

LattesData e a adoção aos Princípios FAIR: uma análise usando a F-UJI Automated FAIR Data Assessment Tool

Adriane Groehs

Universidade do Estado de Santa Catarina, Florianópolis, SC, Brasil;
E-mail: adriane.groehs.bibliotecaria@gmail.com; ORCID: <https://orcid.org/0000-0003-0574-1134>

Crislaine Zurilda Silveira

Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil;
Email: crislaine.silveira@ufsc.br; ORCID: <https://orcid.org/0000-0003-3081-9968>

Elaine Maria de Lima

Universidade do Estado de Santa Catarina, Florianópolis, SC, Brasil;
E-mail: elainelimaedf@gmail.com; ORCID: <https://orcid.org/0000-0002-0831-0519>

Eliane Dittrich

Universidade do Estado de Santa Catarina, Florianópolis, SC, Brasil
E-mail: eliane.dittrich@udesc.br; ORCID: <https://orcid.org/0000-0001-8013-1921>

Igor Yure Ramos Matos

Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil;
E-mail: igor.yure@ufsc.br; ORCID: <https://orcid.org/0000-0001-7230-5100>

Divino Ignacio Ribeiro Junior

Universidade do Estado de Santa Catarina, Florianópolis, SC, Brasil;
E-mail: divino.ribeiro@udesc.br; ORCID: <https://orcid.org/0000-0002-9705-6507>

Resumo: As agências de fomento dos países desenvolvidos começaram a exigir o depósito, em acesso aberto, dos dados de pesquisa, para isso eles precisam cumprir exigências que permitam sua integridade, completude, qualidade e correta descrição dos seus metadados. Surgem assim, os Princípios FAIR - *Findability, Accessibility, Interoperability, and Reusability*, diretrizes para melhorar a reutilização dos dados de pesquisa. Com o objetivo de contribuir para essas discussões, este artigo buscou verificar a adoção dos Princípios FAIR nos *datasets* depositados no repositório LattesData, a partir do uso da ferramenta F-UJI. Essa ferramenta foi desenvolvida para automatizar o trabalho de análise das métricas do FAIR. Do ponto de vista metodológico, esta é uma pesquisa de caráter descritivo, na qual foram analisados 13 *datasets* disponíveis no LattesData. Os resultados obtidos permitem observar que o princípio da Encontrabilidade foi o melhor contemplado nos *datasets*, os princípios da Acessibilidade, da Interoperabilidade e da Reutilização, carecem de atenção, pois nem todas as métricas foram contempladas. Conclui-se que o repositório LattesData tem adotado os Princípios FAIR, ainda que de maneira incipiente. Necessita de políticas e diretrizes que garantam que os dados depositados

possam ser reutilizados. A F-UJI se mostrou uma ferramenta prática e simples de ser utilizada. O uso de ferramentas para a gestão de dados de pesquisa pode facilitar na implementação de boas práticas e no depósito de conjunto de dados com qualidade.

Palavras-chave: dados de pesquisa; princípios FAIR; repositório de dados

1 Introdução

As discussões em torno dos dados de pesquisa vêm ganhando força nos campos científicos motivadas por iniciativas de organizações internacionais, tais como a Organisation for Economic Co-operation and Development (OECD). Essas discussões culminaram na publicação dos “Princípios e Diretrizes para acesso a dados de pesquisa de Financiamento público” (em inglês *Principles and Guidelines for Access to Research Data from Public Funding*), em 2007 (OECD, 2007).

Diversas agências internacionais de fomento à pesquisa passaram a requerer, como critério para a concessão de recursos, o comprometimento por parte dos pesquisadores que os dados derivados das pesquisas fossem gerenciados e arquivados em repositórios, a fim de garantir preservação a longo prazo e maior facilidade de compartilhamento (Bertin; Visoli; Drucker, 2017).

Embora as discussões em torno do compartilhamento dos dados pareçam estar sedimentadas por parte das organizações e agências de financiamento, nas áreas do conhecimento ainda há discussões sobre o que de fato seriam dados de pesquisa.

De acordo com Bertin, Visoli e Drucker (2017, p. 38) “[...] dados de pesquisa, por sua vez – e também de modo simplificado –, são todo o tipo de registro produzido, compilado ou utilizado no decorrer da pesquisa”. Esse conceito tem uma amplitude de significados que varia de acordo com a área do conhecimento, o contexto e o objeto da pesquisa, as metodologias usadas na geração e na coleta e até mesmo na sua destinação ou finalidade (Bertin; Visoli; Drucker, 2017; Sayão; Sales, 2020).

Em busca de um conceito de dados de pesquisa que vise representar suas especificidades em diferentes áreas do conhecimento, Sayão e Sales (2020, p. 32) cunharam a seguinte definição:

[...] dado de pesquisa é todo e qualquer tipo de registro coletado, observado, gerado ou utilizado no âmbito da pesquisa científica, que pode ser interpretado, tratado e **aceito como evidência pela comunidade científica** e necessário para analisar, validar e produzir resultados de pesquisa.

Conforme pontuado por Sayão e Sales (2020), a condição para ser um dado de pesquisa, é que ele precisa ser aceito como **evidência pela comunidade científica**, isso leva a reflexões do tipo: todos os dados devem ou não ser compartilhados? Eles podem ser reutilizados para outras pesquisas em diferentes áreas do conhecimento?

Chauvette, Schick-Makaroff e Molzahn (2019) afirmam que nem todos os dados de pesquisa qualitativa são passíveis de compartilhamento, pois há questões epistemológicas, metodológicas, legais e éticas que podem trazer implicações significativas para a qualidade da análise e interpretação nos estudos.

Para Childs *et al.* (2014) os dados de pesquisa podem ser compartilhados e reusados por outras áreas do conhecimento, mas as decisões sobre o seu compartilhamento, precisam ser tomadas no estágio projeto de pesquisa, pois afetam as escolhas metodológicas e a aprovação ética, portanto, as pessoas mais adequadas para tomar essas decisões sobre quais dados são apropriados para reter e compartilhar, são os próprios pesquisadores em negociação com os participantes da pesquisa.

Apesar dessas discussões, o fato é que as práticas de compartilhamento vêm sendo discutidas e em muitos casos exigidas por periódicos e agências de fomento à pesquisa. Portanto, o alcance da sustentabilidade e da competitividade no sistema científico moderno, perpassa pelo compromisso das instituições e dos pesquisadores em garantir o apropriado gerenciamento e a preservação dos

dados de pesquisa, de modo a possibilitar a verificação futura de resultados e a reutilização dos dados originais (Bertin; Visoli; Drucker, 2017).

Assim como não há uma definição unívoca para dados de pesquisa, também não há uma definição precisa para o termo reutilização. Seu entendimento varia entre disciplinas e até mesmo indivíduos e nenhum padrão comum parece ser aplicado ainda (Van de Sandt *et al.*, 2019). No caso deste artigo a definição de reutilização considera o uso de qualquer recurso de pesquisa independentemente de quando é usado, a finalidade, as características dos dados e seu usuário (Van de Sandt *et al.*, 2019).

Dentre as vantagens para reuso de dados destacam-se economia de tempo na amostragem, coleta e processamento dos dados de pesquisa, aquisição de dados de participantes de difícil acesso, redução da carga de pesquisa sobre os participantes e maximização do valor de suas contribuições para pesquisas em diferentes campos de conhecimento (Chatfield, 2020).

No entanto, a disponibilidade dos dados de pesquisa depende da infraestrutura para torná-los descobertos, recuperáveis, interpretáveis e utilizáveis, da documentação de metadados de proveniência dos dados, do fornecimento de informações contextuais necessárias para a sua interpretação e o uso de licenças podem ser cruciais na superação dos desafios inerentes a sua reutilização (Bhattacharya *et al.*, 2018; Borgman; Scharnhorst; Golshan, 2019; Childs *et al.*, 2014).

Considerando a necessidade de melhorar a infraestrutura que apoia a reutilização de dados de pesquisa, bem como contribuir para que as máquinas possam encontrá-los e usá-los automaticamente, além de possibilitar que os indivíduos os compreendam, foram criados os Princípios FAIR (*Findability, Accessibility, Interoperability, and Reusability*) (Wilkinson *et al.*, 2016). Traduzidos para o português, esses princípios significam que os dados devem ser *Encontráveis, Acessíveis, Interoperáveis e Reutilizáveis*. Esses princípios definem características para ferramentas, vocabulários e infraestruturas que auxiliam na descoberta e reutilização de dados de pesquisa por terceiros (Wilkinson *et al.*, 2016).

No Brasil, mais precisamente na Ciência da Informação, ainda são incipientes os relatos na literatura científica de trabalhos que exploram a adoção dos Princípios FAIR a conjuntos de dados (ou *datasets*) de pesquisa depositados em repositórios de dados de maneira prática, conforme busca realizada na busca Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci)¹.

Os trabalhos encontrados foram de Felipe e Santos (2022), cujo objetivo foi avaliar, com base nos metadados FAIR Data Point, a qualidade dos metadados do Repositório de Dados de Estudos Ecológicos (PPBIO); o trabalho de Bonetti e Arakaki (2022) cujo objetivo foi avaliar a conformidade dos *datasets* depositados no repositório da Universidade Federal de São Carlos (UFSCar) quanto aos 4 princípios FAIR e o trabalho de Rodrigues, Dias e Lourenço (2022) cujo objetivo foi investigar a gestão e curadoria dos conjuntos de dados de pesquisa disponibilizados nos repositórios digitais institucionais sul-americanos à luz dos princípios FAIR.

Nesse sentido, com o objetivo de contribuir para essas discussões, este artigo buscou verificar a adoção dos princípios FAIR nos *datasets* depositados no repositório LattesData, a partir do uso de uma ferramenta automática denominada F-UJI Automated FAIR Data Assessment Tool².

O repositório LattesData foi lançado ainda no primeiro semestre de 2022, foi desenvolvido por meio da parceria entre Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Seu objetivo é “reunir, armazenar e divulgar os conjuntos de dados científicos de pesquisadores beneficiários CNPq” (CNPq, c2022).

Nesse sentido considerando que o lançamento do repositório é recente e considerando que ele busca reunir os dados de pesquisa dos pesquisadores que recebem fomento do CNPq, portanto, dados advindos de diversas áreas do conhecimento e em diversos formatos, entende-se que esse trabalho pode trazer como contribuição um diagnóstico ainda que inicial para os gestores do LattesData.

2 Dados de pesquisa

Os dados de pesquisa são todas as evidências que são coletadas para subsidiar um estudo, testar hipóteses, descobrir novos produtos, materiais, medicamentos, etc. De acordo com Pavão *et al.* (2019, p. 7) definem os dados de pesquisa como “[...] dados coletados, observados ou produzidos durante a pesquisa (números, textos, imagem, som, saídas de equipamentos) para fins de análise e produção de resultados de pesquisa originais”.

Eles estão presentes em todas as áreas do conhecimento e são o combustível para o desenvolvimento de técnicas e tecnologias que beneficiam a vida no planeta, o progresso da humanidade, maior expectativa e qualidade de vida, melhoria dos deslocamentos e formas de transporte, disponibilidade de informações, desenvolvimento de aparelhos tecnológicos, entre outros, ou seja, sem as descobertas científicas o mundo ficaria paralisado:

Dados de pesquisa são os materiais comumente registrados e aceitos na comunidade científica como necessários para validar os resultados de pesquisa [...] e incluem: fatos e estatísticas recolhidas para posterior referência ou análise, documentos (texto, Word), planilhas (Excel, etc.), cadernos de laboratório, cadernos de campo, diários, questionários, transcrições, fitas de áudio, fitas de vídeo, fotografias, filmes, sequências de proteínas ou genéticos, respostas de teste, slides, artefatos, amostras, coleção de objetos digitais adquiridos e gerados durante o processo de pesquisa, conteúdos de banco de dados (vídeo, áudio, texto, imagens), modelos, algoritmos, scripts, arquivos de log, software de simulação, metodologias e fluxos de trabalho, procedimentos operacionais, padrões e protocolos (Dudziak, 2016).

Para Silva (2019, p. 3) os dados são valiosos porque, “[...] se o conhecimento é o motor do avanço científico, os dados são seu combustível”. Em termos de carga semântica, o dado é um elemento desprovido de significado, algo de difícil interpretação quando examinado fora de um contexto. Segundo Schreiber *et al.* (2000, p. 3):

[...] os dados são o sinal não interpretado que alcança nossos sentidos, *zilhões* a cada minuto. Por exemplo, uma luz vermelha, verde ou amarela em um cruzamento. Os computadores estão cheios de dados: os sinais elétricos dos circuitos digitais são representados

por sequências de números, caracteres e outros símbolos que são manuseados em grandes quantidades.

Nessa definição, os autores apresentam o dado como um objeto identificado a partir da percepção humana e que são operáveis cognitivamente ou computacionalmente, com um único fim: produzir informação. Assim, sua utilização requer um processo anterior de agregação de valor semântico, ou seja, um processo que permita a reunião, processamento, contextualização e a consequente interpretação, para que o dado faça sentido e se transforme em informação útil e reutilizável.

Segundo Sayão e Sales (2015) os dados de pesquisa podem incluir: números, imagens, textos, vídeos, áudio, software, algoritmos, equações, animações, modelos, simulações. Nesse contexto, eles deixam de ser simples subprodutos das atividades de pesquisa e se tornam recursos informacionais de primeira grandeza, caracterizando um novo paradigma científico pautado pelo compartilhamento, amplo acesso e reúso (Sayão; Sales, 2016). Portanto, devido ao fenômeno da *e-Science*, que se caracteriza pela intensa produção de dados a partir de instrumentos tecnológicos, os dados de pesquisa passam a ser motores de grande potencial de inovação, com isso, a sua gestão pode favorecer o reúso por diferentes pesquisadores e de diferentes áreas do conhecimento.

2.1 Princípios FAIR

Os princípios FAIR são diretrizes para melhorar a reutilização dos dados de pesquisa. Estes orientam ações que permitem que os dados sejam encontrados, acessíveis, interoperáveis e reusados por máquinas, sistemas e outros softwares, sem necessariamente a intervenção humana:

[...] os princípios FAIR, um acrônimo para Findable, Accessible, Interoperable e Reusable, apresentam-se como elementos norteadores ao processo de gestão de dados de pesquisa. Estes princípios vêm sendo amplamente difundidos por comunidades internacionais nos últimos anos e, recentemente chegando ao Brasil como mais uma vertente de estudos e aplicação por gestores de dados. As discussões têm sido encaminhadas com o intuito de proporcionar o grau máximo de reuso de dados científicos, a partir

da adoção de padrões, metadados, vocabulários controlados, ontologias e identificadores persistentes que proporcionam significado preciso aos dados e aos demais objetos a eles vinculados (Henning *et al.*, 2019, p. 177).

Para que os dados de pesquisa sejam descobertos tanto por softwares de busca de dados (que serão comuns num futuro próximo), quanto por outros pesquisadores, isto é, para que os dados sejam **encontráveis**, eles precisam cumprir alguns requisitos como: o uso de identificadores persistentes como o *Digital Object Identifier* (DOI), o uso de metadados com padrão internacional, que atendam ao protocolo OAI-PMH, descrição correta dos metadados descritivos e a indexação dos repositórios de dados em ferramentas como o *Registry of Research Data Repositories* (Re3data), no quadro 1 estão sintetizados o requisitos relativos a encontrabilidade dos conjuntos de dados.

Quadro 1 - FAIR Findability

FINDABILITY - [ENCONTRÁVEL]	
F1	(Meta)dados são atribuídos a um identificador globalmente exclusivo e persistente
F2	Os dados são descritos com metadados avançados (definidos por R1 abaixo)
F3	Os metadados incluem clara e explicitamente o identificador dos dados que descrevem
F4	(Meta)dados são registrados ou indexados em um recurso pesquisável

Fonte: Elaborado pelos autores com base no Fair Principles ([2022]).

Os requisitos ligados a acessibilidade estão relacionados com os metadados que devem ser recuperáveis utilizando um protocolo de comunicação aberto, gratuito, implementados universalmente, como: o padrão Transmission Control Protocol (TCP), o File Transfer Protocol (FTP) e o Hypertext Transfer Protocol (HTTP) (conforme A1, A.1.1 e A1.2). E exige-se que os metadados estejam acessíveis e abertos mesmo que os dados estejam embargados, ou aguardando a sua publicação ou sejam dados sensíveis e disponíveis só a determinadas pessoas pertencentes a algum órgão/departamento/laboratório

(requisito A2). No quadro 2 estão sintetizados os requisitos relativos a acessibilidade.

Quadro 2 - FAIR Accessibility

ACCESSIBILITY [ACESSÍVEL]	
A1	(Meta)dados são recuperáveis por 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 metadados são acessíveis, mesmo quando os dados não estão mais disponíveis

Fonte: Elaborado pelos autores com base no Fair Principles ([2022]).

A interoperabilidade significa que os metadados podem ser acessados, reconhecidos e compartilhados de forma automática entre diferentes sistemas e softwares, para isso segundo o Fair Principles ([2022]) “[...] é fundamental usar (1) vocabulários controlados, ontologias, tesouros comumente usados (tendo identificadores globalmente únicos e persistentes resolvíveis” (veja F1) e “(2) um bom modelo de dados (uma estrutura bem definida para descrever e estruturar (meta)dados)”. No quadro 3 estão sintetizados os requisitos relativos a interoperabilidade dos dados.

Quadro 3 - FAIR Interoperability

INTEROPERABILITY [INTEROPERÁVEL]	
I1	(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 FAIR
I3	(Meta)dados incluem referências qualificadas a outros (meta)dados

Fonte: Elaborado pelos autores com base no Fair Principles ([2022]).

Para que os dados sejam reutilizados é necessário que os metadados estejam bem descritos e detalhados, entendíveis tanto por homens quanto por máquinas, os metadados devem fornecer informações se o conjunto de dados é útil ou não. Os metadados podem:

[...] incluir os protocolos experimentais, o fabricante e a marca da máquina ou sensor que criou os dados, as espécies usadas, o regime de medicamentos etc. Além disso, o R1 afirma que o editor de dados não deve tentar prever a identidade e as necessidades do consumidor de dados. Escolhemos o termo 'pluralidade' para indicar que o autor dos metadados deve ser o mais generoso possível ao fornecer metadados (Fair Principles, [2022]).

A licença atribuída permite que se diga o que as pessoas podem fazer com os dados, a licença deve estar descrita claramente, para ser entendida por pessoas e máquinas (R1.1 é sobre interoperabilidade legal). Para serem reutilizados, seus dados precisam ser descritos contando a história dos dados, origem, fluxos de coleta, quem coletou, como foi processado (conforme R1.2). Outro critério presente para o reúso é atender as especificações de arquivamento e compartilhamento e padrões da comunidade científica (R1.3). Segundo o Fair Principles (2022, p. 1) é mais fácil reutilizar conjuntos de dados semelhantes: mesmo tipo de dados, organizados de forma padronizada, formatos de arquivo bem estabelecidos e sustentáveis, documentação (metadados) seguindo um modelo comum e usando vocabulário comum:

[...] Além de ser um dos quatro princípios FAIR, o reúso é, ainda, a finalidade dos processos de curadoria de dados. Como a principal característica dos princípios FAIR é a oferta de um conjunto de orientações concisas, de alto nível, que valem para qualquer domínio e que devem ser aplicadas não somente aos dados, mas também aos metadados, aos identificadores, ao software e aos planos de gestão de dados, então eles se apresentam como facilitadores e orientadores do reúso. (Caregnato; Rocha; Gabriel Júnior, 2021, p. 198)

No quadro 4, estão sintetizados os requisitos relativos a reusabilidade dos dados.

Quadro 4 - FAIR Reuse

REUSE [REUSÁVEL]	
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 a proveniência detalhada
R1.3	(Meta)dados atendem aos padrões da comunidade relevantes para o domínio

Fonte: Elaborado pelos autores com base no Fair principles ([2022]).

Para Caregnato, Rocha e Gabriel Junior (2021, p. 200) “[...] todo compartilhamento de dados de pesquisa pressupõe a sua reutilização para benefício da própria ciência, da comunidade científica e da sociedade em geral, ou seja, na perspectiva do ecossistema de pesquisa”. Assim, a adoção dos princípios FAIR permite que esses dados tenham qualidade para o reuso por outros pesquisadores, os princípios são orientações práticas para que os dados ganhem significado, sejam interpretados, descritos com metadados ricos e que agregam valor, qualidade.

3 Procedimentos metodológicos

Trata-se de uma pesquisa de caráter descritivo, cujo objetivo foi verificar a adoção dos princípios FAIR nos *datasets* depositados no repositório LattesData.

O repositório em questão foi desenvolvido usando a plataforma Dataverse, trata-se de uma arquitetura de software livre para a publicação, citação, análise, preservação e reuso dos dados de pesquisa (Araújo; Mardero Arellano; Ferrer, 2018). No caso dessa pesquisa, ela se concentrou em verificar se os conjuntos de dados depositados no LattesData obedecem aos princípios FAIR.

Os propósitos do LattesData são:

[...] incentivar a criação de repositórios de dados de pesquisa, servindo como repositório modelo para tal, além de orientar a aplicação de boas práticas internacionais a repositórios de dados de

pesquisa; funcionar como ferramenta de disponibilização de dados de pesquisa, promover a transparência e possibilitar o reuso desses dados, aumentando assim sua reprodutibilidade e, por consequência, fazer crescer confiança e solidez das pesquisas envolvidas, o que por sua vez ajuda a maximizar o retorno de aplicação de recursos públicos; viabilizar a preservação digital de dados de pesquisa, colaborando para evitar sua perda e/ou má gestão, orientando e incentivando os autores na organização e planejamento do ciclo de vida dos dados de pesquisa, além de possibilitar a aplicação adequada dos princípios FAIR (Findable - Encontrável, Accessible - Acessível, Interoperable - Interoperável, Reusable - Reutilizável); tornar os dados citáveis e possibilitar o uso de métricas, valorizando e dando visibilidade a outros produtos da pesquisa ao contrário de unicamente seu resultado (CNPq, c2022).

Considerando que o LattesData se propõe a ser um repositório de referência para o depósito de *datasets* pelos pesquisadores financiados pelo CNPq e para outras instituições, consideramos importante analisar se os conjuntos de dados já depositados estão cumprindo os princípios FAIR.

Foram coletados 13 *datasets*, em 26 de julho de 2022, os assuntos cobriram as áreas da Medicina, Saúde e Ciências da Vida. Os arquivos de dados vinculados a esses *datasets* somam 55, sendo imagens (na extensão .jpg e .png), documentos (na extensão .xls), dados tabulares (na extensão .tab) e dados textuais (na extensão .csv). Todos os arquivos estavam disponíveis para acesso público.

A análise da aderência aos princípios FAIR foi verificada a partir de relatórios extraídos da F-UJI Automated FAIR Data Assessment Tool, é um serviço *web*, em desenvolvimento, com o objetivo de avaliar automaticamente a adesão ao FAIR de objetos de dados de pesquisa com base em métricas desenvolvidas pelo projeto FAIRsFAIR. (Devaraju; Huber, 2020). A avaliação é baseada em 16 das 17 principais métricas, as quais correspondem a uma parte ou à totalidade de um princípio FAIR. (Devaraju *et al.*, 2022). No quadro 5 na primeira coluna estão os Princípios FAIR, na segunda coluna estão as métricas e na terceira o seu significado.

Quadro 5 - Métricas FAIR

	Métricas	Significado
Encontrabilidade	FsF-F1-01D	Os dados recebem um identificador exclusivo
	FsF-F1-02D	Os dados recebem um identificador persistente
	FsF-F2-01M	Os metadados incluem para localização
	FsF-F3-01M	Os metadados incluem o identificador dos dados
	FsF-F4-01M	Os metadados são recuperados por máquinas.
Acessibilidade	FsF-A1-01M	Os metadados contêm o nível de acesso e as condições de acesso
	FsF-A1-02M	Os metadados são acessíveis por meio de um protocolo de comunicação padronizado
	FsF-A1-03D	Os dados são acessíveis por meio de um protocolo de comunicação padronizado
	FsF-A2-01M*	Os metadados permanecem disponíveis, mesmo que os dados não estejam mais disponíveis
Interoperabilidade	FsF-I1-01M	Os metadados usam linguagem de representação de conhecimento
	FsF-I1-02M	Os metadados usam recursos semânticos.
	FsF-I3-01M	Os metadados incluem links entre os dados e entidades relacionadas
Reusabilidade	FsF-R1-01MD	Os metadados especificam o conteúdo dos dados.
	FsF-R1.1-01M	Os metadados incluem informações de licença para reutilização
	FsF-R1.2-01M	Os metadados incluem informações de proveniência
	FsF-R1.3-01M	Os metadados seguem um padrão recomendado pela comunidade de pesquisa
	FsF-R1.3-02D	Os dados estão disponíveis em um formato de arquivo recomendado pela comunidade

Fonte: Adaptado de Devaraju *et al.* (2022).

Nota: * Essa métrica não é avaliada pela ferramenta F-UJI, pois o tratamento de dados excluídos e registros de metadados depende das políticas e procedimentos do repositório Devaraju *et al.* (2022).

Cada métrica FAIR, possui uma pontuação, que indica seu nível de conformidade. Os níveis de conformidades variam de incompleto (0), inicial (1), moderado (2) e avançado (3). Com isso, esses níveis são calculados pela F-UJI da seguinte forma: é realizado o cálculo para cada princípio FAIR, a partir da pontuação recebida por cada métrica. Caso pelo menos uma métrica de um determinado princípio tenha atingido o nível “inicial”, o princípio correspondente não pode ter um nível inferior, mesmo que todas as outras métricas tenham uma pontuação inferior. Os demais níveis são calculados a partir da média arredondada de todas as pontuações de métricas para um determinado princípio. Portanto, o nível F-UJI FAIR geral para um determinado objeto de dados de pesquisa é calculado de forma análoga como uma média arredondada de todas as pontuações obtidas no cálculo das métricas daquele princípio (Robert *et al.*, 2021).

No repositório LattesData foram identificados os *Digital Object Identifier* (DOI) de cada um dos 13 *datasets* depositados. Em seguida, cada DOI foi inserido individualmente na F-UJI, que de maneira automática, gerou 13 relatórios indicando quais métricas foram contempladas ou não, para cada *dataset*. Desta forma, foi possível identificar quais princípios foram atendidos ou quais precisam ser melhorados.

Os dados coletados foram sintetizados numa planilha do software Microsoft Excel, onde foram registradas as variáveis relacionadas a cada princípio FAIR, os *datasets* e os níveis de conformidade.

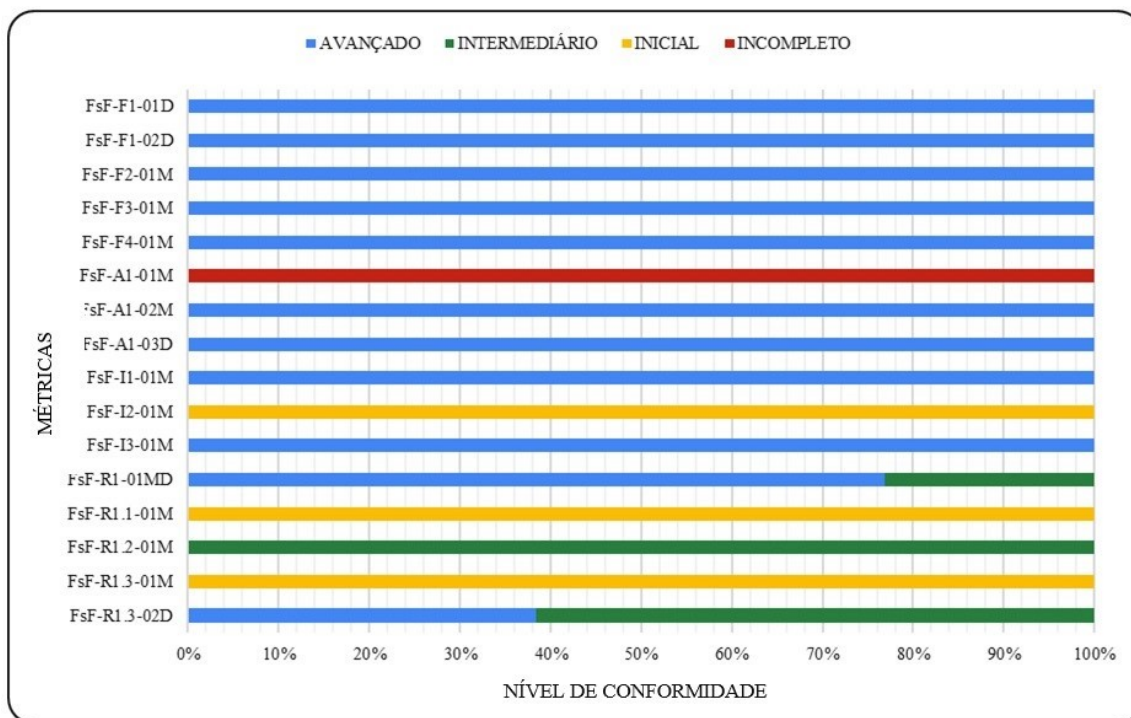
4 Resultados e discussões

A aplicação da ferramenta F-UJI para análise dos *datasets* do repositório LattesData possibilitou identificar que vários princípios FAIR já estão contemplados.

No Gráfico 1, realizou-se uma síntese de cada uma das métricas mencionadas no Quadro 5 com o propósito de apresentar o grau de adequação

dos *datasets*. No eixo vertical estão arroladas as métricas, enquanto o nível de conformidade de cada métrica é apresentado no eixo horizontal.

Gráfico 1 - Níveis de adequação dos Princípios FAIR



Fonte: Dados da pesquisa.

No gráfico 1 é possível observar que as métricas (1)FsF-F1-01D, (2)FsF-F1-02D, (3)FsF-F2-01M, (4)FsF-F3-01M e (5)FsF-F4-01M, que representam o **Princípio da Encontrabilidade**, foram atendidas, por 100% (13) dos *datasets*, demonstrando grau avançado de conformidade.

As métricas relacionadas ao Princípio da Encontrabilidade verificam (1) se os dados são atribuídos a um identificador globalmente exclusivo, (2) se os dados recebem um identificador persistente, (3) se os metadados incluem elementos centrais descritivos (criador, título, identificador de dados, editor, data de publicação, resumo e palavras-chave) para apoiar a localização dos dados, da mesma forma, verifica (4) se os metadados incluem o identificador dos dados que descrevemos e (5) se os metadados são disponibilizados de forma que possam ser recuperados programaticamente (Devaraju; Huber, 2021).

Neste contexto, fica evidente que os *datasets* analisados atingiram um nível avançado de conformidade para o **Princípio da Encontrabilidade**, garantindo a acessibilidade e a identificação eficaz desses dados.

No que diz respeito ao **Princípio da Acessibilidade**, considerando as métricas FsF-A1-01M, que avalia o grau de acesso tanto aos metadados quanto aos dados, observou-se que 100% (13) dos *datasets* apresentaram um nível de conformidade incompleto de acordo com os parâmetros estabelecidos para o princípio. Em contrapartida, quanto às métricas FsF-A1-03D, que verifica a acessibilidade dos dados e FsF-A1-02M, que avalia se os metadados podem ser acessados por meio de um protocolo de comunicação padronizado, 100% (13) dos *datasets* demonstraram um nível avançado de conformidade com esse princípio (Devaraju; Huber, 2021).

Quanto ao **Princípio da Interoperabilidade**, as métricas FsF-I1-01M, que verifica se os metadados utilizam uma linguagem de representação do conhecimento, e a métrica FsF-I3-01M, que verifica se os metadados incluem links entre os dados e as entidades relacionadas, foram atendidas por 100% (13) dos *datasets*, alcançando o nível avançado de conformidade com o FAIR. Contudo, o princípio FsF-I2-01M, que verifica se os metadados usam recursos semânticos como ontologia, tesauro e taxonomias para garantir a interpretação inequívoca dos assuntos e termos utilizados na representação da informação, ainda precisa ser aprimorada nos *datasets*, uma vez que 100% (13), atingiram apenas o nível inicial de conformidade com o Princípio da Interoperabilidade.

Na análise sobre o **Princípio da Reutilização** dos dados, observou-se que ele requer maior atenção, devido às consideráveis disparidades nos níveis de adequação dos *datasets*. Na métrica FsF-R1-01MD, que verifica se os metadados especificam o conteúdo dos dados, 76,9% (10) dos *datasets* obtiveram um nível de conformidade avançado, enquanto os 23% (3) restantes obtiveram um nível intermediário. Nas métricas FsF-R1.1-01M, que avaliam se os metadados incluem informações sobre as licenças para reutilização, e FsF-R1.3-01M, que verificam se os metadados seguem um padrão recomendado pela

comunidade de pesquisa dos dados, 100% (13) dos *datasets* obtiveram um nível inicial de conformidade.

Ainda em relação ao Princípio da Reutilização, a métrica FsF-R1.2-01M, que avalia se os metadados incluem informações de proveniência sobre a criação ou geração de dados, revelou que 100% (13) dos *datasets* avaliados demandam melhorias, uma vez que obtiveram nível intermediário de conformidade.

Por fim, o teste com a métrica FsF-R1.3-02D, que verifica se os dados estão disponíveis em um formato de arquivo recomendado pela comunidade de pesquisa alvo, apresentou os seguintes resultados: 38,46% (5) dos *datasets* se enquadram no nível avançado e 61,53% (8) atingiram o nível intermediário de conformidade com o princípio.

Com isso, podemos observar que os *datasets* depositados no LattesData, demonstram alto grau de conformidade com o Princípio da Encontrabilidade. Em relação ao Princípio da Acessibilidade, de acordo com as 3 métricas avaliadas, 2 atingiram o nível avançado de conformidade, enquanto 1 delas enquadra-se no nível incompleto. Quanto ao Princípio da Interoperabilidade, 2 das 3 métricas alcançaram o nível avançado de conformidade, enquanto 1 está no nível incompleto. Já o Princípio da Reutilização, que possui o maior número de métricas, demanda atenção especial. As 2 métricas encontram-se em um estágio inicial de conformidade, 1 está no nível intermediário e 2 apresentam variação entre os diferentes *datasets*.

Observa-se que os *datasets* analisados tem um desempenho positivo com relação ao Princípio da Encontrabilidade e na maioria das métricas relacionadas aos Princípios da Acessibilidade e Interoperabilidade, contudo ainda existem aspectos específicos relacionados aos metadados que devem ser aprimorados para poder garantir acesso e informações seguras quanto a procedência. No entanto, há desafios significativos em relação ao Princípio da Reutilização, com necessidade de melhorias na especificação de metadados, informação de proveniência e formato de dados e de arquivos. A análise destaca a importância de aprimorar a forma de registro dos metadados e armazenamento dos dados

para garantir sua qualidade, confiabilidade e utilidade máxima para o uso e reúso (Devaraju; Huber, 2021).

No Tabela 1, pode-se identificar o quantitativo total de métricas atendidas, por *dataset*, em cada nível de conformidade dos diferentes princípios. A última coluna apresenta o número de métricas para referência, da mesma forma, os *datasets* que demonstraram pior desempenho estão marcados em cinza claro.

Tabela 1 - Nível de conformidade dos Princípios dos *datasets*

ITEM	LATTES DATA	A	INT	INI	INC	MTR
1	https://doi.org/10.57810/lattedata/QU5LC9	9	3	3	1	16
2	https://doi.org/10.57810/lattedata/L3JK6O	11	1	3	1	16
3	https://doi.org/10.57810/lattedata/TQAD9D	11	1	3	1	16
4	https://doi.org/10.57810/lattedata/ODBRQG	10	2	3	1	16
5	https://doi.org/10.57810/lattedata/51N0TE	10	2	3	1	16
6	https://doi.org/10.57810/lattedata/UWUC2I	11	1	3	1	16
7	https://doi.org/10.57810/lattedata/4YPJII	11	1	3	1	16
8	https://doi.org/10.57810/lattedata/6WUTNU	9	3	3	1	16
9	https://doi.org/10.57810/lattedata/QB4OZ3	9	3	3	1	16
10	https://doi.org/10.57810/lattedata/HSUS6N	11	1	3	1	16
11	https://doi.org/10.57810/lattedata/53D2XM	10	2	3	1	16
12	https://doi.org/10.57810/lattedata/VG2MA6	10	2	3	1	16
13	https://doi.org/10.57810/lattedata/3TKOSU	10	2	3	1	16

Legenda: A - Avançado, INT - Intermediário, INI - Inicial, INC - Incompleto e MTR - N° de Métricas

Fonte: Dados da pesquisa.

Pode-se observar que os *datasets* 1, 8 e 9, possuem os menores níveis de conformidade. Já os *datasets* que apresentam melhores níveis de conformidade são 2, 3, 6, 7 e 10, que atingiram nível avançado em 11 das 16 métricas avaliadas.

O Quadro 6, apresenta um diagnóstico individual dos *datasets*, apresentando os 34 critérios testados referente às 16 métricas analisadas.

Quadro 6 - Diagnóstico dos *datasets*

FAIR	MÉTRICA	CRITÉRIOS TESTADOS	DATASETS - DIAGNÓSTICO/ ITENS IDENTIFICADO COMO DEMANDA DE MELHORIA												
			1	2	3	4	5	6	7	8	9	10	11	12	13
ENCONTRABILIDADE	FsF-F1-01D	O objeto é identificado por um identificador exclusivo (GUID ou IRI) que segue uma sintaxe adequada.	A	A	A	A	A	A	A	A	A	A	A	A	A
		O identificador é acessível pela web (não quebrado).	A	A	A	A	A	A	A	A	A	A	A	A	A
	FsF-F1-02D	Um identificador de dados é especificado com base em um esquema de identificador persistente comumente aceito e adequado para dados de pesquisa.	A	A	A	A	A	A	A	A	A	A	A	A	A
		O identificador é acessível pela web, ou seja, resolve para uma landing page com metadados do objeto de dados.	A	A	A	A	A	A	A	A	A	A	A	A	A
	FsF-F2-01M	Alguns metadados (todos) foram disponibilizados através de padrões comuns (web).	A	A	A	A	A	A	A	A	A	A	A	A	A
		Os metadados básicos mínimos de citação são especificados (criador, título, data de publicação, editor e identificador)	A	A	A	A	A	A	A	A	A	A	A	A	A
		Os metadados descritivos básicos mínimos são especificados (criador, título, editor, data de publicação, resumo, palavras-chave,	A	A	A	A	A	A	A	A	A	A	A	A	A

		identificador) por meio de campos de metadados apropriados.													
	FsF-F3-01M	Os metadados contêm um PID ou URL que indica a localização do conteúdo de dados para download	NA	A	A	A	A	A	A	A	A	A	A	A	A
		Um identificador de dados é incluído nos metadados e corresponde ao identificador fornecido como parte da solicitação de avaliação	A	A	A	A	A	A	A	A	A	A	A	A	A
	FsF-F4-01M	Os metadados do objeto podem ser recuperados programaticamente por meio de pelo menos um dos seguintes métodos: Dados estruturados incorporados na página inicial do objeto de dados e Links digitados de documento de metadados ou links de cabeçalho de sinalização.	A	A	A	A	A	A	A	A	A	A	A	A	A
		Negociação de conteúdo com um serviço provedor de PID (ex.: DataCite).	A	A	A	A	A	A	A	A	A	A	A	A	A
ACESSIBILIDADE	FsF-A1-01M	Os metadados incluem o nível de acesso aos dados (por exemplo, público, embargado, restrito) e suas condições de acesso usando campos de metadados apropriados.	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
		Os metadados de nível de acesso são legíveis por máquina e isso é verificado em relação a vocabulários controlados.	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

INTEROPERABILIDADE	FsF-A1-02M	O esquema do URI de metadados é baseado em um protocolo de aplicação comum.	A	A	A	A	A	A	A	A	A	A	A	A	A
		Os metadados são acessíveis através do identificador fornecido.	A	A	A	A	A	A	A	A	A	A	A	A	A
	FsF-A1-03D	O esquema do URI de dados é baseado em um protocolo de aplicação compartilhado.	NA	A	A	A	A	A	A	A	A	A	A	A	A
		Os dados são acessíveis através do identificador fornecido.	NA	A	A	A	A	A	A	A	A	A	A	A	A
	FsF-A2-01M *	Portanto, este teste só é aplicável para objetos excluídos, substituídos ou obsoletos. É importante ressaltar que o acesso contínuo aos metadados depende da prática de preservação do repositório de dados. Portanto, consideramos que a avaliação da métrica se aplica ao nível de um repositório, e não ao nível de objetos individuais. Por esse motivo, excluimos os detalhes da sua avaliação da implementação do F-UJI.	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
			A	A	A	A	A	A	A	A	A	A	A	A	A
INTEROPERABILIDADE	FsF-I1-01M	Dados estruturados e analisáveis são incorporados na página de destino	A	A	A	A	A	A	A	A	A	A	A	A	A
		Metadados formais e analisáveis (por exemplo, RDF, JSON-LD) são acessíveis por meio de negociação de conteúdo, links digitados ou endpoint sparql.	A	A	A	A	A	A	A	A	A	A	A	A	A

REUTILIZAÇÃO	FsF-I2-01M	Namespaces de recursos semânticos conhecidos estão presentes nos metadados de um objeto; exclua namespaces comuns (por exemplo, rdf, rdfs, xsd, owl) do teste.	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP
	FsF-I3-01M	Os metadados capturam a relação entre um objeto de dados e sua entidade relacionada.	A	A	A	A	A	A	A	A	A	A	A	A	A
		A relação é expressa usando um tipo de relação de acordo com os tipos de relação PROV-O ou DataCite.	A	A	A	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
		As relações são preferencialmente expressas como identificadores únicos ou persistentes acionáveis.	A	A	A	A	A	A	A	A	A	A	A	A	A
	FsF-R1-01MD	Os metadados incluem o tipo do objeto e as propriedades técnicas do seu arquivo de dados, como formato, tamanho, variáveis observadas ou medidas.	A	A	A	A	A	A	A	A	A	A	A	A	A
		Os valores de metadados das propriedades estão em conformidade com o arquivo de dados real.	NA	A	A	A	A	A	A	A	A	A	A	A	A
	FsF-R1.1-01M	Os metadados contêm informações de licença representadas por meio de um elemento de metadados apropriado.	A	A	A	A	A	A	A	A	A	A	A	A	A
		Uma licença padrão legível por máquina é especificada.	A	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

FsF-R1.2-01M	Os metadados incluem propriedades que representam a criação de dados, como criador, contribuidores, datas de criação e modificação e versão, origem e relações que indicam atividades de criação de dados.	A	A	A	A	AP	A	A	A	A	A	A	A	A
	Os metadados de proveniência estão disponíveis numa versão legível por máquina do PROV-O ou PAV.	NA	NA	NA	NA	A	NA	NA	NA	NA	NA	NA	NA	NA
FsF-R1.3-01M	Os metadados estão disponíveis através de pelo menos um dos padrões de metadados de domínio listados no RDA Metadata Standards Catalog	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP
FsF-R1.3-02D	Os dados estão disponíveis em um formato de arquivo de longo prazo, conforme definido na ISO/TR 22299	NA	A	NA	A	A	A	A	A	A	A	A	A	A
	Os dados estão disponíveis em formato aberto (ver, por exemplo, https://en.wikipedia.org/wiki/List_of_open_formats)	NA	A	NA	A	NA	NA	NA	A	A	NA	NA	NA	NA
	Os dados estão disponíveis em um formato de arquivo científico	NA	A	A	NA	NA	A	A	NA	NA	A	NA	NA	A
Legenda: A - Atende, NA - Não Atende, AP - Atende Parcialmente e NSA - Não se aplica														

Fonte: Dados da pesquisa.

A ferramenta realiza a análise abrangente das 16 métricas, aplicando diferentes critérios de teste ou testes práticos para avaliar os *datasets* sob múltiplos aspectos referentes a cada uma das métricas. São 34 critérios de teste, sendo que cada métrica composta por 1 ou mais critérios de análise, que contribuem para constituir o diagnóstico final. Desta forma, é possível identificar com precisão quais medidas precisam ser tomadas para a melhoria do acesso e uso dos dados (Devaraju; Huber, 2021).

Dos 34 critérios propostos, um deles não foi testado pela ferramenta F-UJI pois somente pode ser aplicado quando o *dataset* foi excluído da base de dados, o que não se aplica aos 13 conjuntos de dados (Devaraju; Huber, 2021). Assim sendo, consta na tabela, mas não se aplica a análise, por isso consideramos que são 33 critérios de teste.

De forma geral, 19 dos 33 critérios testados são atendidos por todos os 13 *datasets*. Os demais 14 critérios não são atendidos, total ou parcialmente, por pelo menos 1 dos conjuntos de dados.

Conforme demonstrado no Quadro 6, o dataset número 1 é o conjunto de dados que menos contempla os critérios testados. Além dos critérios já citados, outros 10 testes realizados para compor o diagnóstico, revelaram problemas como a falta de arquivo em formatos aberto, científico e de validade de longo prazo. Em contrapartida, este dataset possui uma licença padrão legível por máquina, o que os demais 12 conjuntos de dados não possuem.

No teste com o *dataset* 8 e *dataset* 9, pôde-se observar que 8 critérios testados não foram atendidos, como é o caso da falta de registro para tamanho dos arquivos nos metadados.

Quanto ao dataset mais completo, o conjunto de dados número 2 contempla o maior número de critérios testados. Esse conjunto de dados atende parcialmente dois critérios e não atende quatro dos 33 critérios testados. Neste caso, o dataset oferece os arquivos nos 3 formatos testados, aberto, científico e de validade de longo prazo, ou seja, os autores tiveram o cuidado preparando e organizando os dados e arquivos de acordo com padrões específicos, visando a

acessibilidade, a preservação e a interoperabilidade de dados científicos, projetado para resistir à obsolescência.

Com relação aos critérios testados referentes ao Princípio de Encontrabilidade, todos foram atendidos por 100% (13) *datasets*, com uma única exceção, o dataset 1 quanto a métrica FsF-F3-01M. (Referente ao critério de teste: Os metadados contêm um PID ou URL que indica a localização do conteúdo de dados para download.

O atendimento de todos os critérios de teste, do Princípio de Encontrabilidade, indica que os dados são de alta qualidade, bem documentados, acessíveis e sustentáveis ao longo do tempo. Desta forma, com relação à qualidade e usabilidade dos dados, o atendimento a todos os aspectos analisados para este princípio, reflete na integridade dos dados, acessibilidade e permanência dos dados, padrões e identificação persistente, metadados de qualidade, acessibilidade dos metadados, localização de conteúdo de dados, recuperação programática de metadados e conformidade com padrões (Devaraju; Huber, 2021).

Neste contexto, os *datasets* da base LattesData oferece dados mais valiosos para a pesquisa, além de possibilitar o compartilhamento e colaboração com a comunidade científica e o público em geral.

Quanto aos critérios testados e não atendidos, cinco testes não foram atendidos por pelo menos nove dos 13 *datasets*. Dentre estes, estão os testes realizados referente a métrica FSF-A1-01M (Princípio da Acessibilidade), que estão relacionados ao nível de acesso, bem como, sobre os campos de metadados apropriados para esse fim.

Os testes demonstraram que estes critérios não são contemplados nos 13 *datasets*, revelando que as informações sobre quem pode acessar os dados, se eles estão disponíveis ao público em geral, ou se existem restrições específicas para acessá-los, não estão claramente especificadas nos metadados dos *datasets*. Esta é uma lacuna significativa, já que dificulta a compreensão de quem pode ter acesso, fazer uso dos dados e sob quais circunstâncias pode utilizá-los.

A análise também revelou a falta de dados estruturados para o Princípio de Reusabilidade. No teste realizado sobre a métrica FsF-R1.2-01M que verifica se existem informações ou metadados que descrevem explicitamente como os dados estão relacionados a outros dados ou recursos por meio de padrões ou vocabulários específicos, mostrou a falta de informações claras sobre os dados referentes à proveniência. A falta de informações legíveis por máquina, podem dificultar a identificação da proveniência, autenticidade ou a conexão dos dados com outros recursos, o que reduz a confiabilidade dos dados para reuso. Neste contexto, o uso de vocabulários controlados, com termos padronizados é a solução para garantir a gestão e o compartilhamento eficaz dos dados. (Alves; Tartarotti; Fujita, 2022).

Outro critério testado que não foi contemplado por 12 dos 13 *datasets*, foi o teste sobre a existência de uma licença padrão legível por máquina especificada. Esta lacuna mostra que os termos de uso e as restrições associadas a estes *datasets* não está definido em formato que possa ser compreendido ou processado por linguagem de máquina. A ausência de uma licença legível por máquina pode gerar incertezas sobre o reuso dos dados, compartilhamento, redistribuição de forma legal.

Na análise das métricas relativas ao Princípios da Reutilização dos dados de pesquisa, indica que se requer maior atenção, pois apresentou um maior número de verificações que não foram contempladas. A análise aponta a falta de informações sobre o tipo e tamanho do arquivo dos dados de pesquisa, da mesma forma, não informa variáveis medidas e tipo de observações.

Além disso, ainda quanto ao Princípio de Reutilização, aos padrões de metadados, não foram encontrados metadados específicos para a comunidade da área. Da mesma forma, não foram encontrados padrões indicados pelo Re3data, além de não atender ao padrão de formatos de arquivo de dados específico indicado pela comunidade científica da área.

Quanto à licença SPDX, a análise apontou a inexistência de registro de licença ou qualquer informação de proveniência que possa conferir conformidade, segurança e confiabilidade aos *datasets* analisados.

Além do formato do arquivo de armazenamento de dados, é necessário analisar o tipo de arquivo escolhido para armazenar as informações. Neste contexto, a escolha poderá refletir na interoperabilidade, longevidade, qualidade, integridade, reprodutibilidade, segurança, facilidade de análise e visualização.

Os 13 *datasets* da LattesData possuem no total 55 arquivos de dados no repositório. Quanto à reutilização dos dados de pesquisa, no que diz respeito aos métodos de codificação de informações digitais ou formato dos arquivos, esses são compostos por diferentes tipos de extensão, como .csv, .png, .jpg, tab e .xl ou.xlsx. O quadro 7 sintetiza, a extensão dos arquivos vinculados aos *datasets*.

Quadro 7 - Extensão dos arquivos de dados de pesquisa

ITEM	CITAÇÃO DE CONJUNTO DE DADOS	EXTENSÃO				
		.csv	.xls/.xlsx	.tab	.png	.jpg
1	https://doi.org/10.57810/lattedata/QU5LC9	.csv	—	.tab	—	—
2	https://doi.org/10.57810/lattedata/L3JK6O	—	.xls	—	—	—
3	https://doi.org/10.57810/lattedata/TQAD9D	—	.xlsx	—	—	—
4	https://doi.org/10.57810/lattedata/ODBRQG	.csv	—	—	—	—
5	https://doi.org/10.57810/lattedata/51N0TE	—	.xls	—	—	.jpg
6	https://doi.org/10.57810/lattedata/UWUC2I	—	.xls	—	.png	—
7	https://doi.org/10.57810/lattedata/4YPJ1I	—	.xls	—	—	.jpg
8	https://doi.org/10.57810/lattedata/6WUTNU	—	—	.tab	—	—
9	https://doi.org/10.57810/lattedata/QB4OZ3	—	—	.tab	—	—
10	https://doi.org/10.57810/lattedata/HSUS6N	—	.xls	—	—	.jpg
11	https://doi.org/10.57810/lattedata/53D2XM	—	.xls	—	—	.jpg
12	https://doi.org/10.57810/lattedata/VG2MA6	—	—	—	.png	.jpg
13	https://doi.org/10.57810/lattedata/3TKOSU	—	.xls	—	—	.jpg

Fonte: Dados da pesquisa.

A ferramenta F-UJI analisa se o formato escolhido pelo autor é um formato de longo prazo, aberto e se o arquivo de dados é um formato científico. Neste sentido, a análise dos arquivos possibilitou identificar que a maioria dos formatos dos arquivos escolhidos pelos autores possibilita o acesso ao seu conteúdo, já que são softwares convencionais, contudo, alguns destes, não são formatos recomendados pela comunidade científica.

A ferramenta possibilitou a ampla análise da aplicação dos princípios FAIR nos *datasets* disponíveis no repositório LattesData, possibilitando identificar quais aspectos da política dos repositórios e diferentes bases devem ser implementadas ou adequadas para viabilizar a adoção de boas práticas a fim de favorecer a localização, acessibilidade, interoperabilidade e reutilização dos dados de pesquisa.

Apesar dos problemas apontados, observa-se resultados positivos, assim sendo, é possível indicar a LattesData como referência para outras bases poderem rever as práticas para elaboração e registro de metadados, entre outros aspectos, relacionados aos Princípios de Encontrabilidade, Acessibilidade e Interoperabilidade. No entanto, alguns aspectos podem ser melhorados na plataforma LattesData, como o registro adequado das informações sobre licença, metadados adequados sobre proveniência, e a conformidade dos padrões, vocabulários controlados e ontologias.

5 Considerações finais

Considerando que este artigo buscou verificar a adoção dos princípios FAIR nos *datasets* depositados no repositório LattesData, a partir do uso da ferramenta F-UJI, pode-se indicar que o referido repositório tem adotado esses princípios.

A partir das análises foi possível verificar que as métricas relacionadas ao Princípio da Encontrabilidade foram todas contempladas, as métricas relacionadas ao Princípio da Acessibilidade foram parcialmente contempladas, já as métricas relacionadas aos Princípios da Interoperabilidade e Reutilização, requerem atenção.

O LattesData ainda necessita de políticas e diretrizes que realmente garantam que os dados ali depositados possam ser reutilizados. Espera-se que esse estudo forneça um diagnóstico desses conjuntos.

O uso de ferramentas tecnológicas para analisar a qualidade dos conjuntos de dados e a gestão de dados num todo é de extrema relevância para que tenhamos repositórios de dados confiáveis. Os repositórios precisam buscar o padrão de qualidade e excelência, para obterem alguma certificação, como a CoreTrustSeal. Assim, os repositórios confiáveis (certificados) garantem a credibilidade para pesquisadores depositarem, a segurança da preservação dos dados e do repositório, a visibilidade deste e da instituição mantenedora, a concorrer em financiamentos externos, entre outros fatores.

A ferramenta F-UJI mostrou-se bastante simples e eficaz em realizar a análise do conjunto de dados. Uma ferramenta que está acessível e disponível para que outros gestores e curadores de repositórios de dados utilizem e consigam verificar se estão cumprindo os princípios FAIR. O uso de ferramentas para a gestão de dados de pesquisa facilita na implementação de boas práticas e conjunto de dados com qualidade.

Recomenda-se que os repositórios verifiquem quais aspectos ainda precisam ser adequados para poder cumprir aos princípios FAIR e poder oferecer ambiente seguro, encontrável, acessível, interoperável e que garanta o reúso dos dados de pesquisa, podendo assim contribuir para o avanço da ciência.

Referências

ALVES, L. S. S.; TARTAROTTI, R. C. D.; FUJITA, M. S. L. **Avaliação do uso de vocabulário controlado em repositórios institucionais. Informação@Profissões**, Londrina, v. 11, n. 1, p. 52–77, 2022. Disponível em: <https://doi.org/10.5433/2317-4390.2022v11n1p52>. Acesso em: 17 set. 2023.

ARAÚJO, L. M. S; MARDERO ARELLANO, M. A.; FERRER, I. D. **Guia para os usuários do repositório Dataverse do Ibict**. Brasília: IBICT, 2018.

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, 2017.

BHATTACHARYA, S. *et al.* ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. **Scientific Data**, London, v. 5, p. 1-9, 2018. Disponível em: <https://doi.org/10.1038/sdata.2018.15>. Acesso em: 16 set. 2023.

BONETTI, L. G.; ARAKAKI, A. C. S. Princípios fair e a avaliação de datasets no repositório institucional da UFSCAR. **Informação & Informação**, Londrina, v. 27, n. 1, p. 485-510, 2022. Disponível em: <https://doi.org/10.5433/1981-8920.2022v27n1p485> Acesso em: 25 jul. 2022.

BORGMAN, C L.; SCHARNHORST, A.; GOLSHAN, M. S. Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. **Journal of the Association for Information Science and Technology**, United States, v. 70, n. 8, p. 888-904, 2019. Disponível em: <https://doi.org/10.1002/asi.24172>. Acesso em: 25 jul. 2022.

CAREGNATO, S. E.; ROCHA, R. P.; GABRIEL JUNIOR, R. F. Reúso de dados: princípios FAIR e o ecossistema de pesquisa. In: SALES, L. F. *et al.* (org). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: IBICT, 2021.

CHATFIELD, S. L. Recommendations for secondary analysis of qualitative data. **The Qualitative Report**, Fort Lauderdale, v. 25, n. 3, p. 833-842, 2020. Disponível em: <https://doi.org/10.46743/2160-3715/2020.4092>. Acesso em: 16 set. 2023.

CHAUVETTE, A.; SCHICK-MAKAROFF, K.; MOLZAHN, A. E. Open data in qualitative research. **International Journal of Qualitative Methods**, United Kingdom, v. 18, p. 1-6, 2019. Disponível em: <https://doi.org/10.1177/1609406918823863>. Acesso em: 22 jul. 2022.

CHILDS, S. *et al.* Opening research data: issues and opportunities. **Records Management Journal**, United Kingdom, v. 24, n. 2, p. 142-162, 2014. Disponível em: <https://doi.org/10.1108/RMJ-01-2014-0005>. Acesso em: 22 jul. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). **LattesData**, Brasília, c2022.

DEVARAJU, A.; HUBER, R. F-UJI: an automated FAIR Data assessment tool. **Zenodo**, Genève, 2020. Disponível em: <https://doi.org/10.5281/zenodo.4063720>. Acesso em: 06 set. 2023.

DEVARAJU, A.; HUBER, R. F-UJI (FAIRsFAIR Research Data Object Assessment Service): Methods. **Fostering Fair Data Practices in Europe - FAIRsFAIR**, [s.l.], 2021.

DEVARAJU, A. *et al.* FAIRsFAIR data object assessment metrics. **Zenodo**, Genève, 2022.

DUDZIAK, E. Dados de Pesquisa agora devem ser armazenados e citados. **ABCD USP**, São Paulo, 2016.

FAIR Principles. **GO FAIR**, Hamburgo, [2022].

FELIPE, C. B. M.; SANTOS, R. F. D. Avaliação de metadados em repositórios de dados de pesquisa sobre biodiversidade. **Em Questão**, Porto Alegre, v. 28, n. 3, p.1-19, 2022. Disponível em: <https://doi.org/10.19132/1808-5245283.117591> Acesso em: 25 jul. 2022.

HENNING, P. C. *et al.* Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, João Pessoa, v. 14, n. 3, p. 175-192, 2019. Disponível em: <https://doi.org/10.22478/ufpb.1981-0695.2019v14n3.46969>. Acesso em: 10 set. 2023.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Principles and guidelines for access to research data from public funding**. France: OECD, 2007.

PAVÃO, C. P. *et al.* Acesso aberto a dados de pesquisa no Brasil: políticas para repositórios de dados de pesquisa. **RDP Brasil - Rede de Dados de Pesquisa**, Porto Alegre, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/1263>. Acesso: 19 set. 2023.

ROBERT, H. *et al.* D4.5 Report on FAIR Data Assessment toolset and badging scheme. **Zenodo**, Genève, 2021. Disponível em: <https://doi.org/10.5281/zenodo.6656444>. Acesso em: 7 set. 2023.

RODRIGUES, M. M.; DIAS, G. A.; LOURENÇO, C. A. Repositórios de dados científicos na América do Sul: uma análise da conformidade com os princípios FAIR. **Em Questão**, Porto Alegre, v. 28, n. 2, p. 295-339, 2022. Disponível em: <https://doi.org/10.19132/1808-524500.%p> . Acesso em: 25 jul. 2022.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados científicos para bibliotecários de pesquisadores**. Rio de Janeiro: CNEN, 2015.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, 2016. Disponível em: <https://doi.org/10.5433/1981-8920.2016v21n2p90> Acesso em: 29 jul. 2022.

SAYÃO, L. F.; SALES, L. F. Afinal, o que é dado de pesquisa? **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 34, n. 2, p. 32-51, 2020. Disponível em: <https://doi.org/10.14295/biblos.v34i2.11875>. Acesso em: 22 jul. 2022.

SCHEREIBER, G. *et al.* **Knowledge engineering and management: the CommonKADS methodology**. Cambridge: MIT Press, 2000.

SILVA, F. C. C. da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019.

VAN DE SANDT, S. *et al.* The definition of reuse. **Data Science Journal**, United Kingdom, v. 18, n. 22, p. 1-19, 2019. Disponível em: <https://doi.org/10.5334/dsj-2019-022>. Acesso em: 22 jul. 2022.

WILKINSON, M. D *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, p. 1-9, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 22 jul. 2022.

LattesData and the adoption of the FAIR Principles: an analysis using the F-UJI Automated FAIR Data Assessment Tool

Abstract: The funding agencies of developed countries have begun to demand the open-access deposition of research data, requiring compliance with standards that ensure data integrity, completeness, quality, and accurate metadata descriptions. As a response to this need, the FAIR principles - Findability, Accessibility, Interoperability, and Reusability have emerged as guidelines to enhance the reusability of research data. With the aim of contributing to these discussions, this article sought to assess the adoption of FAIR principles in *datasets* deposited in the LattesData repository, utilizing the F-UJI tool developed for automating the analysis of FAIR metrics. Methodologically, this is a descriptive research study in which 13 *datasets* available in the LattesData repository were analyzed. The results obtained indicate that the "Findability" principle was the most well-addressed in the *datasets*, suggesting that the data can be easily located and identified. However, the principles of "Accessibility," "Interoperability," and "Reusability" require more attention, as not all metrics associated with these principles were adequately met. In conclusion, it is observed that the LattesData repository has started to adopt the FAIR principles, albeit in a preliminary manner. There is a need for policies and guidelines to ensure that deposited data can be effectively

reused. The F-UJI tool has proven to be a practical and straightforward tool for use in this context. The utilization of such tools for research data management can facilitate the implementation of best practices and the deposition of high-quality *datasets*. The implementation of good practices and the deposit of quality data sets.

Keywords: research data; FAIR principles; data repository

Recebido: 08/02/2023

Aceito: 22/09/2023

Declaração de autoria:

Concepção e elaboração do estudo: Crislaine Zurilda Silveira, Divino Ignacio Ribeiro Júnior, Igor Yure Ramos Matos.

Coleta de dados: Adriane Groehs, Elaine Maria de Lima, Eliane Dittrich.

Análise e interpretação de dados: Adriane Groehs, Elaine Maria de Lima, Eliane Dittrich.

Redação: Adriane Groehs, Crislaine Zurilda Silveira, Igor Yure Ramos Matos.

Revisão crítica do manuscrito: Divino Ignacio Ribeiro Júnior

Como citar:

GROEHS, Adriane; SILVEIRA, Crislaine Zurilda; LIMA, Elaine Maria de; DITTRICH, Eliane; MATOS, Igor Yure Ramos; RIBEIRO JUNIOR, Divino Ignacio. LattesData e a adoção aos Princípios FAIR: uma análise usando a F-UJI Automated FAIR Data Assessment Tool. **Em Questão**, Porto Alegre, v. 29, e-130018, 2023. DOI: <https://doi.org/10.1590/1808-5245.29.130018>



¹ A busca foi realizada no dia 31 agosto de 2023, na qual o termo “FAIR”, foi recuperado em todos os campos. Essa busca resultou na recuperação de 91 trabalhos, os quais foram eliminadas as duplicatas. Com isso, foi realizada a leitura de 77 resumos, tendo em vista identificar relatos práticos de uso dos princípios FAIR em repositórios de dados. Desses trabalhos 31, utilizaram o termo FAIR no resumo ou tinham alguma relação com a temática. Por fim, visando identificar apenas os trabalhos que avaliavam a adesão ao FAIR dos dados depositados em repositórios, foram identificados 3 trabalhos, todos elencados na introdução.

² A ferramenta foi sugerida com base na página FAIRsFAIR. Durante a revisão dos dados deste trabalho foi publicado o primeiro trabalho brasileiro de Bonetti e Arakaki (2022). Os trabalhos posteriores à escrita deste não serão analisados.