

# Artigo à Prova de Futuro

## Jornada de Open Science na Prática

### Home

Página do curso “**Artigo à Prova de Futuro: Jornada de Open Science na Prática**”. Aqui você encontrará informações sobre o programa do curso, materiais para seu acompanhamento e sugestões de leituras sobre a prática da ciência aberta (artigos, notas de aulas, blogs, vídeos, etc.).

Caso você caiu nessa página por acaso ( ), saiba que o poderá se inscrever no curso [aqui](#): independente se ele estiver acontecendo no momento, será convidado a participar da próxima versão.

### Sobre os instrutores

O curso é coordenado e ministrado por Pablo Rogers, doutor em administração pela Universidade de São Paulo (FEA/USP) e professor de finanças e métodos quantitativos desde 2005. Em sua [página de perfil do Github](#) temos informações de seus trabalhos recentes, e no seu [site pessoal](#), detalhes sobre suas formações, competências, trajetória e projetos.

Na sua versão atual o curso também será ministrado por Ricardo Limongi, doutor em administração pela Fundação Getúlio Vargas (FGV-SP) e professor de marketing e métodos quantitativos desde 2008 e atual editor chefe da Brazilian Administration Review (BAR). Em seu [perfil do Instagram](#) é possível acompanhar sua agenda de atividades, cursos e palestras sobre inteligência artificial aplicada aos negócios e pesquisa. Em seu [canal do YouTube](#), é possível encontrar vídeos das suas atividades: congressos, palestras, aulas, etc.

### Sobre o curso

O curso tem objetivo de introduzir os conceitos relacionados com a ciência aberta e a prática da pesquisa reprodutível. O curso aborda temas introdutórios sobre ciência aberta, com foco no ferramental disponível para tornar a pesquisa mais transparente, reprodutível e acessível. O curso é voltado para pesquisadores e estudantes de pós-graduação, mas aberto a qualquer pessoa interessada em aprender sobre a prática da ciência aberta. O protagonista do curso é o pesquisador brasileiro que deseja aprimorar a qualidade e a transparência de sua pesquisa, e que busca ferramentas para torná-la mais eficiente e acessível.

Trata-se de um curso intermitente programado para acontecer em 4 encontros de 4

horas/aula (ou 8 encontros de 2 horas/aula), totalizando 16 horas/aula. Num primeiro momento, a ideia que o curso seja remoto e síncrono para alcançar um número maior de interessados. Ele poderá acontecer mais de uma vez no ano, com datas e horários a serem definidos. Para o calendário atual do curso, consulte a seção [Agenda](#).

O curso é gratuito e com de certificado de extensão pela Universidade Federal de Uberlândia (UFU). As inscrições são feitas por meio de um [formulário](#) intermediado pelo projeto [Psico&Econo\\_METRIA](#). Quando da previsão das datas, uma campanha de e-mail marketing divulgará o link para a inscrição através de coordenações de pós-graduações selecionadas.

As vagas são limitadas e a seleção será feita por ordem de inscrição. Após o preenchimento das vagas, os demais interessados serão inscritos automaticamente numa lista de espera e, tempestivamente, serão avisados sobre a próxima edição do curso. Após selecionados, os inscritos receberão um e-mail com instruções para acesso à plataforma de aulas síncronas e para a realização das atividades prévias ao curso.

## Ementa do curso

Introdução da Ciência Aberta / Repositórios da Ciência Aberta / Gerenciamento de Referências e Bibliotecas / Gestão de Dados e Projetos / Controle de Versão / Documentos Reprodutíveis / Controle de Ambiente (containers) / IA Aplicada à Pesquisa Científica.

## Metodologia

Num primeiro momento, o curso foi concebido para acontecer de forma remota e síncrona, com aulas expositivas e teóricas, porém em grande medida, o conteúdo é essencialmente prático. Algumas aulas poderão ser gravadas e disponibilizadas no [canal do YouTube do projeto Psico&Econo\\_METRIA](#), mas a intenção é que o conteúdo principal seja síncrono, para uma maior interação entre os participantes.

Nesse sentido, o material do curso organizado nessa página refere-se ao roteiro estruturado de tudo que se vê nas aulas síncronas e conteúdos adicionais (bibliografia, notas de aulas, links, etc).

A proposta do curso busca seguir de perto a mensagem de Dogucu and Çetinkaya-Rundel (2022). Nesse artigo as autoras abordam a importância da reprodutibilidade na ciência de dados, tanto na pesquisa quanto no ensino. Elas recomendam que os professores-pesquisadores adotem fluxos de trabalho reprodutíveis em suas pesquisas e ensinem esses fluxos de trabalho aos seus alunos. Elas propõem uma dimensão para as práticas de reprodutibilidade, focada exclusivamente nas ferramentas para o ensino (todos os materiais de ensino devem ser computacionalmente reprodutíveis, bem documentados e abertos).

Artigo à Prova de Futuro: Jornada de Open Science na Prática by Pablo Rogers is licensed under CC BY-NC-SA 4.0

## Pré-requisitos

O curso não exige conhecimento prévio em programação, mas é recomendável que o aluno tenha familiaridade com o uso de computadores (ambiente Windows) e com a

escrita de textos científicos. Nesse sentido, não é necessário ter conhecimento prévio sobre as ferramentas e plataformas que utilizaremos no curso: Zotero, OSF, Zenodo, Git, Github, RStudio, Quarto/RMarkdown, Docker, etc; mas desejável que o aluno já as tenha instalado e/ou cadastro nas plataformas.

Abaixo eu descrevo sucintamente o que é cada uma dessas ferramentas e plataformas, e como você pode se preparar para o curso. Também apresento um vídeo curto sobre a instalação e cadastro em cada uma delas. A ideia é que você já tenha todas as ferramentas e plataformas instaladas e/ou cadastro antes do início do curso, para que possamos focar no conteúdo e prática durante as aulas síncronas. Mas pode ficar tranquilo, pois na primeira aula do curso abordaremos essas tarefas, e caso ainda haja alguma dúvida na instalação e cadastro, dedicaremos algum tempo para saná-las.

Outras soluções que iremos discutir e testar durante o curso, como alguns pacotes do R, e aplicações de IA no último módulo, deixaremos para as aulas remotas. Essas soluções na sua maioria requerem cadastros rápidos, e podem ser feitos de forma instantânea via conta Google/Microsoft/Apple.

#### ! Tip 1: ChatGPT para suas notas de leituras

Os resumos das bibliografias que apresento em cada uma das seções foram elaborados com o auxílio do ChatGPT 4, seja pelo o [webapp da OpenAI](#) ou pelo [Copilot](#) (ou buscador Bing) da Microsoft.

Eu destaco (seleciono através de marca texto no Zotero, por exemplo) as passagens que considero importante do artigo científico, tendo em vista a minha perspectiva e fins no momento da leitura, e posteriormente copio e colo as notas de leitura com a seguinte prompt:

*“Sentences in the text are reading notes, that is, what I found most important and interesting, from a scientific article on the topic open science. I would like you to summarize the notes in a descriptive text and concatenate the arguments highlighted in the notes. Give your answer in Portuguese”*

#### 🔥 Não confie cegamente na IA

Eu simplesmente copieie e coleie os resultados do ChatGPT para compilar essas notas de leituras? Não. Após o resultado do ChatGPT eu reviso o sumário das notas de leituras e faço ajustes, que somente são possíveis porque li o artigo por completo. Apesar do ChatGPT fazer um bom serviço nesse sentido, ele ainda comete muitos deslizes. Deslizes esses que você não pode deixar passar num texto científico, e somente captaria a partir da leitura do artigo ou sendo conhecedor do assunto abordado.

#### i Outra curiosidade...

A [imagem cover desse curso](#) foi gerada por uma IA, com posteriores ajustes (off course!). Existem diversos geradores de imagens que você pode testar gratuitamente, mas eu costumo utilizar o i) [DALL-E](#), que é uma solução da OpenAI que também pode ser utilizada no [Copilot da Microsoft](#); ii) o [PlaygroundAI](#), e iii) o

## Github

Primeiramente, se cadastre no Github: <https://github.com/signup>, pois com ele você poderá acessar o material do curso e interagir com os demais participantes. E com a conta do Github você também poderá se cadastrar em outras plataformas, como o Zenodo, OSF, etc. Algumas features que aprenderemos no curso exigem o vínculo entre as contas. Se for professor ou estudante, você pode solicitar o [GitHub Education](#) e ter acesso, por exemplo, ao Copilot, uma das ferramentas de IA que abordaremos no último módulo. Por isso, é importante que você se cadastre com um e-mail institucional. Use o mesmo e-mail para se cadastrar em todas plataformas.

<https://youtu.be/Nmjh9KsV6eU>

## Git

Github não é a mesma coisa que Git. O Github é uma plataforma, e o Git é uma ferramenta. Instale a versão mais recente do Git: <https://git-scm.com/downloads>. O Git é uma ferramenta de controle de versão, e o Github é uma plataforma que utiliza o Git. O Git é uma ferramenta essencial para a prática da ciência aberta, e é uma das ferramentas mais importantes para o pesquisador que deseja tornar sua pesquisa mais transparente e reprodutível.

<https://youtu.be/XCa6mE0bEI0>

## Zotero

Baixe a versão mais recente do Zotero: <https://www.zotero.org/download/> e cadastre uma conta: <https://www.zotero.org/user/register/>. Vamos discutir sobre o Zotero e diversos plugins que são úteis no dia-a-dia do pesquisador. Atualmente, o Zotero é a ferramenta mais completa para gerenciamento de referências e bibliotecas, e se integra nativamente com o RStudio.

<https://youtu.be/ZSFq6LHaDJ4>

## OSF

Cadastre no Open Science Framework (OSF): <https://osf.io/register/>. Como veremos, essa plataforma é uma das mais importantes para a prática da ciência aberta. Ela está no começo (pré-registro) e no final (repositório de dados e pré-print) do ciclo de vida (workflow) de um projeto de pesquisa.

<https://youtu.be/WQ4O-8O6MwI>

## Zenodo

Apesar do Zenodo cumprir funções similares ao OSF e até mesmo ao Github, ele é mais voltado para a publicação de dados e publicações científicas. Cadastre no Zenodo: <https://zenodo.org/login/> e vincule sua conta com o Github. Isso será útil, principalmente, para geração de DOI de repositórios do Github.

<https://youtu.be/pZaqL3Auxb0>

## RStudio

Baixe a versão mais recente do RStudio: <https://posit.co/download/rstudio-desktop/>. O RStudio é uma Integrated Development Environment (IDE) para a linguagem R. O RStudio é uma ferramenta essencial para a prática da ciência aberta em R, pois integra as principais soluções que abordaremos no curso (Zotero, Quarto, Git/Github, etc.). A empresa RStudio recentemente mudou o nome para Posit, com o objetivo refletir melhor a expansão da empresa para além do desenvolvimento de ferramentas para R, incluindo Python e outras linguagens. Nesse mesmo link você pode baixar o R, que é a linguagem de programação que utilizaremos no curso.

<https://youtu.be/KM2jxaNIEUk>

## Quarto

Baixe a versão mais recente do Quarto: <https://www.quarto.org/>. O Quarto é uma linguagem de marcação que permite a criação de documentos reprodutíveis e dinâmicos. Ele é uma evolução e tende a substituir o RMarkdown, que é a principal linguagem de marcação do R. O Quarto engloba e adiciona diversas outras vantagens ao RMarkdown, tal como a possibilidade de criar documentos reprodutíveis em Python, Julia, etc. Se você já tem algum conhecimento de RMarkdown, não se preocupe, pois o Quarto é uma extensão natural.

<https://youtu.be/-HvOMVkk6I4>

## Docker

Baixe a versão mais recente do Docker: <https://www.docker.com/products/docker-desktop>. Nesse mesmo link você cria uma conta. O Docker é uma plataforma para desenvolvimento, envio e execução de aplicativos. O Docker é uma ferramenta essencial para a prática da ciência aberta, pois permite a criação de ambientes reprodutíveis.

<https://youtu.be/WjXQxhTLlrQ>

## Agenda

Planejamento dos dias ( ) e horários das aulas ( ), conforme a ementa do curso. Na seção de cada uma das aulas temos materiais adicionais para o respectivo conteúdo. Quando disponível, por aqui, poderás acessar os slides utilizados nas aulas ( ), aulas gravadas ou indicações de vídeo ( ) e leituras básica sobre os conteúdos ( ).

| Aula/Conteúdo | Data           | Material Principal | Instrutor       |
|---------------|----------------|--------------------|-----------------|
| Chapter 1     | 04/06/24 19:00 |                    | Ricardo Limongi |
| Chapter 2     | 06/06/24 19:00 |                    | Pablo Rogers    |
| Chapter 3     | 11/06/24 19:00 |                    | Pablo Rogers    |
| Chapter 4     | 13/06/24 19:00 |                    | Pablo Rogers    |
| Chapter 5     | 18/06/24 19:00 |                    | Pablo Rogers    |
| Chapter 6     | 20/06/24 19:00 |                    | Pablo Rogers    |

| Aula/Conteúdo | Data           | Material Principal | Instrutor       |
|---------------|----------------|--------------------|-----------------|
| Chapter 7     | 25/06/24 19:00 |                    | Pablo Rogers    |
| Chapter 8     | 27/06/24 19:00 |                    | Ricardo Limongi |

## 1 Introdução à Ciência Aberta

A pesquisa científica atual<sup>1</sup> enfrenta vários desafios (Munafò et al. 2017). Problemas como o pequeno tamanho da amostra, pequenos tamanhos de efeito, p-hacking e HARKing (viés positivo de publicação), conflitos de interesse e a competição entre cientistas que trabalham isoladamente sem combinar seus esforços, têm sido apontados como catalizadores do que se convencionou chamar de “crise de reprodutibilidade” na ciência (M. Baker 2016; Munafò et al. 2017).

Pesquisas apontam que mais de 70% de pesquisadores que tentaram, falharam em reproduzir os experimentos de outros cientistas, e mais da metade falhou em reproduzir seus próprios experimentos (M. Baker 2016), com estimativa de que 85% dos esforços de pesquisas estejam sendo desperdiçados (Munafò et al. 2017), gerando custos econômicos bilionários (Freedman, Cockburn, and Simcoe 2015).

A despeito daqueles que advogam que não existe essa tal “crise de reprodutibilidade” (Bernard 2023; Fanelli 2018; Protzko et al. 2023), a grande maioria da comunidade científica concorda com sua existência e defende a melhoria da transparência, reprodutibilidade e eficiência na ciência (M. Baker 2016).

Nesse contexto, o movimento da Ciência Aberta (CA) tem ganhado notoriedade e mudado a percepção sobre o cenário científico global (Crüwell et al. 2019). Ele busca tornar o conhecimento científico mais acessível, transparente e colaborativo. Se apresenta como uma coleção de práticas de democratização do conhecimento e ruptura com o formato único de divulgação do conhecimento científico (Crüwell et al. 2019; Heinz and Miranda 2024; Munafò et al. 2017). Ele surge do embate entre aqueles que buscam compartilhar o conhecimento e aqueles que defendem mecanismos de apropriação privada para a produção científica (Heinz and Miranda 2024).

A CA é um termo múltiplo e genérico (Vicente-Saez and Martinez-Fuentes 2018), que representa diversas interpretações, e é considerada um novo modelo de divulgação e produção de resultados científicos por meio do acesso livre e irrestrito ao conhecimento (Heinz and Miranda 2024). A CA não é apenas um conceito, mas uma prática multifacetada que influencia o ciclo de vida da pesquisa, desde a concepção até a disseminação dos resultados (Silva and Silveira 2019).

Existem pelo menos cinco escolas de pensamento dentro da CA. Estas escolas abrangem desde a arquitetura tecnológica necessária para suportar a ciência até a inclusão do público geral na criação de conhecimento, passando pela medição do impacto alternativo, acesso ao conhecimento como um direito humano, e a pesquisa colaborativa como inovação aberta (Silva and Silveira 2019).

<sup>1</sup>O texto apresentado nessa seção é uma compilação de fragmentos do Projeto APQ-01225-24 submetido ao Edital Nº 001/2024 - Demanda Universal, da FAPEMIG, ainda em análise. O detalhamento desse projeto, quando se tornar público, poderá ser acompanhado no [OSF](#): Rogers, P., Limongi, R., & Barboza, F. (2024, May 29). The Practice of Open Science in Brazil. Retrieved from [osf.io/dnrgf](#)

A taxonomia proposta pela **FOSTER** (Facilitate Open Science Training for European Research), e sua releitura revisada e ampliada para o contexto latino americano por Silveira et al. (2023), tendo em vista as recomendações da UNESCO (2021), nos dá uma dimensão da complexidade do assunto (vide ilustração em: <https://doi.org/10.5281/zenodo.7836884>).

Existem vários argumentos que sustentam a importância da CA (Heinz and Miranda 2024). Primeiramente, a CA pode trazer benefícios sociais significativos, pois contribui para o avanço do conhecimento, a inovação, a educação, a transparência e a participação cidadã. Além disso, a CA pode trazer benefícios científicos ao aumentar a qualidade, a reprodutibilidade, a eficiência e o impacto da pesquisa científica. Ela também facilita a colaboração, a comunicação e a interdisciplinaridade entre os pesquisadores. Por fim, a CA pode trazer benefícios éticos ao promover a integridade, a responsabilidade, a equidade e a diversidade na ciência, além de respeitar os direitos dos autores, dos participantes e da sociedade como um todo.

Esses argumentos são fundamentais para legitimar a CA e destacar sua importância no mundo atual (Heinz and Miranda 2024), principalmente, como potencial transformador para reduzir desigualdades existentes em tecnologias de informação e comunicação – reduzir exclusões digitais, tecnológicas e de conhecimento –, e acelerar o progresso rumo à implementação da Agenda 2030 e realização dos Objetivos de Desenvolvimento Sustentável (UNESCO 2021).

O movimento da CA no Brasil está em uma fase transitória (Rezende and Falgueras 2020) – ainda consolidando o acesso aberto – com o governo desempenhando um papel crucial nesse processo. O Brasil tem ganhado destaque por sua abordagem única na implementação da CA. Esta abordagem é moldada por marcos regulatórios que se estendem desde o governo até as instituições e agências de financiamento. Os regulamentos brasileiros, particularmente aqueles que promovem a abertura de dados governamentais, têm um impacto direto na prática científica. Eles incentivam a transparência e facilitam o acesso a dados científicos originados de instituições públicas (Rezende and Falgueras 2020).

A trajetória brasileira rumo à CA inicia com a abertura de dados na esfera governamental entre 2009 e 2016, evoluindo para a criação de um grupo de trabalho em 2017 pelo Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) para desenvolver uma política nacional para a CA. Este esforço concentrou ênfase no reconhecimento dos dados de pesquisa como ativos de desenvolvimento científico, econômico e social, buscando facilitar seu acesso, compartilhamento e reutilização (Rezende and Falgueras 2020).

Talvez por esse motivo, as políticas institucionais brasileiras revelam um cenário ainda muito influenciado pela “via verde” do movimento de acesso aberto, caracterizado pelo depósito de dados em repositórios digitais abertos, e que o comprometimento efetivo do Brasil com a CA ainda é incipiente. As regulamentações atuais favorecem principalmente o acesso aberto, sem abordar de maneira abrangente outros aspectos da CA (Rezende and Falgueras 2020). O Brasil é um dos líderes mundiais no fornecimento de acesso universal às suas pesquisas e estudos (Neto, Willinsky, and Alperin 2016), com crescimento estável de sua produção científica disponível em acesso aberto, principalmente, as áreas de Agricultura e Ciência & Tecnologia (Caballero-Rivero, Sánchez-Tarragó, and Santos 2019).

Em termos de pesquisa acadêmica sobre o tema no Brasil, os estudos são precoces e concentrados na área de Ciência da Informação (Albano, Pedrosa, and Caetano 2023). Apesar da maturidade da CA no Brasil, a importância do tema – materializada na quantidade de produção acadêmica – tem aumentado vertiginosamente (Albano, Pedrosa, and Caetano 2023), e a dispersão de autores e respectivas instituições que publicam sobre o assunto, parece ser a situação predominante.

Apesar de importantes atores nacionais, tais como CAPES, CNPq e Scielo, defenderem o crescimento de iniciativas de CA (Mendes-Da-Silva 2023), o assunto no Brasil parece estar circunscrito em iniciativas de importantes periódicos nacionais sobre dados abertos, capitaneados pelas orientações da Scielo. Não foi encontrada nenhuma pesquisa empírica, sobre a prática da CA no Brasil.

Por prática de CA entende-se a perspectiva micro da CA, relacionadas com as terminologias e conhecimento em torno do fluxo de trabalho do gerador de conhecimento científico aberto (Figure 1), ou seja, o cientista que se propõe tornar sua pesquisa transparente, reprodutível e replicável.



Figure 1: Perspectiva micro da CA. Taxonomia relacionada com terminologias e conhecimento em torno da prática (fluxo de trabalho) do gerador de conhecimento científico aberto. Ilustração disponível em: <https://doi.org/10.5281/zenodo.10835001>.

Exclui-se a perspectiva macro, relacionadas com as ramificações conceituais da CA concernentes às políticas públicas, infraestrutura, envolvimento aberto de atores sociais e diálogo aberto com outros sistemas de conhecimento (Figure 2). Essa última perspectiva está fora do escopo da discussão do curso, que se concentra em algumas das dimensões da perspectiva micro, particularmente, as ferramentas disponíveis para compilação dos produtos científicos que integram a publicação científica (UNESCO 2021).

Apesar de uma verossímil expectativa desabonadora, tendo em vista o contexto da CA no Brasil, o diagnóstico da situação da prática da CA se mostra importante, *per si*, pois:

1. ajuda a compreender a natureza exata do problema, suas causas, efeitos e riscos;
2. auxilia a alocação eficiente de recursos pelos atores envolvidos (e.g., os programas de pós-graduação podem direcionar os recursos para onde eles são mais necessários e onde terão o maior impacto;





e 2) promover o desenvolvimento de uma intervenção educacional sobre a prática (work-flow) e principais ferramentas para compilação dos produtos científicos que integram uma publicação científica aberta.

**i** @klein2018 *Reading Note* (Tip 1)

O artigo é um guia prático para pesquisadores que desejam compartilhar os produtos de sua pesquisa. Os autores argumentam que as práticas de pesquisa transparentes são essenciais para melhorar a credibilidade e a cumulatividade da ciência. O artigo fornece recomendações específicas sobre como compartilhar os seguintes produtos de pesquisa:

- Protocolo de estudo
- Materiais
- Dados e metadados
- Procedimento de análise
- Relatórios de pesquisa

As recomendações gerais dos autores são as seguintes:

- **Torne a transparência um padrão:** isso significa compartilhar o máximo possível de informações sobre sua pesquisa, desde o início do processo.
- **Não deixe que o perfeito seja inimigo do bom:** mesmo que você não possa compartilhar todos os detalhes de sua pesquisa, compartilhar alguma coisa é melhor do que nada.
- **Compartilhe e documente o que puder:** isso ajudará a garantir que sua pesquisa seja reproduzível e confiável.
- **Comece cedo:** começar a compartilhar informações sobre sua pesquisa no início do processo pode ajudá-lo a evitar problemas e economizar tempo.

Os autores também discutem algumas preocupações comuns que os pesquisadores têm sobre as práticas de pesquisa transparentes. Eles argumentam que essas preocupações são geralmente infundadas e que as práticas transparentes têm muitos benefícios.

Os principais benefícios das práticas de pesquisa transparentes são:

- **Melhor credibilidade da pesquisa:** os pesquisadores transparentes são mais propensos a serem vistos como confiáveis e honestos.
- **Maior cumulatividade da ciência:** os pesquisadores transparentes tornam mais fácil para outros pesquisadores construir sobre seu trabalho.
- **Mais oportunidades de colaboração e financiamento:** os pesquisadores transparentes são mais propensos a serem convidados para colaborar com outros pesquisadores e a receber financiamento.
- **Maior eficiência da pesquisa:** as práticas transparentes podem ajudar os pesquisadores a economizar tempo e recursos.

Em conclusão, o artigo fornece informações valiosas para pesquisadores que desejam compartilhar os produtos de sua pesquisa. As recomendações dos autores são baseadas em evidências e podem ajudar os pesquisadores a melhorar a qualidade e a credibilidade de seu trabalho.

Este artigo fornece um guia para ajudar estudantes de pós-graduação e seus orientadores a se envolverem na prática da ciência aberta (CA). A CA é descrita como um termo amplo que se refere a uma variedade de princípios e comportamentos relacionados à transparência, credibilidade, reprodutibilidade e acessibilidade.

O artigo sugere oito práticas de CA (das quais destaco sete) que estudantes de pós-graduação iniciantes podem começar a adotar hoje. Cada comportamento é classificado em termos de dificuldade (fácil, médio, difícil) e apresentado em ordem de adoção sugerida. Em cada prática, eles seguem o formato de o quê, por quê, como e preocupações.

Algumas das práticas sugeridas incluem:

1. **Fluxo de Trabalho do Projeto (Nível Fácil):** isso inclui a estrutura da pasta do arquivo, convenções de nomenclatura de documentos, controle de versão, armazenamento em nuvem e outros detalhes. Ter um sistema de fluxo de trabalho de projeto dedicado ajuda a manter sua pesquisa organizada, melhorando a reprodutibilidade, minimizando erros e facilitando colaborações com outros e com você no futuro.
2. **Preprints (Nível Fácil):** postar um manuscrito antes de submetê-lo a uma revista permite um feedback mais amplo do que o que é proporcionado pela revisão por pares e pode ajudar a melhorar um artigo antes da submissão, identificando quaisquer falhas importantes.
3. **Código Reprodutível (Nível Médio):** o código reprodutível para análise de dados e visualizações (por exemplo, tabelas, figuras) refere-se a uma versão detalhada e escrita do seu código que permitiria a outra pessoa (ou a você no futuro) gerar a mesma saída relatada em seu manuscrito.
4. **Compartilhamento de Dados (Nível Médio):** compartilhar dados refere-se a tornar o conjunto de dados desidentificado usado para um projeto disponível para outros pesquisadores.
5. **Escrita de Manuscrito Transparente (Nível Médio):** Para escrever um manuscrito transparente, claro e reprodutível, é útil seguir as diretrizes ou padrões de escrita de manuscritos.
6. **Pré-registro (Nível Médio):** o pré-registro refere-se à postagem de um esboço cronometrado das perguntas de pesquisa, hipóteses, método e plano de análise para um projeto específico antes da coleta de dados e/ou análise.
7. **Relatório Registrado (Nível Difícil):** os Relatórios Registrados envolvem um processo de submissão em duas partes, onde os autores primeiro enviam uma proposta de Estágio 1, que inclui a introdução, método e plano de análise - tudo antes que a coleta de dados e/ou análise tenha sido feita.

O artigo enfatiza que se envolver em uma prática de CA é melhor do que nenhuma e que a CA é apenas uma boa ciência. Além disso, sugere-se que se construa sobre o trabalho que já foi feito em vez de reinventar a roda. A maioria das práticas sugeridas se concentra no uso do [Open Science Framework](#).

## 2 Repositórios da Ciência Aberta

No contexto da ciência aberta, existem diversos repositórios disponíveis, cada um com suas funções e propósitos específicos. Esses repositórios são essenciais para promover

a transparência, acessibilidade e colaboração na pesquisa científica. Eles assumem um papel crucial na democratização do conhecimento e na promoção da colaboração científica. Cada qual com suas particularidades, oferecem aos pesquisadores ferramentas para armazenar, compartilhar e gerenciar dados, publicações e outros materiais de pesquisa, ou se preferir, todo o ciclo de vida da pesquisa.

- **Zenodo**: é um repositório gerido pelo CERN em colaboração com o projeto **OpenAIRE** da União Europeia. Oferece armazenamento gratuito e seguro para dados de pesquisa, com a capacidade de gerar DOIs para facilitar a citação dos dados.
- **Figshare**: é um repositório comercial que permite aos pesquisadores armazenar, compartilhar e descobrir dados de pesquisa. Oferece ferramentas para visualização de documentos, gráficos e outros tipos de dados diretamente no navegador, além de gerar DOIs para os projetos.
- **Mendeley Data**: é um repositório de dados de pesquisa da Elsevier, permitindo o armazenamento, compartilhamento e citação de conjuntos de dados. Ele suporta uma ampla gama de tipos de dados e está integrado com a plataforma de referência Mendeley.
- **Harvard Dataverse**: é uma rede de repositórios que permite aos pesquisadores compartilhar, armazenar e citar dados de pesquisa. Ele oferece ferramentas avançadas para a gestão de dados, incluindo controle de versões e metadados ricos, essenciais para o gerenciamento do ciclo de vida da pesquisa.
- **arXiv**: é um repositório de pré-impressões de artigos científicos em física, matemática, ciência da computação e outras áreas. Ele permite aos pesquisadores compartilhar seus trabalhos antes da revisão por pares, facilitando o acesso à pesquisa em estágios iniciais.
- **Github**: é uma plataforma de desenvolvimento colaborativo baseada em Git, amplamente utilizada por pesquisadores para compartilhar código, documentos e outros materiais de pesquisa. Ele oferece controle de versões, rastreamento de problemas e integração com outras ferramentas de desenvolvimento.

Além desses exemplos, poderíamos citar outras soluções que cumprem papéis semelhantes ou focado em certas disciplinas: **Databrary**, **DataverseNO**, **DataONE**, **DataCite**, **DataHub**, **DataMed**, **DataShare**, **DataVerse**, **Dryad**, **EarthChem**, **EUDAT**, **European Nucleotide Archive (ENA)**, **GenBank**, **Google Dataset Search**, **HathiTrust Research Center**, **ICPSR**, **JSTOR Data for Research**, **National Center for Biotechnology Information (NCBI)**, **National Institutes of Health (NIH) Data Sharing Repositories**, **National Oceanographic Data Center (NODC)**, **PLOS ONE**, **PubMed Central**, **Research Data Australia** e **UK Data Service**; e em última instância, as redes sociais acadêmicas como **Academia.edu**, **Google Scholar**, **ORCID** e **ResearchGate**, também podem ser usadas para compartilhar e descobrir pesquisas.

A despeito de todas essas opções, vamos focar na plataforma **Open Science Framework (OSF)** para a realização do nosso curso. O **OSF** é uma plataforma de código aberto para colaboração em pesquisa, que oferece uma estrutura para conectar os fluxos de trabalho de pesquisa, desde a concepção do projeto até a publicação. O **OSF** é mantido pelo **Center for Open Science (COS)**, uma organização sem fins lucrativos com sede nos Estados Unidos. O **OSF** é um dos principais produtos do COS e é usado por pesquisadores de todo o mundo para colaborar em projetos de pesquisa.

O **OSF** oferece uma série de recursos para ajudar os pesquisadores a gerenciar seus projetos de pesquisa, incluindo:

- **Criar projetos de pesquisa:** organizar seus estudos, incluindo metadados, datasets, materiais de pesquisa e publicações.
- **Carregar e publicar dados:** armazenar e compartilhar seus dados de forma segura e acessível.
- **Colaboração em equipe:** convidar colaboradores para participar do projeto, atribuir tarefas e acompanhar o progresso.
- **Integração com outras ferramentas:** conectar a armazenamentos nas nuvens (Box, DropBox, Google Drive e OneDrive), gerenciadores de referências (Zotero e Mendeley) e outros repositórios (Dataverse, Github, figshare, etc).

O **OSF** tem um foco mais amplo em todo o ciclo de vida da pesquisa, desde a concepção da ideia até a publicação dos resultados (Figure 3). Já algumas das soluções citadas foca principalmente no compartilhamento de dados e publicações. O **OSF** oferece ferramentas mais robustas para colaboração em equipe, como wikis, painéis de discussão e ferramentas de gerenciamento de tarefas, e principalmente, possui uma comunidade mais ativa de pesquisadores e colaboradores.



Figure 3: OSF Research Lifecycle

## 2.1 Open Science Framework (OSF)

O material utilizado nesse módulo do curso segue de perto a proposta de Olson et al. (2022), um [projeto oficial do COS](#) que possui recursos, modelos e práticas para ajudar os pesquisadores a iniciar sua jornada **OSF**. Claro que ele foi adaptado para nossos fins, principalmente, em decorrência do tempo destinado ao módulo.

### Bifurcando ou duplicando um projeto

Você sabia que é possível executar um “*forking*” (criar uma cópia do projeto existente) ou “*duplicate as template*” (duplicar apenas a estrutura do projeto e seus componentes) de um projeto público no OSF?

Você que se interessa em iniciar seu próprio projeto OSF com um modelo, pode criar sua própria duplicata do projeto Olson et al. (2022) para começar!

Neste [projeto](#), existem templates e recursos básicos para diversos casos de uso encontrados no OSF; coordenação de equipes de pesquisa, planejamento de gerenciamento de dados, documentos reprodutíveis e até mesmo gerenciamento de cursos.

Para os alunos que desejam uma leitura sobre o **OSF** na prática, indico os artigos de Sullivan, DeHaven, and Mellor (2019) e Soderberg (2018). Apesar de o leitor poder encontrar *prints* das telas da plataforma desatualizadas, esses dois artigos podem ser um bom começo para entender a lógica da plataforma. E *off course*, recomendo fortemente você dar uma olhada no [supORTE do OSF](#), onde podemos encontrar vídeos introdutórios excelentes.

Esse curso poderia ter sido concebido e gerenciado dentro do **OSF**, no entanto, devido a proposta de apresentarmos também o [Git/Github](#) e sua integração com documentos reprodutíveis no [RStudio/Quarto](#) (como esse que está lendo), optamos por priorizar o repositório do Github. Por isso, que também nesse módulo passamos pelo [Zenodo](#), que integra com o Github e tem a capacidade de gerar DOIs para as versões dos repositórios.

### @sullivan2019 *Reading Note* (Tip 1)

O artigo apresenta um protocolo para a implementação de práticas de Ciência Aberta (CA), com foco no uso do Open Science Framework (OSF). As principais ideias do texto são as seguintes:

- A CA é um movimento que promove a transparência, a reprodutibilidade e a acessibilidade dos resultados de pesquisa;
- As práticas de CA podem contribuir para a melhoria da qualidade da pesquisa científica, tornando-a mais confiável e robusta;
- O OSF é uma plataforma gratuita e de código aberto que pode ser usada para implementar práticas de CA;

O protocolo apresentado no texto fornece instruções passo a passo para as seguintes práticas de CA:

- **Planejamento de gerenciamento de dados:** O planejamento de gerenciamento de dados é essencial para garantir que os dados de pesquisa sejam armazenados, organizados e gerenciados de forma eficiente e eficaz. O OSF fornece ferramentas para ajudar os pesquisadores a planejar e implementar seus planos de gerenciamento de dados.
- **Pré-registro de estudos:** O pré-registro de estudos é uma prática que consiste em publicar um plano de pesquisa antes de iniciar o estudo. Isso ajuda a garantir que o estudo seja realizado de forma objetiva e transparente. O OSF fornece um recurso para pré-registrar estudos.
- **Controle de versão:** O controle de versão é uma prática que consiste em rastrear as alterações feitas em arquivos de texto. Isso ajuda a garantir

que os resultados de pesquisa sejam reprodutíveis e que as alterações feitas nos dados sejam rastreáveis. O OSF fornece ferramentas para gerenciar o controle de versão de arquivos de pesquisa.

- **Compartilhamento de dados e materiais:** O compartilhamento de dados e materiais de pesquisa é uma prática importante para aumentar a transparência e a reprodutibilidade da pesquisa. O OSF fornece um repositório para compartilhar dados e materiais de pesquisa.
- **Publicação de pré-impressões:** As pré-impressões são versões preliminares de artigos científicos que são publicadas online antes de serem revisados por pares. As pré-impressões podem ajudar a acelerar a divulgação da pesquisa e a promover o debate científico. O OSF fornece um repositório para publicar pré-impressões.

O artigo fornece informações valiosas para os pesquisadores que desejam implementar práticas de CA. O protocolo apresentado pode ser usado como um guia para implementar essas práticas de forma eficaz.

### 3 Gerenciamento de Referências e Bibliotecas

### 4 Gestão de Dados e Projetos

### 5 Controle de versão

### 6 Documentos Reprodutíveis

### 7 Controle de Ambiente (Containers)

### 8 IA Aplicada à Pesquisa Científica

## Referências

- Albano, Cláudio Sonáglio, Paula de Oliveira Pedroso, and Doriedson Oliveira Caetano. 2023. “Ciência Aberta: Um Panorama Sobre as Publicações No Cenário Brasileiro.” *Saber Científico* 12 (1): 1–12.
- Baker, Daniel H, Mareike Berg, Kirralise Hansford, Bartholomew Patrick Anselm Quinn, Federico Gabriele Segala, and Erin English. 2023. “ReproduceMe: Lessons from a Pilot Project on Computational Reproducibility.” Preprint. PsyArXiv. <https://doi.org/10.31234/osf.io/k8d4u>.
- Baker, Monya. 2016. “1,500 Scientists Lift the Lid on Reproducibility.” *Nature* 533 (7604): 452–54. <https://doi.org/10.1038/533452a>.
- Bernard, Christophe. 2023. “Stop Reproducing the Reproducibility Crisis.” *Eneuro* 10 (2): ENEURO.0032–23.2023. <https://doi.org/10.1523/ENEURO.0032-23.2023>.
- Bezjak, Sonja, April Clyburne-Sherin, Philipp Conzett, Pedro Fernandes, Edit Görögh, Kerstin Helbig, Bianca Kramer, et al. 2018. *Open Science Training Handbook*. Zenodo. <https://doi.org/10.5281/ZENODO.1212496>.
- Caballero-Rivero, Alejandro, Nancy Sánchez-Tarragó, and Raimundo Nonato Macedo dos Santos. 2019. “Práticas de Ciência Aberta da comunidade acadêmica brasileira:



- estudo a partir da produção científica.” *Transinformação* 31 (November): e190029. <https://doi.org/10.1590/2318-0889201931e190029>.
- Chopik, William J., Ryan H. Bremner, Andrew M. Defever, and Victor N. Keller. 2018. “How (and Whether) to Teach Undergraduates About the Replication Crisis in Psychological Science.” *Teaching of Psychology* 45 (2): 158–63. <https://doi.org/10.1177/0098628318762900>.
- Committee on Reproducibility and Replicability in Science, Board on Behavioral, Cognitive, and Sensory Sciences, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Nuclear and Radiation Studies Board, Division on Earth and Life Studies, Board on Mathematical Sciences and Analytics, et al. 2019. *Reproducibility and Replicability in Science*. Washington, D.C.: National Academies Press. <https://doi.org/10.17226/25303>.
- Crüwell, Sophia, Johnny Van Doorn, Alexander Etz, Matthew C. Makel, Hannah Moshontz, Jesse C. Niebaum, Amy Orben, Sam Parsons, and Michael Schulte-Mecklenbeck. 2019. “Seven Easy Steps to Open Science: An Annotated Reading List.” *Zeitschrift für Psychologie* 227 (4): 237–48. <https://doi.org/10.1027/2151-2604/a000387>.
- Dogucu, Mine, and Mine Çetinkaya-Rundel. 2022. “Tools and Recommendations for Reproducible Teaching.” *Journal of Statistics and Data Science Education* 30 (3): 251–60. <https://doi.org/10.1080/26939169.2022.2138645>.
- European Commission. Directorate General for Research and Innovation. 2017. *Providing Researchers with the Skills and Competencies They Need to Practise Open Science*. LU: Publications Office. <https://data.europa.eu/doi/10.2777/121253>.
- Fanelli, Daniele. 2018. “Is Science Really Facing a Reproducibility Crisis, and Do We Need It To?” *Proceedings of the National Academy of Sciences* 115 (11): 2628–31. <https://doi.org/10.1073/pnas.1708272114>.
- Freedman, Leonard P., Iain M. Cockburn, and Timothy S. Simcoe. 2015. “The Economics of Reproducibility in Preclinical Research.” *PLOS Biology* 13 (6): e1002165. <https://doi.org/10.1371/journal.pbio.1002165>.
- Heinz, Michele, and Miranda Miranda. 2024. “Ciência Aberta: Argumentos e Desafios Para Sua Legitimação Científica.” *Em Questão* 30. <https://doi.org/10.1590/1808-5245.30.135618>.
- Janz, Nicole. 2015. “Bringing the Gold Standard into the Classroom: Replication in University Teaching.” *International Studies Perspectives*, March, n/a–. <https://doi.org/10.1111/insp.12104>.
- McAleer, Phil, Niamh Stack, Heather Woods, Lisa Marie DeBruine, Helena Paterson, Emily Nordmann, Carolina Ellen Kuepper-Tetzel, and Dale J. Barr. 2022. “Embedding Data Skills in Research Methods Education: Preparing Students for Reproducible Research.” Preprint. PsyArXiv. <https://doi.org/10.31234/osf.io/hq68s>.
- Mendes-Da-Silva, Wesley. 2023. “What Lectures and Research in Business Management Need to Know About Open Science.” *Revista de Administração de Empresas* 63 (4): e0000–0033. <https://doi.org/10.1590/s0034-759020230408x>.
- Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. “A Manifesto for Reproducible Science.” *Nature Human Behaviour* 1 (1): 0021. <https://doi.org/10.1038/s41562-016-0021>.
- Neto, Silvio Carvalho, John Willinsky, and Juan Pablo Alperin. 2016. “Measuring, Rating, Supporting, and Strengthening Open Access Scholarly Publishing in Brazil.”



- Education Policy Analysis Archives* 24 (May): 54–54. <https://doi.org/10.14507/epaa.24.2391>.
- Olson, Eric, Nicole Pfeiffer, Mark Call, and Daniel Steger. 2022. “Getting Started on OSF,” August. <https://doi.org/10.17605/OSF.IO/YAQE8>.
- Protzko, John, Jon Krosnick, Leif Nelson, Brian A. Nosek, Jordan Axt, Matt Berent, Nicholas Buttrick, et al. 2023. “High Replicability of Newly Discovered Social-Behavioural Findings Is Achievable.” *Nature Human Behaviour*, November. <https://doi.org/10.1038/s41562-023-01749-9>.
- Rezende, Laura Vilela Rodrigues, and Ernest Abadal Falgueras. 2020. “Estado Da Arte Dos Marcos Regulatórios Brasileiros Rumo à Ciência Aberta.” *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência Da Informação* 25 (September): 01–25. <https://doi.org/10.5007/1518-2924.2020.e71370>.
- Silva, Fabiano Couto Corrêa Da, and Lúcia Da Silveira. 2019. “O Ecossistema Da Ciência Aberta.” *Transinformação* 31: e190001. <https://doi.org/10.1590/2318-0889201931e190001>.
- Silveira, Lúcia Da, Nivaldo Calixto Ribeiro, Remedios Melero, Andrea Mora-Campos, Daniel Fernando Piraquive-Piraquive, Alejandro Uribe Tirado, Priscila Machado Borges Sena, et al. 2023. “Taxonomia Da Ciência Aberta: Revisada e Ampliada.” *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência Da Informação* 28 (June). <https://doi.org/10.5007/1518-2924.2023.e91712>.
- Soderberg, Courtney K. 2018. “Using OSF to Share Data: A Step-by-Step Guide.” *Advances in Methods and Practices in Psychological Science* 1 (1): 115–20. <https://doi.org/10.1177/2515245918757689>.
- Sullivan, Ian, Alexander DeHaven, and David Mellor. 2019. “Open and Reproducible Research on Open Science Framework.” *Current Protocols Essential Laboratory Techniques* 18 (1): e32. <https://doi.org/10.1002/cpet.32>.
- Toelch, Ulf, and Dirk Ostwald. 2018. “Digital Open Science—Teaching Digital Tools for Reproducible and Transparent Research.” *PLOS Biology* 16 (7): e2006022. <https://doi.org/10.1371/journal.pbio.2006022>.
- UNESCO. 2021. “UNESCO Recommendation on Open Science.” UNESCO. <https://doi.org/10.54677/MNMH8546>.
- Vicente-Saez, Ruben, and Clara Martinez-Fuentes. 2018. “Open Science Now: A Systematic Literature Review for an Integrated Definition.” *Journal of Business Research* 88 (July): 428–36. <https://doi.org/10.1016/j.jbusres.2017.12.043>.