

PRINCÍPIOS FAIR E A AVALIAÇÃO DE DATASETS NO REPOSITÓRIO INSTITUCIONAL DA UFSCAR

FAIR PRINCIPLES AND EVALUATION OF DATASETS IN THE UFSCAR INSTITUTIONAL REPOSITORY

Letícia Guarany Bonetti^a

Ana Carolina Simionato Arakaki^b

RESUMO

Objetivo: Inserido no contexto do potencial dos dados de pesquisa para o avanço científico, o presente artigo busca avaliar a conformidade dos *datasets* depositados no repositório da Universidade Federal de São Carlos (UFSCar) quanto aos 4 princípios FAIR. **Metodologia:** A pesquisa é de natureza quantitativa, com caráter descritivo e exploratório. A amostra contempla os 12 *datasets* depositados no repositório institucional da UFSCar. Como instrumento de coleta de dados, utilizou-se a ferramenta auxiliar F-UJI, que permite a avaliação automática dos *datasets* e fornece relatórios de *feedback*. **Resultados:** A média de pontuação geral quanto à aderência dos *datasets* aos princípios foi de 30%. As pontuações individuais para cada um dos quatro princípios também foram baixas, com as maiores notas sendo: 4/7 para encontrável, 1,5/3 para acessível, 1/4 para interoperável e 2/4 para reusável. **Conclusões:** Para que o repositório da UFSCar siga as diretrizes internacionais de boas práticas, recomenda-se à instituição que considere a implementação gradual dos princípios FAIR como uma melhoria do serviço, especialmente em relação ao princípio interoperável, um dos aspectos essenciais quando se fala em repositórios de dados de pesquisa.

Descritores: Dados de pesquisa. Repositório institucional. Princípios FAIR. Certificação.

1 INTRODUÇÃO

No contexto atual dos avanços tecnológicos, pesquisadores têm produzido quantidades cada vez maiores de dados para suas pesquisas, muitos dos quais são poucos explorados em seu potencial para os avanços científicos

^a Mestranda em Ciência da Informação pela Universidade Federal de São Carlos (UFSCar). E-mail: leticiagbonetti@gmail.com

^b Doutora em Ciência da Informação pela Universidade Estadual Paulista (UNESP). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos (UFSCar). E-mail: simionato.ac@gmail.com

(BERTIN; VISOLI; DRUCKER, 2017). Nesse cenário, os dados de pesquisa criados a partir da dinâmica científica, deixam de ser meros subprodutos para ganhar papel de destaque mundialmente (SAYÃO; SALES, 2016).

Somado a isso, segundo Simionato (2017, p. 01), os dados produzidos pelos pesquisadores “[...] em sua grande parte, nem sempre estão estruturados de forma coerente para serem disponíveis aos usuários, como o que acontece com o fenômeno *Big Data*”. A falta dessa estrutura dos dados de pesquisa dificulta o acesso e o uso, criando obstáculos para o consumo (SANTARÉM SEGUNDO; SILVA; MARTINS, 2019). Há, portanto, a necessidade de uma gestão dos dados de pesquisa, buscando adicionar contexto e valor, de forma a permitir a preservação, acesso e usabilidade (SAYÃO; SALES, 2016).

Nesse cenário, segundo Monteiro e Sant’Ana (2018), as instituições acadêmicas como as universidades passam a ter maiores responsabilidades quanto à gestão dos dados de pesquisa que são produzidos por seus pesquisadores. Em concordância, Tenopir *et al.* (2017) afirmam que não é surpreendente que as bibliotecas universitárias tenham um papel pioneiro na gestão dos dados de pesquisa, uma vez sua função é providenciar acesso ao que é produzido pelas instituições acadêmicas em suas diversas formas.

Para que os dados de pesquisa possam ser gerenciados e preservados para futuros usos é necessário que sejam armazenados. Segundo Curty (2019), a expansão de políticas voltadas para o compartilhamento dos dados de pesquisa, principalmente os financiados com recursos públicos, tem resultado em um crescimento acelerado dos repositórios de dados de pesquisa de acesso aberto. Sayão e Sales (2016) argumentam sobre a importância dessas infraestruturas informacionais formalizadas, uma vez que, diferente das publicações acadêmicas, os dados de pesquisa não falam por si só.

Sayão e Sales (2016, p. 91) também destacam que esses conjuntos de dados demandam ações específicas que levem em conta todo o seu ciclo de vida, uma vez que isso os permite “[...] revelarem e transmitirem conhecimento no tempo e no espaço, e a partir daí serem interpretadas, sintetizadas e reanalisadas em contextos diversos”, o que traz inúmeras vantagens para a

ciência como a economia de recursos, a maior transparência e visibilidade institucional e a agilidade no ciclo científico.

Em âmbito nacional, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), uma das principais agências de fomento à pesquisa científica e tecnológica do país, já reconhece a importância da gestão adequada dos dados de pesquisa como parte essencial das boas práticas na ciência. Para tanto, “[...] considera necessário que os dados resultantes de projetos financiados pela Fundação sejam gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico e tecnológico” (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2020).

Nesse contexto, há que se considerar os diversos aspectos que dizem respeito aos dados que são gerenciados, uma vez que “[...] para tornar possível o acesso aberto a dados de pesquisa de qualidade, faz-se necessária a publicação de tais dados seguindo princípios que orientam a adoção de uma estrutura semântica legível por humanos e máquinas” (SILVA; SANTARÉM SEGUNDO; SILVA, 2018, p. 5197), como é o caso das orientações dos princípios FAIR (do inglês *findable, accessible, interoperable e reusable*).

Wilkinson *et al.* (2016) destacam que uma boa gestão de dados é amplamente subjetiva e indefinida, ficando, geralmente, a cargo do proprietário dos dados ou do repositório definir. Os princípios FAIR buscam trazer clareza em torno dos objetivos de uma boa gestão e administração de dados e definir diretrizes simples para informar aqueles que publicam e/ou preservam dados de pesquisa (WILKINSON *et al.*, 2016). Isso é importante porque há uma grande demanda por novas técnicas e métodos para melhor apoiar os estudiosos na localização de dados relevantes para suas pesquisas em repositórios de dados (LÖFFLER *et al.*, 2021).

Apesar da importância dada aos dados de pesquisa mundialmente, com várias instituições científicas apoiando e incentivando boas práticas, como a *National Science Foundation* (NSF), a *National Institute of Health* (NIH), a Comissão Europeia, a FAPESP e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), ainda são escassos os estudos que focam no tema de dados de pesquisa e os repositórios de dados, demandando maiores

investigações (TARTAROTTI; DAL´EVEDOVE; FUJITA, 2019). Dias, Anjos e Rodrigues (2019, p. 185) corroboram ao afirmar que “Poucos são os trabalhos que avaliam efetivamente a aplicação do FAIR em casos concretos”.

Sendo assim, este estudo tem como objetivo avaliar o cenário dos dados de pesquisa depositados no repositório institucional da Universidade Federal de São Carlos (UFSCar) frente aos princípios FAIR. Busca identificar em qual nível de conformidade os conjuntos de dados se encontram e proporcionar maiores debates no tema. A aderência a esses princípios é fundamental para acompanhar o cenário internacional, onde organizações como a Comissão Europeia já divulgam amplamente ações como o *Turning FAIR data into reality*¹.

O artigo foi estruturado em uma contextualização sobre a temática na Introdução e por uma fundamentação teórica sobre os dados de pesquisa e os princípios FAIR. Após esse referencial, detalha-se sobre os procedimentos metodológicos, e conseguinte, a apresentação e discussão dos resultados e as conclusões do estudo.

2 DADOS DE PESQUISA E OS PRINCÍPIOS FAIR

Os dados de pesquisa, segundo a *Organisation for Economic Co-operation and Development* (OECD), são “Registros factuais usados como fontes primárias na pesquisa científica, e que são geralmente aceitos na comunidade científica como sendo necessários para validar os resultados de pesquisa” (OECD, 2007, p. 13, tradução nossa).

Entretanto, os dados de pesquisa podem variar de acordo com os pesquisadores que os produzem e as áreas de conhecimento, possuindo características diversas. O “[...] reconhecimento dessa idiossincrasia torna-se crucial quando se estabelecem as opções gerenciais e tecnológicas para o arquivamento persistente e para a curadoria digital” (SAYÃO; SALES, 2013).

Os dados de pesquisa são heterogêneos, e suas particularidades precisam ser levadas em conta no momento da gestão, principalmente em repositórios multidisciplinares, que gerenciam dados de vários tipos, como é o

¹ Disponível em: https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf.

caso do repositório da UFSCar.

Segundo Pinfield (2009, p. 165, tradução nossa) o repositório é “[...] um conjunto de sistemas e serviços que facilitam a tarefa, armazenamento, gerenciamento, recuperação, exibição e reutilização de objetos digitais”. Os repositórios garantem uma maior visibilidade da produção científica, contribuindo para o aumento do impacto acadêmico (PAVÃO *et al.*, 2012).

Em concordância, Monteiro e Sant’Ana (2018, p.161) descrevem os repositórios de dados como “[...] ambientes digitais implementados nas universidades com infraestrutura para dar suporte aos pesquisadores na gestão e na disponibilização de dados científicos”. Ainda segundo os autores, os repositórios de dados estão, em sua maioria, vinculados às instituições de pesquisa e universidades. Todavia, o estudo recente de Rodrigues (2020) aponta que a implementação de repositórios de dados no Brasil ainda está em estágio inicial, nas fases de estudos e discussões.

Um dos aspectos que precisam ser levados em conta para a implementação de repositórios de dados de pesquisa são as diretrizes e padrões internacionais, que buscam a credibilidade e certificação deles. É o caso dos princípios FAIR, um acrônimo para: encontráveis (*findable*), acessíveis (*accessible*), interoperáveis (*interoperable*), e reutilizáveis (*reusable*). Segundo Henning *et al.* (2019), esses princípios são os norteadores do processo de gestão de dados de pesquisa e vêm sendo debatidos e difundidos amplamente no cenário mundial.

Em 2016, o ‘*FAIR Guiding Principles for scientific data management and stewardship*’ foi publicado na *Scientific Data*. As orientações pretendiam fornecer diretrizes para melhorar a encontrabilidade, o acesso, a interoperabilidade e a reutilização de objetos digitais. Os princípios FAIR, segundo Wilkinson *et al.* (2016), enfatizam o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente. No caso, os princípios FAIR “[...] não têm o objetivo de constituir padrões ou especificar formatos ou tecnologias, mas, sim, contextualizar e apontar para direção de maior utilidade e melhores serviços de dados” (HENNING *et al.*, 2019, p. 179).

Os aspectos dos princípios FAIR são independentes e podem ser

implementados em diferentes combinações, apesar de estarem relacionados (WILKINSON *et al.*, 2016). De acordo com os autores o “[...] resultado final, quando implementado, será uma gestão e administração mais rigorosas desses valiosos recursos digitais, para o benefício de toda a comunidade acadêmica” (WILKINSON *et al.*, 2016, p. 7, tradução nossa).

Percebe-se, portanto, que a sistematização de verificações de controle de qualidade dos conjuntos de dados depositados é de extrema relevância no cenário dos repositórios (TARTAROTTI; DAL´EVEDOVE; FUJITA, 2019). Os objetivos propostos pelos princípios FAIR já se constituem como expectativas de agências e editoras internacionalmente (WILKINSON *et al.*, 2016), revelando a importância de sua aplicação no cenário dos repositórios de dados nacionais.

3 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa é de natureza quantitativa, do tipo descritiva e exploratória, e busca avaliar os níveis de aderência dos dados de pesquisa depositados no repositório institucional da UFSCar com relação aos quatro princípios FAIR. Os dados coletados durante as avaliações foram compilados em planilhas para a análise dos resultados, possibilitando comparações e sistematização, além da preservação para futuros reúsos.

Na data de coleta e análise dos dados (31/05/2021), o repositório contava com 13 *datasets*² depositados, que podem ser visualizados na coleção “*datasets*”. O repositório também inclui em sua coleção dissertações, teses e Trabalhos de Conclusão de Curso (TCC) e, portanto, não é um repositório exclusivo para dados de pesquisa. Para este estudo, serão considerados apenas os *datasets*, que são o objeto de análise da pesquisa. Um dos *datasets* apresenta restrições de visualização, então não pôde ser avaliado no trabalho. A lista detalhada dos 12 *datasets* da amostra se encontra no Apêndice A.

Para a verificação da aderência dos *datasets* aos princípios FAIR foi utilizada uma ferramenta auxiliar, a F-UJI *Automated FAIR Data Assessment*

² Disponível em: <https://repositorio.ufscar.br/discover>.

*Tool*³, desenvolvida pela *Fostering Fair Data Practices in Europe - FAIRsFAIR*⁴. A ferramenta desenvolveu um papel fundamental nesta pesquisa, uma vez que os princípios FAIR são concebidos como diretrizes e não especificam regras de implementação, o que torna difícil verificar a prática desses princípios manualmente.

Logo, o projeto *FAIRsFAIR* visa fornecer soluções práticas para o uso dos princípios de dados FAIR ao longo do ciclo de vida dos dados de pesquisa, incentivando a adoção de boas práticas. O *FAIRsFAIR* desenvolve padrões globais para a certificação FAIR de repositórios e os dados depositados neles, sendo, portanto, um projeto de extrema relevância no contexto deste estudo e no cenário internacional (DEVARAJU *et al.*, 2021).

O F-UJI é um serviço baseado em REST para avaliação programática do "FAIRness" dos conjuntos de dados de pesquisa e foi financiado pela *European Union's Horizon 2020*⁵. A avaliação F-UJI é baseada em 16 das 17 métricas básicas de avaliação dos objetos FAIR. O código-fonte da ferramenta está disponível sob uma licença gratuita no *Github*⁶. O serviço, portanto, automatiza a avaliação dos conjuntos de dados de pesquisa quanto à aderência aos princípios FAIR, permitindo, de forma visual e detalhada, identificar o nível de "FAIRness" que os dados se encontram.

A ferramenta ainda fornece um resumo quanto a cada um dos quatro princípios: *findable*, *accessible*, *interoperable* e *reusable*. Além do resumo com as pontuações gerais obtidas, o serviço fornece também um relatório com maiores detalhes sobre quais aspectos foram ou não aderidos em cada um dos quatro princípios. Isso permite analisar os pontos fracos e fortes dos conjuntos de dados quanto aos princípios FAIR, facilitando a identificação de aspectos que precisam ser trabalhados. Além de indicar quais tecnologias e padrões não foram encontrados durante a avaliação, dando um guia para futuras melhorias que levariam à aderência dos aspectos avaliados.

³ Disponível em: <https://www.f-uji.net/index.php>.

⁴ Disponível em: <https://www.fairsfair.eu>.

⁵ Disponível em: <https://ec.europa.eu/programmes/horizon2020/en/home>.

⁶ Disponível em: <https://github.com/pangaea-data-publisher/fuji/issues>.

O objetivo é, portanto, analisar o nível de aderência dos *datasets* da amostra quanto aos princípios FAIR, identificando os pontos fracos e fortes de acordo com a encontrabilidade, acesso, interoperabilidade e reusabilidade. De tal forma será possível apontar aspectos a serem trabalhados pela UFSCar e seu repositório, garantindo dados mais FAIR no cenário brasileiro.

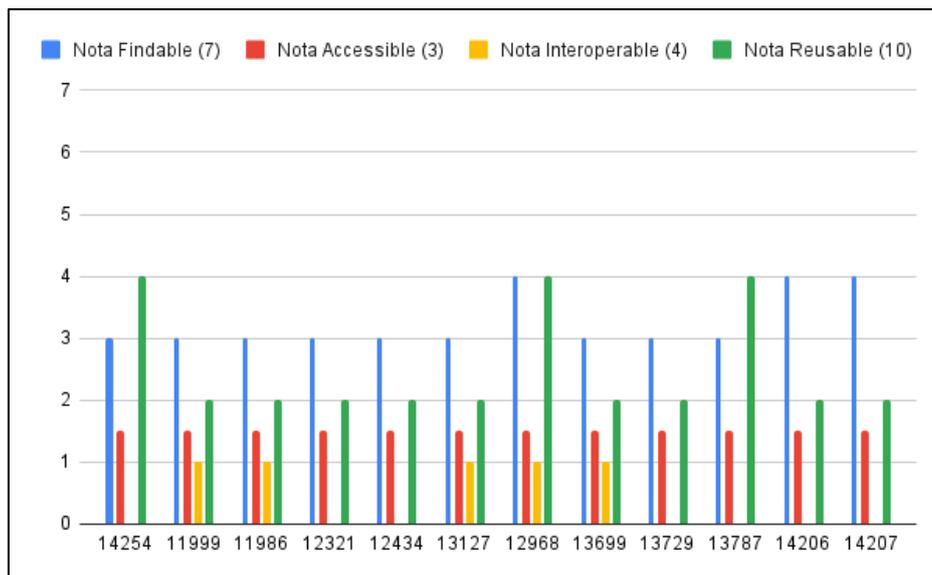
4 ANÁLISE DOS DATASETS DEPOSITADOS NO REPOSITÓRIO INSTITUCIONAL DA UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCAR)

Os 12 *datasets* foram individualmente testados e analisados por meio da ferramenta F-UJI. Na parte de resumo da ferramenta foi possível identificar uma porcentagem geral do nível FAIR do conjunto de dados (0-100%) e no relatório foi possível analisar em maiores detalhes em quais aspectos os dados não aderem aos princípios FAIR. De forma geral, pode-se verificar a pontuação dos *datasets* da amostra em cada um dos quatro princípios no Gráfico 1.

Cada princípio possui uma nota máxima, conforme o número indicado na legenda: *findable* (máximo sete), *accessible* (máximo três), *interoperable* (máximo quatro) e *reusable* (máximo dez). Os números no eixo horizontal correspondem aos códigos dos *datasets*, apontados no Apêndice A.

É possível observar pelo gráfico 1 que nenhum dos *datasets* conseguiu atingir nota máxima, o que se aplica a todos os quatro princípios. A maior nota obtida em *findable* foi 4/7, em *accessible* 1,5/3, em *interoperable* 1/4 e em *reusable* foi 4/10. Os *datasets*, de acordo com o levantamento, obtiveram baixas pontuações quanto aos princípios, principalmente em interoperabilidade, revelando um problema a ser trabalhado.

Gráfico 1 - Pontuação para os quatro princípios FAIR



Fonte: Elaborado pelas autoras.

Os *datasets* também foram avaliados com notas gerais numa escala de 0 a 100% com relação à aderência aos quatro princípios no geral. A média de pontuação foi 31,6%, com apenas um *dataset* atingindo a pontuação mais alta de 43%. Essa porcentagem geral é outro indicativo da falta de aderência dos *datasets* da amostra aos princípios FAIR.

No relatório com maiores detalhes que é dado pela própria ferramenta, o princípio interoperabilidade (*interoperable*) é dividido em três aspectos: 1) Os metadados são representados usando uma linguagem de representação de conhecimento formal (FsF-I1-01M); 2) Metadados usam recursos semânticos (FsF-I1-02M) e 3) Os metadados incluem *links* entre os dados e suas entidades relacionadas (FsF-I3-01M). Assim, os *datasets* da UFSCar foram testados em todos eles e o resultado individual de aderência pode ser visto na Tabela 1.

Tabela 1 - Aderência quanto ao princípio *interoperable*

Código	FsF-I1-01M	FsF-I1-02M	FsF-I3-01M
14254	Não	Não	Não
11999	Não	Não	Sim
11986	Não	Não	Sim
12321	Não	Não	Não
12434	Não	Não	Não
13127	Não	Não	Sim

12968	Não	Não	Sim
13699	Não	Não	Sim
13729	Não	Não	Não
13787	Não	Não	Não
14206	Não	Não	Não
14207	Não	Não	Não

Fonte: Elaborado pelas autoras.

Como indicado pelos princípios FAIR, a interoperabilidade é alcançada quando os metadados utilizam uma linguagem formal, acessível, compartilhada e aplicável para a representação e identificação dos *datasets*. Bem como utilizam vocabulários que seguem os princípios FAIR e incluem referências qualificadas a outros dados e metadados.

Nenhum dos conjuntos de dados foi capaz de aderir ao primeiro aspecto, que diz respeito a uma linguagem de representação de conhecimento formal. Os *datasets* foram testados quanto à existência de metadados estruturados e analisáveis (JSON-LD, RDFa) incorporados ao código XHTML/HTML da página de destino, além de analisáveis por meio de negociação de conteúdo, links digitados ou SPARQL. Segundo Löffler *et al.* (2021), a ideia que vem sendo difundida é adicionar informações descritivas aos dados estruturados, como XML ou HTML, a fim de aumentar a encontrabilidade e interoperabilidade dos dados de pesquisa.

Segundo Silva, Martins e Siqueira (2019, p. 103), o protocolo *Simple Protocol and RDF Query Language* (SPARQL) “[...] se trata de uma linguagem de consulta, que utiliza um conjunto de especificações que fornecem linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web”. Os *datasets* não atenderam a esses requisitos, revelando uma falta de estruturação dos metadados quanto aos padrões necessários para se publicar dados FAIR.

O *Resource Description Framework* (RDF) é um padrão da *World Wide Web Consortium* (W3C) para a descrição conceitual e representação das informações sobre os dados no ambiente Web, fornecendo uma estrutura comum destinada a representar metadados (SILVA; MARTINS; SIQUEIRA, 2019). Segundo Berners-Lee (2009), é um padrão fundamental para que os

dados sejam ligados na Web de forma interoperável e processável por máquina.

O mesmo cenário se repetiu quanto ao segundo aspecto de interoperabilidade: uso de recursos semânticos. Segundo Dias, Anjos e Rodrigues (2019) é importante a existência de instrumentos que padronizem semanticamente os sistemas, permitindo uma efetiva interoperabilidade entre os conjuntos de dados.

A ferramenta buscou por uso de *namespaces*. Segundo Silva, Martins e Siqueira (2019) o uso de ontologias em repositórios é necessário para que o significado pretendido pelo publicador do *dataset* seja o mesmo do entendido pelo usuário dos dados. Além disso, os autores refletem que “Cada ontologia é descrita por um documento apontado por uma URI ou *namespace* e as classes e as propriedades são construídas a partir da concatenação da URI com o nome da classe ou propriedade desejada” (SILVA; MARTINS; SIQUEIRA, 2019, p. 102). Os *datasets* da amostra, portanto, não atenderam aos quesitos expostos, demonstrando uma lacuna quanto à semântica dos dados e metadados.

A semântica é de suma importância uma vez que, conforme dito anteriormente, os princípios FAIR buscam enfatizar o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente. Mas, diferente dos humanos que conseguem interpretar dados por um senso intuitivo de semântica (o significado ou a intenção de um objeto digital), as máquinas precisam de auxílio por meio de aplicação de tecnologias e padrões de interoperabilidade (KALINAUSKAITĖ, 2017).

Quanto ao último aspecto da interoperabilidade, que diz respeito a ligações entre os dados e suas entidades relacionadas, o resultado se mostrou um pouco melhor: cinco dos 12 *datasets* aderem. Nesses cinco *datasets* a ferramenta F-UJI foi capaz de encontrar recursos relacionados que são mencionados explicitamente nos metadados e que são indicados por links ou identificadores legíveis por máquina.

Apesar do resultado incipiente e da baixa aderência quanto ao princípio interoperável, os autores Henning *et al.* (2019) afirmam que, de fato, a interoperabilidade é o elemento mais desafiador das práticas FAIR. A necessidade de habilidades técnicas e de entendimento quanto aos padrões que

devem ser adotados pode ser um dos fatores que contribui para a dificuldade de se alcançar a interoperabilidade desejada em repositórios institucionais.

Quanto à encontrabilidade (*findable*), os *datasets* necessitam de uma referência inequívoca, descrita de forma única e persistente para que sejam identificadas e pesquisáveis. Assim, os *datasets* foram testados em cinco aspectos: 1) Um identificador globalmente exclusivo é atribuído aos dados (FsF-F1-01D); 2) Um identificador persistente é atribuído aos dados (FsF-F1-02D); 3) Os metadados incluem elementos essenciais descritivos (criador, título, identificador de dados, editor, data de publicação, resumo e palavras-chave) para oferecer suporte à encontrabilidade dos dados (FsF-F2-01M); 4) Os metadados incluem o identificador dos dados que descrevem (FsF-F3-01M) e 5) Os metadados são oferecidos de forma que possam ser recuperados programaticamente (FsF-F4-01M). Os resultados para cada um dos aspectos podem ser observados na Tabela 2.

Tabela 2 - Aderência quanto ao princípio *findable*

Código	FsF-F1-01D	FsF-F1-02D	FsF-F2-01M	FsF-F3-01M	FsF-F4-01M
14254	Sim	Não	Sim	Não	Sim
11999	Sim	Não	Sim	Não	Sim
11986	Sim	Não	Sim	Não	Sim
12321	Sim	Não	Sim	Não	Sim
12434	Sim	Não	Sim	Não	Sim
13127	Sim	Não	Sim	Não	Sim
12968	Sim	Sim	Sim	Não	Sim
13699	Sim	Não	Sim	Não	Sim
13729	Sim	Não	Sim	Não	Sim
13787	Sim	Não	Sim	Não	Sim
14206	Sim	Sim	Sim	Não	Sim
14207	Sim	Sim	Sim	Não	Sim

Fonte: Elaborado pelas autoras.

Todos os *datasets* aderiram aos aspectos 1, 3 e 5, conforme mostra a Tabela 2. Ou seja, os *datasets* possuem identificadores exclusivos, os metadados incluem elementos essenciais para a descrição dos objetos digitais e são oferecidos de forma que possam ser recuperados programaticamente. Segundo Henning *et al.* (2019, p. 178) os *datasets* “[...] devem possuir

metadados ricos o suficiente para que, uma vez indexados para um mecanismo de busca, esses metadados possam ajudar o usuário dos dados a encontrá-los”.

Dentre os elementos verificados no Dublin Core, o único não encontrado foi o resumo, que poderia passar a ser preenchido como parte dos elementos essenciais na instituição. Todavia, a validação da existência de metadados essenciais, de acordo com a análise da ferramenta F-UJI, é um bom indicador. Recomenda-se, entretanto, estudos sobre a aplicação de metadados JSON-LD, que diz respeito a um método de codificação de dados vinculados usando JSON, que também foi testado pela ferramenta, mas não identificado.

Com relação ao aspecto FsF-F4-01M, os *datasets* foram testados quanto a metadados fornecidos de uma maneira que os principais mecanismos de pesquisa podem inseri-los em seus catálogos (JSON-LD, Dublin Core, RDFa). O Dublin Core foi identificado, mas a ferramenta não encontrou os *datasets* na plataforma *Datacite.org*⁷, organização líder global do fornecimento dos *Digital Object Identifier* (DOI) para dados de pesquisa. Os identificadores persistentes são amplamente debatidos e cobrados internacionalmente na gestão de dados de pesquisa, o que aponta uma questão a ser trabalhada pelo repositório.

Segundo Neumann e Brase (2014), o uso de *Uniform Resource Locators* (URLs) não parece ser uma boa solução para a preservação e localização devido às inúmeras vezes que páginas ou objetos não podem mais ser encontrados devido às alterações de nome ou servidor. Por isso a importância de identificadores persistentes como os DOIs. De acordo com Dias, Anjos e Rodrigues (2019, p.182) para “[...] maximizar a encontrabilidade dos dados, é muito importante que sejam atribuídos aos conjuntos de dados identificadores persistentes, únicos e globais”.

O “DOI consiste em uma sequência de caracteres única que identifica uma entidade em um ambiente digital - em outras palavras, identifica o próprio objeto e não o local onde está localizado” (NEUMANN; BRASE, 2014, p. 1036, tradução nossa). Dessa forma, mesmo que o objeto (*datasets*) seja movido para outro local (o que leva à mudança da URL), ainda será possível identificar e localizar o objeto de forma persistente.

⁷ Disponível em: <https://datacite.org>.

Além disso, o uso de identificadores persistentes para identificar conjuntos de dados permite que a comunidade científica vá além de periódicos e livros e torne mais conteúdo científico e técnico visível, disponível e pesquisável (NEUMANN; BRASE, 2014), o que também favorece a citação dos dados quando reutilizados por outros pesquisadores. A possibilidade de citá-los e garantir os direitos autorais é de suma importância na popularização dessa prática, incentivando os pesquisadores a adotarem.

Outro aspecto que foi testado, mas não encontrado, diz respeito a “metadados contêm informações relacionadas ao conteúdo de dados (nome do arquivo, tamanho, tipo)”. De acordo com a ferramenta, os metadados não contêm identificadores persistentes que indicam a localização do conteúdo dos dados para *download*. Nenhum elemento de “tamanho”, por exemplo, foi recuperado, o que atrapalha a interpretação automática dos dados pelas máquinas.

É importante frisar que, muitas vezes, o motivo pelo qual é necessário semanas ou meses de esforço técnico para encontrar os dados de interesse para uma pesquisa não é a falta de tecnologia apropriada (KALINAUSKAITĖ, 2017). O maior impedimento é a falta de cuidado adequado com os dados, que não são tratados e depositados de maneira adequada, ou seja, seguindo princípios que melhorem sua encontrabilidade e acesso.

Em relação ao princípio de acesso (*accessible*), os *datasets* devem ser recuperáveis pelo seu identificador e devem utilizar um protocolo aberto, gratuito e universalmente implementável, para sua autenticação e autorização, quando necessário. Isto é, eles devem ser acessíveis, mesmo quando os dados não estão mais disponíveis. Nesse quesito, os dados foram testados em três aspectos: 1- Os metadados contêm o nível de acesso e as condições de acesso aos dados (FsF-A1-01M); 2- Os dados são acessíveis por meio de um protocolo de comunicação padronizado (FsF-A1-03D) e 3- Os metadados são acessíveis por meio de um protocolo de comunicação padronizado (FsF-A1-02M). Segundo Dias, Anjos e Rodrigues (2019, p.183) o “Princípio *accessible* está relacionado ao fato de que os dados deveriam poder ser acessados por protocolos-padrões”.

Nota-se, pela Tabela 3, que todos os *datasets* aderiram ao primeiro e terceiro aspecto. No aspecto FsF-A1-01M o critério “informações sobre

restrições de acesso a direitos podem ser identificadas em metadados” foi verificado e localizado. Segundo Dias, Anjos e Rodrigues (2019, p. 184) “O que pode ser feito com os dados e quem pode usá-los deve ser especificado de forma bastante clara em um termo de uso.” Os *datasets* do repositório institucional da UFSCAR indicam as licenças *Creative Commons* adotadas para cada conjunto de dados. Ou seja, é possível identificar se o acesso é público, embargado ou restrito. Mas, de acordo com o relatório da ferramenta F-UJI, as informações de acesso aos dados não são legíveis por máquina, um aspecto fundamental para a aderência aos princípios FAIR.

Tabela 3 - Aderência quanto ao princípio *accessible*

Código	FsF-A1-01M	FsF-A1-03D	FsF-A1-02M
14254	Sim	Não	Sim
11999	Sim	Não	Sim
11986	Sim	Não	Sim
12321	Sim	Não	Sim
12434	Sim	Não	Sim
13127	Sim	Não	Sim
12968	Sim	Não	Sim
13699	Sim	Não	Sim
13729	Sim	Não	Sim
13787	Sim	Não	Sim
14206	Sim	Não	Sim
14207	Sim	Não	Sim

Fonte: Elaborado pelas autoras.

Em relação ao aspecto FsF-A1-02M, era esperado que os *datasets* aderissem, uma vez que diz respeito aos dados serem recuperados usando um protocolo de comunicação padrão, como o HTTP. Todos os links da página de destino são baseados em protocolos de comunicação da web padronizados, ou seja, os *datasets* são identificados por links que começam com “http://”.

Por fim, quanto ao princípio reutilizável (*reusable*) os *datasets* deveriam ser descritos em um nível de granularidade significativo para que detalhes e atributos sejam representados, bem como, esta descrição deveria ser precisa, clara e relevante, tanto para a comunidade quanto para o domínio de cada repositório de dados. Os dados foram testados em cinco aspectos: 1) Metadados

especificam o conteúdo dos dados (FsF-R1-01MD); 2) Os metadados incluem informações de licença sob as quais os dados podem ser reutilizados (FsF-R1.1-01M); 3) Os metadados incluem informações de proveniência sobre a criação ou geração de dados (FsF-R1.2-01M); 4) Os metadados seguem um padrão recomendado pela comunidade de pesquisa-alvo dos dados (FsF-R1.3-01M) e 5) Os dados estão disponíveis em um formato de arquivo recomendado pela comunidade de pesquisa-alvo (FsF-R1.3-02D).

É possível observar, pela Tabela 4, que os únicos aspectos que todos os *datasets* conseguiram aderir foram a inclusão de informações de proveniência sobre a criação ou origem de dados e o uso de metadados que seguem um padrão recomendado pela comunidade de pesquisa-alvo. Quanto ao aspecto dos metadados especificarem o conteúdo dos dados, uma constatação importante pode ser feita: os *datasets* indicam o tamanho dos arquivos na opção de *download*, mas não em um elemento adequado dos metadados. Isso acabou por impedir a pontuação no aspecto, uma vez que os princípios FAIR buscam enfatizar o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente, o que não seria possível nessa situação.

Tabela 4 - Aderência quanto ao princípio *reusable*

Código	FsF-R1-01MD	FsF-R1.1-01M	FsF-R1.2-01M	FsF-R1.3-01M	FsF-R1.3-02D
14254	Não	Sim	Sim	Sim	Não
11999	Não	Não	Sim	Sim	Não
11986	Não	Não	Sim	Sim	Não
12321	Não	Não	Sim	Sim	Não
12434	Não	Não	Sim	Sim	Não
13127	Não	Não	Sim	Sim	Não
12968	Não	Sim	Sim	Sim	Não
13699	Não	Não	Sim	Sim	Não
13729	Não	Não	Sim	Sim	Não
13787	Não	Sim	Sim	Sim	Não
14206	Não	Não	Sim	Sim	Não
14207	Não	Não	Sim	Sim	Não

Fonte: Elaborado pelas autoras.

Já a questão do tipo de dado, também contemplado dentro do aspecto de metadados especificarem o conteúdo dos dados, há outro conflito: o valor do

elemento “*dc.type*” é “*dataset*”, o que informa apenas que o arquivo é um conjunto de dados. Mas esse valor não informa sobre os tipos de dados contidos no *dataset*, como, por exemplo, “dados observacionais” ou “computacionais” (SAYÃO; SALES, 2015). Pondera-se que esses déficits poderiam ser sanados caso o metadado “*dc.description*”, por exemplo, fosse preenchido com mais elementos.

A contextualização dos dados, fornecendo o máximo de informação possível sobre as especificações, é de suma importância para agregar valor e permitir sua reutilização por outros pesquisadores. Isso porque, segundo Curty (2019, p.179) “[...] a reusabilidade dos dados depende, em grande parcela, da qualidade associada à documentação dos dados disponibilizados para reuso”.

Outro aspecto importante a ser observado é “o formato do arquivo de dados é um formato aberto”. Após o *download* de cada um dos *datasets* foi possível observar que a maioria se trata de formatos proprietários (.xlsx e .docx), ou seja, elaborados em softwares pagos e, portanto, não estão disponíveis sem barreiras de acesso e custo (DEZ ANOS..., 2012).

Para viabilizar a reutilização de dados, Berners-Lee (2009) fez uso de um sistema que classifica em até cinco estrelas a disponibilização dos dados abertos ligados (*Linked Open Data*). A terceira estrela diz respeito ao uso de formatos não proprietários, fundamental para fornecer e democratizar o acesso aos dados de pesquisa, possibilitando seu reuso. Essa recomendação, portanto, não foi aderida pelos *datasets* da amostra.

Quanto à adoção de um padrão recomendado, a ferramenta tentou recuperar informações sobre os metadados no diretório re3data.org⁸, mas o repositório não se encontra mapeado no mesmo. Isso acarreta perdas para a instituição, uma vez que, segundo Sayão e Sales (2015), esses diretórios fornecem informações atualizadas para as comunidades interessadas, além de trazer maior visibilidade para as instituições e suas publicações.

Por fim, há o aspecto de inclusão de informações de licença sob as quais os dados podem ser reutilizados, o que não foi validado em todos os *datasets* pela ferramenta F-UJI. As licenças no repositório da UFSCar são especificadas

⁸ Disponível em: <https://www.re3data.org>.

no elemento “*dc.rights*”. Entretanto, os *datasets* que tiveram o aspecto FsF-R1.1-01M validado utilizaram-se da licença *Creative Commons* em inglês como, por exemplo, “<https://creativecommons.org/publicdomain/zero/1.0/>”. Já os *datasets* que não foram validados pela ferramenta, utilizaram-se da licença *Creative Commons* em português do Brasil, como pode-se ver neste exemplo: “<https://creativecommons.org/licenses/by-nc-nd/3.0/br/>”. Ou seja, recomenda-se que o repositório padronize esta questão. E para que a licença seja validada é necessário optar pelo uso da versão original em inglês. O próprio usuário pode, após ser direcionado pelo link em inglês da licença, mudar o idioma no site *Creative Commons*, permitindo a tradução dos termos de uso, mas sem afetar a validação da licença no repositório pelas ferramentas automáticas.

Todos os *datasets* obtiveram notas baixas no princípio *reusable*, prejudicando o potencial de serem reutilizados por outros pesquisadores. Essa lacuna abre espaço para a discussão de um ponto importante: para que dados de pesquisa sejam realmente interoperáveis e reutilizáveis, “[...] não devem apenas ser devidamente licenciados, mas os métodos para acessá-los e/ou baixá-los também devem ser bem descritos e, de preferência, totalmente automatizados e usando protocolos bem estabelecidos” (KALINAUSKAITĖ, 2017, p. 22). Além disso, segundo Curty (2019, p. 191) “[...] o compartilhamento de dados de pesquisa de modo sistemático e comprometido com o potencial de reúso futuro deixa de ser opcional e passa a ser imperativo”.

Sendo assim, os resultados obtidos por meio da ferramenta demonstram uma falta de aderência aos quatro princípios, com foco em relação à interoperabilidade, onde todas as notas variaram entre 0 e 1 (sendo o máximo 4). Sete *datasets* não aderiram a nenhum dos aspectos de interoperabilidade, o que indica um ponto fraco do repositório institucional da UFSCar. Com relação ao acesso, todos os *datasets* da amostra foram pontuados com a mesma nota: 1,5 de 3 (acesso). Quanto ao princípio encontrável, nove *datasets* receberam pontuação 3 de 7 e apenas três *datasets* obtiveram a nota mais alta 4 de 7. Por fim, quando à reutilização, apenas três *datasets* obtiveram a maior nota de 4 de 10, enquanto todos os outros foram pontuados com 2 de 10. Isso demonstra que o potencial de reutilização desses dados está prejudicado, e recomenda-se que

o repositório da UFSCar estude melhorias para este princípio.

A falta de adoção de um identificador persistente, como o DOI, que é amplamente divulgado internacionalmente, é uma das lacunas que precisam ser trabalhadas. De tal forma, o recomendado é que a UFSCar busque se atualizar sobre a adoção dos princípios FAIR em seu repositório de dados enquanto a coleção ainda é pequena e pode ser mais facilmente aperfeiçoada, evitando assim problemas futuros principalmente de interoperabilidade e acesso aos dados de pesquisa dentro do cenário cada vez mais difundido do *Linked Open Data*. O aperfeiçoamento quanto aos princípios FAIR é um passo inicial para que o repositório possa almejar certificações internacionais de qualidade, trazendo maior visibilidade para a instituição e seus pesquisadores.

5 CONSIDERAÇÕES FINAIS

Os dados de pesquisa não são mais meros subprodutos da pesquisa, ganhando um papel de destaque na ciência contemporânea. Quando bem geridos em repositórios, os dados podem trazer inúmeros benefícios como a economia de recursos, a maior transparência e visibilidade institucional, e a agilidade no ciclo científico. É essencial que os repositórios busquem aderir a princípios e boas práticas, como é o caso dos princípios FAIR, amplamente difundidos por instituições internacionais como a Comissão Europeia.

A média de pontuação quanto à aderência dos *datasets* aos princípios foi de 31,6%, com apenas um *dataset* avaliado em 43% (maior pontuação). Quanto às pontuações individuais para cada um dos quatro princípios, as notas também foram baixas, com as maiores pontuações sendo: 4/7 para *findable*, 1,5/3 para *accessible*, 1/4 para *interoperable* e 4/10 para *reusable*.

Sugere-se, portanto, que o repositório institucional da UFSCar trabalhe no estudo e implementação gradual dos princípios FAIR, podendo dar maior urgência à questão de interoperabilidade, uma vez que este é um dos aspectos mais importantes a se considerar em um repositório de dados de pesquisa.

Além disso, outra prioridade poderia ser a implementação de identificadores persistentes para toda a coleção, garantindo assim o acesso e a citação adequada dos *datasets*, trazendo maior visibilidade para a instituição. A

adoção de identificadores persistentes também é importante para que o repositório siga as tendências e diretrizes internacionais. Outro aspecto importante seria a inserção do repositório em diretórios internacionais como o re3data.org, como foi citado nesta pesquisa. A indexação em grandes diretórios é de suma importância para a visibilidade e reconhecimento acadêmico.

Em trabalhos futuros, propõe-se a avaliação de um número maior de repositórios institucionais de dados de pesquisa, possibilitando comparações e análises de pontos fracos e fortes que podem ser comuns no cenário brasileiro. O aumento da amostra poderia trazer novos *insights* quanto aos aspectos a serem trabalhados pelas instituições que fazem a gestão dos repositórios de dados de pesquisa. Com essas análises os repositórios poderão almejar certificações internacionais, dando maior destaque para os serviços nacionais.

REFERÊNCIAS

- BERNERS-LEE, T. **Linked Data**. 2009. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 18 ago. 2020.
- BERTIN, P.; VISOLI, M.; DRUCKER, D. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de p&d. **PontodeAcesso**, Salvador, v. 11, n. 2, p. 34-48, 2017. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/21449>. Acesso em: 04 abr. 2020.
- CURTY, R. Abordagens de reúso e a questão da reusabilidade dos dados científicos. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/4777>. Acesso em: 17 ago. 2021.
- DEVARAJU, A.; MOKRANE, M.; CEPINSKAS, L.; HUBER, R.; HERTERICH, P.; de VRIES, J.; AKERMAN, V.; L'HOURS, H.; DAVIDSON, J.; DIEPENBROEK, M. From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. **Data Science Journal**, London, v. 20, n. 1, p. 1-14, 2021. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2021-004/>. Acesso em: 15 jun. 2021.
- DEZ ANOS da iniciativa de Budapeste em acesso aberto: a abertura como caminho a seguir. **Budapest Open Access Initiative**, [S. l.: s. n.], 2012. Disponível em: <https://aberta.org.br/dez-anos-da-iniciativa-de-budapeste-em-acesso-aberto-a-abertura-como-caminho-a-seguir/>. Acesso em: 4 mar. 2019.

DIAS, G. A.; ANJOS, R. L.; RODRIGUES, A. A. Os Princípios FAIR: viabilizando o reuso de dados científicos. *In*: DIAS, G. A.; OLIVEIRA, B. M. F. (org.). **Dados científicos: perspectivas e desafios**. Paraíba: Editora UFPB, 2019. p. 177-187. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/359>. Acesso em: 10 ago. 2021.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Gestão de dados**. 2020. Disponível em: <https://fapesp.br/gestaodedados>. Acesso em: 8 abr. 2020.

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, L. R.; SANTOS; L. O. B. da S. 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://periodicos.ufpb.br/index.php/pbcib/article/view/46969/27455>. Acesso em: 10 jun. 2021.

KALINAUSKAITĖ, D. **To be findable, accessible, interoperable and reusable**: language data and technology infrastructure for supporting the FAIR data approach. 2017. Disponível em: <http://ceur-ws.org/Vol-1856/p04.pdf>. Acesso em: 10 jun. 2021.

LÖFFLER, F.; WESP, V.; KÖNIG-RIES, B.; KLAN, F. Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs?. **PLOS ONE**, San Francisco, v. 16, n. 3, p. 1-36, 2021. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0246099>. Acesso em: 20 jun. 2021.

MONTEIRO, E. C. de S. de A.; SANT'ANA, R. C. G. Plano de gerenciamento de dados em repositórios de dados de universidades. **Encontros Bibli: Revista eletrônica de Biblioteconomia e Ciência da informação**, Florianópolis, v. 23, n. 53, p. 160-173, 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p160>. Acesso em: 20 jun. 2021.

NEUMANN, J.; BRASE, J. DataCite and DOI names for research data. **Journal of Computer-Aided Molecular Design**, v. 28, n. 10, p. 1035-1041, 2014. Disponível em: <https://link.springer.com/article/10.1007/s10822-014-9776-5#citeas>. Acesso em: 20 jun. 2021.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Principles and Guidelines for Access to Research Data**. [S.l.: s.n.], 2007. Disponível em: <http://www.oecd.org/sti/sci-tech/38500813.pdf>. Acesso em: 12 out. 2020.

PAVÃO, C. G.; COSTA, J. S. B da; HOROWITZ, Z; FERREIRA, M. K.; CAREGNATO, S. E. Contribución del acceso abierto a la visibilidad de la literatura científica en una institución de educación superior. **E-colabora**, [S.l.], v. 2, n. 3, p. 48-66, 2012. Disponível em: <https://lume.ufrgs.br/handle/10183/37188>. Acesso em: 10 ago. 2020.

PINFIELD, S. Journals and repositories: an involving relationship? **Learned Publishing**, [S.l.], v. 22, n. 3, p. 165-175, 2009. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1087/2009302/epdf>. Acesso em: 05 jan. 2017.

RODRIGUES, M. M. **Repositórios de dados científicos na América do Sul: uma análise da conformidade com os Princípios FAIR**. 2020. 110 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2020. Disponível em: <https://repositorio.ufmg.br/handle/1843/34774>. Acesso em: 20 abr. 2021.

SANTARÉM SEGUNDO, J. E.; SILVA, M. F.; MARTINS, D. L. Revisitando a interoperabilidade no contexto dos acervos digitais. **Informação & Sociedade: Estudos**, João Pessoa, v. 29, n. 2, p. 61-84, 2019. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/38107>. Acesso em: 27 abr. 2022.

SAYÃO, L. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, [S.l.], v. 6, n. 1, 2013. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/119469>. Acesso em: 04 nov. 2019

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015. Disponível em: <http://carpedien.ien.gov.br:8080/handle/ien/1624>. Acesso em: 15 jan. 2021.

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: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 01 nov. 2019.

SILVA, M. F.; MARTINS, D. L.; SIQUEIRA, J. Web semântica em repositórios: ontologia para representação de bibliotecas digitais. **Ciência da Informação em Revista**, Maceió, v. 6, n. 1, p. 99-113, 2019. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/pbcib/article/view/49529>. Acesso em: 10 jun. 2021.

SILVA, L. C.; SANTARÉM SEGUNDO, J. E.; SILVA, M. F. Princípios fair e melhores práticas do linked data na publicação de dados de pesquisa. **Informação & Tecnologia**, Marília/João Pessoa, v. 5, n. 2, p. 81-103, 2018.

Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/44812>.
Acesso em: 07 abr. 2021.

SIMIONATO, A. C. Mapeamento dos metadados para dados científicos. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB)*, 18., 2017. **Anais [...]** Marília: UNESP, 2017. Disponível em: <http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/563/874>. Acesso em: 6 jun. 2018.

TARTAROTTI, R. C. D.; DAL´EVEDOVE, P. R.; FUJITA, M. S. L. Biblioteconomia de dados em repositórios de pesquisa: perspectivas para a atuação bibliotecária. **Informação & Informação**, Londrina, v. 24, n. 3, p. 207-226, 2019. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/pbcib/article/view/53615>. Acesso em: 09 jun. 2021.

TENOPIR, C.; TALJA, S.; HORSTMANN, W.; LATE, E.; HUGHES, D.; POLLOCK, D.; SCHMIDT, B.; BAIRD, L.; SANDUSKY, R.; ALLARD, S. Research Data Services in European Academic Research Libraries. **LIBER Quarterly: The Journal of the Association of European Research Libraries**, [S.l.], v. 27, n. 1, p. 23-44, 2017. Disponível em: <https://liberquarterly.eu/article/view/10706>. Acesso em: 15 ago. 2021.

WILKINSON, M.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.-W.; SANTOS, L. B. da S.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; C´T HOEN, P. A.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; SCHAİK, R. van; SANSONE, S.-A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; LEI, J. van der; MULLIGEN, E. van; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, [S.l.], v. 3, n. 160018, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 20 jan. 2021.

APÊNDICE A - LISTA DE *DATASETS* DA UFSCAR AVALIADOS

Código	Título	Autores	URL
11999	Fotoluminescência em anéis quânticos: oscilações Aharonov-Bohm em partículas neutras	Marques, Gilmar Eugenio; Teodoro, Marcio Daldin; Campo Junior, Vivaldo Leiria; Richard, Victor Lopez; Marega Junior, Euclides; Gobato, Yara Galvão; Iikawa, Fernando; Brasil, Maria Jose Santos Pompeu; AbuWaar, Ziad Y.; Dorogan, Vitaliy G.; Mazur, Yuriy I.; Benamara, Mourad; Salamo, Gregory J.	https://repositorio.ufscar.br/handle/ufscar/11999

11986	Aplicação de estratégias combinatórias para mapeamento de fases quasicristalinas em ligas do sistema Al-Cu-Fe	Kiminami, Claudio Shyinti; Wolf, Witor; Bolfarini, Claudemiro; Botta, Walter	https://repositorio.ufscar.br/handle/ufscar/11986
12321	Data of Cerrado's tree crown networks	Prado, Carlos Henrique Britto de Assis	https://repositorio.ufscar.br/handle/ufscar/12321
12434	Crown network <i>datasets</i> of Caatinga's tree Cenostigma pyramidale (Tul.) Gagnon & G.P.Lewis (Leguminosae - Caesalpinoideae)	Trovão, Dilma Maria de Brito Melo; Costa, Gyslayne Gomes da Silva; Prado, Carlos Henrique Britto de Assis	https://repositorio.ufscar.br/handle/ufscar/12434
13127	Efeito do exercício de alta intensidade na oxigenação muscular cerebral, respiratória e periférica de pacientes com IC e DPOC-IC	Goulart, Cássia da Luz; Arêas, Guilherme Peixoto Tinoco; Caruso, Flávia Rossi; Araújo, Adriana Sanches Garcia de; Tonello, Sílvia Cristina Garcia de Moura; Catai, Aparecida Maria; Beltrame, Thomas; Carvalho Junior, Luiz Carlos Soares de; Santos, Polliana Batista dos; Roscani, Meliza Goi; Mendes, Renata Gonçalves; Arena, Ross; Borghi-Silva, Audrey	https://repositorio.ufscar.br/handle/ufscar/13127
12968	The value of cardiopulmonary exercise testing in determining severity in patients with both systolic heart failure and COPD	Goulart, Cássia da Luz; Caruso, Flávia Rossi; Arêas, Guilherme Peixoto Tinoco; Marinho, Renan Shida; Camargo, Patricia de Faria; Alexandre, Tiago da Silva; Oliveira, Claudio R.; Mendes, Renata Gonçalves; Roscani, Meliza Goi; Borghi-Silva, Audrey; Santos, Polliana Batista dos; Silva, Andréa Lúcia Gonçalves da	https://repositorio.ufscar.br/handle/ufscar/12968
13699	Dataset: improving automatic data extraction from financial statements with clustering analysis	Naldi, Murilo; Ferraz, Victor; Olivato, Gabriel; Magollo, Igor	https://repositorio.ufscar.br/handle/ufscar/13699
13729	Dataset: scalable MapReduce approach for clustering distributed data stream batches with k estimation	Naldi, Murilo; Cândido, Paulo; Silva, Jonathan; Faria, Elaine	https://repositorio.ufscar.br/handle/ufscar/13729
13787	Répteis da região do campus Lagoa do Sino da Universidade Federal de São Carlos, Buri, SP, Brasil	São Pedro, Vinícius Avelar; Nehemy, Ibrahim Kamel Rodrigues	https://repositorio.ufscar.br/handle/ufscar/13787
14206	Can respiratory muscle training change the blood pressure levels in hypertension? A systematic review with meta-analysis	Silva, Claudio Donisete da; Abreu, Raphael Martins de; Santos, Patrícia Rehder dos; Noronha, Marcos de; Catai, Aparecida Maria	https://repositorio.ufscar.br/handle/ufscar/14206
14207	Oxylipin response to acute and chronic exercise: a systematic review	Signini, Étore De Favari; Nieman, David Christopher; Silva, Claudio Donisete da; Sakaguchi, Camila Akemi; Catai, Aparecida Maria	https://repositorio.ufscar.br/handle/ufscar/14207
14254	Níveis de TSH e maturação do sistema auditivo em RN a termo: dados de pesquisa	Germano, Carla Maria Ramos; Pallone, Letícia Valerio; Melo, Debora Gusmão; Gonçalves, Gleice Aline; Navarra, Laura Carvalho; Ferreira, Rodrigo Alves	https://repositorio.ufscar.br/handle/ufscar/14254

FAIR PRINCIPLES AND EVALUATION OF *DATASETS* IN THE UFSCAR INSTITUTIONAL REPOSITORY

ABSTRACT

Objective: This study seeks to evaluate the 12 datasets deposited in the institutional repository of the Federal University of São Carlos (UFSCar) regarding their adherence to the FAIR principles. **Methodology:** The research is quantitative in nature, and in relation to the work method, this research is classified as descriptive and exploratory. For data collection, the auxiliary tool F-UJI was used, which allows the automatic evaluation of datasets and provides feedback reports. **Results:** The overall average score for adherence of datasets to the principles was 30%. Individual scores for each of the four principles were also low, with the highest scores being: 4/7 for findable, 1.5/3 for accessible, 1/4 for interoperable, and 2/4 for reusable. **Conclusions:** The institution is recommended to consider the implementation of the FAIR principles as an improvement of its repository, especially in relation to the interoperable principle, one of the essential aspects for research data repositories.

Descriptors: Research data. Institutional repository. FAIR Principles. Certification.

PRINCIPIOS FAIR Y EVALUACIÓN DE CONJUNTOS DE DATOS EN EL REPOSITORIO INSTITUCIONAL DE UFSCAR

RESUMEN

Objetivo: Este estudio busca evaluar los 12 conjuntos de datos depositados en el repositorio institucional de la Universidad Federal de São Carlos (UFSCar) con respecto a su adherencia a los principios FAIR. **Metodología:** La investigación es de naturaleza cuantitativa, y con relación al método de trabajo, esta investigación se clasifica en descriptiva y exploratoria. Para la recolección de datos se utilizó la herramienta auxiliar F-UJI, que permite la evaluación automática de conjuntos de datos y brinda informes de retroalimentación. **Resultados:** El puntaje promedio general para la adherencia de los conjuntos de datos a los principios fue del 30%. Los puntajes individuales para cada uno de los cuatro principios también fueron bajos, siendo los puntajes más altos: 4/7 para encontrar, 1.5 / 3 para accesible, 1/4 para interoperable y 2/4 para reutilizable. **Conclusiones:** Se recomienda que la institución considere la implementación de los principios FAIR como una mejora de su repositorio, especialmente en relación con el principio de interoperación, uno de los aspectos esenciales para los repositorios de datos de investigación.

Descriptor: Datos de investigación. Repositorio institucional. Principios FAIR. Certificación.

Recebido em: 03.09.2021

Aceito em: 30.04.2022