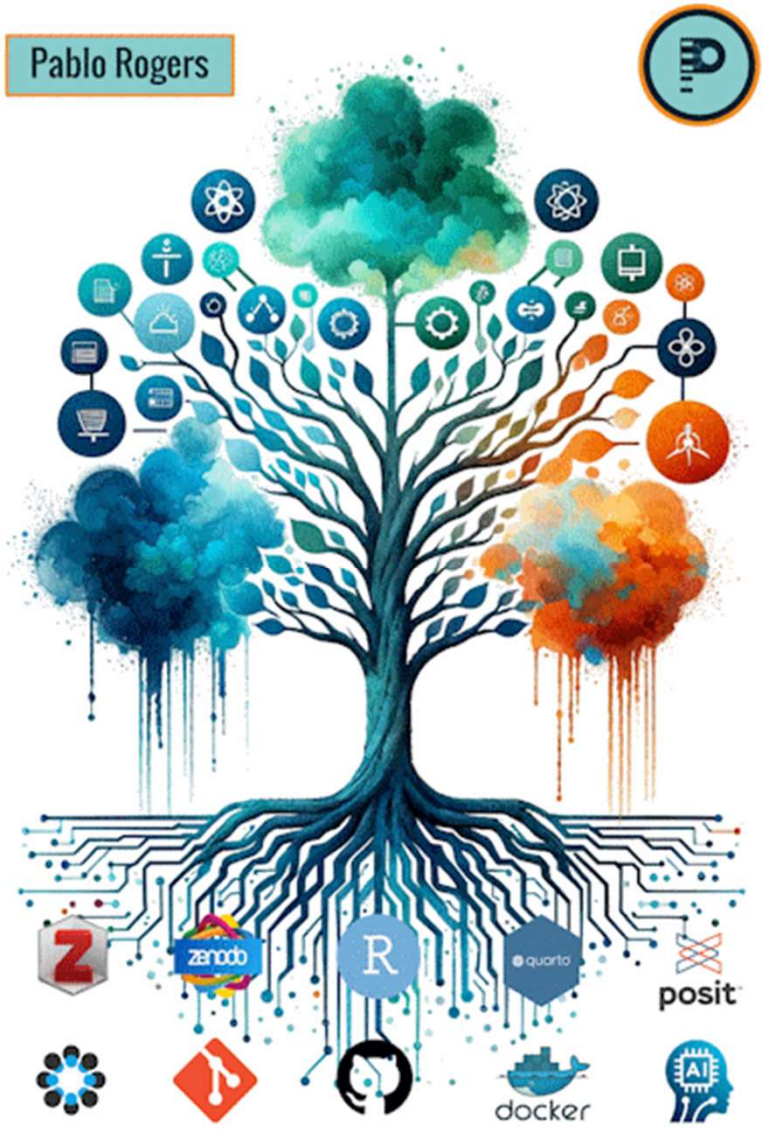
Gestão de Dados e Projetos



PERSPECTIVE

Good enough practices in scientific computing

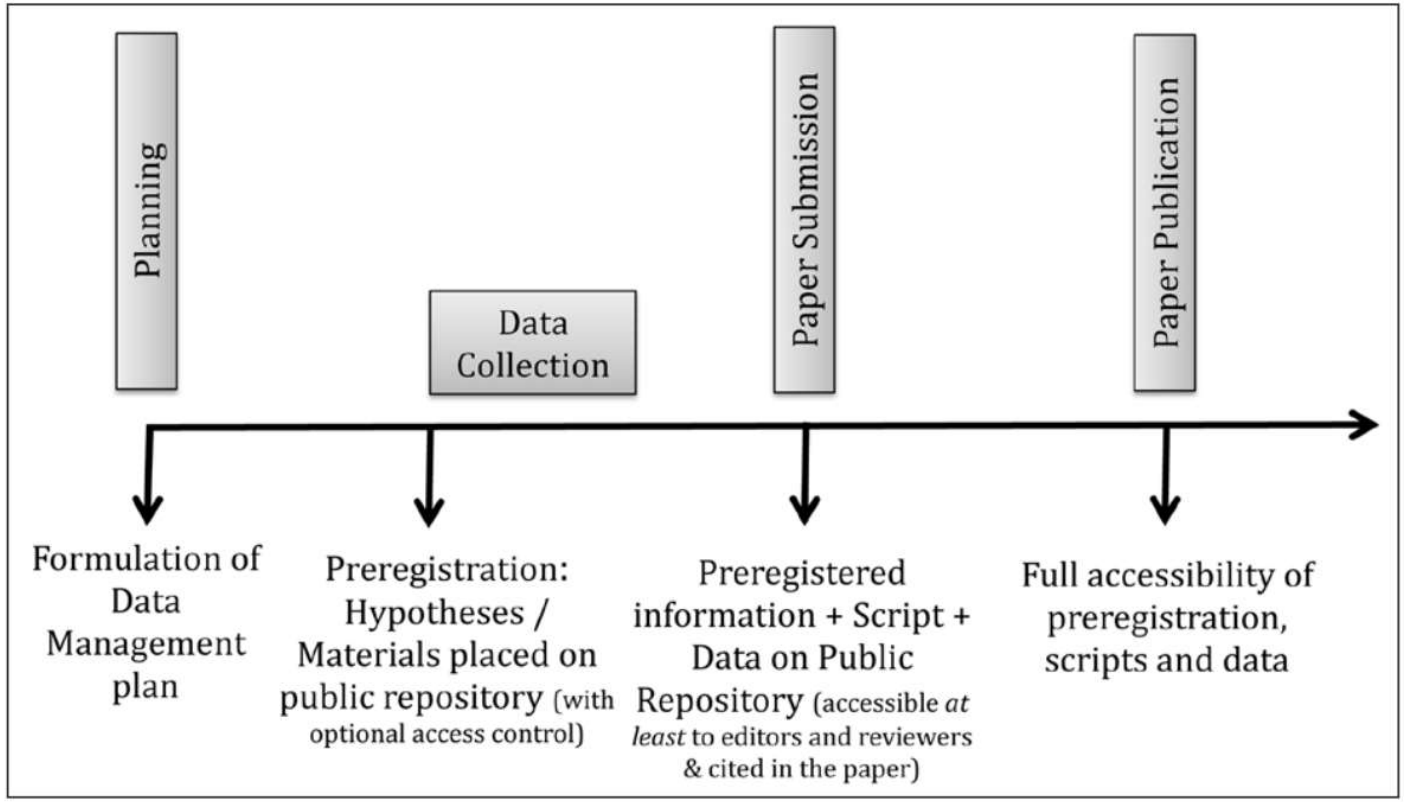
Greg Wilson[1☯]*, Jennifer Bryan[2☯], Karen Cranston[3☯], Justin Kitzes[4☯], Lex Nederbragt[5☯], Tracy K. Teal[6☯]

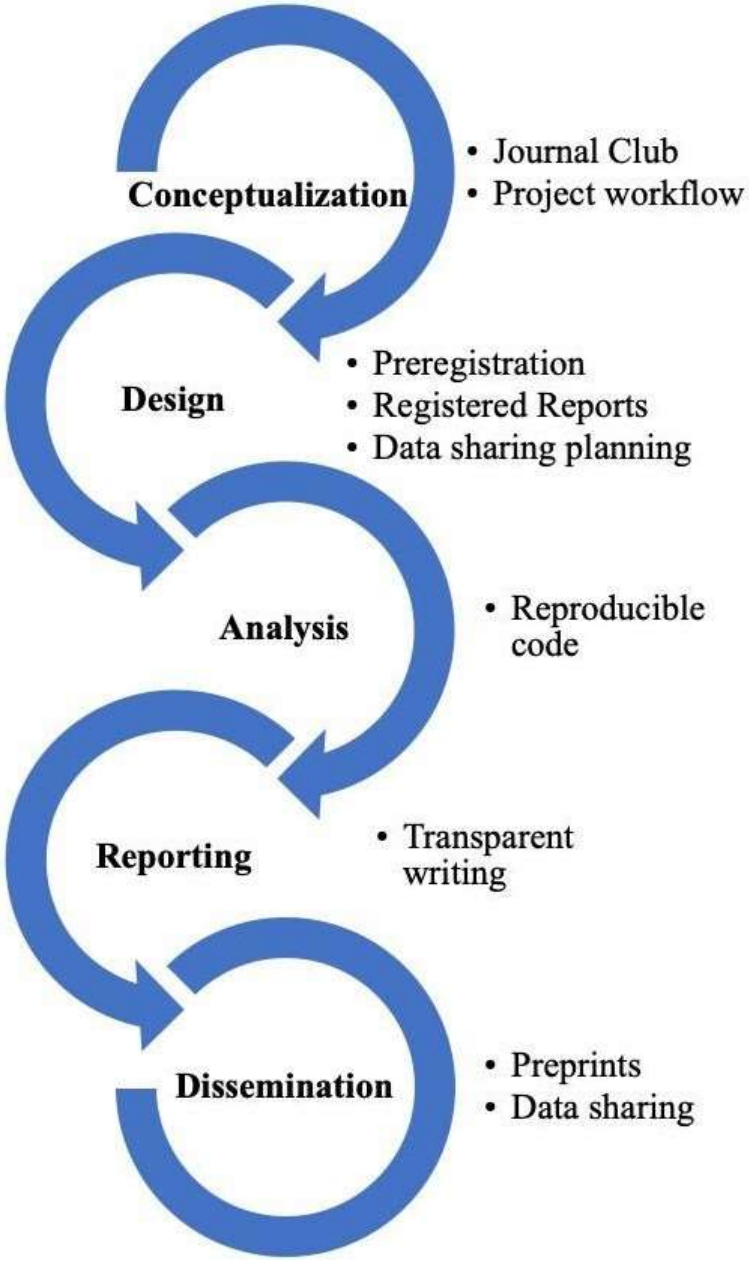
1 Software Carpentry Foundation, Austin, Texas, United States of America, **2** RStudio and Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada, **3** Department of Biology, Duke University, Durham, North Carolina, United States of America, **4** Energy and Resources Group, University of California, Berkeley, Berkeley, California, United States of America, **5** Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway, **6** Data Carpentry, Davis, California, United States of America

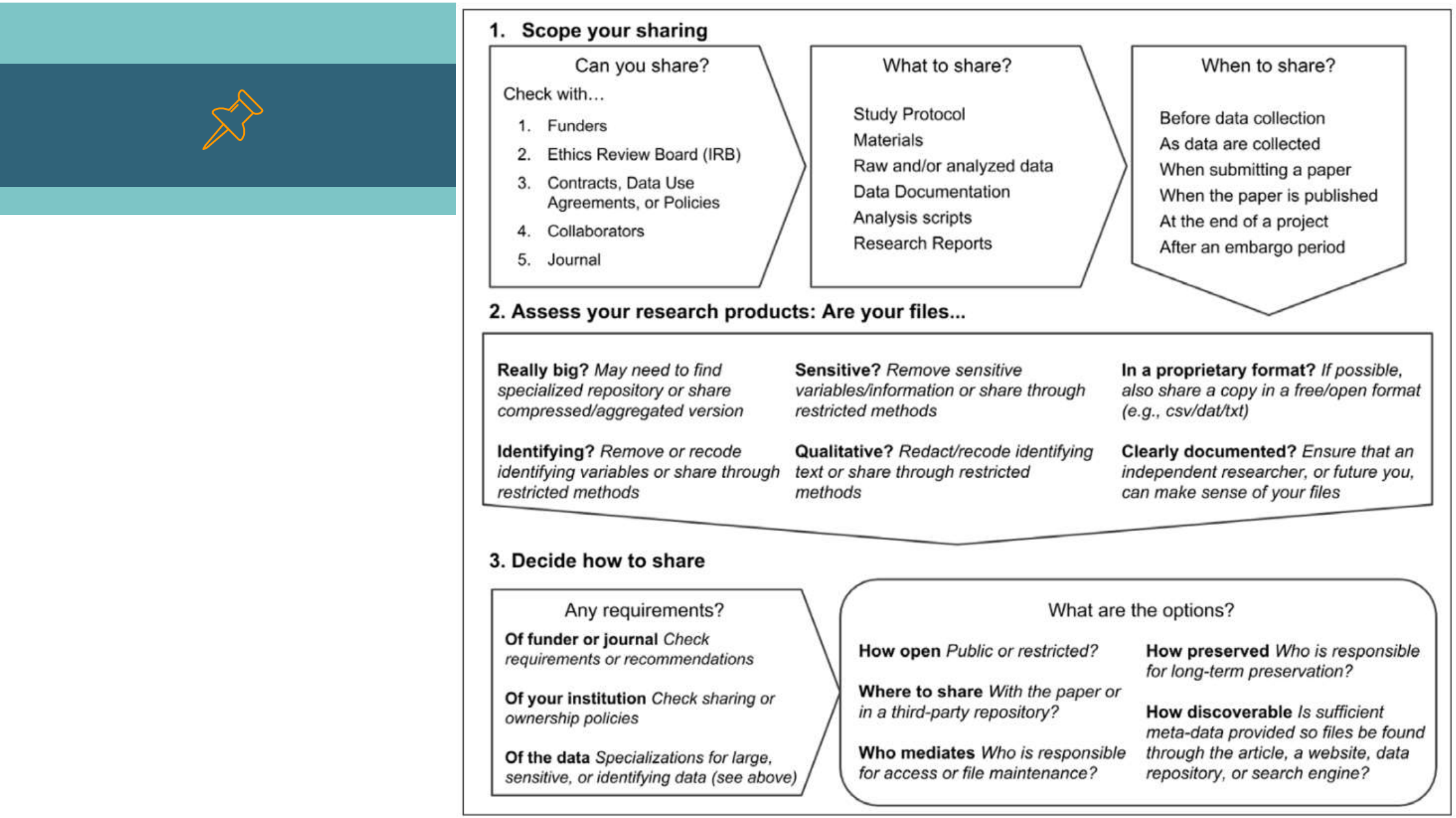
☯ These authors contributed equally to this work.
* gvwilson@software-carpentry.org

Author summary

Computers are now essential in all branches of science, but most researchers are never taught the equivalent of basic lab skills for research computing. As a result, data can get lost, analyses can take much longer than necessary, and researchers are limited in how effectively they can work with software and data. Computing workflows need to follow the same practices as lab projects and notebooks, with organized data, documented steps, and the project structured for reproducibility, but researchers new to computing often don't know where to start. This paper presents a set of good computing practices that every researcher can adopt, regardless of their current level of computational skill. These practices, which encompass data management, programming, collaborating with colleagues, organizing projects, tracking work, and writing manuscripts, are drawn from a wide variety of published sources from our daily lives and from our work with volunteer organizations that have delivered workshops to over 11,000 people since 2010.


Planning
Data Collection
Paper Submission
Paper Publication
Formulation of Data Management plan
Preregistration: Hypotheses / Materials placed on public repository (with optional access control)
Preregistered information + Script + Data on Public Repository (accessible at least to editors and reviewers & cited in the paper)
Full accessibility of preregistration, scripts and data


Conceptualization
• Journal Club
• Project workflow
Design
• Preregistration
• Registered Reports
• Data sharing planning
Analysis
• Reproducible code
Reporting
• Transparent writing
Dissemination
• Preprints
• Data sharing


1. Scope your sharing
Can you share?
Check with…
1. Funders
2. Ethics Review Board (IRB)
3. Contracts, Data Use Agreements, or Policies
4. Collaborators
5. Journal
What to share?
Study Protocol
Materials
Raw and/or analyzed data
Data Documentation
Analysis scripts
Research Reports
When to share?
Before data collection
As data are collected
When submitting a paper
When the paper is published
At the end of a project
After an embargo period
2. Assess your research products: Are your files...
Really big? May need to find specialized repository or share compressed/aggregated version
Sensitive? Remove sensitive variables/information or share through restricted methods
In a proprietary format? If possible, also share a copy in a free/open format (e.g., csv/dat/txt)
Identifying? Remove or recode identifying variables or share through restricted methods
Qualitative? Redact/recode identifying text or share through restricted methods
Clearly documented? Ensure that an independent researcher, or future you, can make sense of your files
3. Decide how to share
Any requirements?
Of funder or journal Check requirements or recommendations
Of your institution Check sharing or ownership policies
Of the data Specializations for large, sensitive, or identifying data (see above)
What are the options?
How open Public or restricted?
Where to share With the paper or in a third-party repository?
Who mediates Who is responsible for access or file maintenance?
How preserved Who is responsible for long-term preservation?
How discoverable Is sufficient meta-data provided so files be found through the article, a website, data repository, or search engine?

GERENCIAMENTO DE DADOS

a. Salve os dados brutos;
b. Backup dos dados brutos;
c. Transformar os dados brutos em formatos mais fáceis de trabalhar;
d. Adotar princípios de "dados arrumados";

GERENCIAMENTO DE DADOS

e. Registrar todas as etapas do processamento de dados;
f. Use identificadores únicos para registros em múltiplas tabelas;
g. Disponibilizar dados em repositórios com DOI;

SOFTWARE

a. Comentários iniciais em scripts;
b. Decompor rotinas em funções menores;
c. Evitar a duplicação de código;
d. Buscar e utilizar bibliotecas existentes;
e. Testar bibliotecas antes de utilizá-las;
f. Dar nomes significativos a funções e variáveis;

SOFTWARE

f. Documentar dependências e requisitos;
g. Disponibilizar conjuntos de dados de exemplo ou testes;
h. Enviar código para repositórios com DOI.

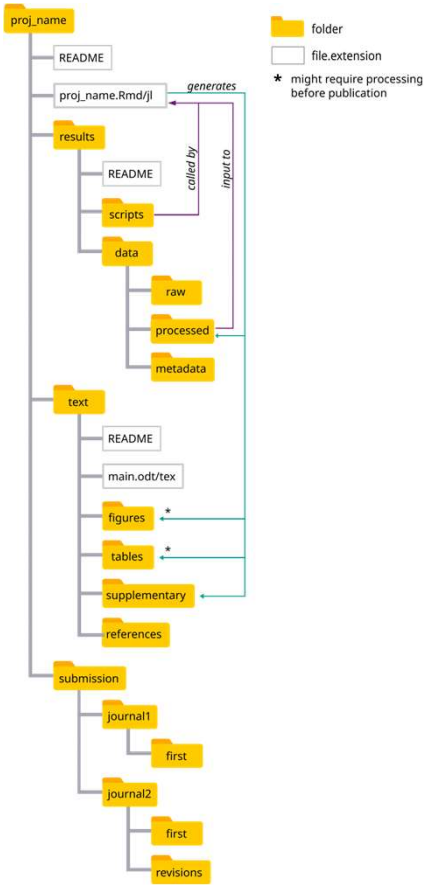
COLABORAÇÃO

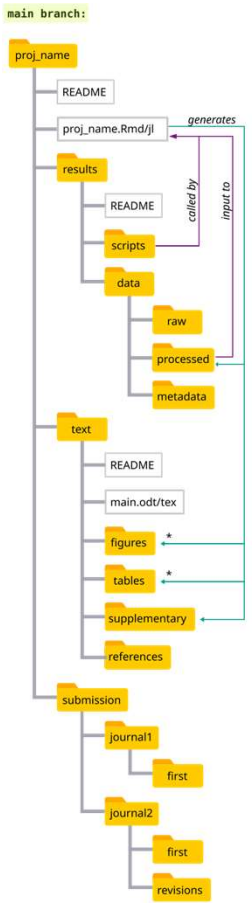
a. Desenvolver um resumo do projeto;
b. Manter uma lista de tarefas acessível a todos os colaboradores;
c. Definir e documentar estratégias de comunicação;
d. Especificar claramente a licença do projeto;
e. Tornar o projeto citável;

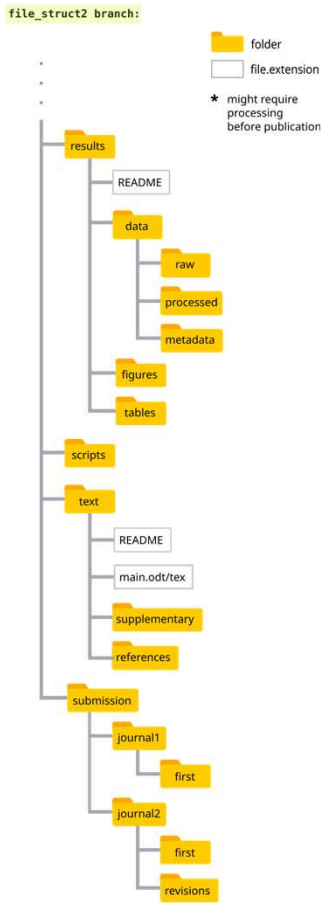
ORGANIZAÇÃO DO PROJETO

a. Cada projeto deve ter seu próprio diretório;
b. Colocar documentos de texto no diretório doc;
c. Colocar dados brutos e metadata no diretório data;
d. Centralizar as rotinas numa pasta de scripts;
e. Nomear arquivos para refletir seu conteúdo ou função;

ORGANIZAÇÃO DO PROJETO







ACOMPANHAMENTO DE MUDANÇAS

a. Fazer backup de tudo que foi criado;
b. Manter as mudanças pequenas e gerenciáveis;
c. Compartilhar mudanças regularmente;
d. Criar e usar uma checklist para gerenciar mudanças;
e. ~~Adicionar um arquivo CHANGELOG.txt ao projeto;~~
f. ~~Copiar o projeto inteiro sempre que uma mudança significativa for feita;~~
g. Uso controle de versão

MANUSCRITO

a. ~~Escrever manuscritos usando ferramentas online~~;
b. Escrever o manuscrito em um formato de texto simples que permita controle de versão.

😉 BORA COMEÇAR! 🧑

Table 1. Strategies for fostering open science in psychology and beyond

Basic strategies: becoming an open science user	
Strategy	***Concrete actions***
Experiment with integrating open science practices and policies into daily workflows	Start small: use open-source software, publish a preprint, pre-register a study or sign a peer-review (i.e., open peer reviews). As you get more comfortable with open science practices, share lab notebooks, deposit data and code in public repositories, or write a registered report.
Become familiar with national and institutional open science policies	Create spaces to read, share, and discuss open science mandates developed by government agencies and other institutions.
Analyze and share successful cases of implementation of open science practices	Become familiar with and inspired by cases of successful implementation of open science practices. Discuss them with others. Identify, compare and eventually implement strategies to spur open science at your institution.
Introduce open science practices in your courses	Incorporate open science practices in course content. Allow students to come into contact and experiment with the open research culture before carrying out major projects.
Embrace open educational resources (OERs)	Use and/or create OERs, such as open courses, syllabi, lectures, assignments, and textbooks. If you create OERs, publish them under a Creative Commons license through open repositories.

😊 BORA COMEÇAR! 🧑

Advanced strategies: from open science user to advocate	
Strategy	***Concrete actions***
Collaborate with others using open tools	Exploit collaborations to rethink, question, and overcome embedded closed practices. Try writing an article with others using R Markdown or Quarto, sharing data using open formats, or keeping a version history of research files using GitHub or similar platforms.
Develop networks of open collaboration	Establish collaborative networks with others that advocate for open research and work on related topics. Develop standards and guidelines to wrangle, analyze, and visualize data, to comment and proofread code, and to create accompanying documentation.
Voice your opinion	Write testimonials, reports, glossaries, or declarations of support for open science. Outline principles, values, and considerations relevant to your community.
Rethink and promote changes in the assessment of scholarly production	Take ownership of the discussion on research assessment. Collaborate in building spaces to discuss and promote alternatives to the traditional measurement of academic performance (e.g., impact factor, h-index). Advocate for the recognition of a variety of research outputs (e.g., open datasets, open peer reviews).
Create opportunities for people to specialize in open science	If you have or can pursue funding to hire personnel, create open-science related job opportunities (e.g., PhD scholarships and research assistantships, research manager positions, postdoctoral positions).