

REPRODUTIBILIDADE EM E-SCIENCE: UMA VISÃO GERAL DOS CONCEITOS RELACIONADOS E DAS FERRAMENTAS DE SUPORTE MAIS CITADAS

Reproducibility in e-Science: An overview of the related concepts and the most cited support tools

Manuela Klanovicz Ferreira

Universidade Federal do Rio Grande do Sul, PPGCOM, Porto Alegre, RS, Brasil

manuelakf@cpd.ufrgs.br

<https://orcid.org/0000-0002-9089-7725> 

Samile Andrea de Souza Vanz

Universidade Federal do Rio Grande do Sul, PPGCOM, Porto Alegre, RS, Brasil

samile.vanz@ufrgs.br

<https://orcid.org/0000-0003-0549-4567> 

A lista completa com informações dos autores está no final do artigo 

RESUMO

Objetivo: Apresentar conceitos relacionados à reproduzibilidade de pesquisas, como repetitividade e replicabilidade, além de reúso, e também explorar como as ferramentas de suporte à pesquisa citadas na literatura podem ser utilizadas para promover a reproduzibilidade das pesquisas em e-Science.

Método: Pesquisa bibliográfica nas bases *Web of Science*, *Scopus* e *Google Acadêmico*, além de pesquisa documental nos manuais das ferramentas investigadas.

Resultado: Apresentaram-se os conceitos para reproduzibilidade, repetitividade e replicabilidade com discussão de suas origens, sendo também explorado o quanto o reúso se distingue do uso dos dados. A revisão bibliográfica também revelou ferramentas que dão suporte ao desenvolvimento das pesquisas em e-Science e promovem a reproduzibilidade das pesquisas científicas e o reúso de seus dados. As quatro ferramentas mais citadas foram, por ordem, a ferramenta de versionamento de código-fonte *Git Hub*, a ferramenta de *containers Docker*, a ferramenta de gestão de projetos de pesquisa *Open Science Framework (OSF)* e a ferramenta *Jupyter Notebook* para a descrição interativa de programas de processamento de dados.

Conclusões: A ciência precisa ir além de compartilhar seus resultados por meio da literatura, é preciso compartilhar os artefatos utilizados nas pesquisas para alcançar a reproduzibilidade e reúso, conceitos esses que requerem uma definição clara e consensual. Esses artefatos devem ser compartilhados com uma documentação completa e sem erros, essencial para o entendimento do seu significado. Para as áreas de e-Science, que precisam de uma infraestrutura computacional para processar seus grandes volumes de dados, é possível encontrar ferramentas que podem ser utilizadas para facilitar esta documentação. O contexto atual de pesquisa exige que as equipes gestoras da informação e de repositórios de dados conheçam tais conceitos e ferramentas utilizadas pelos pesquisadores.

PALAVRAS-CHAVE: Reproduzibilidade de pesquisas. Reúso de dados. Gestão de dados de pesquisa.

ABSTRACT

Objective: To present concepts related to research reproducibility, such as repeatability and replicability, as well as reuse, and explore how the most cited research support tools in the literature can be used to promote the reproducibility of research in e-Science.

Methods: A bibliographic search was conducted in the Web of Science, Scopus, and Google Scholar databases, along with a document analysis of the manuals of the investigated tools.

Results: Concepts of reproducibility, repeatability, and replicability were presented, with a discussion of their origins, and an exploration of how reuse differs from data use. The literature review also revealed tools that support the development of research in e-Science and promote the reproducibility of scientific research and data reuse, with the four most cited tools were, in order: the source code versioning tool Git Hub, the container

tool Docker, the research project management tool Open Science Framework (OSF), and the Jupyter Notebook tool for interactive data processing program descriptions.

Conclusions: Science must go beyond sharing its results through literature; it is necessary to share the artifacts used in research to achieve reproducibility and reuse, concepts that require clear and consensual definitions. These artifacts must be shared with complete and error-free documentation, which is essential for understanding their meaning. For e-Science fields that require computational infrastructure to process large amounts of data, tools are available to facilitate this documentation. The current research context requires that information management teams and data repository managers understand such concepts and tools used by researchers.

KEYWORDS: Research reproducibility. Data reuse. Research data management.

1 INTRODUÇÃO

Com o aumento da complexidade das pesquisas na *e-Science*, os processos de descrição e reprodução de experimentos vêm tornando-se tarefas cada vez mais intrincadas, pois utilizam-se inúmeras ferramentas ao longo dos estudos. Devido ao grande volume de dados produzidos por estas pesquisas, a disponibilidade de infraestrutura torna-se fundamental. O conjunto de ferramentas necessário inclui serviços computacionais, de manipulação e análise colaborativa e multidisciplinar de dados, além de códigos, softwares e ambientes de pesquisa virtuais (UNESCO, 2021) que permitam o processamento, a análise e a curadoria deste grande volume de dados.

Nesse cenário, o baixo índice de reproduzibilidade das pesquisas foi constatado em enquete realizada pela revista *Nature* em 2016 (Baker, 2016), com 1.576 pesquisadores de áreas como Química, Físicas e Engenharias, Biologia, Medicina e Meio Ambiente. Este estudo questiona se existe uma “crise de reproduzibilidade” na ciência e conclui que mais de 70% dos respondentes falharam em reproduzir experimentos de outros cientistas e mais da metade falhou em repetir seus próprios experimentos. Entre os motivos apontados estavam: o relato seletivo dos resultados, mantendo ocultos resultados discrepantes que tornem as conclusões das pesquisas menos unâimes; a reexecução insuficiente no laboratório de origem da pesquisa; e a metodologia, o código e os dados brutos não disponíveis.

Descrever as pesquisas de forma suficiente para que sejam reproduzíveis por outros pesquisadores e, até mesmo, pelos próprios autores, aumenta o custo, seja monetário, seja em tempo (Baker, 2016; Peng; Hicks, 2021). Também é preciso priorizar para não despendar grandes esforços com dados que não tenham o mínimo de requisitos necessários para serem reutilizados (Vanz *et al.*, 2021). Entretanto, ter estas pesquisas bem descritas, documentadas e preservadas é essencial para permitir que elas sejam



verificadas e que os dados possam realmente ser reutilizados (Peng; Hicks, 2021; Moreau; Wiebles; Boettiger, 2023).

É importante salientar, todavia, que a disponibilização dos dados não promove apenas a reproduzibilidade das pesquisas, ela possibilita o seu reúso, onde o valor atribuído à pesquisa está diretamente relacionado ao potencial de seus dados serem reinterpretados em outras áreas e contextos diferentes daqueles onde originalmente foram gerados, estabelecendo novos padrões de socialização e de trabalho cooperativo independente de barreiras geográficas ou disciplinares (Sayão; Sales, 2014). Contudo, simplesmente abrir os dados não é o suficiente para garantir que eles sejam reusados (Chen *et al.*, 2019), pois a falta de documentação destes dados e de como eles foram utilizados no estudo original pode impedir sua reutilização.

Garantir que os dados sejam reusáveis exige uma ampla e adequada documentação, a qual tem o objetivo de descrever o contexto em que a pesquisa foi executada de forma que seja possível compreender como os resultados foram obtidos e porque a pesquisa foi desenvolvida de determinada maneira. Nesta documentação da pesquisa estão incluídas anotações de laboratório, procedimentos de coleta, limpeza e normalização dos dados, execução de experimentos, inclusive os que não obtiveram sucesso junto aos motivos do insucesso (Munafò *et al.*, 2017), sendo necessário documentar os procedimentos de obtenção e de tratamento dos dados de pesquisa desde a sua coleta (Chen *et al.*, 2019; Peng; Hicks, 2021). Toda esta documentação com informações da pesquisa é importante para permitir o reúso dos dados em outras pesquisas científicas e até mesmo no âmbito do ensino.

O profissional de Ciência da Informação ou bibliotecário, sendo o responsável pela gestão da informação acadêmica produzida na instituição, desempenha um papel relevante na conscientização dos pesquisadores sobre a importância da preservação e do compartilhamento dos dados de pesquisa, além de mostrar como esses dados podem contribuir para o aumento da visibilidade e do impacto da pesquisa institucional. Deve, ainda, promover atividades instrutivas, atuando como um provedor de informações sobre as melhores práticas para a comunicação das pesquisas e de seus dados, bem como facilitador de acesso.

O bibliotecário pode, por exemplo, auxiliar o pesquisador a selecionar o repositório mais indicado para a disponibilização dos seus dados de pesquisa, conforme a área e suas necessidades. Mas, para isso, é desejável que o profissional de Ciência da Informação



acompanhe as tendências nas práticas de gestão de dados, devendo conhecer as principais ferramentas utilizadas para este fim e saber onde obter conselhos de especialistas. No desenvolvimento desta pesquisa, contatou-se que até mesmo conceitos de reprodução e replicação de pesquisas científicas podem divergir, dependendo da área em que são utilizados, ficando evidente a necessidade de uma fonte de informação segura que pode ser unificada na figura do bibliotecário.

Em um contexto em que a *e-Science* está despontando, princípios e padrões estão sendo criados para aumentar a qualidade da comunicação das pesquisas e a tendência é que estratégias e ferramentas evoluam para facilitar que os pesquisadores criem e compartilhem dados que atendam a estes padrões. Em paralelo, as práticas de documentação da pesquisa vêm se ampliando, bem como as ferramentas disponíveis para isso. Neste cenário, surgem algumas perguntas de pesquisa: ferramentas de documentação de pesquisas científicas, comumente citadas na literatura, atendem efetivamente às necessidades de bibliotecários, gestores de informação e equipe responsável pelo repositório de dados? Assim como pesquisadores que lidam com *e-Science*, o profissional da informação minimamente sabe como estas ferramentas funcionam? Este profissional entende a diferença entre os conceitos de reprodução e replicação de pesquisas?

Este estudo tem como objetivo apresentar conceitos relacionados à reproduzibilidade de pesquisa, como repetitividade, replicabilidade e reúso, e também explorar como as ferramentas de suporte à pesquisa citadas na literatura estão sendo utilizadas para promover a reproduzibilidade das pesquisas em *e-Science*, para que profissionais da área de informação possam conhecê-las e ter uma visão geral de como estas ferramentas são aplicadas pelo pesquisador.

Na próxima seção, abordam-se as opções metodológicas desta pesquisa. Em seguida, apresentam-se os conceitos de reproduzibilidade de pesquisa e reúso de dados. Depois, é desenvolvido um abrangente levantamento das ferramentas utilizadas atualmente para o suporte ao desenvolvimento das pesquisas e de sua documentação, particularmente em *e-Science*. Então, tem-se uma seção de discussão, seguida das considerações finais com as principais contribuições desta pesquisa.



2 PROCEDIMENTOS METODOLÓGICOS

Trata-se de uma pesquisa exploratória que, mediante pesquisa bibliográfica e documental, apurou a contribuição das ferramentas disponíveis que compõem a infraestrutura de suporte à pesquisa, particularmente as utilizadas pela e-*Science*, para a reproduzibilidade das pesquisas e reúso dos dados ao longo do tempo.

A primeira etapa deste trabalho englobou a realização de uma busca das pesquisas mais atuais publicadas na literatura, a fim de investigar o estado da arte das ferramentas utilizadas atualmente para dar suporte à documentação da pesquisa científica e de seus dados, com o propósito de promover o reúso e a reproduzibilidade das mesmas. Para tanto, procurou-se a seguinte expressão de busca nas bases *Web of Science* (WoS) e *Scopus*: (*reproducibility OR reproducible OR reuse OR “secundary use”*) AND “*research data*”, delimitada a um período de até cinco anos, contando, então, com publicações de 2018 a 2022 que tenham sido indexadas até 25 de março deste último, quando a busca foi realizada. Foram aplicados sucessivos ciclos de triagem nos resultados, inicialmente eliminando os duplicados, seguido da análise de títulos, palavras-chave e resumo, de forma a selecionar os resultados relacionados à importância de documentar as pesquisas e, principalmente, que descrevessem a utilização de estratégias ou ferramentas de suporte à documentação das pesquisas e seus dados. As referências citadas nas publicações encontradas também foram avaliadas, resultando na identificação e inclusão nesta pesquisa de publicações mais antigas, porém bastante relevantes, por fornecerem o embasamento para o estado da arte dos últimos anos. Ao longo do desenvolvimento desta pesquisa, entretanto, novos trabalhos abordando o contexto acima descrito foram sugeridos e, quando pertinentes, referenciados neste estudo.

Com a leitura dos textos selecionados no levantamento bibliográfico, verificou-se a falta de uma definição clara dos conceitos de reproduzibilidade e uma abordagem distinta para o conceito de reúso, levando a uma discussão sobre estes conceitos e de como eles se relacionam entre si, apresentada na seção 3 deste artigo.

A revisão bibliográfica também permitiu a identificação de ferramentas de suporte à documentação e gestão da pesquisa científica, sendo feito posteriormente uma classificação da quantidade de citações de cada uma destas ferramentas no Google Acadêmico em combinação com pelo menos um dos seguintes termos: “*reproducible research*”, “*reproducible science*”, “*research reproducibility*” ou “*scientific reproducibility*”.



Essas consultas foram realizadas em outubro de 2023 por meio da interface do programa *Publish or Perish* (Harzing, 2021). Os termos foram inseridos na janela de busca do programa e selecionou-se o Google Acadêmico como base de dados para a pesquisa. Os resultados foram salvos e classificados segundo o ano de publicação dos documentos retornados pela busca, permitindo gerar o Gráfico 1.

Neste ponto, as ferramentas mais citadas no levantamento bibliográfico tiveram seu funcionamento investigado em profundidade, sendo elas: *Docker*, *Singularity*, *Docker Hub*, *Git Hub*, *Jupyter Notebook* e *Jupyter Hub*, *Google Colab*, *Code Ocean*, *Whole Tale* e *Open Science Framework* (OSF). O procedimento aplicado para esta análise foi a instalação para o caso do *Jupyter Notebook* e *Jupyter Hub* e criação de cadastro para as demais ferramentas analisadas. Além disso, foram elaborados testes, como, por exemplo, o *Jupyter Notebook* no *Google Colab*¹ e o projeto no OSF², simulando documentação de pesquisas, com o intuito de compreender como as ferramentas funcionam. Estas ferramentas estão descritas na seção quatro, sendo que as informações foram coletadas entre setembro e outubro de 2023.

3 REPRODUTIBILIDADE E REÚSO

Quando o assunto é reproduzibilidade, muitos autores apontam para uma atual crise (*reproducibility crisis*) tendo em vista que a maioria dos pesquisadores encontra dificuldade em reproduzir resultados de publicações (Baker, 2016; Barba, 2018; Figueiredo Filho *et al.*, 2019). Nestes trabalhos, é interessante notar que a confiabilidade das pesquisas é, geralmente, apontada como a principal preocupação em relação a esta crise de reproduzibilidade, alegando-se que pesquisas que podem ser reproduzidas são mais confiáveis, excetuando-se casos em que a reprodução total não seria possível devido a situações únicas como nos casos de estudos epidemiológicos referentes, por exemplo, à pandemia da COVID-19³. Mas até mesmo nestes casos, a reprodução parcial é defendida como uma forma de aumentar o crédito dos resultados.

Entretanto, o uso moderno do termo “pesquisa reproduzível” (*reproducible research*) foi originalmente aplicado por Claerbout e Karrenbach (1992) fazendo alusão à

¹ Teste Google Colab com *Jupyter Notebook* - <https://colab.research.google.com/drive/1bK68KWj6SXhBQMNeMEE1YNnyV15ljd6?usp=sharing>

² Projeto de pesquisa de teste no OSF - <http://dx.doi.org/10.17605/OSF.IO/U3N5G>

³ Pandemia de COVID-19 - https://pt.wikipedia.org/wiki/Pandemia_de_COVID-19



transparência (Goodman; Fanellis; Ioannidis, 2016; Barba, 2018), a fim de promover um maior entendimento por parte dos leitores das publicações científicas. Claerbout e Karrenbach (1992, p. 2), no contexto da sismologia, defendiam que “O principal objetivo de uma publicação científica é ensinar novos conceitos, mostrar as implicações resultantes destes conceitos em uma ilustração e prover detalhes suficientes para tornar o trabalho reproduzível”.

Posteriormente Peng, Dominici e Zeger (2006, p. 1) fizeram uma distinção do termo replicar, alegando que “A evidência científica é fortalecida quando descobertas importantes são replicadas por vários investigadores independentes usando dados, métodos analíticos, laboratórios e instrumentos independentes”.

Além de replicabilidade, existem outros termos relacionados à reproduzibilidade, como a repetibilidade, sendo que, dependendo da área de pesquisa, o significado destes termos pode variar (Barba, 2018). A definição mais difundida de reproduzibilidade das pesquisas está relacionada à definição de reúso de dados, pois é mediante o acesso aos dados bem documentados que uma pesquisa pode ser reproduzida.

A Association for Computing Machinery (ACM, 2020)⁴ apresenta a seguinte definição:

Reproduzibilidade (Equipes diferentes, mesma configuração de experimentos): A medição pode ser obtida com precisão por uma equipe diferente usando o mesmo procedimento de medição, o mesmo sistema de medição, sob as mesmas condições de operação, no mesmo ou em um local diferente em várias tentativas. Para experimentos computacionais, isso significa que um grupo independente pode obter o mesmo resultado usando os próprios artefatos do autor (ACM, 2020, tela 2, tradução nossa).

Além de reproduzibilidade, a ACM também conceitua os termos replicabilidade e repetibilidade, com base na origem da configuração dos experimentos e na equipe que os realiza:

Repetibilidade (mesma equipe, mesma configuração de experimentos): A medição pode ser obtida com precisão pela mesma equipe usando o mesmo procedimento de medição, o mesmo sistema de medição, sob as mesmas condições de operação, no mesmo local em várias tentativas. Para experimentos computacionais, isso significa que um pesquisador pode repetir com segurança seu próprio experimento (ACM, 2020, tela 2, tradução nossa, grifo nosso).

Replicabilidade (Equipes diferentes, diferentes configurações de experimentos): A medição pode ser obtida com precisão por uma equipe diferente, um sistema de medição diferente, em um local diferente em várias tentativas. Para experimentos computacionais, isso significa que um grupo

⁴ Aprovada em 24 de agosto de 2020.

independente pode obter o mesmo resultado usando artefatos que eles desenvolvem de forma totalmente independente (ACM, 2020, tela 2, tradução nossa, grifo nosso).

O Quadro 1 ilustra, resumidamente, as características de cada uma das definições apresentadas acima. Nele, fica explícito que a reproduutibilidade é aquela que é executada por uma equipe diferente daquela onde o experimento ou pesquisa original ocorreu, entretanto, utilizando os mesmos dados, os quais precisam ter sido compartilhados de forma o mais inteligível possível.

Quadro 1 - Comparativo entre Repetibilidade, Reproduutibilidade e Replicabilidade

	Repetibilidade	Reproduutibilidade	Replicabilidade
Equipe	Mesma	Diferente	Diferente
Artefato	Mesmo	Mesmo	Diferente

Fonte: Elaborado pelas autoras.

Cabe ressaltar que os conceitos de reproduutibilidade, replicabilidade e repetibilidade não são unâimes dentro da ciência (Barba, 2018; National Academies of Sciences, Engineering, and Medicine, 2019). A própria ACM originalmente definiu, em junho de 2016⁵, conceitos equivocados de reproduutibilidade e replicabilidade, conforme nota publicada junto às definições destes conceitos atualizadas em agosto de 2020 (ACM, 2020, tela 3, tradução nossa), onde afirma que para harmonizar sua terminologia com aquelas usadas na comunidade de pesquisa científica mais ampla, a Associação concordou com a recomendação de trocar os termos “reproduutibilidade” e “replicação”.

Um estudo de Barba (2018) apontou a discrepância da antiga definição da ACM para os termos replicabilidade e reproduutibilidade em relação a diversas áreas do conhecimento. A autora classificou em três as correntes de definição:

A) não faz distinção entre os termos **reproduutibilidade** e **replicabilidade**, caso da Ciência Política e da Economia;

B1) considera **reprodutível** a pesquisa em que o autor fornece todos os dados e códigos-fonte para uma nova execução das análises, recriando os resultados; e **replicação** quando um estudo chega aos mesmos resultados partindo de novos dados e executando

⁵ Situação da página das definições da ACM em julho de 2016 recuperada no *Web Archive*, onde consta 8 de junho de 2016 como data de divulgação da primeira versão das definições

<https://web.archive.org/web/20160702213802/http://www.acm.org/publications/policies/artifact-review-badging>

novas análises; sendo B1 adotada pelas áreas de Processamento de Sinal, Computação Científica, Epidemiologia, Estudos Clínicos, Clínica Médica, Fisiologia (Neurofisiologia), Biologia Computacional, Biomedicina e Estatística;

B2) considera **reprodutibilidade** quando outros pesquisadores alcançam os mesmos resultados utilizando outros dados e métodos, enquanto a **replicabilidade** ocorre quando pesquisadores alcançam os mesmos resultados por meio dos artefatos originais; com B2 adotada, à época, apenas pela área de Microbiologia e Imunologia, que aderiu à definição da Federation of American Societies for Experimental Biology (FASEB), e de Ciência da Computação, com a antiga definição da ACM. Assim como a ACM, a FASEB trocou suas definições de replicabilidade e repetibilidade originais (FASEB, 2016), revendo sua posição em 2019⁶, quando adotou as definições conforme B1.

Sendo assim, a partir da revisão de literatura, atualmente a maioria das áreas adota os conceitos de B1, onde reproduzir corresponde ao momento em que outro pesquisador alcança os mesmos resultados com os mesmos artefatos e replicar é quando outro pesquisador alcança os mesmos resultados com novos artefatos.

Cabe salientar que reproduzir um experimento não quer dizer que ele está correto e é verdadeiro (Peng; Hicks, 2021), pois parte-se dos mesmos artefatos utilizados na pesquisa original com o objetivo de encontrar os mesmos resultados. Para Rougier *et al.* (2017), o objetivo da reprodução é verificar que o protocolo computacional que leva aos resultados foi corretamente registrado. Já na definição de replicação, quando se chega a um mesmo resultado, ou seja, a resposta de uma mesma questão de pesquisa, a partir de outro método, dados e análise independente, a validade do resultado é maior, pois o resultado independente tem menores chances de compartilhar qualquer vício ou viés com a pesquisa original (Peng; Dominici; Zeger, 2006; Goodman; Fanellis; Ioannidis, 2016).

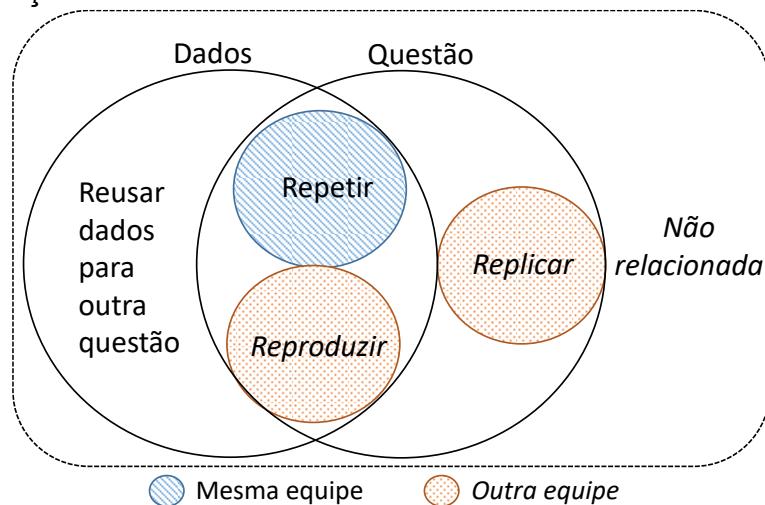
Independente das definições para reprodução e replicação adotadas por cada autor, é interessante notar os termos utilizados nessas definições e averiguar quais deles fazem parte da mesma dimensão: i) equipe (usuário); ii) artefatos, dados ou método (código ou descrição do método) (ACM, 2020) e iii) resultado (propósito, questão) (Rougier *et al.*, 2017).

Em Chen *et al.* (2019), Van de Sandt (2019), Pasquetto, Randles e Borgman (2017) e Caregnato, Rocha e Gabriel Junior (2021) ocorre uma intersecção entre as terminologias

⁶ National Academies Releases Report on Reproducibility and Replicability in Science - <https://www.faseb.org/journals-and-news/washington-update/national-academies-releases-report-on-reproducibility-and-replicability-in-science>

relacionadas à pesquisa reproduzível e ao reúso dos dados. Considerando as três dimensões citadas anteriormente, para as definições de reúso é acrescentado o tempo, resultando em um espectro que varia desde a repetição, reprodução e replicação da pesquisa, perpassando o reúso dos dados até o nível de uma pesquisa não relacionada, quando uma pesquisa nova não utiliza nem dados ou métodos de pesquisas anteriores e também possui uma questão distinta. Este relacionamento das definições referentes à reproduzibilidade e ao reúso com as dimensões utilizadas nestas definições é ilustrado na Figura 1.

Figura 1 – Relação entre reúso de dados e conceitos referentes à reproduzibilidade



Apesar do reúso de dados estar presente em diferentes modelos do ciclo de vida dos dados de pesquisa como resultado esperado da gestão destes ativos científicos, sendo frequentemente mencionado na literatura, não é sempre que o termo reúso de dados é definido explicitamente. O levantamento sistemático feito por Van de Sandt *et al.* (2019) de obras que contivessem o termo reúso (*reuse*) ou “uso secundário” (*secondary use*) de dados de pesquisa mostrou que apenas 20 dos 65 trabalhos encontrados mencionavam uma definição de reúso, seja por citação de outro autor, seja pela proposta de uma definição própria.

A partir das definições mencionadas nos trabalhos recuperados, os autores identificaram quatro dimensões comumente utilizadas para distinguir o uso do termo reúso de dados (Van de Sandt *et al.*, 2019): caráter dos dados, usuário, propósito e tempo. Para cada uma destas dimensões, foram exemplificadas situações nas quais as dimensões não poderiam ser utilizadas para distinguir uso de reúso, surgindo questões como: quando um

subconjunto dos membros originais de um grupo de pesquisa (dimensão de usuário) faz uso dos mesmos dados do grupo original, seria reúso? Ou para os dados coletados por sensores sem um propósito inicial, seria apropriado falar de reúso devido à mudança de propósito, se este não estava determinado inicialmente? Ou quando um pesquisador publica antes do autor original dos dados, devido aos trâmites de publicação e até mesmo à complexidade das pesquisas, quem estaria reusando os dados neste caso? Desta forma, os autores definiram que (re)uso é o uso de qualquer artefato da pesquisa, independentemente de quando é usado, do propósito para o qual é usado, das características dos dados ou de seu usuário, ou seja, concluíram que o reúso deve ser considerado uso e que o impacto da autocitação dos dados deve ser o mesmo daquele medido na citação dos dados por terceiros.

Olhando pelo aspecto da documentação das pesquisas, considerar reúso como uso implicaria em documentar as pesquisas em desenvolvimento tão bem como as pesquisas que foram finalizadas e estão sendo compartilhadas, ou seja, ter identificador único dos dados de entrada utilizados, antes mesmo de finalizadas as análises, descrever os procedimentos da pesquisa detalhadamente de forma compartilhável durante o decorrer da mesma e não apenas no final da pesquisa ou no momento em que for gerada uma publicação. Porque, mesmo que esta pesquisa venha a ser mantida restrita, sua completa documentação é importante para que até mesmo os próprios autores possam comprehendê-la no futuro.

A discussão sobre a definição de reprodutibilidade e reúso, suas vantagens e desvantagens, é essencial para que possamos identificar, de forma clara, em que circunstâncias a documentação dos dados de forma completa e sem erros é importante para a descrição, validação e evolução das pesquisas. Na seção seguinte, serão apresentadas as ferramentas de suporte às pesquisas, que foram identificadas na revisão bibliográfica deste trabalho, e como elas podem ser utilizadas para viabilizar uma melhor documentação das pesquisas.

4 FERRAMENTAS DE SUPORTE À PESQUISA

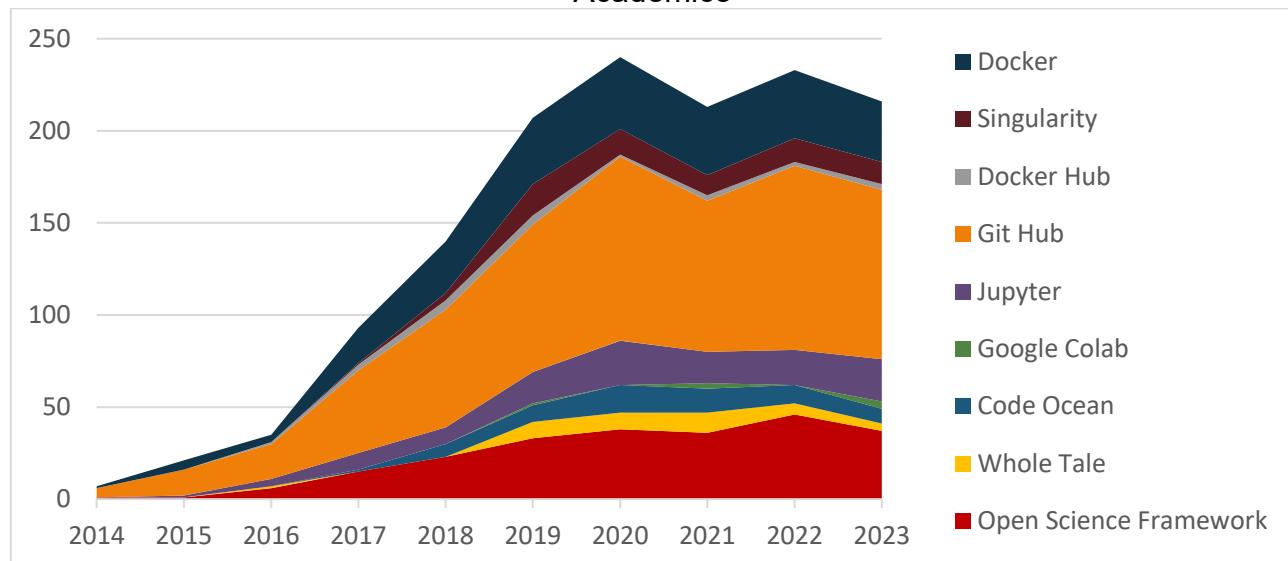
Apesar da crescente conscientização pela documentação das pesquisas para permitir sua reprodutibilidade e o reúso dos dados, a efetiva aplicação destas práticas é considerada uma tarefa dispendiosa (Baker, 2016; Peng; Hicks, 2021). Para facilitar este



processo, existem ferramentas de suporte à pesquisa utilizadas tanto no planejamento como na documentação de fluxos de trabalho e contribuições da pesquisa. Também encontram-se na literatura as chamadas “ferramentas de reproduzibilidade” (Trisovic *et al.*, 2020), estas mais voltadas à documentação do ambiente computacional de desenvolvimento dos experimentos. Este conjunto de ferramentas tem o objetivo de facilitar a documentação das pesquisas durante o seu desenvolvimento, utilizando, para isso, diversas estratégias.

O Gráfico 1 mostra a evolução das menções às ferramentas de suporte à documentação das pesquisas no Google Acadêmico, resultando das buscas descritas nos procedimentos metodológicos.

Gráfico 1 – Evolução das menções a ferramentas de suporte às pesquisas no Google Acadêmico



Fonte: Elaborado pelas autoras com a ferramenta *Publish or Perish* (Harzing, 2021). Dados coletados em outubro de 2023. *Jupyter* contém menções a *Jupyter Notebook* e *Jupyter Hub*.

Conforme evidenciado no respectivo Gráfico, as menções destas ferramentas, acompanhadas de termos relacionados à reproduzibilidade das pesquisas, tiveram um grande aumento a partir de 2016, alcançando o seu ápice e o que pode ser considerado um platô, a partir de 2020. As referências às ferramentas nos trabalhos recuperados vão desde o contexto de discussão sobre sua utilidade na reproduzibilidade das pesquisas até sua concreta utilização como fonte do compartilhamento dos dados dos respectivos trabalhos. O Gráfico 1 mostra que estas ferramentas estão sendo adotadas na reproduzibilidade de pesquisas científicas de forma mais ampla há cerca de oito anos apenas.

A seguir, as nove ferramentas citadas no Gráfico 1 são descritas em quatro seções, agrupadas conforme suas semelhanças de funcionamento.

4.1 CONTAINERS: PRESERVANDO O AMBIENTE COMPUTACIONAL DE PESQUISA: *DOCKER, SYNGULARITY E DOCKER HUB*

Para preservar um ambiente computacional, é preciso encapsular os softwares instalados, de forma que seja possível executar este mesmo ambiente computacional em outra máquina, ou seja, outro hardware.

A conteinerização dos softwares permite esse encapsulamento (Mouat, 2016), sendo que as tecnologias de *containers* evoluíram na última década e entre as mais difundidas estão o *Docker*⁷ e o *Singularity*⁸, ambas disponíveis de forma gratuita. Devido ao fato do *Docker* ter suporte em computadores Windows e MacOS, além de computadores Linux, para o qual ambas as tecnologias citadas foram originalmente desenvolvidas, o *Docker* possui uma maior quantidade de usuários, o que permitiu a criação de um ecossistema maior que conta com mais ferramentas para a sua utilização e que o suportam. Por este motivo, o *Docker* será a tecnologia referida neste trabalho ao citar *containers*.

Pode-se citar três principais vantagens dos *containers* para a reprodução de ambientes computacionais (Moreau; Wiebels; Boettiger, 2023; Mouat, 2016): a) **documentação de camadas em histórico no formato texto**, a partir do histórico é possível, em teoria, recriar o *container*, reexecutando os passos para a criação de cada camada de software que foi adicionada ao ambiente computacional original, mas os softwares precisam estar disponíveis externamente; b) **encapsulamento em imagens, sem dependência externa**, pois a imagem que pode ser salva contém todos os softwares instalados, variáveis de ambiente configuradas, arquivos e organização de pastas em um arquivo de imagem de *container*; c) **utiliza imagem preexistente e acrescenta camadas**, onde o conceito de camadas, que embasa a criação dos *containers*, culminou com a criação de ferramentas de versionamento de imagens, como o *Docker Hub*⁹, um portal de utilização gratuita por meio da qual é possível criar um *container* a partir de uma imagem pública disponível neste repositório de imagens, facilitando a criação do *container*. Esta ferramenta

⁷ Docker (lançado em 2013) - <https://www.docker.com>

⁸ Singularity (lançado em 2015) - <https://sylabs.io/guides/3.5/user-guide/introduction.html>

⁹ Docker Hub (lançado em 2013) - <https://hub.docker.com>



também permite salvar a imagem com todo o ambiente computacional em que um pesquisador desenvolveu seus experimentos a fim de que outro pesquisador obtenha esta imagem depositada e consiga reexecutar o *container* contendo os experimentos do pesquisador autor, podendo, inclusive, submeter uma nova versão da imagem no *Docker Hub*, contendo suas próprias alterações nos dados e experimentos do pesquisador autor.

Os *containers* estão sendo utilizados para aumentar a transparência, portabilidade e reproduzibilidade das pesquisas em algumas áreas como ciência de dados (Boettiger, 2015) e pesquisas em engenharia de software (Cito; Gall, 2016). Executar as análises em um *container* aumenta a confiabilidade do fluxo de trabalho, visto que é possível executar os experimentos independentemente das configurações e dependências disponíveis no computador do autor.

4.2 *GIT HUB: VERSIONANDO CÓDIGO E GERENCIANDO TAREFAS*

O versionamento de código-fonte é um instrumento imprescindível entre os desenvolvedores de software. Ao longo da sua evolução, o versionamento foi implementado utilizando diversas abordagens e, atualmente, existem algumas ferramentas que dão suporte a esta atividade, sendo o *Git*¹⁰ e o *Subversion (SVN)*¹¹ bastante difundidos e gratuitos.

O *Git Hub*¹² é uma plataforma que permite uma visão gráfica, por meio de um navegador, do repositório de fontes criado com a ferramenta *Git*, além de hospedar este repositório de forma gratuita. Outra funcionalidade disponível na plataforma do *Git Hub* é a de gestão de tarefas, onde é possível criar, atribuir, documentar e controlar o andamento de tarefas entre o grupo de membros de um projeto (Perez-Riverol *et al.*, 2016). Desta forma, é possível acompanhar quem está executando cada papel dentro do projeto, facilitando também o resgate da contribuição de cada um.

Conforme ilustrado no Gráfico 1, o *Git Hub* é a ferramenta mais citada no Google Acadêmico associada à reproduzibilidade de pesquisas. Apesar de não ter sido originalmente desenvolvida para atender a demanda de reproduzibilidade, suas funcionalidades mostram-se úteis nas pesquisas científicas. Figueiredo Filho *et al.* (2019) recomendam o uso do *Git Hub* como uma ferramenta de reproduzibilidade e Rougier *et al.*

¹⁰ *Git* - <https://git-scm.com/>

¹¹ *SVN* - <https://subversion.apache.org/>

¹² *Git Hub* (lançado em 2008) - <https://github.com/>

(2017) descrevem como o periódico *ReScience* utiliza o *Git Hub* para compartilhar cada implementação dos estudos computacionais publicados.

4.3 FERRAMENTAS JUPYTER, GOOGLE COLAB, WHOLE TALE E CODE OCEAN: PRESERVANDO O FLUXO DE EXPERIMENTOS

Para permitir uma descrição amigável dos passos de uma pesquisa, de modo que estejam acompanhados dos comandos utilizados para coletar, processar e analisar os dados, bem como para gerar os gráficos e visualizações dos resultados, temos os *Jupyter Notebook*¹³, um interpretador interativo gratuito para diversas linguagens de programação. Por meio desta tecnologia, é disponibilizado um portal, acessível através do navegador, onde é possível criar um arquivo .ipynb e inserir texto formatado e imagens, por exemplo, descrevendo um experimento, junto a trechos de código que podem ser executados nesta mesma página, através do navegador, utilizando um botão semelhante ao botão de “play”. Desta forma, um pesquisador pode visualizar a execução e a saída de um código-fonte, seja ela a coleta e o processamento de dados ou a geração de gráficos de visualização. A evolução desta ferramenta é o *Jupyter Lab*¹⁴, que pode ser instalado localmente ou ser acessado de forma *on-line*.

A plataforma *Google Colab*¹⁵ é semelhante ao *Jupyter Lab*, mas apenas disponível *on-line*. O *Google Colab* é um serviço com opções gratuitas e pagas para a criação de experimentos implementados e documentados por meio de *Jupyter Notebook*. Ele inicia seu ambiente a partir de um arquivo .ipynb salvo no *Google Drive*.

As versões gratuitas e *on-line* do *Jupyter Lab* e *Google Colab* não guardam os arquivos de entrada e nem os de saída, possivelmente gerados pelos experimentos, preservando apenas o arquivo do *Jupyter Notebook*.

Existem ferramentas que permitem o salvamento permanente tanto dos *Jupyter Notebook* como dos arquivos de dados, seja de entrada ou saída dos experimentos, como o *Whole Tale*¹⁶ (Cheifet, 2021) e o *Code Ocean*¹⁷ (Brinckman *et al.*, 2019), os quais são serviços *on-line* bastante semelhantes e com possibilidades de uso gratuito. Estes serviços

¹³ *Jupyter Notebook* (lançado em 2014) - <https://jupyter.org/try-jupyter/retro/notebooks/?path=notebooks/Intro.ipynb>

¹⁴ *Jupyter Lab* (lançado em 2018) - <https://jupyter.org/try-jupyter/lab?path=notebooks%2FIntro.ipynb>

¹⁵ *Google Colab* (lançado em 2017) - <https://colab.research.google.com/>

¹⁶ *Whole Tale* (lançado em 2017) - <https://whotentale.org/>

¹⁷ *Code Ocean* (lançado em 2017) - <https://codeocean.com/>



são classificados como RaaS (*Reproducibility as a Service*), pois fornecem acesso a recursos e ferramentas para que pesquisadores possam desenvolver suas pesquisas em ambientes controlados que garantam o rastreamento da procedência, atribuição de identificador único, ao mesmo tempo em que gerenciam a infraestrutura (Hernández; Colom, 2023). Como exemplo de utilização de ferramentas de RaaS, Hernández e Colom (2023) citam a disponibilização do código-fonte dos artigos disponíveis no Institute of Electrical and Electronics Engineers (IEEE) no *Code Ocean*¹⁸, onde esses códigos podem ser executados, modificados e descobertos, pois passam a possuir *Digital Object Identifier* (DOI). Estes serviços permitem a execução do ambiente dos experimentos salvo e compartilhado apenas em sua própria plataforma *on-line* (Malik, 2023).

4.4 OPEN SCIENCE FRAMEWORK: FERRAMENTA DE GESTÃO DE PROJETOS DE PESQUISA

A ferramenta *Open Science Framework*¹⁹ (OSF) é um exemplo de ferramenta de gestão de projeto de pesquisa com potencial de servir a qualquer área do conhecimento (Potterbusch; Lotrecchiano, 2018). Diferentemente das ferramentas de reproduzibilidade, esta assemelha-se a uma ferramenta gratuita de gestão de projetos que pode conectar-se com outras ferramentas gratuitas como *Git Hub*, *Dataverse*²⁰, *Google Drive*, entre outras. Além disso, o OSF permite o cadastramento de um time de pesquisa utilizando o *Open Researcher and Contributor ID* (ORCID)²¹ e a delegação de tarefas, ao mesmo tempo que documenta qual colaborador da pesquisa executou determinada tarefa, ao criar uma linha do tempo de atividades no projeto de pesquisa dentro da OSF e nas ferramentas que podem ser conectadas com ele.

Adicionalmente, esta ferramenta permite a criação de um DOI para referenciar inequivocamente o projeto de pesquisa cadastrado, sendo mais um argumento para que toda a documentação relacionada à pesquisa, inclusive o projeto, sejam preservados e compartilhados.

¹⁸ *Code Ocean* do IEEE - <https://codeocean.com/signup/ieee>

¹⁹ *Open Science Framework* (lançado em 2013) - <https://osf.io>

²⁰ Projeto *Dataverse* - <https://dataverse.org/>

²¹ Identificador único de pesquisadores ORCID - <https://orcid.org/>



5 DISCUSSÃO

O compartilhamento e reúso de dados, assim como a reproduzibilidade das pesquisas, têm alcançado uma crescente relevância na ciência, sendo o entendimento destes conceitos fundamental para que os pesquisadores e profissionais da informação conheçam as diversas possibilidades de aplicação dos dados de pesquisa. Para o reúso acontecer, os dados devem ser de boa qualidade, dispor de metadados e documentação para auxiliar em sua compreensão. Nesse sentido, o conhecimento acerca das ferramentas de documentação também é fundamental para o profissional da informação que atua no contexto dos repositórios de dados de pesquisa.

Ferramentas gratuitas como *GitHub*, *Docker*, *Open Science Framework* (OSF) e as ferramentas *Jupyter*, por exemplo, tiveram um grande aumento de suas citações em trabalhos de pesquisa associados à reproduzibilidade das pesquisas e reúso dos dados. O *GitHub* facilita a colaboração entre pesquisadores, ao permitir o versionamento do código-fonte e o controle das mudanças ao longo do tempo, promovendo a transparência no desenvolvimento e implementação dos experimentos (Figueiredo Filho *et al.*, 2019). Já o *Docker*, ao criar *containers* que preservam o ambiente computacional, permite que os pesquisadores reproduzam experimentos com alta fidelidade, inclusive preservando as versões exatas dos programas e dependências necessárias (Cito; Gall, 2016; Moreau; Wiebels; Boettiger, 2023). O OSF, por sua vez, oferece uma plataforma robusta para o gerenciamento de projetos de pesquisa, onde é possível armazenar, compartilhar e documentar os dados de forma integrada, proporcionando um ambiente colaborativo que facilita o acesso e reutilização dos dados por outros pesquisadores (Potterbusch; Lotrecchiano, 2018). As ferramentas *Jupyter*, que permitem o uso interativo de notebooks com código executável, texto e visualizações, oferecem uma maneira acessível de documentar e compartilhar a execução de experimentos de forma interativa, o que facilita a compreensão e o reúso desses dados por outros pesquisadores.

Conhecer as ferramentas descritas neste artigo e outras disponíveis no mercado é fundamental no contexto atual, tendo em vista que elas são de uso corriqueiro pelos pesquisadores da e-Science e vem ganhando espaço em outras áreas. A documentação da pesquisa tem sido demandada até mesmo nas Ciências Sociais e Humanidades, área reconhecida por protocolos de pesquisa mais qualitativos e individualizados (Schaefer; Campos; Cândido, 2023). Este conhecimento é importante, portanto, para que o



bibliotecário compreenda o conteúdo contido na documentação e solicite o depósito deste arquivo junto aos arquivos de dados, nos casos em que o pesquisador tenha documentado. Em áreas onde o uso de ferramentas ainda não é uma prática, o conhecimento é importante para que o bibliotecário apoie a difusão destas ferramentas visando facilitar processos e enriquecer a descrição dos dados de pesquisa depositados em repositórios.

Para apoiar o movimento de compartilhamento e reúso de dados dentro de uma instituição, o bibliotecário deve ser capaz de identificar as melhores práticas para o armazenamento, preservação e compartilhamento de dados, bem como estar familiarizado com os requisitos legais, éticos e técnicos que envolvem esses processos. Além disso, o profissional da Ciência da Informação deve entender as necessidades de documentação associadas ao reúso de dados, como a criação de metadados e a adoção de padrões de formatação que assegurem que os dados sejam comprehensíveis e reutilizáveis a longo prazo (Caregnato; Rocha; Gabriel Junior, 2021; Pasquetto; Randles; Borgman, 2017). Sem a orientação adequada os pesquisadores podem não conseguir tirar o melhor proveito das facilidades oferecidas pelas ferramentas de suporte à documentação das pesquisas. Desta forma, o bibliotecário tem a oportunidade de estar preparado para educar os pesquisadores sobre as vantagens e desafios do reúso de dados, promovendo uma cultura de ciência aberta e compartilhamento responsável dentro da instituição.

Quanto aos repositórios de dados, é necessário que eles garantam não apenas o armazenamento dos dados, mas também sua acessibilidade e integridade ao longo do tempo, e que suportem o acesso a grandes volumes de informações, tendo uma infraestrutura projetada para lidar com o crescimento contínuo dos dados, considerando aspectos como o armazenamento escalável e a preservação a longo prazo. Primeiramente, um repositório deve ter uma estrutura que permita o versionamento dos dados, garantindo que as versões anteriores possam ser acessadas e compreendidas, sem perder o contexto de suas alterações. O versionamento é um aspecto fundamental, pois assegura que os pesquisadores possam acompanhar a evolução dos dados e dos métodos, além de permitir a reproduzibilidade dos experimentos em diferentes momentos (Mouat, 2016).

Os repositórios também devem ser capazes de hospedar não apenas os dados, mas também a documentação associada, como metadados, scripts e descrições detalhadas dos métodos utilizados. Os repositórios devem prover meios para que os artefatos de pesquisa por ele armazenados estejam de acordo como os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) (Wilkinson *et al.*, 2016), os quais



estabelecem diretrizes claras para a gestão e compartilhamento de dados de pesquisa. O princípio “*Findable*” refere-se à capacidade de localizar os dados facilmente, enquanto “*Accessible*” assegura que os dados possam ser acessados de forma transparente e sem restrições, dentro dos limites legais e éticos. O princípio “*Interoperable*” garante que os dados possam ser integrados e utilizados por diferentes sistemas, e o princípio “*Reusable*” requer que os dados sejam descritos de forma clara, permitindo seu uso por outros pesquisadores em contextos diferentes.

Os dados de pesquisa constituem uma nova tipologia de documentos em bibliotecas e serviços de informação. Diferentemente dos documentos mais tradicionais, como o livro e o artigo, os dados demandam uma atenção distinta por parte do profissional da Ciência da Informação. O contexto de produção do dado é tão importante quanto o próprio dado, o que exige, além da documentação de pesquisa, a descrição de metadados mais detalhada. O bibliotecário é o profissional habilitado para modelar metadados adequados (descritivos, técnicos, de preservação, entre outros) e aplicá-los eficientemente de modo a gerar e disponibilizar elementos qualificados que permitam o reúso dos dados.

6 CONSIDERAÇÕES FINAIS

Este estudo deteve-se em apresentar conceitos relacionados à reproduzibilidade de pesquisas e reúso de dados, além de explorar quais ferramentas estão sendo utilizadas, atualmente, para dar suporte à pesquisa científica e sua contribuição para que as pesquisas sejam reproduzíveis e para o aumento do reúso de seus dados.

A partir da revisão de literatura, foram definidos os conceitos de reproduzibilidade, replicabilidade e repetibilidade a fim de encontrar as definições mais consensuais.

O estudo realizou um abrangente levantamento, por meio da revisão bibliográfica e documental, das ferramentas disponíveis atualmente para a documentação das pesquisas, particularmente em *e-Science*. As ferramentas mais citadas no levantamento bibliográfico tiveram seu funcionamento investigado em profundidade, com instalação ou criação de cadastro, conforme o caso, e elaboração de exemplos de teste simulando documentação de pesquisas, totalizando nove ferramentas descritas na seção 4. Entre estas, as quatro mais citadas foram, por ordem, a ferramenta de versionamento de código-fonte *Git Hub*, a ferramenta de *containers Docker*, a ferramenta de gestão de projetos de



pesquisa *Open Science Framework* (OSF) e a ferramenta *Jupyter Hub* para a descrição interativa de programas de processamento de dados.

A Ciência precisa ir além de compartilhar seus resultados por meio da literatura, é preciso compartilhar os artefatos utilizados nas pesquisas e esses dados devem ter sua compreensão facilitada por meio da adoção de padrões de formato e conteúdo, bem como de uma documentação completa e sem erros, essencial para o entendimento do significado de dados que se tornam cada vez mais complexos. Mas, além dos dados, as pesquisas têm muito mais artefatos para compartilhar: cadernos de laboratório, programas de automatização de experimentos, metodologias de experimentos que não obtiveram sucesso junto das respectivas decisões a que essas falhas levaram.

Para as áreas de *e-Science*, que precisam de uma infraestrutura computacional para processar suas grandes quantidades de dados, é possível encontrar ferramentas que podem ser utilizadas para facilitar esta documentação, como os *Jupyter Notebook*, que permitem a utilização, no mesmo documento, de textos formatados e imagens, bem como de trechos de código-fonte executáveis com um “*play*”; o *Git Hub* para desenvolvimento, gestão e compartilhamento do código-fonte de experimentos; ou os *container Docker* para a preservação do ambiente computacional de processamento dos experimentos, incluindo o código-fonte e as dependências com suas versões precisas, assim como os dados de entrada, ou uma amostra. A experiência da *e-Science* descrita neste artigo pode servir como inspiração para pesquisadores oriundos de outras áreas de pesquisa, como as Humanas e Sociais, que vêm trabalhando em processos de registro e documentação de suas coletas de dados a partir de protocolos próprios. As ferramentas que automatizam processos na *e-Science* podem, quem sabe, auxiliar pesquisadores de outras áreas.

Como pesquisas futuras, pretende-se investigar como a adoção de selos com níveis de reproduzibilidade de pesquisas atrelados à disponibilidade de seus artefatos, conforme sugerido por Feger *et al.* (2021) e Trisovic *et al.* (2020), poderiam influenciar a reutilização destes artefatos por outros pesquisadores. Outra opção sugerida, seria analisar o quanto as ferramentas de suporte citadas atendem ou não aos princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) (Wilkinson *et al.*, 2016), visto que estes foram criados para guiar o aumento da reutilização dos dados.

Disponibilizar dados de pesquisa para reúso demanda infraestrutura tecnológica e de gestão. O reúso dos dados de pesquisa só se torna possível se estes forem recuperados, acessados, e se mantiverem legíveis e estáveis a longo prazo. Neste contexto, o



profissional de Ciência da Informação ou bibliotecário precisa conhecer muito bem o tema e as ferramentas disponíveis para auxiliar no processo de desenvolvimento de competências e habilidades dos pesquisadores, apoiando o processo de documentação dos procedimentos envolvidos no ciclo de vida dos dados.

REFERÊNCIAS

ACM. **Artifact Review and Badging**. [S. I.: s. n.], 2020. Disponível em: <https://www.acm.org/publications/policies/artifact-review-and-badging-current>. Acesso em: 18 out. 2023.

BAKER, Monya. 1,500 scientists lift the lid on reproducibility. **Nature**, [s. I.], v. 533, p. 452-454, maio 2016. Disponível em: <https://doi.org/10.1038/533452a>. Acesso em: 29 abr. 2022.

BARBA, Lorena A. Terminologies for reproducible research. **arXiv**. [S. I.: s. n.], 2018. arXiv:1802.03311. Disponível em: <https://doi.org/10.48550/arXiv.1802.03311>. Acesso em: 28 set. 2023.

BOETTIGER, Carl. An introduction to Docker for reproducible research. **ACM SIGOPS Operating Systems Review**, [s. I.], v. 49, n. 1, p. 71-79, Jan. 2015. Disponível em: <http://doi.acm.org/10.1145/2723872.2723882>. Acesso em: 21 fev. 2022.

BRINCKMAN, Adam; CHARD, Kyle; GAFFNEY, Niall; HATEGAN, Mihael; JONES, Matthew B.; KOWALIK, Kacper; KULASEKARAN, Sivakumar; LUDÄSCHER, Bertram; MECUM, Bryce D.; NABRZYSKI, Jarek; STODDEN, Victoria; TAYLOR, Ian J.; TURK, Matthew J.; TURNER, Kandace. Computing environments for reproducibility: capturing the "Whole Tale". **Future Generation Computer Systems**, [s. I.], v. 94, p. 854-867, 2019. Disponível em: <https://doi.org/10.1016/j.future.2017.12.029>. Acesso em: 13 jan. 2023.

CAREGNATO, Sônia Elisa; ROCHA, Rafael Port da; GABRIEL JUNIOR, Rene Faustino. Reúso de dados: princípios FAIR e o ecossistema de pesquisa. In: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.).

Princípios FAIR aplicados à gestão de dados de pesquisa. Rio de Janeiro: Ibitc, 2021. p. 187-200. Disponível em: https://ridi.ibict.br/bitstream/123456789/1182/2/IBICT_Principios%20FAIR%20aplicados%20a%20gest%C3%A3o%20de%20dados%20de%20pesquisa_2021.pdf. Acesso em: 15 abr. 2023.

CHEIFET, Barbara. Promoting reproducibility with code ocean. **Genome Biology**, [s. I.], v. 22, n. 1, p. 1-2, 2021. Disponível em: <https://doi.org/10.1186/s13059-021-02299-x>. Acesso em: 13 jan. 2023.

CHEN, Xiaoli; DALLMEIER-TIESSEN, Sünje; DASLER, Robin; FEGER, Sebastian; FOKIANOS, Pamfilos; GONZALEZ, Jose Benito; HIRVONSAKO, Harri; KOUSIDIS, Dinos; LAVASA, Artemis; MELE, Salvatore; RODRIGUEZ, Diego Rodriguez; ŠIMKO, Tibor; SMITH, Tim; TRISOVIC, Ana; TRZCINSKA, Anna; TSANAKTSIDIS, Ioannis;



ZIMMERMANN, Markus; CRANMER, Kyle; HEINRICH, Lukas; WATTS, Gordon; HILDRETH, Michael; IGLESIAS, Lara Lloret; LASSILA-PERINI, Kati; NEUBERT, Sebastian. Open is not enough. **Nature Physics**, [s. l.], v. 15, p. 113-119, Feb. 2019. Disponível em: <https://doi.org/10.1038/s41567-018-0342-2>. Acesso em: 24 maio 2022.

CITO, Jürgen; GALL, Harald C. Using Docker Containers to improve reproducibility in software engineering research. In: IEEE/ACM INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING COMPANION (ICSE-C), 38., 2016, Austin, TX, USA. **Proceedings** [...]. Austin: IEEE, 2016. p. 906-907. Disponível em: <http://dx.doi.org/10.1145/2889160.2891057>. Acesso em: 12 maio 2025.

CLAERBOUT, Jon F.; KARENBACH, Martin. Electronic documents give reproducible research a new meaning. In: SEG TECHNICAL PROGRAM EXPANDED, 1992, [s. l.]. **Abstracts** [...]. [S. l.]: Society of Exploration Geophysicists, 1992. p. 601-604. Disponível em: <https://doi.org/10.1190/1.1822162>. Acesso em: 28 nov. 2023.

FASEB. **Enhancing research reproducibility**. Bethesda: FASEB, 2016. Disponível em: https://www.faseb.org/FASEB/media/PDF/News/Washington%20Update/FASEB_Enhancing-Research-Reproducibility_1.pdf. Acesso em: 23 nov. 2023.

FEGER, Sebastian Stefan; WOŹNIAK, Paweł W.; NIESS, Jasmin; SCHMIDT, Albrecht. Tailored science badges: enabling new forms of research interaction. In: ACM DESIGNING INTERACTIVE SYSTEMS CONFERENCE, 21., 2021, New York. **Proceedings** [...]. New York: Association for Computing Machinery, 2021. p. 576-588. Disponível em: <https://doi.org/10.1145/3461778.3462067>. Acesso em: 03 jan. 2022.

FIGUEIREDO FILHO, Dalson; LINS, Rodrigo; DOMINGOS, Amanda; JANZ, Nicole; SILVA, Lucas. Seven reasons why: a user's guide to transparency and reproducibility. **Brazilian Political Science Review**, [s. n.], v. 13, n. 2, e0001, 2019. Disponível em: <https://doi.org/10.1590/1981-3821201900020001>. Acesso em: 20 nov. 2023.

GOODMAN, Steven N.; FANELLI, Daniele; IOANNIDIS, John P. A. What does research reproducibility mean? **Science translational medicine**, [s. n.], v. 8, n. 341, p. 341ps12-341ps12, 2016. Disponível em: <https://doi.org/10.1126/scitranslmed.aaf5027>. Acesso em: 12 set. 2023.

HARZING, Anne-Wil. **Publish or perish**. Versão 8.8.4384, 2021. Software. Disponível em: <https://harzing.com/resources/publish-or-perish/command-line>. Acesso em: 23 fev. 2023.

HERNÁNDEZ, José Armando; COLOM, Miguel. Repeatability, Reproducibility, replicability, reusability (4R) in Journals' Policies and Software/Data Management in Scientific Publications: a survey, discussion, and perspectives. **arXiv**. [S. l.: s. n.], 2023. arXiv:2312.11028. Disponível em: <https://doi.org/10.48550/arXiv.2312.11028>. Acesso em: 13 mar. 2024.

MALIK, Tanu. Reproducible eScience: the data containerization challenge. In: IEEE INTERNATIONAL CONFERENCE ON E-SCIENCE (E-SCIENCE), 19., 2023, Limassol. **Proceedings** [...]. Limassol: IEEE, 2023. p. 1-5. Disponível em: <https://doi.org/10.1109/e-Science58273.2023.10254837>. Acesso em: 9 out. 2023.



MOREAU, David; WIEBELS, Kristina; BOETTIGER, Carl. Containers for computational reproducibility. **Nature Reviews Methods Primers**, [s. n.], v. 3, n. 1, p. 1-16, 2023. Disponível em: <https://doi.org/10.1038/s43586-023-00236-9>. Acesso em: 9 set. 2023.

MOUAT, Adrian. **Using docker**: developing and deploying software with containers. Boston: O'Reilly, 2016. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=wpYpCwAAQBAJ&oi=fnd&pg=PP1&dq=container+history+docker&ots=QhL5yLnT9Q&sig=7PVD3sSBTX9ab0glq5lrMlqkVps#v=onepage&q&f=false>. Acesso em: 20 nov. 2022.

MUNAFÒ, Marcus R.; NOSEK, Brian A.; BISHOP, Dorothy V. M.; BUTTON, Katherine S.; CHAMBERS, Christopher D.; PERCIE DU SERT, Nathalie; SIMONSOHN, Uri; WAGENMAKERS, Eric-Jan; WARE, Jennifer J.; IOANNIDIS, John P. A. A manifesto for reproducible science. **Nature human behaviour**, [s. n.], v. 1, n. 1, p. 1-9, 2017. Disponível em: <http://dx.doi.org/10.1038/s41562-016-0021>. Acesso em: 29 fev. 2024.

NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE. **Reproducibility and replicability in science**. Washington: The National Academies Press, 2019. Disponível em: <https://doi.org/10.17226/25303>. Acesso em: 23 nov. 2023.

PASQUETTO, Irene V.; RANDLES, Bernadette M.; BORGMAN, Christine L. On the reuse of scientific data. **Data Science Journal**, [s. n.], v. 16, n. 8, p. 1-9, 2017. Disponível em: <https://doi.org/10.5334/dsj-2017-008>. Acesso em: 15 abr. 2023.

PENG, Roger D.; DOMINICI, Francesca; ZEGER, Scott L. Reproducible epidemiologic research. **American Journal of Epidemiology**, [s. n.], v. 163, n. 9, p. 783-789, 2006. Disponível em: <https://doi.org/10.1093/aje/kwj093>. Acesso em: 20 nov. 2023.

PENG, Roger D.; HICKS, Stephanie. C. Reproducible research: a retrospective. **Annual Review of Public Health**, [s. n.], v. 42. p. 79-93, 2021. Disponível em: <https://doi.org/10.1146/annurev-publhealth-012420-105110>. Acesso em: 24 abr. 2022.

PEREZ-RIVEROL, Yasset; GATTO, Laurent; WANG, Rui; SACHSENBERG, Timo; USZKOREIT, Julian; LEPREVOST, Felipe da Veiga; FUFEZAN, Christian; TERNENT, Tobias; EGLEN, Stephen J.; KATZ, Daniel S.; POLLARD, Tom J.; KONOVALOV, Alexander; FLIGHT, Robert M.; BLIN, Kai; VIZCAÍNO, Juan Antonio. Ten simple rules for taking advantage of Git and GitHub. **PLoS Computational Biology**, [s. n.], v. 12, n. 7, e1004947, 2016. Disponível em: <https://doi.org/10.1371/journal.pcbi.1007142>. Acesso em: 15 nov. 2023.

POTTERBUSCH, Megan; LOTRECCHIANO, G. R. Shifting paradigms in information flow: an open science framework (OSF) for knowledge sharing teams. **Informing Science**, [s. n.], v. 21, p. 179, 2018. Disponível em: <https://doi.org/10.28945/4031>. Acesso em: 14 jan. 2023.

ROUGIER, Nicolas P.; HINSEN, Konrad; ALEXANDRE, Frédéric; ARILDSEN, Thomas; BARBA, Lorena A.; BENUREAU, Fabien C.Y.; BROWN, C. Titus; DE BUYL, Pierre; CAGLAYAN, Ozan; DAVISON, Andrew P.; DELSUC, Marc-André; DETORAKIS, Georgios; DIEM, Alexandra K.; DRIX, Damien; ENEL, Pierre; GIRARD, Benoît; GUEST, Olivia; HALL, Matt G.; HENRIQUES, Rafael N.; HINAUT, Xavier; JARON, Kamil S.; KHAMASSI, Mehdi; KLEIN, Almar; MANNINEN, Tiina; MARCHESI, Pietro; MCGLINN,



Daniel; METZNER, Christoph; PETCHEY, Owen; PLESSER, Hans Ekkehard; POISOT, Timothée; RAM, Karthik; RAM, Yoav; ROESCH, Etienne; ROSSANT, Cyrille; ROSTAMI, Vahid; SHIFMAN, Aaron; STACHELEK, Jemma; STIMBERG, Marcel; STOLLMEIER, Frank; VAGGI, Federico; VIEJO, Guillaume; VITAY, Julien; VOSTINAR, Anya E.; YURCHAK, Roman; ZITO, Tiziano. Sustainable computational science: the ReScience initiative. **PeerJ Computer Science**, [s. n.], v. 3, e142, 2017. Disponível em: <https://doi.org/10.7717/peerj-cs.142>. Acesso em 23 nov. 2023.

SAYÃO, Luis F.; SALES, Luana F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS**, Rio de Janeiro, v. 8, n. 2, p. 76-92, jun. 2014. Disponível em: <https://www.arca.fiocruz.br/handle/icict/17102>. Acesso em: 21 fev. 2022.

SCHAEFER, B.; CAMPOS, L. A.; CANDIDO, M. R. Revista DADOS cria editoria especializada em replicabilidade. In: SCIELO. **SciELO em Perspectiva**, 20 out. 2023. Disponível em: <https://blog.scielo.org/blog/2023/10/20/revista-dados-cria-editoria-especializada-em-replicabilidade/>. Acesso em: 05 fev. 2025.

TRISOVIC, Ana; DURBIN, Philip; SCHLATTER, Tania; DURAND, Gustavo; BARBOSA, Sonia; BROOKE, Danny; CROSAS, Mercê. Advancing computational reproducibility in the dataverse data repository platform. In: INTERNATIONAL WORKSHOP ON PRACTICAL REPRODUCIBLE EVALUATION OF COMPUTER SYSTEMS, 3., 2020, Stockholm. **Proceedings** [...]. New York: Association for Computing Machinery, 2020. p. 15–20. Disponível em: <https://doi.org/10.1145/3391800.3398173>. Acesso em: 8 jun. 2022.

UNESCO. **Recommendation on open science**. [S.I.: s.n.], 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>. Acesso em: 27 maio 2024.

VAN DE SANDT, Stephanie; DALLMEIER-TIESSEN, Sünje; LAVASA, Artemis; PETRAS, Vivien. The definition of reuse. **Data Science Journal**, [s. I.], v. 18, n. 1, p. 1–19, 2019. Disponível em: <http://dx.doi.org/10.5334/dsj-2019-022>. Acesso em: 28 nov. 2022.

VANZ, Samile Andrea de Souza; PAVÃO, Caterina Marta Groposo; CAREGNATO, Sônia Elisa; PASSOS, Paula Caroline Schifino Jardim; MOURA, Ana Maria Mielniczuk de; BORGES, Eduardo Nunes; GABRIEL JUNIOR, Rene Faustino; ROCHA, Rafael Port da. Diretrizes para o estabelecimento de um checklist para curadoria de dados de pesquisa. **Informação em Pauta**, [s. I.], v. 6, n. 00, p. 1-18, 26 out. 2021. Disponível em: <http://www.periodicos.ufc.br/informacaoempauta/article/view/68088/197501>. Acesso em: 22 out. 2022

WILKINSON, Mark D.; DUMONTIER, Michel; AALBERSBERG, IJsbrand Jan; APPLETON, Gabrielle; AXTON, Myles; BAAK, Arie; BLOMBERG, Niklas; BOITEN, Jan-Willem; BONINO DA SILVA SANTOS, Luiz; BOURNE, Philip E.; BOUWMAN, Jildau; BROOKES, Anthony J.; CLARK, Tim; CROSAS, Mercê; DILLO, Ingrid; DUMON, Olivier; EDMUNDS, Scott; EVELO, Chris T.; FINKERS, Richard; GONZALEZ-BELTRAN, Alejandra; GRAY, Alasdair J.G.; GROTH, Paul; GOBLE, Carole; GRETHE, Jeffrey S.; HERINGA, Jaap; 'T HOEN, Peter A.C.; HOOFT, Rob; KUHN, Tobias; KOK, Ruben; KOK, Joost; LUSHER, Scott J.; MARTONE, Maryann E.; MONS, Albert; PACKER, Abel L.; PERSSON, Bengt; ROCCA-SERRA, Philippe; ROOS, Marco; VAN SCHAIK, Rene; SANSONE, Susanna-Assunta; SCHULTES, Erik; SENGSTAG, Thierry; SLATER, Ted; STRAWN, George; SWERTZ, Morris A.; THOMPSON, Mark; VAN DER LEI, Johan; VAN



MULLIGEN, Erik; VELTEROP, Jan; WAAGMEESTER, Andra; WITTENBURG, Peter; WOLSTENCROFT, Katherine; ZHAO, Jun; MONS, Barend. Comments: The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, [s. l.], v. 3, n. 160018, Mar. 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 21 fev. 2022.

NOTAS

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: M. K. Ferreira, S. A. S. Vanz

Coleta de dados: M. K. Ferreira

Análise de dados: M. K. Ferreira, S. A. S. Vanz

Discussão dos resultados: M. K. Ferreira, S. A. S. Vanz

Revisão e aprovação: S. A. S. Vanz

FINANCIAMENTO

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

CONSENTIMENTO DE USO DE IMAGEM

Não se aplica.

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

CONFLITO DE INTERESSES

Não se aplica.

LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Edgar Bisset Alvarez, Patrícia Neubert, Genilson Geraldo, Camila De Azevedo Gibbon, Jônatas Edison da Silva, Luan Soares Silva, Marcela Reinhardt e Daniela Capri.

HISTÓRICO

Recebido em: 07-10-2024 – Aprovado em: 26-05-2025 – Publicado em: 09-06-2025

