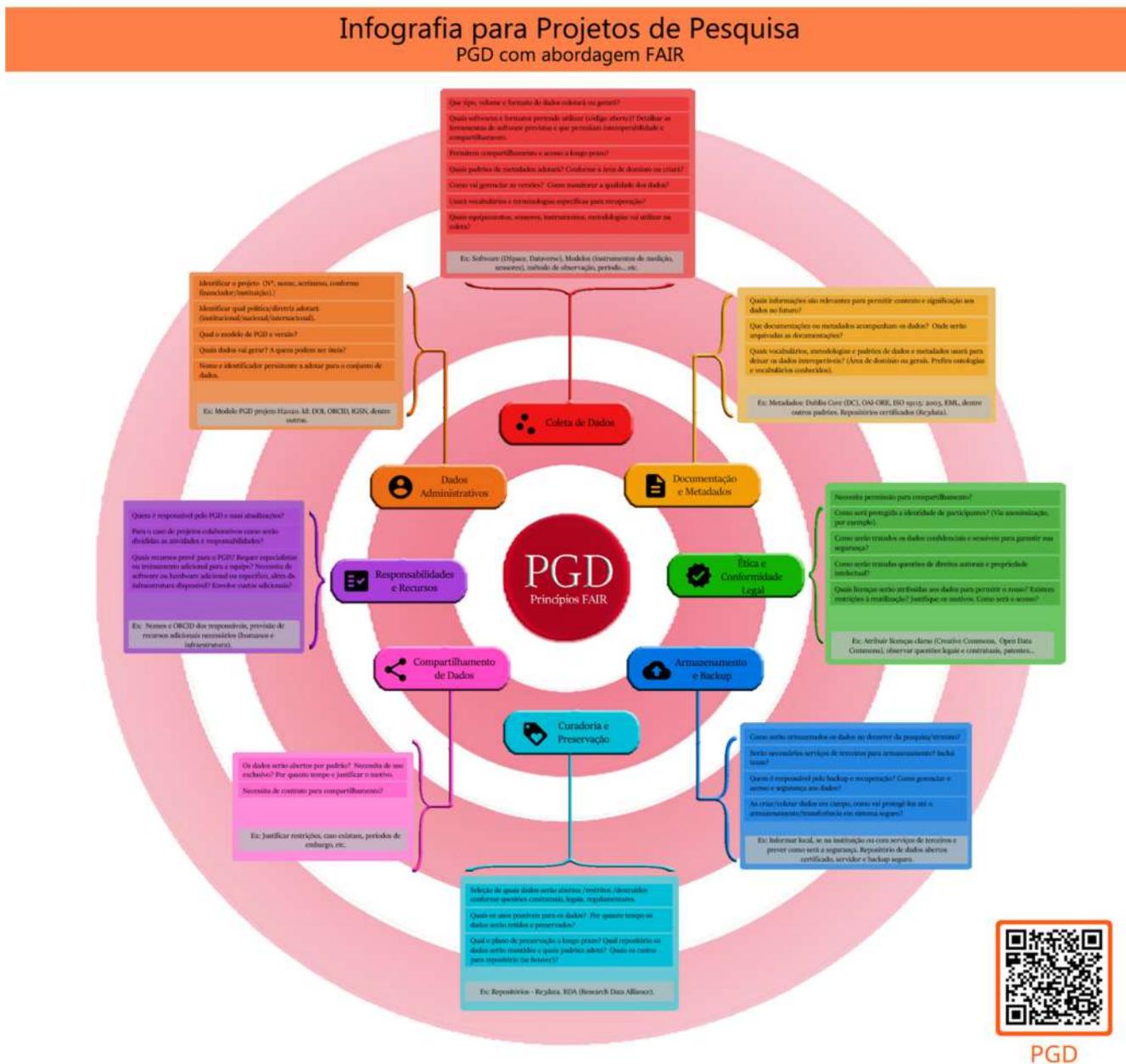
A infografia adaptou o modelo do DCC e do Programa Horizonte 2020, junto às informações decorrentes das análises.

Os itens da infografia não tem uma ordem estrita, porém, recomenda-se a iniciar o detalhamento pelos Dados administrativos.

As ferramentas citadas ou recomendadas pelas Agências estudadas não representam as únicas opções disponíveis, visto que podem ser adaptadas e geradas novas constantemente, porém servem como referencial para elaborar PGD a partir dessa pesquisa.

Todos os exemplos são meros norteadores, devem ser analisados conforme as áreas de domínio.

Figura 14 - Infografia para Projetos de Pesquisa - PGD com abordagem FAIR



Fonte: Dados da pesquisa (2020).

6 CONSIDERAÇÕES FINAIS

Toda pesquisa deve ser considerada a partir de seus recortes ou delimitações (abrangência, seu *corpus* de análise da pesquisa, rol documental, etc) conforme as opções e procedimentos metodológicos adotados no desenvolvimento do estudo. Destaca-se que devido à temática ser emergente e ter estreita ligação com uso de tecnologias, plataformas, e contextos dinâmicos fortemente relacionados às particularidades das diferentes áreas de domínio científico, os documentos sofrem constantes atualizações. Dessa maneira, o cenário representado se relaciona diretamente com as Agências de Fomento selecionadas para o estudo, aos documentos analisados e sua metodologia de análise, não permitindo generalizações. Outras configurações de pesquisa podem apresentar diferentes resultados.

Observou-se que as políticas sobre gestão de dados científicos abertos ainda estão em fase de estruturação e implementação, algumas em nível de país e outras por parte das Agências de Fomento. Independente da abrangência dessas políticas/orientações/diretrizes, todas são pautadas em boas práticas que estimulam a colaboração científica, o compartilhamento e reuso dos dados.

As Agências Nacionais selecionadas foram: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo – (FAPESP); e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). E as Internacionais são: European Union (EU), Fundação para a Ciência e a Tecnologia de Portugal (FCT) e National Research Foundation (NRF) todas com a característica de fomento em áreas multidisciplinares.

Observou-se que as Agências ainda estão em diferentes estágios na atenção aos oito critérios analisados para o PGD, sendo eles: Dados administrativos; Coleta de dados; Documentação e Metadados; Ética e conformidade legal; Armazenamento e Backup; Seleção e Preservação; Compartilhamento de dados e Responsabilidades e Recursos.

Ao abordar o cenário nacional para o PGD, das três Agências analisadas a FAPESP (atende totalmente/majoritariamente a todos os critérios), já adota políticas inclusive de caráter mandatório para alguns projetos, incentivando a abertura e colaboração científica com os dados gerados nos projetos. O CNPq (atende parcialmente aos oito critérios) e a FAPEMIG (atende parcialmente a três critérios: dados administrativos, Seleção e Preservação; Compartilhamento de dados), ambas as Agências apesar de não terem um PGD instituído para seus projetos, atendem a boas práticas no tratamento de dados científicos e apresentam iniciativas importantes no contexto da ciência aberta.

No cenário internacional, dentre as Agências estudadas a União Europeia (atende totalmente/majoritariamente a todos os critérios) é a organização que já apresenta um ecossistema para PGD e Princípios FAIR mais robusto, com documentações e orientações mais detalhadas. Portugal, por meio da FCT (atende totalmente/majoritariamente a seis critérios e parcialmente a dois: Seleção e Preservação e Responsabilidades e Recursos) para o PGD, também denota políticas consolidadas e adesão a boas práticas para manter e ampliar o escopo de ações em dados abertos. O país tem uma Política Nacional de Ciência Aberta que serve de guia para que as instituições agreguem condutas e práticas neste contexto. A NRF atende parcialmente a sete critérios, não sendo identificada informação para Armazenamento e Backup.

Dentre as seis Agências estudadas duas (EU e FAPESP) já adotam o PGD como elemento obrigatório (exigência) para solicitar apoio às pesquisas. As demais (CNPq, FAPEMIG, FCT e NRF), atendem aos critérios analisados de forma parcial, por meio de diretrizes para a gestão de dados científicos abertos em documentos relacionados às chamadas ou via diretrizes específicas das Agências.

Em atenção aos Princípios FAIR, somente foi observada sua adoção em uma Agência Internacional, a European Union (EU), com atendimento total/majoritário a todos os princípios FAIR e subprincípios. Em todas as demais Agências pesquisadas não foi identificada informação. Nas análises observou-se o que a literatura menciona que por vezes as exigências ou recomendações são postas, mas seu efetivo cumprimento é prejudicado por falta de conhecimento dos requisitos e experiência em implementar o que é solicitado. É uma área de pesquisa ainda emergente que busca se consolidar e instituir ferramentais, incentivar programas de treinamento e sensibilização para o reuso de dados científicos como uma dimensão da ciência aberta. Os Princípios FAIR estão em fase de familiarização junto à comunidade científica, o que pode ser apontado como um motivo para o resultado da pesquisa.

Por fim, elaborou-se uma infografia baseada nos critérios do DCC e no modelo do Horizonte 2020, com questões e breves exemplos para elaboração do PGD com abordagem FAIR, no intuito de ser mais uma ferramenta para apoio na gestão de dados científicos abertos. Não é modelo exaustivo, portanto, outras abordagens podem ser concebidas e apresentar novos cenários.

Avalia-se que um ecossistema de dados robusto com gestão bem definida, por meio do PGD alinhado aos Princípios FAIR se torna um propulsor do progresso científico de forma mais veloz e potencializa o investimento público (pois acelera a inovação, facilita à

colaboração, evita esforços duplicados, permite validar resultados e o reuso dos dados, envolve os indivíduos e a sociedade com maior abertura do processo científico).

As ferramentas e infraestruturas, repositórios, padrões de dados e metadados, linguagens, ontologias, e demais adotadas nos processos de curadoria e gestão de dados científicos, devem ser selecionadas conforme as áreas de domínio devido suas características e especificidades. Esta pesquisa citou modelos genéricos disponíveis e critérios que servem de roteiro para iniciar atividades na gestão de dados científicos, numa perspectiva de que elaborar PGD está cada vez mais próximo do cotidiano de todos os envolvidos com a produção de conhecimento científico.

A pesquisa denota que as Agências nacionais de fomento já têm importantes iniciativas no contexto da gestão de dados científicos, porém, seguem com o processo em construção, em que despontam iniciativas alinhadas as práticas internacionais, mas que ainda necessitam ampliar sua abrangência. Talvez, isso ocorra porque implementar a gestão de dados científicos de forma ampla requer/demanda infraestruturas seguras e adequadas, curadoria e tratamento dos dados, repositórios específicos que permitam interoperar, padrões de metadados, além de equipes multidisciplinares e técnicas com conhecimento nas áreas de domínio e também das ferramentas para a gestão dos dados durante todo seu ciclo (antes de iniciar as pesquisas, durante o desenvolvimento e após sua conclusão). Por isso o PGD é um documento dinâmico que deve ser atualizado constantemente para acompanhar a evolução das pesquisas e suas intercorrências durante todo o processo.

As áreas científicas estão cada vez mais imbrincadas e complementares, portanto, compartilhar e disponibilizar os dados científicos com qualidade é uma forma de cooperar com o bem estar social de forma ampla e globalizada.

Dados científicos abertos e confiáveis podem ser aliados estratégicos para soluções de problemas cotidianos cada vez mais complexos, no momento, isso é vivenciado pela Pandemia do COVID19, e a relevância da ciência e colaboração entre os pesquisadores de todas as nações se mostra claramente para o mundo.

REFERÊNCIAS

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência aberta: correntes interpretativas e tipos de ação. **Liinc em revista**, v. 10, n. 2, p. 434-450, 2014. Disponível em: <http://revista.ibict.br/liinc/article/view/3593/3072>. Acesso em: 25 jan. 2019.

ARAÚJO, Carlos A. Á. Teorias e tendências contemporâneas da Ciência da Informação. **Inf. Pauta**, Fortaleza, CE, v. 2, n. 2, p. 10-34, jul./dez. 2017. Disponível em: <http://www.periodicos.ufc.br/informacaoempauta/article/view/20162/71590>>. Acesso em: 20 mar. 2018.

AVENTURIER, P. **Princípios FAIR**: critérios de qualidade para dados de pesquisa. 2017. Disponível em: <https://publicient.hypotheses.org/1456>. Acesso em: 04 nov. 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, agosto 2017. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/168396/1/A-gestao-de-dados-de-pesquisa....pdf>. Acesso em: 27 mar. 2018.

BRAGA, K. S. Aspectos relevantes para a seleção da metodologia adequada à pesquisa social em Ciência da Informação. In: MUELER, Suzana P. M. **Métodos para pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007. Cap. 1, p.17-38.

BROCK, J. "A love letter to your future self": What scientists need to know about FAIR data. Familiarity with FAIR Principles. 2019. Disponível em: <https://www.natureindex.com/news-blog/what-scientists-need-to-know-about-fair-data>. Acesso em: 10 maio 2020.

CAMPELLO, B. S. Organizações como fontes de informação. In: CAMPELLO, B. S.; CÉNDON, B. V.; KREMER, J. M. **Fontes de informação para pesquisadores**. Belo Horizonte: UFMG, 2003.

CAVALCANTI, M. T.; SALES, L. F.; SAYÃO, L. F. Critérios para avaliação e seleção de dados de pesquisa no contexto da curadoria digital. In: ENANCIB, 17, 2016, Salvador, BA. Anais do Encontro Nacional de Pesquisa em Ciência da Informação, 17. **Anais...** Salvador: PPGCI – UFBA, 2016. p. 5386-5392.

CAVALCANTI, M. T.; SALES, L. F. Gestão de dados de pesquisa: um panorama da atuação da união europeia. **BIBLOS** – Rev. Instituto de Ciências Humanas e da Informação, v. 31, n. 1, p. 73-98, ago. 2017. Disponível em: <https://periodicos.furg.br/biblos/article/view/5789>. Acesso em: 29 mar. 2019.

CIÊNCIA ABERTA. **Sobre ciência aberta**. Disponível em: <https://www.ciencia-aberta.pt/sobre-ciencia-aberta>. Acesso em: 10 maio 2019.

CIÊNCIA ABERTA. **Política Nacional de Ciência Aberta**. Disponível em: <https://www.ciencia-aberta.pt/pnca>. Acesso em: 08 jun. 2020.

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico. **RN-009/2016 - Política de Dados do Programa de Pesquisas Ecológicas de Longa Duração - PELD.** 2016. Disponível em: http://www.cnpq.br/web/guest/view/-/journal_content/56_INSTANCE_0oED/10157/4621110. Acesso em: 10 jul. 2020

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico. **O CNPq.** Disponível em: http://www.cnpq.br/web/guest/apresentacao_institucional/. Acesso em: 01 jun. 2020a.

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico. **Bolsas e auxílios.** Disponível em: <http://www.cnpq.br/web/guest/apresentacao-bolsas-e-auxilios/>. Acesso em: 01 jun. 2020b.

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico. **Chamada CNPq/MCTI/CONFAP-FAPs/PELD N° 21/2020** - Programa: Programa de Pesquisa Ecológica de Longa Duração – PELD. Disponível em: http://www.cnpq.br/web/guest/chamadas-publicas?p_p_id=resultadosportlet_WAR_resultadoscnpqportlet_INSTANCE_0ZaM&id=47-1576-7049&detalha=chamadaDetalhada&filtro=abertas. Acesso em: 01 jul. 2020c.

CORRÊA, E. S.; BERTOCCHI, D. O papel do curador num cenário de curadoria algorítmica de informação. IN: CORRÊA, Elisabeth Saad (Org.) **Curadoria digital e o campo da comunicação.** São Paulo: ECA USP, 2012. Disponível em: https://issuu.com/grupo-eausp.com/docs/ebook_curadoria_digital_usp. Acesso em: 13 mar. 2019.

COSTA, M. M. **Diretrizes para uma política de gestão de dados científicos no Brasil.** 18/08/2017. 288 f. Doutorado (Ciências Da Informação). - Instituição de Ensino: Universidade De Brasília, Brasília Biblioteca Depositária: Biblioteca Central da UnB. 2017.

COSTA, M.; LEITE, F. C. L. 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. Disponível em: <http://www.redalyc.org/pdf/4656/465649115005.pdf>. Acesso em: 29 jan. 2019.

COMISSÃO EUROPEIA. **Horizon 2020 em breves palavras:** o programa-quadro de investigação e inovação da EU. Disponível em: https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/H2020_PT_KI0213413PTN.pdf. Acesso em: 08 maio 2019.

COMISSÃO EUROPEIA. **O que é e o que faz.** Luxemburgo: Serviço das Publicações da União Europeia, 2020a. Disponível em: <https://op.europa.eu/webpub/com/eu-what-it-is/pt/>. DOI:10.2775/041764.

COSTA, M. M.; CUNHA, M. B. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação.** Belo Horizonte. v. 19, n. 3, Jul./Set., p.189-206, 2014. Disponível em: <http://www.scielo.br/pdf/pci/v19n3/a10v19n3.pdf>. Acesso em: 26 set. 2018.

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.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. In: ENANCIB, 18, 2017, Marília, SP. Anais do Encontro Nacional de Pesquisa em Ciência da Informação, 18. **Anais...** Marília: UNESP, 23 a 27 outubro de 2017.

Disponível em:

<http://enancib.marilia.unesp.br/index.php/xviiiencib/ENANCIB/paper/viewFile/468/820>. Acesso em: 04 jan. 2019.

DCC - DIGITAL CURATION CENTRE. **What is Digital Curation?** Edinburgh: University of Edinburgh, c2004-2016a. Disponível em: <http://www.dcc.ac.uk/digital-curation/what-digitalcuration>. Acesso em: 10 mar. 2019.

DCC - DIGITAL CURATION CENTRE. **Checklist for a Data Management Plan**. v.4.0.

Edinburgh: Digital Curation Centre. 2013. Disponível em:

<http://www.dcc.ac.uk/resources/data-management-plans>. Acesso em: 06 abr. 2019.

DUMONTIER, M.; WESLEY, K. Advancing Discovery Science with FAIR Data Stewardship: Findable, Accessible, Interoperable, Reusable. **Serials Librarian**, [s. l.], v. 74, n. 1–4, p. 39–48, 2018. Disponível em: <https://doi.org/10.1080/0361526X.2018.1443651>.

Acesso em: 30 jan. 2019.

EUROPEAN COMMISSION. **EOSC Declaration**. 2017. Disponível em:

https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf. Acesso em: 26 ago. 2020

EUROPEAN COMMISSION. **Turning FAIR into reality: final report and action plan from the european commission expert group on fair data**. Luxembourg: Publications Office of the European Union, 2018. Disponível em:

https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf. Acesso em: 01 abr. 2020. DOI: 10.2777/1524.

EUROPEAN COMMISSION. **Guidelines on FAIR Data Management in Horizon 2020**. 2016. Disponível em:

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Acesso em: 26 jun. 2020

EUROPEAN COMMISSION. **Guidelines to the Rules on Open Access to Scientific Publications And Open Access to Research Data in Horizon 2020**. 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: 20 jun. 2020a.

EUROPEAN COMMISSION. **Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics: Guidelines for open access to publications, data and other research outputs**. Version 1.1 April 24, 2020. Disponível em:

https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-guidelines-oa-covid-19_en.pdf. Acesso em: 10 ago. 2020b

EUROPEAN COMMISSION. **Horizon 2020 - Work Programme 2018-2020**. 13. Europe in a changing world – Inclusive, innovative and reflective societies. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-societies_en.pdf. Acesso em: 20 maio 2020.

EUROPEAN RESEARCH COUNCIL – ERC. **Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020**. 2017. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf. Acesso em: 10 ago. 2020.

EUROPEAN RESEARCH COUNCIL – ERC. Scientific Council. **Open Research Data and Data Management Plans: Information for ERC Grantees**. 2019. Disponível em: https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf. Acesso em: 10 jul. 2020.

EUROPEAN SCIENCE FOUNDATION. **Plan S**. Making full and immediate Open Access a reality. Disponível em: https://www.coalition-s.org/plan_s_principles/. Acesso em: 10 ago. 2020.

FANE, Briony; TREADWAY, Jon. What is the State of Open Data in 2018? In: **Digital Science Report**. The State of Open Data 2018. Out. 2018. DOI: <https://doi.org/10.6084/m9.figshare.7195058>. Acesso em: 25 jun. 2020.

FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais. **Quem somos**. Disponível em: <https://fapemig.br/pt/menu-institucional/quem-somos>. Acesso em: 10 mar. 2020a.

FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais. **Chamada FAPEMIG 08/2019 - Programa PIBITI** – Empresa FAPEMIG/CNPq/ MCTIC/SEMPLI. Disponível em: Acesso em: 10 maio 2020b

FAPEMIG- Fundação de Amparo à Pesquisa do Estado de Minas Gerais. **Minas Faz Ciência**. Disponível em: <https://fapemig.br/pt/minas-faz-ciencia/>. Acesso em: 10 jun. 2020c

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.

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. **A FAPESP**. Disponível em: <http://www.fapesp.br/sobre/>. Acesso em: 10 jun. 2020a

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. **Evolução das receitas**. Disponível em: <http://www.fapesp.br/9250>. Acesso em: 10 jun. 2020b

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. **Auxílio à Pesquisa - Projeto Temático**. Disponível em: <http://fapesp.br/176#docs14>. Acesso em: 10 jul. 2020c

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. **Portaria CTA nº 01/2019**. Institui a política para acesso aberto às publicações resultantes de auxílios e bolsas FAPESP. Disponível em: <http://www.fapesp.br/12632>. Acesso em: 20 maio 2020d.

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. **Open Science @ FAPESP**. Disponível em: <http://www.fapesp.br/openscience/>. Acesso em: 20 jun. 2020e

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. **A FAPESP e os Dados Abertos**. Disponível em: <http://www.fapesp.br/openscience/>. Acesso em: 20 jun. 2020f

FCT - Fundação para a Ciência e a Tecnologia de Portugal. **Sobre a FCT**. Disponível em: <https://www.fct.pt/fct>. Acesso em: 10 jun. 2020a

FCT - Fundação para a Ciência e a Tecnologia de Portugal. **Sobre a FCT - Principais linhas de ação**. Disponível em: <https://www.fct.pt/linhasatividadefct>. Acesso em: 10 jun. 2020b

FCT - Fundação para a Ciência e a Tecnologia de Portugal. **Política sobre Acesso Aberto a Publicações Científicas resultantes de Projetos de I&D Financiados pela FCT**. (Adotada em 5 de maio de 2014). Disponível em: https://www.fct.pt/documentos/PoliticaAcessoAberto_Publicacoes.pdf. Acesso em: 10 jun. 2020

FCT - Fundação para a Ciência e a Tecnologia de Portugal. **Acesso Aberto Modelos, Políticas e Custos de Acesso**. 2017. Disponível em: <https://www.fct.pt/acessoaberto/docs/modelosacessoaberto.pdf>. Acesso em: 10 jun. 2020.

FCT - Fundação para a Ciência e a Tecnologia de Portugal. **Memorando de entendimento entre Ministério da Ciência, Tecnologia e Ensino Superior da República Portuguesa (MCTES) e o Ministério da Ciência, Inovações e Comunicações da República Federativa do Brasil (MCTIC) para a cooperação no âmbito de ciência aberta**. 2018. Disponível em: https://www.fct.pt/dsi/docs/Memorando_MCTIC-MCTES_Ciencia_Aberta_junho_2018.pdf. Acesso em: 10 jun. 2020

FERNANDES, H. D. H.; OLIVEIRA, A. F. Gestão da Preservação digital em repositórios de dados de pesquisa. **Revista Ibero-Americana de Ciência da Informação**, [S.l.], v. 11, n. 1, p. 255-273, fev. 2018. ISSN 1983-5213. Disponível em: <http://periodicos.unb.br/index.php/RICI/article/view/27902>. Acesso em: 08 maio 2019.

FOSTER. **Open Science Definition**. Projeto da União Europeia para a promoção da implementação prática do Open Science no Horizonte 2020. Disponível em: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>. Acesso em: 10 jan. 2019. [Tradução nossa].

FIOCRUZ – Fundação Oswaldo Cruz. **Curso introdução à ciência aberta**. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie1/curso1/aula1.html>. Acesso em: 10 jan. 2019a.

FIOCRUZ – Fundação Oswaldo Cruz. **Panorama histórico da ciência aberta**. Cenário brasileiro. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie1/curso2/aula2.html>. Acesso em: 10 maio 2019b.

FIOCRUZ – Fundação Oswaldo Cruz. **Glossário do acesso aberto**. Disponível em: <https://portal.fiocruz.br/glossario-acesso-aberto/V>. Acesso em: 20 maio de 2019c.

FIOCRUZ – Fundação Oswaldo Cruz. **Termo de referência** - Gestão e Abertura de Dados para Pesquisa na Fiocruz. Disponível em: https://www.arca.fiocruz.br/bitstream/icict/26803/2/termo_referencia_lancamento.pdf. Acesso em: 20 maio de 2019d.

FIOCRUZ. **Curso – Dados abertos**. Gestão, compartilhamento e abertura de dados para pesquisa: uma nova cultura no fazer científico. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie3/curso2/aula1.html>. Acesso em: 15 jun. 2020

FIOCRUZ. **Glossário**: Princípios FAIR. Disponível em: <https://portal.fiocruz.br/glossario>. Acesso em: 25 mar. 2020a.

FIOCRUZ. **Acesso Aberto**. Disponível em: <https://portal.fiocruz.br/glossario>. Acesso em: 28 mar. 2020b.

FIOCRUZ. **O que é**. Disponível em: <https://portal.fiocruz.br/o-que-e>. Acesso em: 28 mar. 2020c.

FIOCRUZ. **Curso – Dados abertos**. Gestão ativa de dados, uma mudança cultural. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie3/curso2/aula2.html>. Acesso em: 15 jun. 2020d.

FIOCRUZ. **Curso – Dados abertos**. Políticas dos financiadores, requisitos das revistas científicas e repositórios de dados em saúde. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie3/curso2/aula6.html>. Acesso em: 15 jun. 2020e.

GO FAIR. **FAIRification Process**. Disponível em: <https://www.go-fair.org/fair-principles/fairification-process/>. Acesso em: 09 jul.2019.

GO FAIR. **Declaração GO FAIR Brasil (BR)**. Disponível em: https://www.go-fair.org/wp-content/uploads/2019/06/Declaration_GO-FAIR-Brazil_Jun2019.pdf. Acesso em: 20 jul. 2020.

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, PR: UEL, 22 a 26 de outubro de 2018. p.5047-5067.

HENNING, P. *et al.* GO FAIR e os Princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 389-412, maio/ago. 2019. DOI: <http://dx.doi.org/10.19132/1808-5245252.389-412>.

IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia. **Diretrizes para repositórios de dados e de publicações científicas**. 2020. Disponível em: <https://ibict.br/sala-de-imprensa/noticias/item/2490-ibict-disponibiliza-documentos-relacionados-a-diretrizes-para-repositorios-de-dados-e-de-publicacoes-cientificas>. Acesso em: 22 out. 2020.

LAKATOS, E. M.; MARCONI, M. A. **Metodologia científica**. 5. ed. São Paulo: Atlas, 2007.

MCTES - Ministério da Ciência, Tecnologia e Ensino Superior (Portugal). **Segundo relatório da Comissão Executiva do Grupo de Trabalho - Política Nacional de Ciência Aberta**. 2016. Disponível em: <https://www.ciencia-aberta.pt/documentos>. Acesso em: 10 jun. 2020

MEDEIROS, J. S. Uma investigação sobre a autoria de dados científicos: teias de uma rede em construção. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 14, n. 2, p. 298-317, maio 2016. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8644015>. Acesso em: 23 jan. 2019. DOI:<https://doi.org/10.20396/rdbci.v14i2.8644015>.

MICHENER, 2015, W. K. Ten Simple Rules for Creating a Good Data Management Plan. **PLOS Computational Biology**, v. 11, n. 10, e1004525, 2015. Disponível em: <https://doi.org/10.1371/journal.pcbi.1004525>. Acesso em: 10 jun. 2019.

MINAYO, M. C. S. (Org). **Pesquisa social: teoria, método e criatividade**. 30. ed. Petrópolis, RJ: Vozes, 2011.

MEDEIROS, J. S.; CAREGNATO, S. N. E. Compartilhamento de dados e e-science: explorando um novo conceito para a comunicação científica? Data sharing and e-science: exploring a new concept for scientific communication. **Liinc em revista**, v. 8, n. 2, 2012. Disponível em: <http://www.brapci.inf.br/v/a/22808>. Acesso em: 29 abr. 2018.

MONTEIRO, G. **Mapeamento e Análise das Políticas Institucionais de Financiadores da Pesquisa Brasileira: cenário dos dados científicos abertos**. 2019. 146 p. Dissertação (Mestrado) - Universidade do Estado de Santa Catarina, Centro de Ciências Humanas e da Educação, Programa de Pós-Graduação em Gestão de Unidades de Informação, Florianópolis, 2019a.

MONTEIRO, G. **Mapeamento e Análise das Políticas Institucionais de Financiadores da Pesquisa Brasileira: cenário dos dados científicos abertos**. [Data set]. ZENODO, 2019b. DOI: [10.5281/zenodo.3372826](https://doi.org/10.5281/zenodo.3372826).

NRF - National Research Foundation. **Statement on Open Access to Research Publications from the National Research Foundation (NRF) - Funded Research**. 2015. Disponível em: https://www.nrf.ac.za/sites/default/files/documents/oastatement_2015.pdf. Acesso em 08 jun.2020.

NRF - National Research Foundation. **The African Open Science Platform: The Future of Science and Science for the Future**. 2018. Disponível em: <https://www.nrf.ac.za/sites/default/files/documents/AOSP%20Strategy%20Final%20HR.pdf>. Acesso em: 10 jun. 2020.

NRF - National Research Foundation. **About NRF:** corporate overview. Disponível em: <https://www.nrf.ac.za/about-nrf.pdf>. Acesso em: 10 jun. 2020a

NRF - National Research Foundation. **Annual Performance Plan 2019/20–2021/22.** [2019]. Disponível em: <https://www.nrf.ac.za/sites/default/files/documents/NRF%20APP%202019-2022.pdf>. Acesso em: 10 jun. 2020

ODI - Open Data Institute. **The Data Spectrum.** Disponível em: <https://theodi.org/about-the-odi/>. Acesso em: 10 maio 2020

OECD - Organization for Economic Cooperation and Development. **Principles and guidelines for access to research data from public funding.** Paris, OECD, 2007. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 31 jul. 2020

PAVÃO, C. G.; et al. **Acesso aberto a dados de pesquisa no Brasil:** repositórios brasileiros de dados de pesquisa: relatório 2018. Disponível em: https://www.rnp.br/sites/default/files/rdpbrasil-acessoabertodadospesquisabrasil_repositoriosbrasileiros.pdf. Acesso em: 29 jul. 2019.

PAVÃO, C. M. G., et al. Metadados e repositórios institucionais: uma relação indissociável para a qualidade da recuperação e visibilidade da informação. **Ponto de Acesso**, v. 9, n. 3, p. 103-116, 2015. DOI: 10.9771/1981-6766rpa.v9i3.15163. Acesso em: 21 jul. 2020.

PPBIO - Programa de Pesquisa em Biodiversidade. **Repositório:** repositório de dados de estudos ecológicos. Disponível em: <https://ppbio.inpa.gov.br/repositorio/dados>. Acesso em: 02 maio 2020

RE3DATA. **Repositórios brasileiros.** Disponível em: <https://www.re3data.org/search?query=&countries%5B%5D=BRA>. Acesso em: 05 ago. 2020

RIOS, F. P.; LUCAS, E. R. O.; AMORIM, I. S. Manifestos do movimento de acesso aberto: análise de domínio a partir de periódicos brasileiros. **RBBB - Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 15, n. 1, p. 148-169, jan. 2019. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1152/1103>. Acesso em: 30 jul. 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. DOI: <http://dx.doi.org/10.9771/rpa.v11i3.24945>

RODRIGUES, E. **A nova geração de repositórios e a transição para o acesso aberto: a perspectiva da COAR.** 2018. Disponível em: http://www.capes.gov.br/encontro-ciencia-aberta/apresentacoes/2018-09-18_1-Encontro-de-Ciencia-Aberta_EloyRodrigues1_COAR.pdf. Acesso em: 25 abr. 2019.

RODRIGUES, E.; SARAIVA, R. **Os Repositórios de dados científicos:** estado da arte. 2010. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/23806/2/44632.pdf>. Acesso em: 29 nov. 2018.

SANTOS, T. N. C. **Curadoria digital: o conceito no período de 2000 a 2013**. 2014. 165 f. Dissertação (Mestrado em Ciência da Informação - Faculdade de Ciência da Informação da Universidade de Brasília), Brasília, 2014. Disponível em: http://repositorio.unb.br/bitstream/10482/17324/1/2014_ThayseNataliaCantanhedeSantos.pdf. Acesso em: 20 mar. 2018.

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.

SAYÃO, L. F. **Desafios da ciência invisível: dados da cauda longa da ciência**. Disponível em: <https://www.unifesp.br/reitoria/bibliotecas/images/bibliotecas/LuisSayao-.pdf>. Acesso em: 05 jun. 2020.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.: Estudos**, João Pessoa, v. 22, n.3. p.179-191, set./dez. 2012. Disponível em: <https://bit.ly/2YJjNLe>. Acesso em: 10 mar. 2018.

SAYÃO, L. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 1, 2013. Disponível em: <http://carpedien.ien.gov.br/bitstream/ien/646/1/dados%20de%20pesquisa.pdf>. Acesso em: 27 mar. 2019.

SAYÃO, L. F.; SALES, L. F. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 20 jun. 2019.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, [S.l.], v. 21, n. 2, p. 90-115, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939/20122>. Acesso em: 20 jul. 2019. doi:<http://dx.doi.org/10.5433/1981-8920.2016v21n2p90>.

SAYÃO, L. F.; SALES, L. F. A Ciência Invisível: revelando os dados da cauda longa da pesquisa. In: Encontro Nacional de Pesquisa em Ciência da Informação, ENANCIB, 19. **Anais...** Londrina, 2018, PR: UEL, 22 a 26 de outubro de 2018. p.4180-4199.

SAYÃO, L. F.; SALES, L. F. Curadoria de dados de pesquisa em repositórios. In: Encontro da Rede Sudeste de Repositórios Institucionais, 1., 2019, Rio de Janeiro. **Anais...** Rio de Janeiro: Fiocruz/Icict/UFRJ, 2019. 80 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/33337>. Acesso em: 20 jun. 2019.

SAYÃO; L. F.; SALES, L. F. A Ciência Invisível: os dados da cauda longa da pesquisa científica. In: DIAS, G. A.; OLIVEIRA, B. M. J. F (Org.). **Dados científicos: perspectivas e desafios**. João Pessoa: UFPB, 2019a.

SEMELER, A. R. **Ciência da Informação em Contextos de E-Science: Bibliotecários de Dados em Tempos de Data Science**. 168 f. Doutorado (Ciência da Informação) - Instituição de Ensino: Universidade Federal de Santa Catarina, Florianópolis Biblioteca Depositária: Biblioteca Universitária (UFSC). 2017.

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.

SILVA, F. C. C. O papel dos bibliotecários na gestão de dados científicos. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 14, n. 3, p. 387-406, set. 2016. ISSN 1678-765X. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646333>. Acesso em: 28 jul. 2018. DOI:<https://doi.org/10.20396/rdbci.v14i3.8646333>.

SILVA, F.C. C. **Gestão de Dados Científicos**. Rio de Janeiro: Interciência, 2019.

SILVA GRAÇA, A. C. *et al.* O curador da arte à informação. **Revista Científica Ciência em Curso**. Palhoça, SC, v. 5, n.1, p.67-78, jan./jun. 2016. Disponível em: <http://linguagem.unisul.br/paginas/ensino/pos/linguagem/ciencia-em-curso/0501/050107.pdf>. Acesso em: 15 mar. 2018.

SILVA, L. C.; SANTAREM SEGUNDO, J. E.; SILVA, M. F. Princípios de FAIR e melhores práticas do *linked data* na publicação de dados de pesquisa. In: Encontro Nacional de Pesquisa em Ciência da Informação, ENANCIB, 19, 2018. **Anais...** Disponível em: <http://www.brapci.inf.br/index.php/res/v/103810>. Acesso em: 25 jan. 2019.

UNINETT Sigma2. **About Sigma2**. Disponível em: <https://www.sigma2.no/about-sigma2>. Acesso em: 25 jun. 2020

UZWYSHYN, Ray. Research data repositories: the what, when, why, and how. **Computers in Libraries**, Westport, v. 36, n. 3, Apr. 2016. Disponível em: <http://www.infotoday.com/cilmag/apr16/Uzwyshyn--Research-Data-Repositories.shtml>. Acesso em: 15 jul. 2019.

VANZ, S. A. S., *et al.* **Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório 2018**. Disponível em: <https://hdl.handle.net/20.500.11959/5>. Acesso em: 20 maio 2019.

VIANA, C. L. M; MÁRDERO ARELLANO, M. A.; SHINTAKU, M. Repositórios institucionais em Ciência e Tecnologia: uma experiência de customização do Dspace. In: **PROCEEDINGS SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS**, 3., 2005, São Paulo. Anais... São Paulo. Disponível em: <http://eprints.rclis.org/7168/1/viana358.pdf>. Acesso em: 15 jun. 2020.

VISOLI, M. C., *et al.* Acesso aberto na Embrapa: breve histórico, avanços recentes e desafios. **RECIIS - Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v. 11, p. 1-6, nov. 2017. DOI: 10.29397/reciis.v11i0.1399. Disponível em: <https://www.arca.fiocruz.br/handle/icict/23712>. Acesso em: 25 maio 2020.

WILKINSON, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n.1, p. 160018, 2016. DOI:10.1038/sdata.2016.18.

APÊNDICES

APÊNDICE A – Resultado gerado pela Ferramenta Python – Incidência de termos

outside20200925-142606 - Bloco de Notas  Nome do arquivo com data e horário. Gerado pela ferramenta Python.

Arquivo Editar Formatar Exibir Ajuda

FCT/ACESSO_ABERTO_RELATORIO2.PDF  Nome da pasta/nome do documento analisado.

```
plano de gestão: 1
FAIR data management: 0
FAIR data: 0
Princípios FAIR: 1
FAIR principles: 0
interoperabilidade: 2
repository: 1
repositório: 3
repositórios de dados: 0
armazenamento: 0
coleta de dados: 0
responsabilidade legal: 0
disponibilização de dados: 0
plano de gerenciamento de dados: 0
plano de gestão de dados: 1
DMP: 0
data management plan: 0
research management plan: 0
gestão de dados: 3
metadados: 0
metadata: 1
identificador persistente: 0
dados abertos: 6
open data: 0
gerenciamento de dados: 0
data reuse: 0
reuso: 0
reutilização: 1
research data management: 0
```

Lista de termos e sua respectiva incidência no documento.

FCT/AVISO_ABERTURA_DATASCIENCE_2020_ANEXO_II_FAIR.PDF  Nome do documento analisado.

```
plano de gestão: 4
FAIR data management: 0
FAIR data: 0
```

APÊNDICE B- Modelo da planilha de análise dos documentos de cada Agência – PGD.

Agência de Fomento	
Documento analisado	
Data da Publicação	
Critérios DCC	Dados administrativos
	Coleta de dados
	Documentação e Metadados
	Ética e conformidade legal
	Armazenamento e Backup
	Seleção e Preservação
	Compartilhamento de dados
	Responsabilidades e Recursos
Legenda	
AT - Atende Totalmente/Majoritariamente	
AP - Atende parcialmente	
NI - Não identificada	

Fonte: Dados da pesquisa (2020) adaptado de Monteiro (2019b).

