

ANÁLISE DA ADERÊNCIA AOS PRINCÍPIOS FAIR EM DADOS ABERTOS DO GOVERNO FEDERAL BRASILEIRO COM O FAIRDATABR

ANALYSIS OF ADHERENCE TO FAIR PRINCIPLES IN OPEN DATA FROM THE BRAZILIAN FEDERAL GOVERNMENT WITH FAIRDATABR

Marckson Roberto Ferreira de Sousa^a
Arthur Campos^b
Samyr Delfino^c

RESUMO

Objetivo: Este artigo objetiva analisar a aderência aos princípios FAIR por meio do uso da ferramenta FairDataBR no contexto dos dados abertos do Governo Federal. **Metodologia:** Possui característica qualitativa ao passo que procura averiguar problemas e indicar soluções para o Portal Brasileiro de Dados Abertos por meio da ferramenta FairDataBR. Utiliza como fontes de informação o Portal de Periódicos da Capes, a Base de Dados em Ciência da Informação (Brapci) e o Google Acadêmico para embasamento bibliográfico e teórico. Busca nesses ambientes por termos relacionados aos princípios FAIR, aos dados abertos, aos objetos digitais e ao ciclo de vida dos dados não marcando um período exato para recuperação dos materiais. A ferramenta FairDataBR é utilizada para verificar a aderência FAIR dos dados abertos do referido portal. **Resultados:** Conforme a análise, a ferramenta FairDataBR gera a média de 7,95 para aderência ao FAIR, tendo aderência maior ao princípio *reusable* (8,80) e a menor para o princípio *findable* (7,00). Para os princípios *accessible* e *interoperable*, a ferramenta gera respectivamente as notas 7,50 e 8,50. **Conclusões:** a ferramenta FairDataBR, utilizada para a análise de adaptabilidade do Portal Brasileiro de Dados Abertos do Governo Federal aos princípios FAIR, proporciona uma avaliação estruturada e sistemática dos metadados e dados disponíveis. Atribui que a FairDataBR é uma aliada na promoção de transparência e de boas práticas no compartilhamento de dados públicos, contribuindo para o fortalecimento da governança de dados no Brasil e incentivando a implementação dos princípios FAIR em diferentes domínios.

Descritores: Princípios FAIR. Dados Abertos. Ferramenta FairDataBR. Portal Brasileiro

^a Doutor em Engenharia Elétrica pela Universidade Federal da Paraíba (UFPB). Professor da Universidade Federal da Paraíba (UFPB). João Pessoa, Brasil. Email: marckson.dci.ufpb@gmail.com.

^b Doutor em Ciência da Informação pela Universidade Federal da Paraíba (UFPB). João Pessoa, Brasil. arthurfcampos94@gmail.com.

^c Doutor em Ciência da Informação pela Universidade Federal da Paraíba (UFPB). Professor da Universidade Federal da Paraíba (UFPB). João Pessoa, Brasil. Email: samyr@d@gmail.com.

de Dados Abertos.

1 INTRODUÇÃO

Atribuir valor a algo que não seja necessariamente um número podendo ser um conceito, uma representação ou a característica de algo, entre outras observações gerais, pode ser uma definição não reducionista para os dados. Dados representam um tipo de registro relacionado a algo maior que pode ter sentido para ser medido ou mensurado por um tipo de dispositivo, ou também pode ser arbitrário dependendo do contexto de criação.

Qualquer pessoa ou organização pode gerar dados desde um cadastro na *web* para criação de um perfil em rede social, uma transação *e-commerce*, um contrato entre organizações, o arquivamento de um determinado documento etc. Os dados podem ser abertos ou fechados dependendo de sua finalidade. Conforme a *Open Knowledge Foundation* ([s.d.], *online*), os dados abertos podem ser livremente acessados, utilizados, modificados e compartilhados por qualquer pessoa, estando sujeitos, no máximo, a exigências que visem preservar a sua proveniência e abertura para o público interessado.

Para que os dados possam atingir a todas essas características, os princípios FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) são atributos para que os dados sejam localizáveis, acessíveis, interoperáveis e reutilizáveis. Sua criação, segundo Wilkinson *et al.* (2016), teve origem num grupo de pessoas que constatou a necessidade vigente de se melhorar a infraestrutura para o compartilhamento dos dados de pesquisa e de seu posterior reuso. Assim, os princípios FAIR servem como diretrizes para auxiliar no compartilhamento e reuso de dados de pesquisa, dados abertos no geral, conjuntos de dados, dados arquivados em repositórios de dados, entre outros.

Para isso, esta pesquisa verifica a ferramenta FairDataBr¹ mediante a análise do Portal Brasileiro de Dados Abertos do Governo Federal² no âmbito da adoção dos princípios FAIR. A questão de pesquisa que inicia o estudo é em que

¹ Disponível em: <https://wrcu.ufpb.br/fair/index.html>

² Disponível em: <https://dados.gov.br/home>

medida os Dados Abertos do Governo Federal atendem aos princípios FAIR, segundo avaliação do FairDataBr? Desse modo, o objetivo desta pesquisa é analisar a aderência aos princípios FAIR por meio do uso da ferramenta FairDataBr no contexto dos Dados Abertos do Governo Federal.

A justificativa científica está em fomentar pesquisas de análise da ferramenta FairDataBr, criada por pesquisadores da Universidade Federal da Paraíba (UFPB). Além disso, procura-se estudar e indicar holisticamente as finalidades do Portal Brasileiro de Dados Abertos do Governo Federal. Também busca estimular a discussão sobre os princípios FAIR na comunidade científica brasileira, protagonizando a necessidade de compartilhamento e reuso de dados. Como justificativa social, reunir os princípios Fair, a ferramenta FairDataBr e o referido portal contribui para aguçar estudos que demonstrem para a sociedade o que são os dados, sua importância e o que a Ciência da Informação está construindo nesse sentido.

2 CICLO DE VIDA DOS DADOS E A NECESSIDADE DE COMPARTILHAMENTO E REÚSO

A investigação sobre os dados e, conseqüentemente, sobre o seu ciclo de vida apresenta-se como um assunto recorrente no campo da Ciência da Informação. Isso porque os dados são grandes fontes de informação para tomada de decisão possibilitando análises diversas com comparações e recorrências de fenômenos. Além disso, é uma chave favorável à população para acesso à informação e aos seus direitos no contexto das políticas públicas.

Conforme Sant'Ana (2019), o campo informacional pode ser resultante da interação de ciclos de vida dos dados (CVD), o que é suportado pelas tecnologias da informação; essas vão além da coleta e armazenamento, participando também do processo de filtragem, seleção e geração de novos conteúdos. Face a isso, Castells (2016) ressalta que as tecnologias devem agir sobre a informação e não apenas o contrário, alimentando esse constante fluxo informacional.

Desse modo, há uma preocupação sobre os CVD e a interação entre esses ciclos com pessoas ou organizações, surgindo-se na maioria das vezes

novos ciclos de vida dos dados a partir disso (Sant'Ana, 2019). Dados também são considerados a nova fonte de capital (Lau *et al.*, 2016), isto é, um 'pagamento' comum cobrado por organizações que operam na *web* (Silva Segundo; Araújo, 2019).

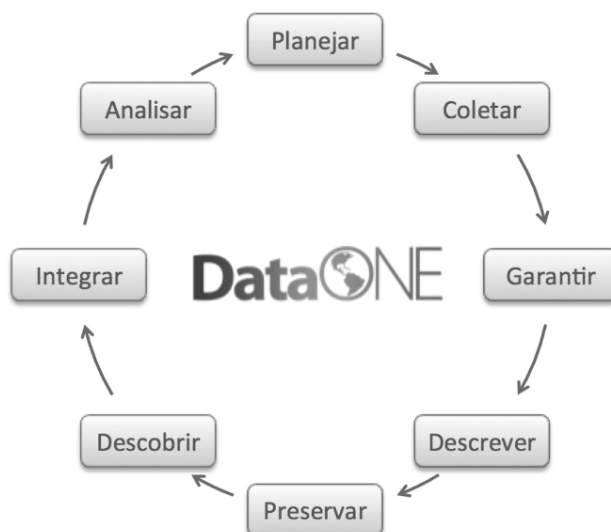
A ideia de 'pagamento' abarca o ato de compartilhamento de dados pelos usuários de um serviço quando querem fazer parte de algo como, por exemplo, compartilhar dados numa rede social para possuir um perfil naquele espaço. Isso deixa um 'rastro' fazendo com que o usuário tenha uma presença naquele espaço por meio do seu compartilhamento de dados pessoais (nome completo, data de nascimento, nome da mãe, nome do pai, escolaridade, endereço postal, cpf, entre outros). Nesse sentido, por mais que seja 'gratuito' criar um perfil numa rede social na *web*, isso funciona a partir da alimentação de um banco de dados pessoais.

Essa coleta de dados pessoais favorece bastante as preferências do usuário de determinada rede social, fomentando o algoritmo a trabalhar mediante sugestões de compras, por exemplo. Categoricamente, segundo Castells (2016), a sociedade em rede está alicerçada nos computadores em rede e na internet e cada rede é um conjunto de 'nós' interconectados dentro da comunicação em rede, isto é, a interação. Para isso, Campos (2024, p. 43) evidencia que

A comunicação em rede, proeminente na sociedade em rede, possui sujeição ao compartilhamento, a disseminação, a divulgação e ao alcance de conteúdos para os sujeitos que se inserem nessa rede. Esse processo atribui interdependência podendo os conteúdos e/ou componentes fazerem parte de duas ou mais redes, formando redes de relações que se intercomunicam e interconectam.

A comunicação e o compartilhamento dentro desse fluxo informacional também contribuem para os ciclos de vida dos dados. A *Data Observation Network for Earth*, de acrônimo *DataONE*, também atribui um CVD dentro das seguintes fases: planejar, coletar, garantir, descrever, preservar, descobrir, integrar e analisar (DataONE, 2017). A Figura 1 ilustra esse ciclo.

Figura 1 – Ciclo de vida dos dados



Fonte: Extraído de Córdula e Araújo (2019, p. 198).

Mediante a Figura 1, pode-se perceber que o CVD “é um processo contínuo de desenvolvimento, manipulação, gerenciamento e estágios de armazenamento de dados” (Córdula; Araújo, 2019, p. 198). O CVD inicia-se pelo planejamento que, sendo um documento formal denominado planejamento de gerenciamento de dados (PGD), delimita o tratamento dos dados durante o ciclo, garantindo a preservação e a utilidade a longo prazo (Córdula; Araújo, 2019).

A *DataONE* procura disponibilizar o acesso a dados armazenados em repositórios parceiros, sendo um projeto comunitário que viabiliza pesquisas voltadas a descobertas de dados ambientais e da terra. Conforme Anjos e Dias (2019), as etapas do CVD da *DataONE* são interdependentes, sendo realizadas conforme as necessidades de cada contexto inserido (comunidade científica, pesquisador individual, grupo de pesquisadores, entre outros).

Iniciando-se na etapa de **planejar**, realiza-se um planejamento de como o processo vai ocorrer redigindo um documento formal denominado planejamento do gerenciamento de dados (PGD) (Córdula; Araújo, 2019), delimitando as etapas e metodologias necessárias para iniciar o ciclo (Strasser *et al.*, 2012). A **coleta** destina-se a pensar em onde e como os dados primários serão coletados. A etapa de **garantir** procura aderir a critérios que promovam a qualidade e validação dos dados, com o intuito de manter a autenticidade gerenciamento a longo prazo (Anjos; Dias, 2019).

A etapa de **descrever** ressalta o processo de descrição dos dados com base em padrões de metadados, procurando permitir a interoperabilidade pelos metadados permitindo a descoberta e o reuso desses dados (Anjos; Dias, 2019). A etapa de **preservar** torna necessário um plano de preservação dos dados a curto e longo prazo (Anjos; Dias, 2019). De acordo com Anjos e Dias (2019), a curto prazo destina-se ao estoque e preservação da coleta dos dados primários e a longo prazo procura viabilizar o reuso desses dados.

Na etapa de **descobrir**, é identificado outros conjuntos, dados ou repositórios que possam agregar valor no reuso por meio da interoperabilidade e descrição dos metadados (Anjos; Dias, 2019). A etapa de **integrar** acontece quando os dados são compatíveis e comparáveis, assim como a integração de conjuntos de dados, integrando o conjunto de dados primário do determinado CVD a outros recursos de dados (Strasser *et al.*, 2012). Essa integração, nesse sentido, possibilita novas investigações e análises. A última etapa, de **análise**, viabiliza a exploração, análise e visualização dos dados (Anjos; Dias, 2019).

Desse modo, o CVD possui um viés favorável ao compartilhamento e ao reuso dos dados mediante aos descritores, interoperabilidade, escolha de metadados e preservação. A importância disso vai ao encontro do potencial econômico voltado a esse compartilhamento e reuso (Silveira; Lacerda; Dias, 2024) visto que,

A ideia de que a reutilização de dados de pesquisa possibilita a economia de custos e esforços de pesquisa, começou a ganhar visibilidade com as discussões lideradas pela Organização para Cooperação e Desenvolvimento Econômico (OCDE), no Reino Unido, em meados da década de 1990, que recomendou o compartilhamento e o acesso aberto a dados de pesquisa com financiamento público dentro e entre comunidades de pesquisa (Silveira; Lacerda; Dias, 2024, p. 3).

Além da discussão voltada aos dados de pesquisa, o compartilhamento e reuso de dados também acontece com dados pessoais que são disponibilizados por pessoas ou por instituições em portais de transparência ou dados abertos disponibilizados pelo Governo Federal, por exemplo. O gov.br, por exemplo, “é um portal que reúne, em um só lugar, serviços para o cidadão e informações sobre a atuação do Governo Federal” (GOV.BR, [s. d.], *online*), sendo o Portal Brasileiro de Dados Abertos do Governo Federal. Esta pesquisa

se debruça a analisar o referido ambiente com base nos princípios FAIR utilizando a ferramenta FairDataBR.

A Ciência Aberta, oriunda do Movimento de Acesso Aberto, procura democratizar os conteúdos científicos viabilizando o acesso livre, sem embargos e de código aberto. Conforme a Rede Nacional de Ensino e Pesquisa (RNP, 2021, *online*), “Ciência Aberta é um termo guarda-chuva que abrange as práticas de abertura de dados seguindo os princípios da transparência e colaboração na Ciência”.

Com essa abertura de dados, os objetos digitais na *web* podem ser armazenados em Repositórios de Dados, os quais são “uma base de dados digital voltada para suprir a necessidade de armazenar, organizar e disponibilizar objetos digitais” (RNP, 2021, *online*). Esses Repositórios de Dados, dessa maneira, são abertos possuindo papel na democratização do conhecimento que, conforme Campos (2024, p. 45), é voltado para a “Inserção e inclusão de sujeitos na sociedade em rede quando pensa-se em acesso a conteúdo informacionais científicos”.

Sobre os dados abertos, é necessário que os detentores de dados os publiquem em plataformas de acesso aberto, levando isso para a sociedade em geral. Segundo o Jornal da Unesp (2022, *online*), a Ciência Aberta deve englobar o conhecimento científico aberto, as infraestruturas para essa abertura, o engajamento aberto dos atores sociais e o diálogo aberto com outros sistemas de conhecimento.

No Brasil, as iniciativas para a Ciência Aberta são também voltadas para os dados de pesquisa abertos. Porém, quando se pensa em dados disponibilizados pelo governo federal, no portal gov.br ou em dados disponibilizados pelo Portal da Transparência, entende-se que essa transparência de dados abertos de fácil acesso vai além dos dados de pesquisa. Este estudo se direciona a esses dados abertos em particular.

Para isso, os Princípios FAIR apresentam-se como uma das maiores iniciativas para tornar os dados visíveis e dando continuidade ao ciclo de vida dos dados. A habilitação do FAIR faz parte de um projeto da *European Commission* (2018), isto é, a Declaração de compromisso do projeto habilitando

FAIRData na Terra, no Espaço e nas Ciências Ambientais para depósito e compartilhamento de dados³.

O acrônimo FAIR indica **F**indable, **A**ccessible, **I**nteroperable e **R**eusable que, em tradução significam tornar os dados localizáveis, acessíveis, interoperáveis e reutilizáveis. Para o princípio *Findable*, a *European Commission* (2018, p. 19) desmembra em F1 - (meta) dados são atribuídos a um identificador globalmente exclusivo e persistente; F2 - os (meta)dados são descritos com metadados ricos (definidos por R1 no princípio *Reusable*); F3 - (meta)dados de forma clara e explícita incluem o identificador dos dados que descreve; F4 - (meta)dados são registrados ou indexados em um recurso pesquisável.

Conforme a *European Commission* (2018, p. 19), o princípio *Accessible* indica tem-se A1 - (meta) dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado; A1.1 - o protocolo é aberto, gratuito e universalmente implementável; A1.2 - o protocolo permite um procedimento de autenticação e autorização, quando necessário; A2 - os (meta) dados estão acessíveis, mesmo quando os dados não estão mais disponíveis. Para *Interoperable*, tem-se I1 - Os (meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento; I2 - (meta) dados usam vocabulários que seguem os princípios do FAIR; I3 - Os (meta) dados incluem referências qualificadas a outros (meta) dados (European Commission, 2018, p. 19).

Por último, para o princípio *Reusable*, tem-se R1 – meta (dados) são ricamente descritos com uma pluralidade de atributos precisos e relevantes; R1.1 - (meta) dados são liberados com uma licença de uso de dados clara e acessível; R1.2 - (meta) dados estão associados à proveniência detalhada; R1.3 - (meta) dados atendem aos padrões da comunidade relevantes ao domínio (European Commission, 2018, p. 19).

Segundo Campos, Dias e Sousa (2023), a adoção dos princípios FAIR facilita os avanços na ciência aberta, favorecendo a otimização dos serviços de informação, assim como para o gerenciamento de conjuntos de dados. Desse

³ Texto original: “*Enabling FAIR Data Project’s Commitment Statement in the Earth, Space and Environmental Sciences for depositing and sharing data*”.

modo, este estudo abarca a utilização dos Princípios FAIR, especificamente da Ferramenta FairDataBR para analisar os dados do Portal Brasileiro de Dados Abertos do Governo Federal.

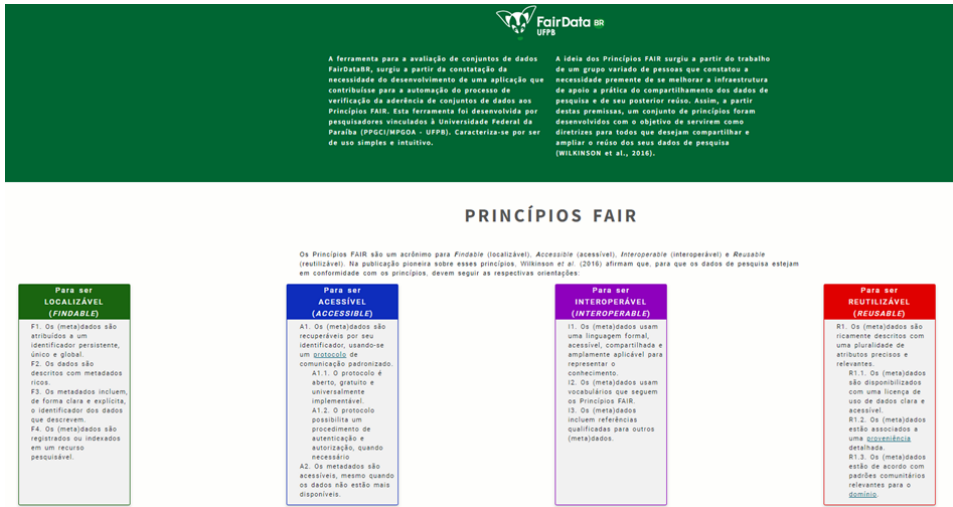
3 FERRAMENTA FAIRDATABR: UMA FERRAMENTA PARA A AVALIAÇÃO DE CONJUNTOS DE DADOS

Em conformidade com o objetivo deste estudo, que tem como foco principal a análise da adaptabilidade do conjunto de dados fornecido pelo Portal Brasileiro de Dados Abertos do Governo Federal, torna-se fundamental investigar o nível de alinhamento entre esse conjunto de dados e os princípios FAIR.

Para conduzir essa investigação, é necessária a utilização de uma ferramenta que permita medir esse alinhamento, de modo a identificar pontos de melhoria, os quais serão apresentados nos resultados deste estudo. Para esse propósito, foi selecionada a aplicação FairDataBR, que, de acordo com FairDataBR (2021), é uma ferramenta destinada à avaliação de conjuntos de dados, contribuindo para a automação do processo de verificação da aderência aos princípios FAIR.

O FairDataBR distingue-se por sua simplicidade e interface intuitiva, tendo sido desenvolvido por pesquisadores vinculados ao Programa de Pós-graduação em Ciência da Informação e ao Programa de Pós-graduação em Gestão nas Organizações Aprendentes da Universidade Federal da Paraíba (PPGCI/MPGOA - UFPB). O processo de avaliação realizado pela ferramenta envolve a verificação da conformidade do conjunto de dados com cada um dos princípios estabelecidos pelos princípios FAIR, conforme ilustrado na Figura 2.

Figura 2 – Itens analisados pela ferramenta FairDataBR



Fonte: Extraído de FairDataBR (2024)

Essa verificação ocorre de maneira individual, considerando cada princípio dos grupos estabelecidos pelo FAIR. Para viabilizar essa avaliação, é necessário preencher um questionário, cujas respostas são pontuadas ao final, resultando em um índice de aderência do conjunto de dados analisado aos princípios FAIR.

É relevante destacar que a ferramenta FairDataBR abrange questionamentos importantes para a compreensão do nível de adaptabilidade de uma estrutura de dados aos princípios FAIR. O questionário inclui tanto perguntas diretamente relacionadas aos princípios FAIR quanto outras questões igualmente relevantes para a adaptabilidade da base de dados analisada.

Ainda no que diz respeito à estrutura da ferramenta, é importante mencionar que o FairDataBR dispõe de uma versão padrão e uma versão estendida, esta última acessível por meio de um link "FairDataBR+" na tela inicial. A versão estendida contém um número maior de perguntas; no entanto, após análise do conteúdo de ambas as versões, constatou-se que as questões da versão padrão estão mais alinhadas aos objetivos deste estudo.

Como a versão padrão foi a escolhida, convém destacar que a numeração das questões pode não seguir uma sequência lógica, uma vez que algumas são exibidas apenas na versão estendida. Assim, na seção 4.1 são abordadas questões relacionadas ao uso da ferramenta FairDataBR no contexto do

conjunto de dados fornecido pelo Portal Brasileiro de Dados Abertos do Governo Federal.

4 PROCEDIMENTOS METODOLÓGICOS

Nesta seção, são apresentados os procedimentos utilizados para a construção deste estudo com base nas etapas de construção do conhecimento científico. Para a construção do referencial teórico, a pesquisa bibliográfica no Portal de Periódicos da CAPES⁴, na Base de Dados em Ciência da Informação (Brapci)⁵ e no Google Acadêmico⁶. O Portal de Periódicos da CAPES ajudou na recuperação de estudos em bases de dados internacionais principalmente e as demais fontes de informação ajudaram na recuperação de estudos nacionais, em periódicos científicos, livros e anais de eventos.

Não foi utilizado um período para buscar os materiais, visto que, a temática é corrente nos campos da Ciência da Informação e da Ciência dos Dados resultando majoritariamente em materiais publicados após os anos 2010. Os termos de busca utilizados foram: “Ciência Aberta AND Dados Abertos”; “Objetos digitais AND Princípios FAIR”; “Ciclo de Vida dos Dados”; “Princípios FAIR”; “*Open Science AND Open Data*”; “*FAIR Principles*”; “*Data Life Cycle AND FAIR Principles AND Open Data*”.

Este estudo possui o caráter exploratório visto que utiliza a Ferramenta FairDataBR para analisar o Portal Brasileiro de Dados Abertos do Governo Federal. Para Moresi (2003), a pesquisa exploratória é conduzida em uma área onde existe pouco conhecimento consolidado e organizado e, nesse contexto, ainda não foi realizado nenhum outro estudo que analisasse o referido portal com base na ferramenta supradita.

A natureza desta pesquisa é qualitativa destinando-se a averiguar problemas e investigar soluções no Portal Brasileiro de Dados Abertos do Governo Federal por meio da Ferramenta FairDataBR, voltada a avaliar

⁴ Disponível em: <https://www.periodicos.capes.gov.br>.

⁵ Disponível em: <https://brapci.inf.br/#/>.

⁶ Disponível em: <https://scholar.google.com>.

conjuntos de dados. A ferramenta opera por meio dos Princípios FAIR demonstrando as orientações informadas neste artigo e foi desenvolvida por pesquisadores da Universidade Federal da Paraíba.

4.1 Portal Brasileiro de Dados Abertos do Governo Federal

O Portal Brasileiro de Dados Abertos do Governo Federal é uma plataforma digital que centraliza e disponibiliza dados públicos de diversos órgãos e entidades governamentais, de forma acessível e aberta ao público.

De acordo com o Portal Brasileiro de Dados Abertos (2021), o portal foi criado em conformidade com a Política de Dados Abertos do Governo Federal, instituída pelo Decreto nº 8.777/2016, que estabelece a obrigatoriedade da abertura de dados governamentais com o objetivo de promover a transparência, estimular a participação social e incentivar o desenvolvimento de soluções tecnológicas e de pesquisa com base nas informações públicas disponíveis.

Ainda de acordo com Portal Brasileiro de Dados Abertos (2021), sua relevância consiste na sua capacidade de democratizar o acesso a informações governamentais, facilitando o acompanhamento e a fiscalização das ações do governo pela sociedade civil. Além de fomentar a transparência pública, o portal promove a inovação ao permitir que os dados sejam reutilizados para diferentes finalidades, como a criação de novos serviços, o desenvolvimento de pesquisas científicas e a análise de políticas públicas. Com isso, o portal contribui para o fortalecimento da governança pública e para o aumento da eficiência na administração dos recursos públicos.

Dentre os objetivos do portal evidencia-se, principalmente, a ampliação da transparência governamental, garantindo que informações públicas sejam acessíveis de forma aberta, possibilitando seu uso e redistribuição por qualquer interessado. O portal também contribui na participação social, por meio da disponibilização de um ambiente em que cidadãos, pesquisadores e empresas possam utilizar os dados para desenvolver soluções que gerem valor para a sociedade, além de promover a inovação tecnológica e o uso de dados abertos como base para a construção de ferramentas e serviços que atendam a demandas públicas e privadas.

5 RESULTADOS

Para a realização da análise de adaptabilidade do conjunto de dados disponibilizados pelo Portal Brasileiro de Dados Abertos do Governo Federal, aqui denominado de "dados abertos – gov.br", foi utilizada, como amostra, uma base de dados relativa aos aposentados sob o regime jurídico único do Poder Executivo Federal Civil. O portal de dados abertos distingue entre metadados e dados. A estrutura dos metadados é disponibilizada por meio de um arquivo em formato PDF, denominado "Dicionário de Dados" e os dados propriamente ditos são disponibilizados em formato CSV, identificados por título, mês e ano da coleta.

Para a presente análise, considerando que a ferramenta FairDataBR tem como foco principal a estrutura de metadados, o Dicionário de Dados da referida base e os elementos presentes na tela de cadastro do conjunto de dados no Portal Brasileiro de Dados Abertos do Governo Federal são utilizados como referência para a avaliação por meio dessa ferramenta. Espera-se, ao final do estudo, identificar o nível de adaptabilidade do conjunto de dados abertos do portal do governo federal aos princípios FAIR.

Ao iniciar a análise de adaptabilidade do conjunto de dados, "dados abertos – gov.br" aos princípios FAIR, o primeiro princípio a ser abordado é o *Findable* (Localizável), com os seguintes questionamentos:

1. Os conjuntos de (meta) dados possuem um identificador único, global e persistente?
2. Os conjuntos de dados são descritos com metadados?
3. O identificador está incluído em todos os registros/arquivos de metadados que descrevem os dados?
4. O recurso digital pode ser encontrado por mecanismos de busca na web?
5. O conjunto de (meta) dados está publicado em um repositório?

Em relação aos itens do princípio "Localizável", e considerando as opções de resposta da ferramenta FairDataBR, para o item 1, a resposta mais adequada é "Endereço Web". Isso se justifica pelo fato de o conjunto de dados utilizar um identificador único (*Uniform Resource Locator* - URL) para todos os elementos,

embora não seja um identificador persistente, como o *Digital Object Identifier*, de acrônimo DOI.

Quanto ao item 2, o conjunto de dados descreve cada campo com detalhes básicos sobre formato e tipo de dado, sem incluir metadados ricos. A descrição é direta e bem alinhada a aspectos técnicos, o que indica que a resposta adequada para o item 2 é "Metadados Estruturado".

No item 3, não foram encontradas evidências de um identificador em todos os registros ou arquivos de metadados. Como o documento não inclui identificadores persistentes nos dados descritos, a resposta mais adequada é "Não".

Quanto ao item 4, que analisa a capacidade de localização dos dados e metadados por ferramentas de busca na web, considerando que o conjunto de dados está no portal de dados abertos do governo, é possível afirmar que ele pode ser encontrado via mecanismos de busca na web. Portanto, a resposta mais apropriada é "Sim".

Por fim, o item 5, que analisa a publicação dos dados em um repositório público, conclui que, como o conjunto de dados está disponível no portal de dados abertos do governo, um repositório público acessível a todos, a resposta será "Público Geral".

Ao concluir a análise dos itens do princípio "Localizável", o nível de adaptabilidade do conjunto de dados "dados abertos – gov.br" com os princípios FAIR atingiu a nota de 7,00.

O próximo princípio a ser analisado é o *Accessible* (Acessível), cujos questionamentos são os seguintes:

1. Quão acessíveis são os (meta) dados?
2. Os (meta) dados estão acessíveis online sem a necessidade de intermediação de protocolos ou ferramentas especializadas, uma vez permitido o acesso?
3. É possível acessar o conjunto de (meta) dados por meio do identificador fornecido?
4. Os metadados estão disponíveis mesmo quando o conjunto de dados não estiver mais acessível?

5. O protocolo utilizado (ex.: HTTP, SAML, OAI-PMH) é aberto, gratuito e universalmente implementável?
6. É possível realizar o download dos conjuntos de (meta) dados?

O item 1 indaga sobre a acessibilidade dos metadados. Considerando que os dados estão disponíveis publicamente no portal de dados abertos, a resposta adequada é "Acessível ao Público".

O item 2 questiona se os (meta) dados estão acessíveis online sem a necessidade de ferramentas especializadas. Como os dados podem ser baixados em formato CSV/ZIP no portal, a resposta apropriada é "Download de arquivo online".

O item 3, que trata do acesso por meio de um identificador fornecido, tem como resposta "Não", uma vez que não há identificador persistente.

O item 4 examina a recuperação dos metadados após a exclusão dos dados. Como a exclusão dos dados não implica na indisponibilidade dos metadados, a resposta mais adequada é "Sim".

Os itens 5 e 6, que avaliam o protocolo utilizado e a possibilidade de download, têm como resposta "Sim", uma vez que o portal adota o padrão HTTP e permite o download dos dados.

Ao final da análise dos princípios de acessibilidade, a adaptabilidade do conjunto de dados, "dados abertos – gov.br" aos princípios FAIR obteve a nota 7,50.

O próximo princípio a ser analisado é o *Interoperable* (Interoperável), com os seguintes questionamentos:

1. Os conjuntos de (meta) dados estão disponíveis em formatos preferenciais?
2. Os dados estão estruturados a partir de um esquema de metadados aprovado pela comunidade?
3. Os (meta) dados estão vinculados a outros (meta) dados por meio de identificadores?
4. Foram utilizados recursos de controle terminológico para a representação dos dados?

Para o item 1, como o formato CSV é amplamente utilizado para

interoperabilidade, a resposta será "Sim".

No item 2, que analisa a estrutura de metadados, a resposta é "Comunidade Padrão", considerando a documentação oficial disponível. Essa escolha leva em consideração o padrão de metadados disponibilizado e amplamente utilizado para o cadastro de diversas bases de dados no âmbito dos portais de dados públicos do governo federal.

O item 3, que investiga a vinculação entre metadados, conclui que a melhor resposta é "Links Internos e Externos", pois há vínculos explícitos entre diversas bases de dados disponibilizadas no ambiente do Portal Brasileiro de Dados Abertos do Governo Federal.

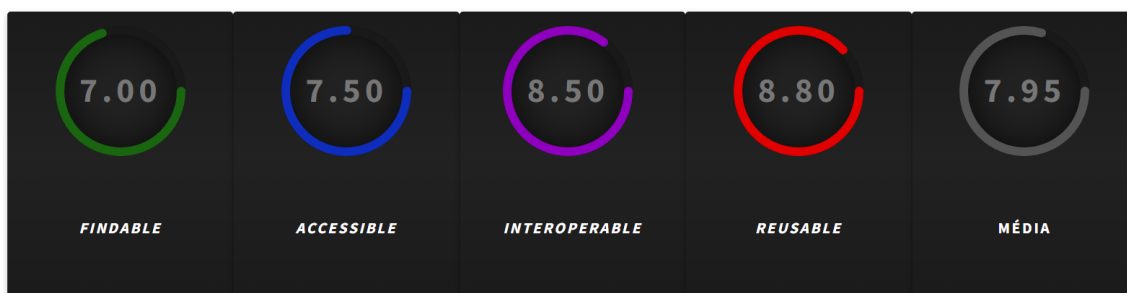
O item 4, que analisa o uso de controle terminológico, tem como resposta "Nenhum recurso/instrumento", uma vez que não há evidências do uso de tais ferramentas. A nota de adaptabilidade para os princípios de interoperabilidade foi de 8,50. Por fim, o princípio Reusable (Reutilizável) será analisado com base nos seguintes questionamentos:

1. Os (meta) dados estão licenciados?
2. Qual é a licença de uso dos (meta) dados?
3. Os (meta) dados possuem proveniência detalhada?
4. Há uma descrição da origem do fluxo de trabalho que produziu os (meta) dados?
5. Os (meta) dados seguem padrões relevantes para o domínio?

Para o item 1, a resposta é "Sim", pois os dados possuem licenças abertas, como as do Creative Commons. O item 2 confirma que a licença de uso é "aberta". O item 3, que trata da proveniência detalhada, tem como resposta "Sim", assim como o item 4, que questiona o rastreamento de proveniência. O item 5, que investiga o cumprimento de padrões relevantes, conclui que os metadados seguem "Padrões Genéricos". A nota de adaptabilidade para os princípios de reutilização foi de 8,80, a mais alta entre todos os princípios.

Concluindo, a análise geral de adaptabilidade do conjunto de dados "dados abertos – gov.br" aos princípios FAIR resultou na nota final de 7,95 conforme ilustrado na Figura 3.

Figura 3 – Média final gerada pela ferramenta FairDataBR



Fonte: Extraído de FairDataBR (2024)

Diante das análises realizadas para cada um dos princípios FAIR, foi possível observar o nível de conformidade dos dados abertos disponibilizados pelo Portal Brasileiro de Dados Abertos do Governo Federal em relação a esses padrões. A base de dados utilizada, relativa aos aposentados sob o regime jurídico único do Poder Executivo Federal Civil, demonstrou níveis variados de adaptabilidade, com resultados consistentes para os princípios *Findable* e *Accessible*, atingindo notas de 7,00 e 7,50, respectivamente. No entanto, obteve melhor desempenho nos princípios *Interoperable* e *Reusable*, com pontuações de 8,50 e 8,80.

A análise do princípio *Findable* revelou que, embora os dados possuam identificadores únicos como URLs, a ausência de identificadores persistentes, como o DOI, limita a localização global e a sustentabilidade de longo prazo dos dados. No princípio *Accessible*, constatou-se que o formato de disponibilização (CSV/ZIP) e o protocolo HTTP garantem acessibilidade pública e aberta, sem necessidade de ferramentas especializadas, embora a inexistência de identificadores persistentes tenha impactado negativamente esse quesito.

No princípio *Interoperable*, o uso do formato CSV como padrão amplamente aceito para a interoperabilidade de dados e o vínculo com outros conjuntos de dados via links internos e externos contribuíram para uma alta adaptabilidade. Contudo, não foi identificado o uso de controle terminológico, o que poderia melhorar a interoperabilidade semântica. Por fim, o princípio *Reusable* apresentou o maior grau de conformidade, com a adoção de licenças abertas e a descrição da proveniência dos dados, atendendo aos requisitos de reutilização e rastreamento.

Considerando os resultados obtidos, a nota final de adaptabilidade da base de dados analisada foi de 7,95, demonstrando que, embora existam oportunidades de melhoria, especialmente no que diz respeito à persistência dos identificadores e à implementação de controles terminológicos, o conjunto de dados analisado apresenta uma boa aderência aos princípios FAIR, contribuindo para o acesso, a interoperabilidade e a reutilização de dados públicos no Brasil.

6 CONSIDERAÇÕES FINAIS

O uso da ferramenta FairDataBR para a análise de adaptabilidade do Portal Brasileiro de Dados Abertos do Governo Federal aos princípios FAIR demonstrou-se altamente vantajoso, proporcionando uma avaliação estruturada e sistemática dos metadados e dados disponíveis. Entre os principais benefícios, destaca-se a capacidade da ferramenta em oferecer uma análise detalhada de cada um dos quatro princípios FAIR, permitindo identificar pontos fortes e oportunidades de melhoria na adequação dos conjuntos de dados públicos a esse padrão.

A FairDataBR permitiu, de maneira clara, identificar a utilização de identificadores únicos (como URLs), a acessibilidade via protocolos amplamente adotados (como o HTTP) e a existência de licenças abertas, aspectos fundamentais para assegurar a visibilidade, acessibilidade e reutilização dos dados. Além disso, a ferramenta foi essencial para evidenciar a ausência de identificadores persistentes e controles terminológicos, itens que poderiam aumentar a robustez e interoperabilidade dos dados analisados.

Outro ponto de destaque é a facilidade com que a FairDataBR permitiu classificar o grau de conformidade em cada princípio, gerando uma pontuação objetiva. Esse resultado foi crucial para quantificar a adaptabilidade e fornecer subsídios para futuras melhorias. Ao longo da análise, a ferramenta revelou-se uma importante aliada na promoção de transparência e de boas práticas no compartilhamento de dados públicos, contribuindo para o fortalecimento da governança de dados no Brasil e incentivando a implementação dos princípios FAIR em diferentes domínios.

Como sugestão de continuação do estudo, poderão ser feitas análises

com a ferramenta FairDataBR em outros ambientes que possuem dados abertos ou conjuntos de dados. É também possível analisar dados de pesquisa armazenados em Repositórios de Dados mediante a sua adaptabilidade aos princípios FAIR. Outra possibilidade é utilizar a versão estendida da referida ferramenta, isto é, a FairDataBR+ que trabalha com os princípios FAIR e com pressupostos em *findability*.

REFERÊNCIAS

ANJOS, Renata Lemos dos; DIAS, Guilherme Ataíde. Atuação dos profissionais da informação no ciclo de vida dos dados – DataONE: um estudo comparado. **Informação & Informação**, Londrina, v. 24, n. 1, 2019. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/31085>. Acesso em: 24 set. 2024.

CAMPOS, Arthur Ferreira. **Contribuições teóricas e práticas da Arquitetura da Informação para a gestão e avaliação de Repositórios Institucionais**. 2024. 146 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-graduação em Ciência da Informação, Universidade Federal da Paraíba, João Pessoa, 2024. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/31584>. Acesso em: 24 set. 2024.

CAMPOS, Arthur Ferreira; DIAS, Guilherme Ataíde; SOUSA, Marckson Roberto Ferreira de. Princípios FAIR, Encontrabilidade e Dados: por um diálogo teórico e uma aplicação prática. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [S. l.], v. 28, p. 1–21, 2023. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/88060>. Acesso em: 8 out. 2024.

CASTELLS, Manuel. **A sociedade em rede**. 17. ed. São Paulo: Paz e Terra, 2016.

CÓRDULA, Flávio Ribeiro; ARAÚJO, Wagner Junqueira de. O compartilhamento dos dados científicos na era do e-science. In: Dias, Guilherme Ataíde; Oliveira, Bernardina Maria Juvenal Freire de. (Organizadores). **Dados científicos: perspectivas e desafios**. João Pessoa: EDUFPB, 2019. Disponível em: <https://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/359>. Acesso em: 24 set. 2024.

EUROPEAN COMMISSION. **Turning FAIR into reality: Final Report and Action Plan on FAIR Data**. Luxembourg: Publications Office of the European Union, 2018. Disponível em:

https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf.
Acesso em: 08 out. 2024.

FAIRDATABR. **FairDataBR: uma ferramenta para a avaliação de conjuntos de dados**. [S. l.], 2021. Disponível em: <https://wrcu.ufpb.br/fair/index.html>.
Acesso em: 13 out de 2022.

GOV.BR. **O que é gov.br?** Portal GOV.BR, [2024]. Disponível:
<https://www.gov.br/pt-br/sobre#:~:text=É%20um%20portal%20que%20reúne,cidadão%20participa%20do%20gov.br%3F>. Acesso em: 24 set. 2024.

Docentes da Unesp colaboram com plano nacional para estimular Ciência Aberta no Brasil. **Jornal da UNESP**, Marília, 2022. Disponível em:
<https://jornal.unesp.br/2022/11/03/docentes-da-unesp-colaboram-com-plano-nacional-para-estimular-ciencia-aberta-no-brasil/>. Acesso em: 02 out. 2024.

LAU, Raymond; ZHAO, J. Leon; CHEN, Guoging; GUO; Xunhua. Big data commerce. **Information & Management**, [s.l.], n. 53, p. 929-933, 2016. Disponível em:
<https://www.sciencedirect.com/science/article/abs/pii/S0378720616300829>. Acesso em 24 set. 2024.

MORESI, Eduardo. (Org.). **Metodologia da pesquisa**. Brasília: Universidade Católica de Brasília, 2003. Disponível em:
<http://www.inf.ufes.br/~pdcosta/ensino/2010-2-metodologia-de-pesquisa/MetodologiaPesquisa-Moresi2003.pdf>. Acesso em: 08 out. 2024.

OPEN KNOWLEDGE FOUNDATION. **What is open?**. [2024]. Disponível em:
<https://okfn.org/en/library/what-is-open/>. Acesso em: 17 out. 2024.

REDE NACIONAL DE ENSINO E PESQUISA. **O que é Ciência Aberta e como ela pode facilitar a vida de cientistas**. [S. l.], 2021. Disponível em:
<https://www.rnp.br/noticias/o-que-e-ciencia-aberta-e-como-ela-pode-facilitar-vida-de-cientistas>. Acesso em: 02 out. 2024.

SANT'ANA, Ricardo César Gonçalves. Campo informacional resultante da interação de ciclos de vida dos dados. In: Dias, Guilherme Ataíde; Oliveira, Bernardina Maria Juvenal Freire de. (Org.). **Dados científicos: perspectivas e desafios**. João Pessoa: EDUEPB, 2019. Disponível em:
<https://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/359>. Acesso em: 24 set. 2024.

SILVA SEGUNDO, Sanderli José da; ARAÚJO, Wagner Junqueira de. Curadoria e ciclo de vida dos dados. In: Dias, Guilherme Ataíde; Oliveira, Bernardina Maria Juvenal Freire de. (Organizadores). **Dados científicos: perspectivas e desafios**. João Pessoa: EDUEPB, 2019. Disponível em:
<https://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/359>. Acesso em: 24 set. 2024.

SILVEIRA, Crislaine Zurilda; LACERDA, Rogério Tadeu de Oliveira; DIAS, Thiago Magela Rodrigues. Reuso de dados de pesquisa: uma revisão bibliométrica. **Palabra Clave (La Plata)**, [S. l.], v. 13, n. 2, p. e219, 2024. DOI: 10.24215/18539912e219. Disponível em: <https://www.palabraclave.fahce.unlp.edu.ar/article/view/PCe219>. Acesso em: 24 set. 2024.

STRASSER, Carly; COOK, Robert; MICHENER, William; BUDDEN, Amber. **Primer on Data Management**: What you always wanted to know. Califórnia: CDL, 2012. Disponível em: <http://escholarship.org/uc/item/7tf5q7n3#page-1>. Acesso em: 24 set. 2024.

WILKINSON, Mark *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Revista Nature**, California, v. 3, n. 1, p. 1-9, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 17 out. 2024.

ANALYSIS OF ADHERENCE TO FAIR PRINCIPLES IN OPEN DATA FROM THE BRAZILIAN FEDERAL GOVERNMENT WITH FAIRDATABR

ABSTRACT

Objective: This article aims to analyze adherence to the FAIR principles through the use of the FairDataBR tool in the context of the Federal Government's open data. **Methodology:** It has a qualitative characteristic while seeking to investigate problems and indicate solutions for the Brazilian Open Data Portal through the FairDataBR tool. It uses as sources of information the Capes Periodicals Portal, the Information Science Database (Brapci) and Google Scholar for bibliographic and theoretical support. Searches in these environments for terms related to the FAIR principles, open data, digital objects and the data life cycle, without setting an exact period for retrieving the materials. The FairDataBR tool is used to verify the FAIR adherence of the open data of the aforementioned portal. **Results:** According to the analysis, the FairDataBR tool generates an average of 7.95 for adherence to FAIR, with greater adherence to the reusable principle (8.80) and the lowest for the findable principle (7.00). For the accessible and interoperable principles, the tool generates scores of 7.50 and 8.50, respectively. **Conclusions:** the FairDataBR tool, used to analyze the adaptability of the Brazilian Open Data Portal of the Federal Government to the FAIR principles, provides a structured and systematic evaluation of the metadata and data available. It attributes FairDataBR as an ally in promoting transparency and good practices in the sharing of public data, contributing to the strengthening of data governance in Brazil and encouraging the implementation of the FAIR principles in different domains.

Descriptors: FAIR Principles. Open Data. FairDataBR Tool. Brazilian Open Data Portal.

ANÁLISIS DE ADHESIÓN A PRINCIPIOS JUSTOS EN DATOS ABIERTOS DEL GOBIERNO FEDERAL BRASILEÑO CON FAIRDATABR

RESUMEN

Objetivo: Este artículo tiene como objetivo analizar la adhesión a los principios FAIR mediante el uso de la herramienta FairDataBR en el contexto de datos abiertos del Gobierno Federal. **Metodología:** Tiene una característica cualitativa ya que busca investigar problemas e indicar soluciones para el Portal Brasileño de Datos Abiertos a través de la herramienta FairDataBR. Utiliza como fuentes de información el Portal de Periódicos de la Capes, la Base de Datos en Ciencias de la Información (Brapi) y Google Scholar para base bibliográfica y teórica. Busque en estos entornos términos relacionados con los principios FAIR, los datos abiertos, los objetos digitales y el ciclo de vida de los datos, sin marcar un período exacto para la recuperación de materiales. La herramienta FairDataBR se utiliza para comprobar la adherencia FAIR de los datos abiertos en el citado portal. **Resultados:** Según el análisis, la herramienta FairDataBR genera un promedio de 7,95 de adherencia a FAIR, con mayor adherencia al principio reutilizable (8,80) y menor al principio encontrable (7,00). Para los principios accesible e interoperable, la herramienta genera puntuaciones de 7,50 y 8,50 respectivamente. **Conclusiones:** la herramienta FairDataBR, utilizada para analizar la adaptabilidad del Portal de Datos Abiertos del Gobierno Federal a los principios FAIR, proporciona una evaluación estructurada y sistemática de los metadatos y los datos disponibles. Atribuye que FairDataBR es un aliado en la promoción de la transparencia y las buenas prácticas en el intercambio de datos públicos, contribuyendo a fortalecer la gobernanza de datos en Brasil y fomentando la implementación de los principios FAIR en diferentes dominios.

Descriptores: Principios FAIR. Datos abiertos. Herramienta FairDataBR. Portal brasileño de datos abiertos.

Recebido em: 30.10.2024

Aceito em: 22.11.2024