

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 tornar-lá 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).

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.

! ChatGPT para suas notas de leituras

Os resumos das bibliografias que apresento nas seções seguintes 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. A despeito do ChatGPT fazer um bom serviço nesse sentido, ele ainda comete muitos deslizos. Deslizos 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 [Gemini](#) do Google.

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/WjXQxhTLrQ>

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 básicas 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
Chapter 7	25/06/24 19:00		Pablo Rogers
Chapter 8	27/06/24 19:00		Ricardo Limongi

1 Introdução à Ciência Aberta

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. 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.



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*

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

- Olson, Eric, Nicole Pfeiffer, Mark Call, and Daniel Steger. 2022. “Getting Started on OSF,” August. <https://doi.org/10.17605/OSF.IO/YAQE8>.
- 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>.