

Computational notebooks for more openness, reproducibility, and productivity in research

A minimal setup to get you started

Dr. Ludmilla Figueiredo
Data & Code Curator (iBID, iDiv)
ECEM 2023
03.09.2023

Agenda

- Round of introductions
- Motivation
- The kit
- (Break)
- Demo/Try out time
- Q&A

Round of introductions

- Who are you
- What are you working on
- How much data and code do you work with
- Which tools do you use: programming languages, editors, version control

Round of introductions



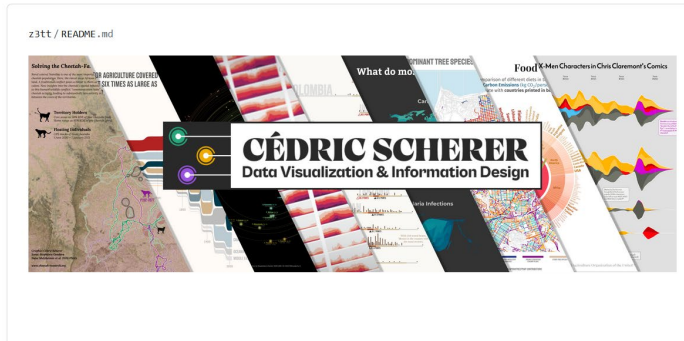
Dr. Ludmilla Figueiredo
iDiv Data & Code Curator



Prof. Juliano Sarmiento Cabral, University of Birmingham
The road to integrate climate change projections with regional land-use-biodiversity models
Thursday, 7th September, Hall 1 CD, S 30a



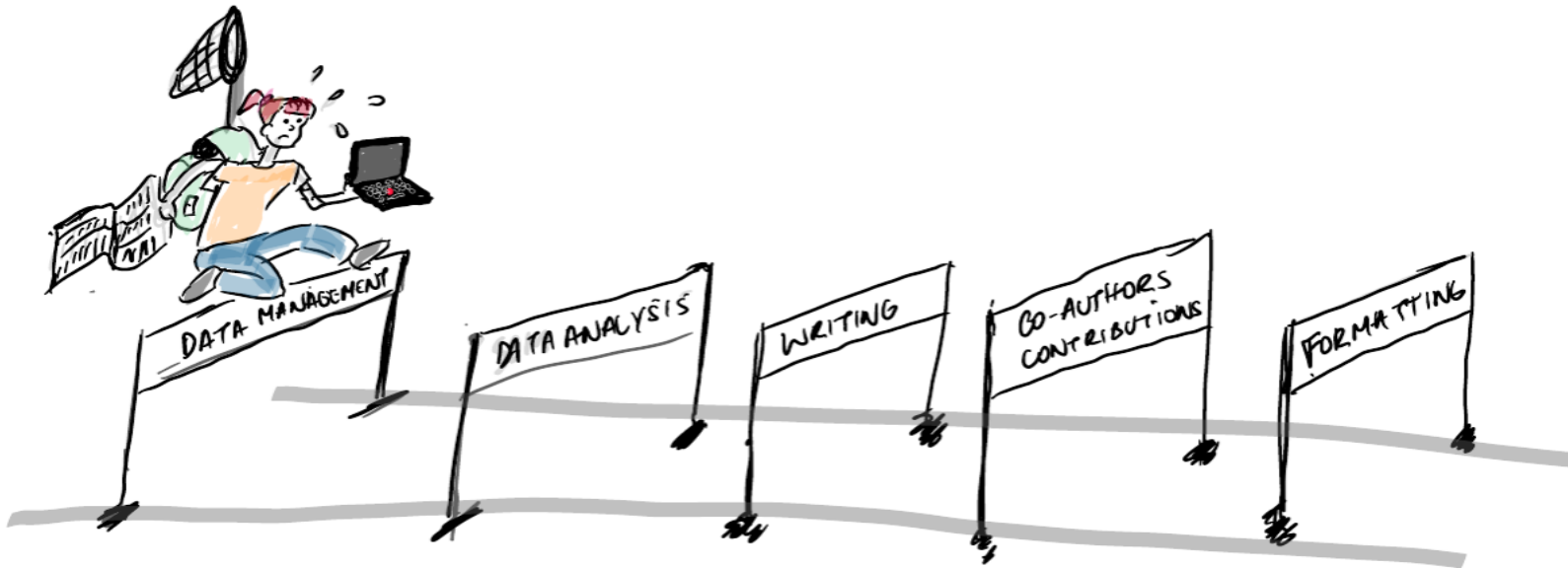
Cédric Scherer
z3tt



Round of introductions

- Who are you
- What are you working on
- How much data and code do you work with
- Which tools do you use: programming languages, editors, version control

Facilitate traceability and reproducibility



CC-BY Ludmilla Figueiredo



Open science



Imming & Tennant (2018)



Open science



Imming & Tennant (2018)

- = Open data
- = Open access
- = Open source
- = Pre-registration
- = Open peer reviews
- = Open standards
- = Open protocols



Open science



Imming & Tennant (2018)

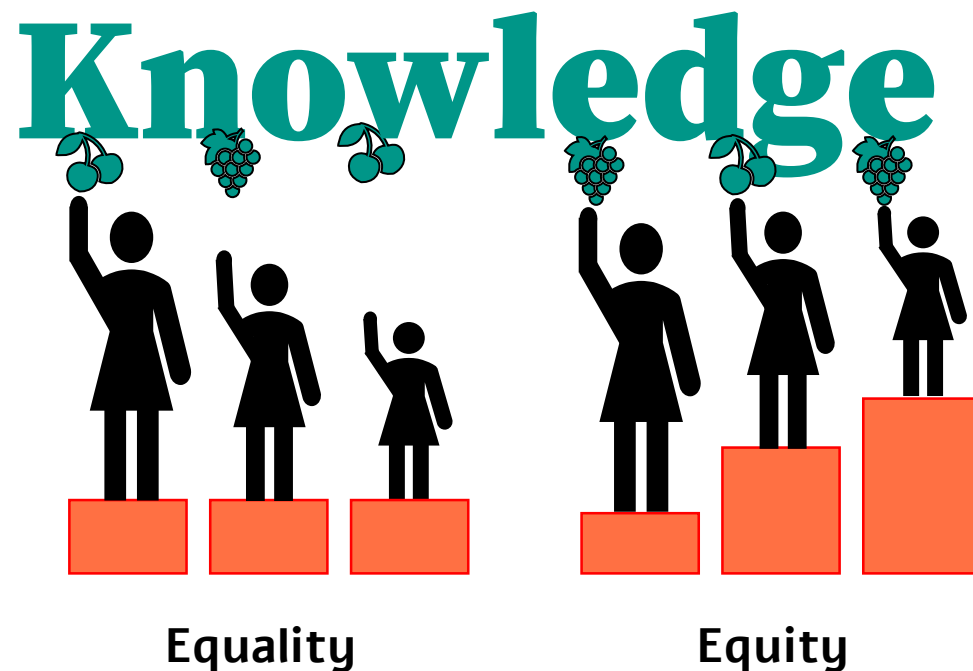
Requirements for FAIR/O:

- Journals
- Institutions
- Funding agencies:
 - European Research Council
 - Horizon Europe
 - DFG



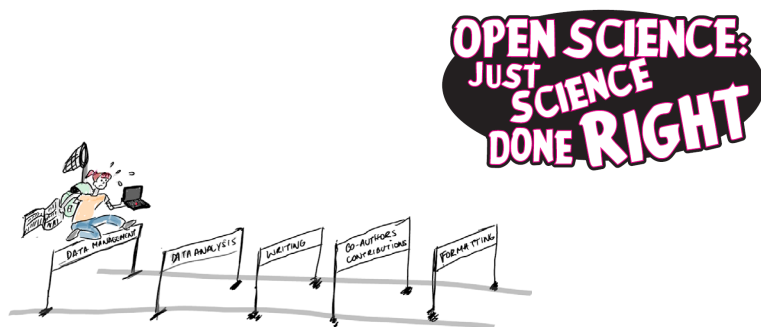
Knowledge equity

- Done is better than perfect
- Ease of use despite computational background



Facilitate traceability and reproducibility

💡 Knowledge equity



HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)






<https://xkcd.com/927>

➡ “minimal” setup

PLOS COMPUTATIONAL BIOLOGY

EDUCATION

A simple kit to use computational notebooks for more openness, reproducibility, and productivity in research

Ludmilla Figueiredo ^{1a,b,*}, **Cédric Scherer** ², **Juliano Sarmiento Cabral** ^{1,3}

1 Ecosystem Modeling, Center for Computational and Theoretical Biology (CCTB), University of Würzburg, Würzburg, Germany, **2** Department of Ecological Dynamics, Leibniz Institute for Zoo and Wildlife Research (IZW), Berlin, Germany, **3** Biodiversity Modelling and Environmental Change, School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom

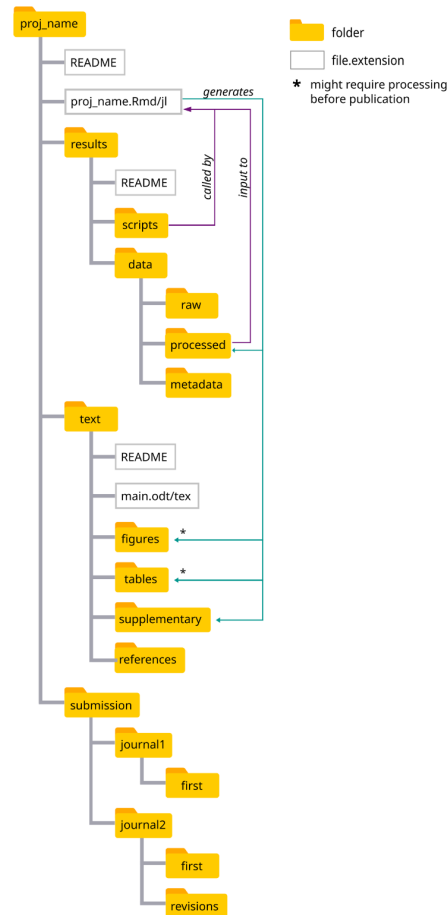
<https://doi.org/10.1371/journal.pcbi.1010356>

Contents

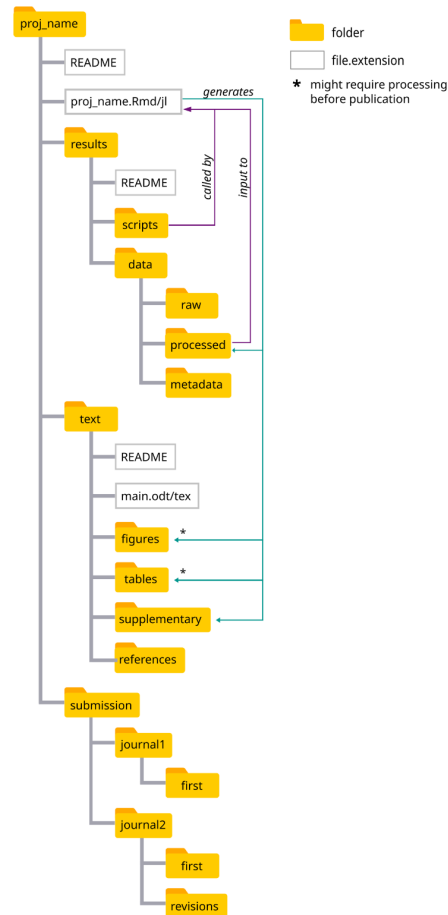
Direct download

→ `set_kit.R`

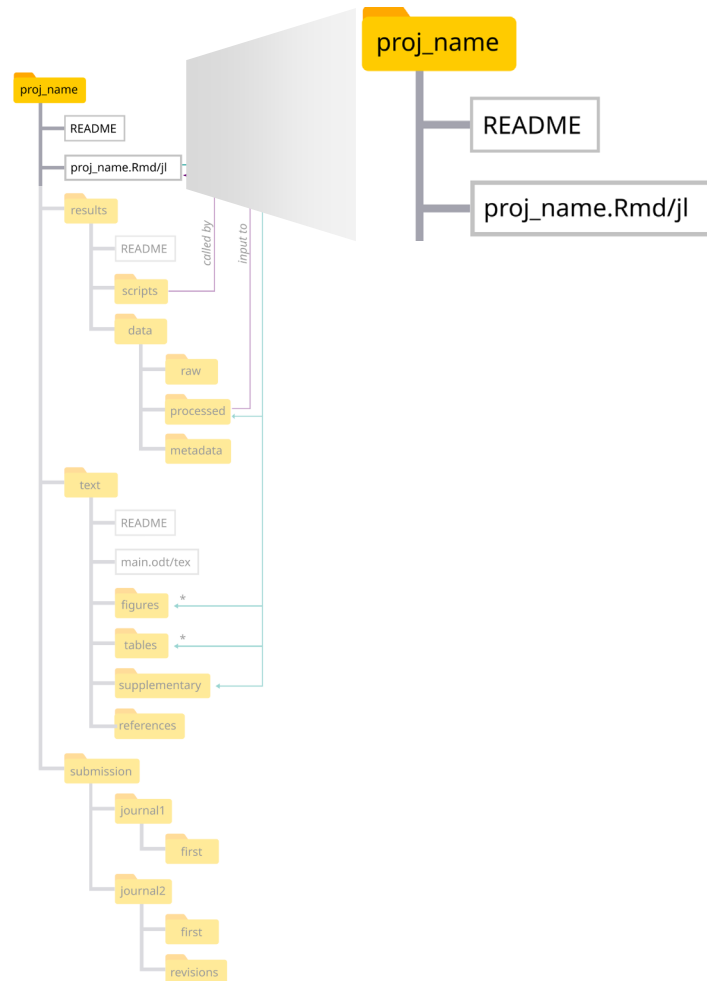
→ `set_kit.jl`



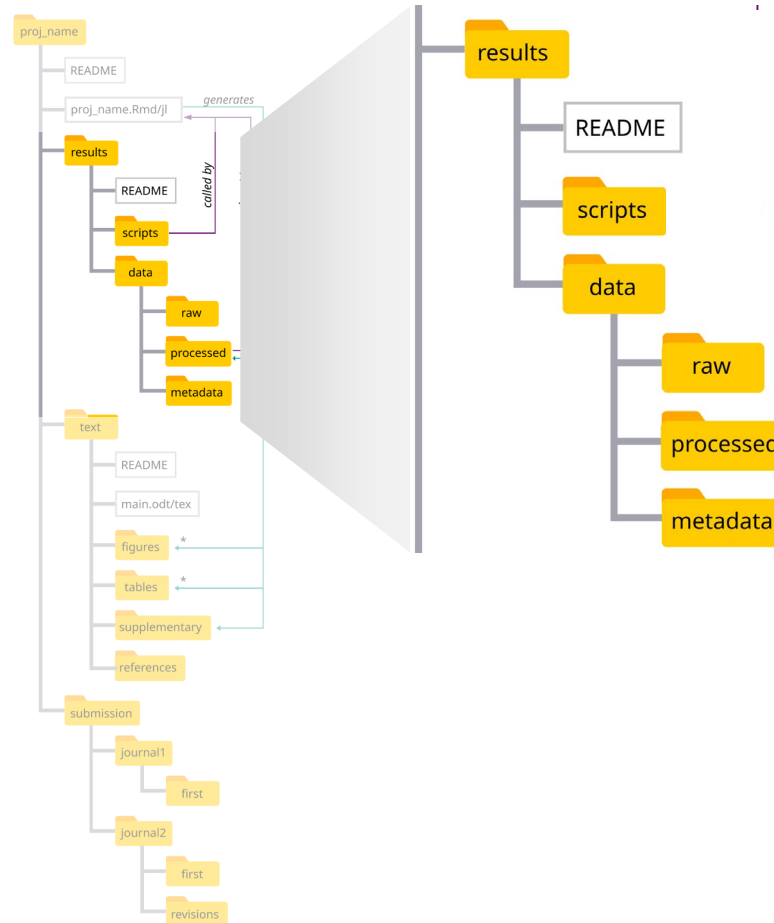
File structure



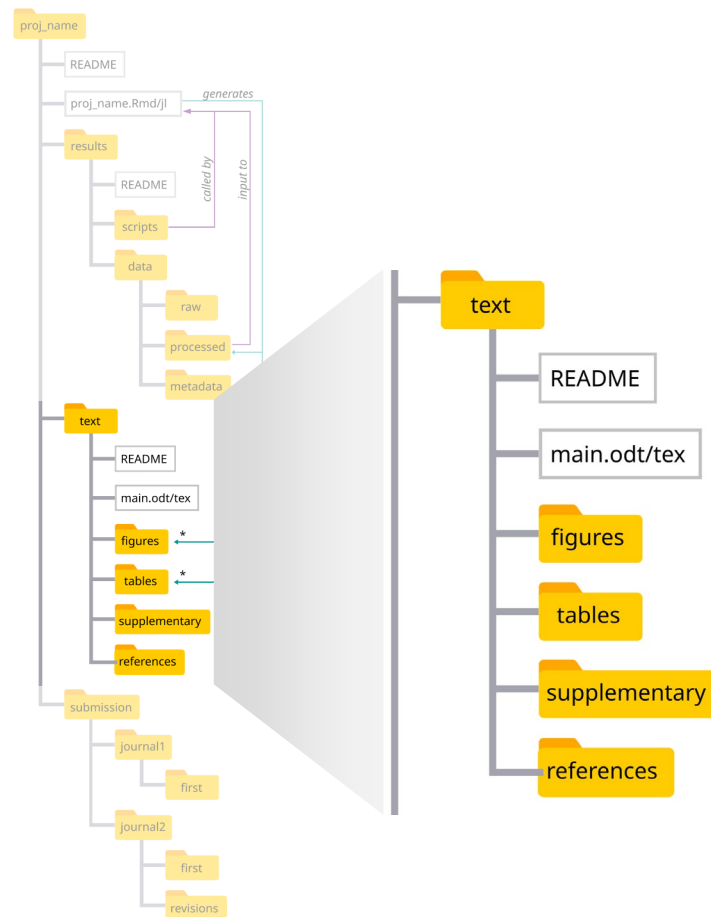
File structure



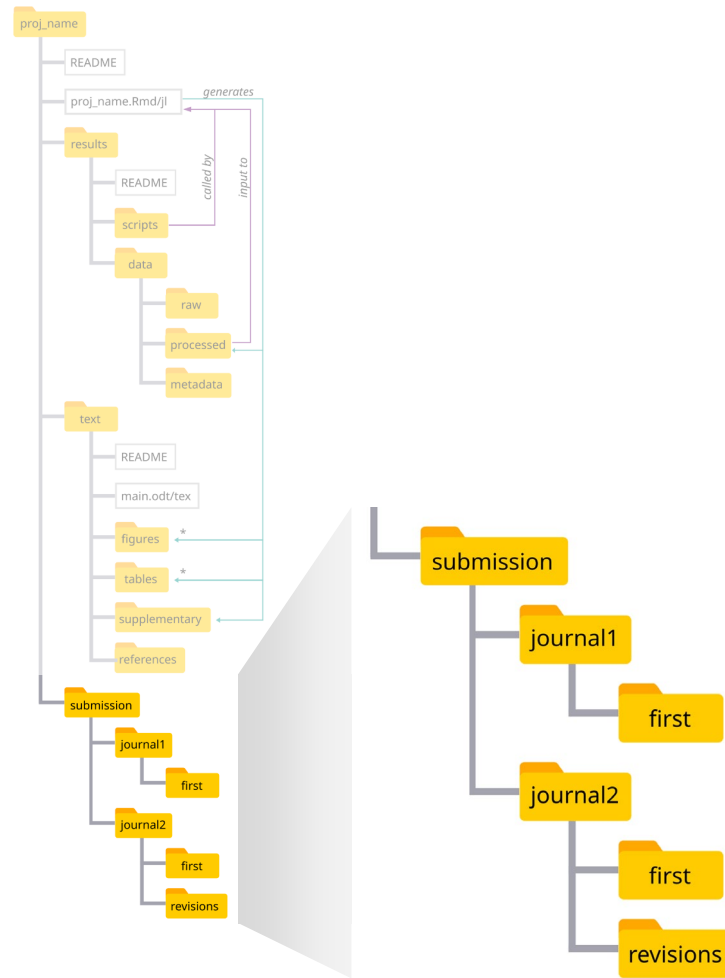
File structure



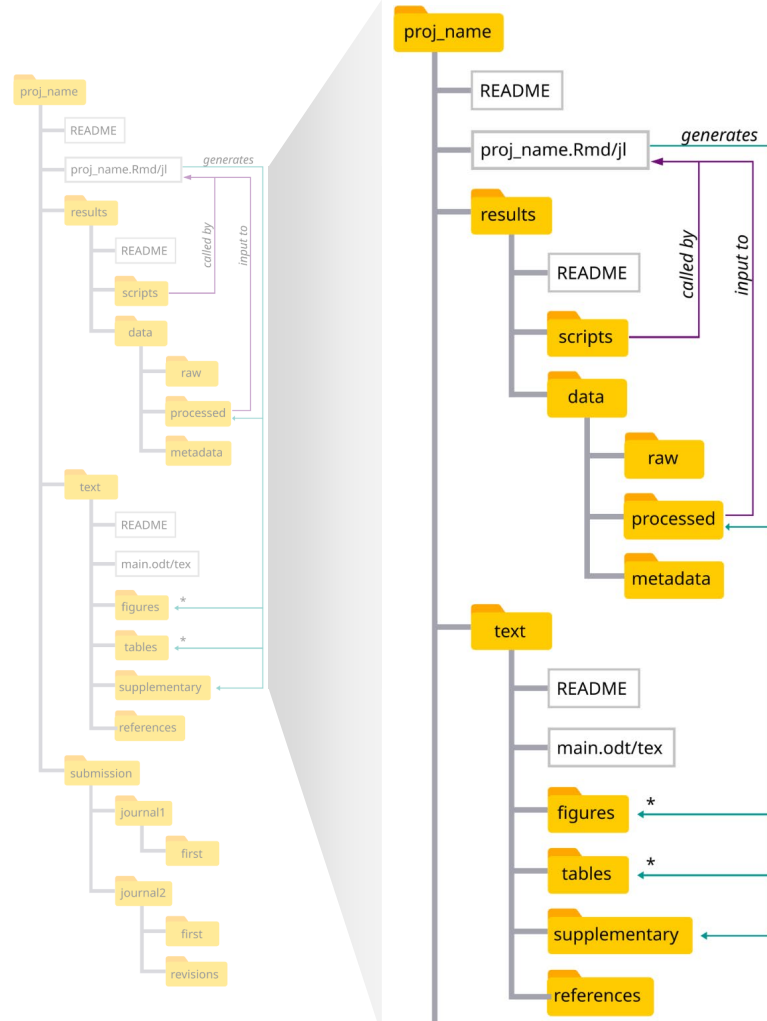
File structure



File structure

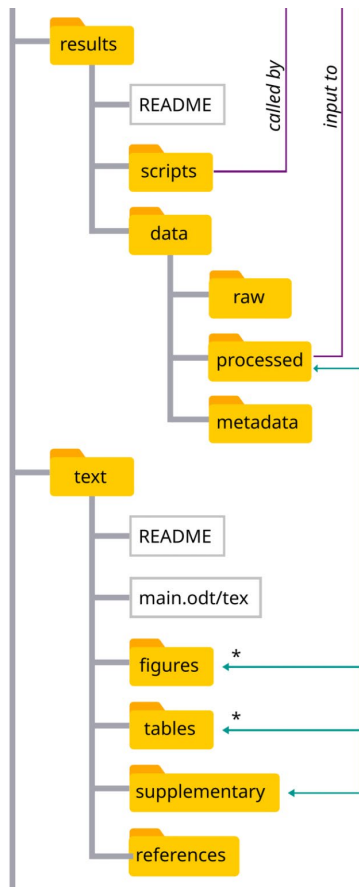


File structure

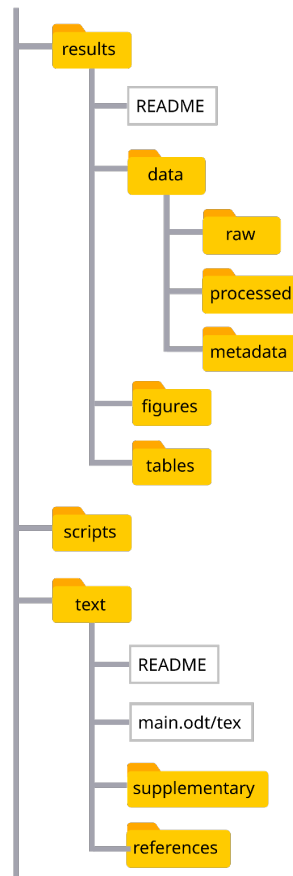


New file structure

main branch:



file_struct2 branch:



File structures

Notebook is adapted to handle both

main branch:

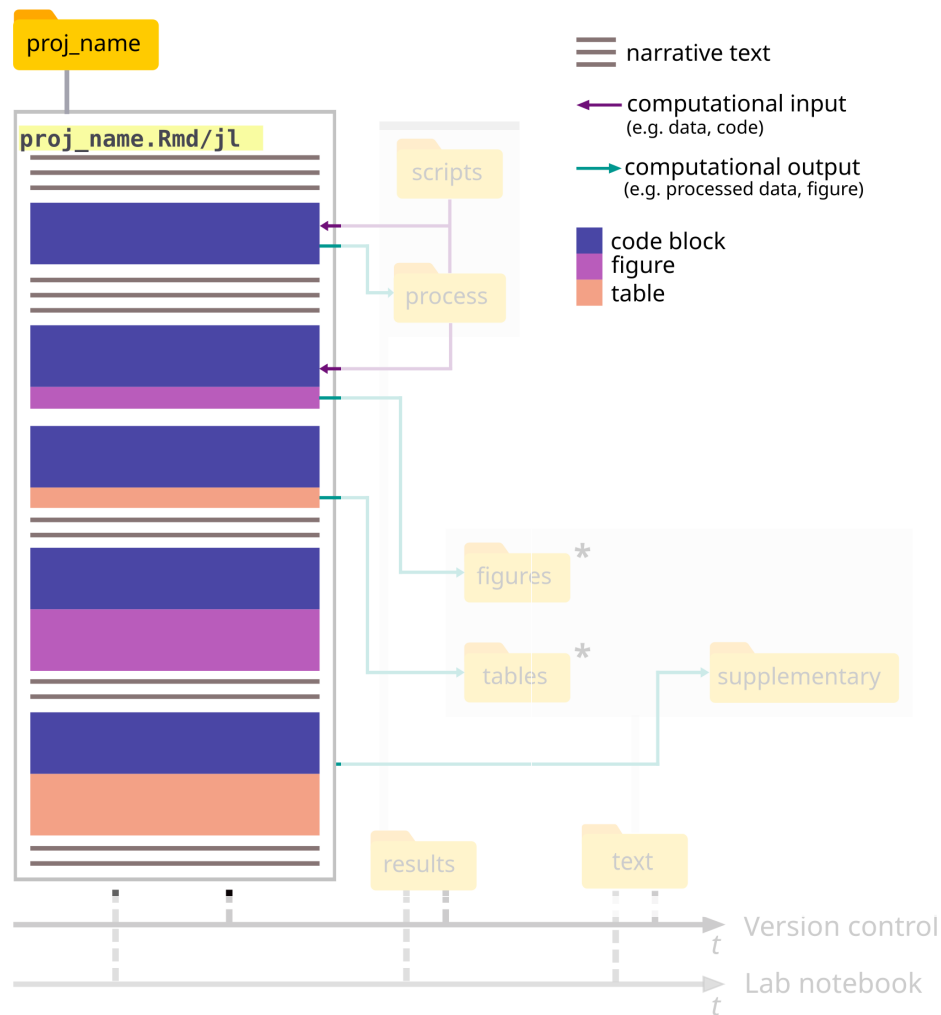
```
data_dir <- file.path("results", "data", "processed")
scripts_dir <- file.path("results", "scripts")
suppl_dir <- file.path("results", "supplementary")
figs_dir <- file.path("text", "figures")
tabs_dir <- file.path("text", "tables")
```

file_struct2 branch:

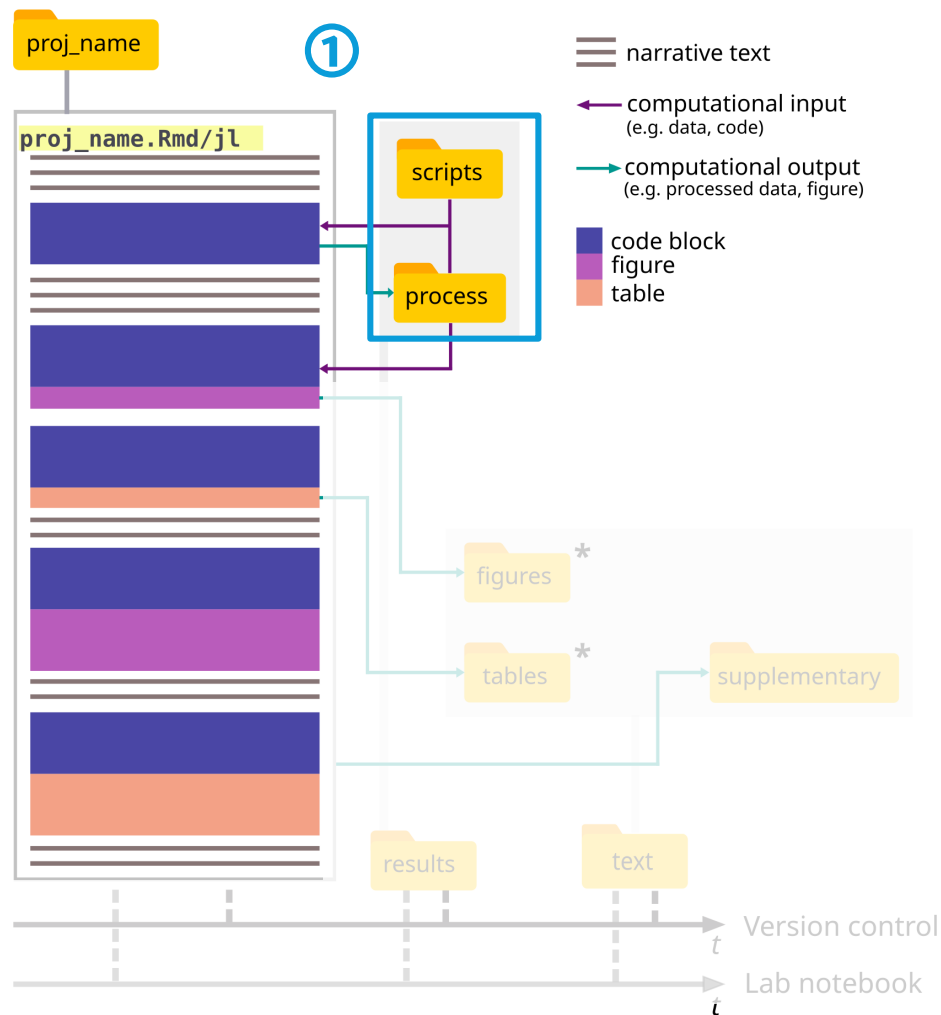
```
data_dir <- file.path("results", "data", "processed")
scripts_dir <- file.path("scripts")
figs_dir <- file.path("results", "figures")
tabs_dir <- file.path("results", "tables")
suppl_dir <- file.path("text", "supplementary")
```



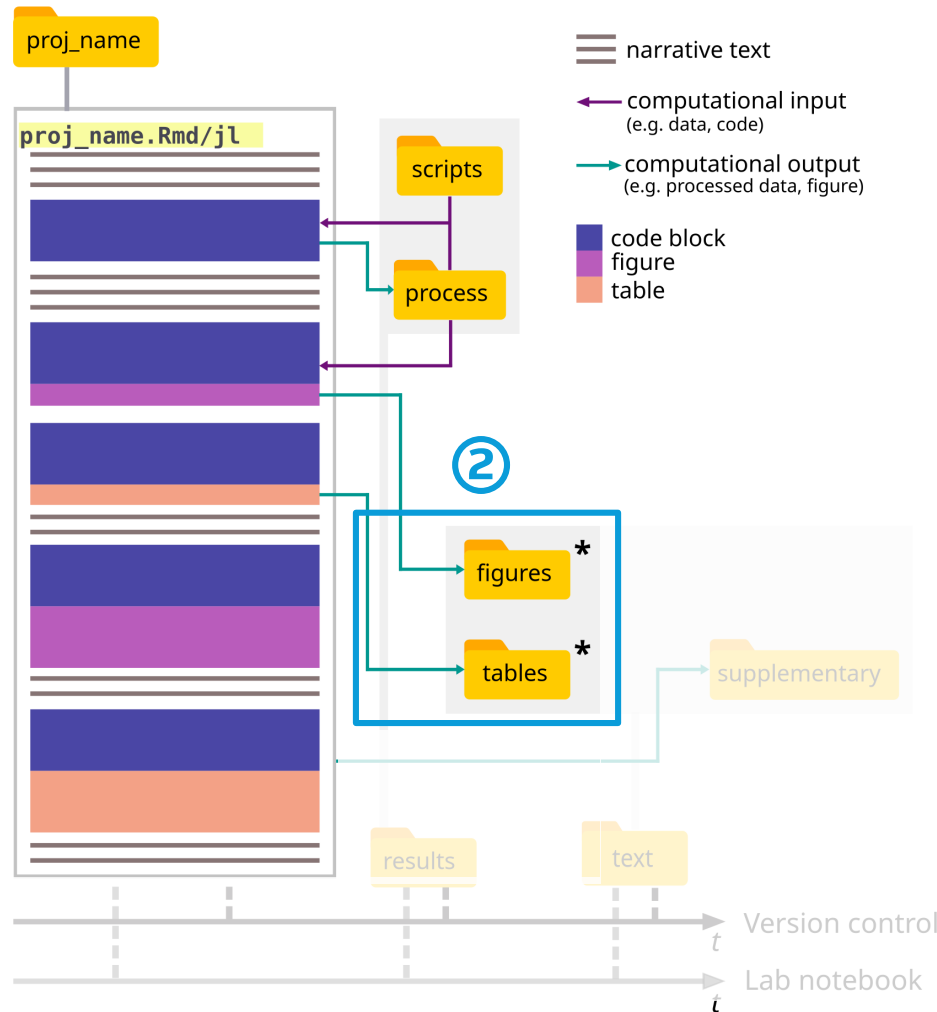
One notebook to rule them all



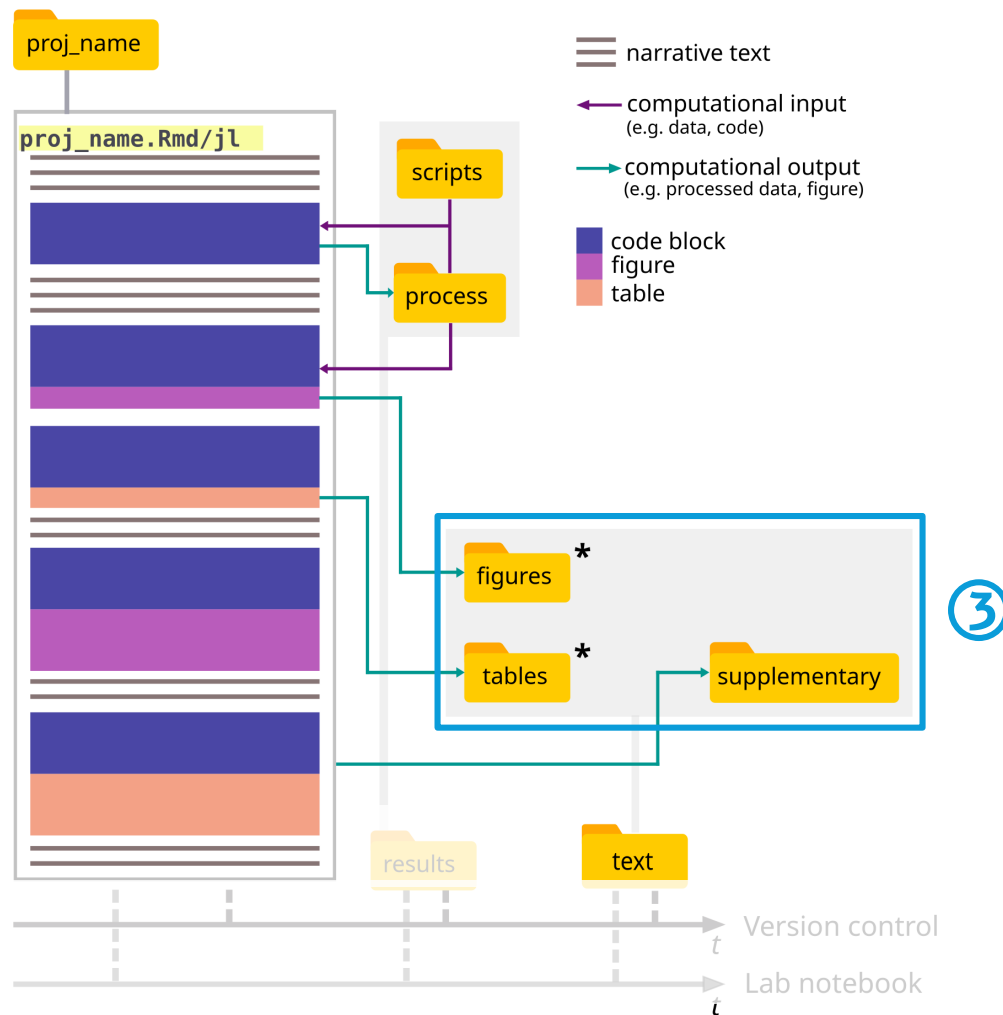
One notebook to rule them all



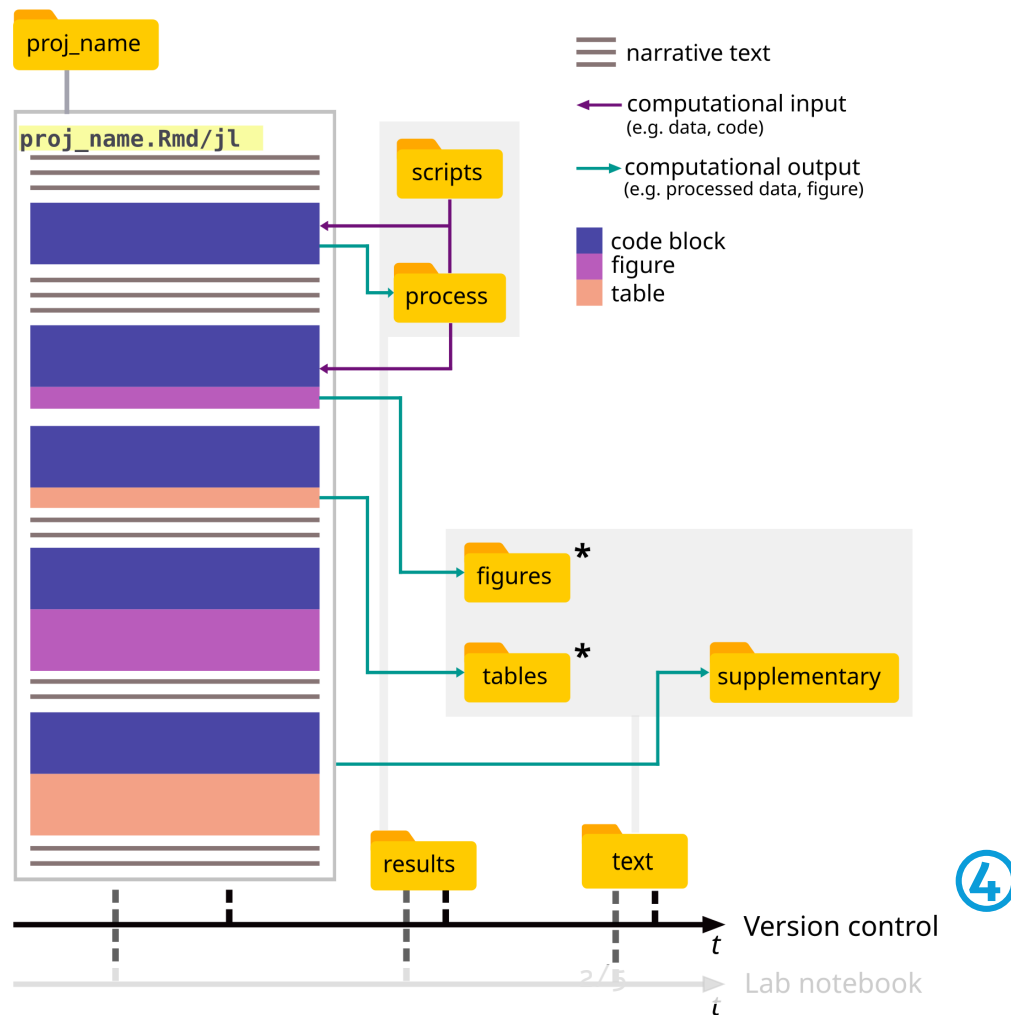
One notebook to rule them all



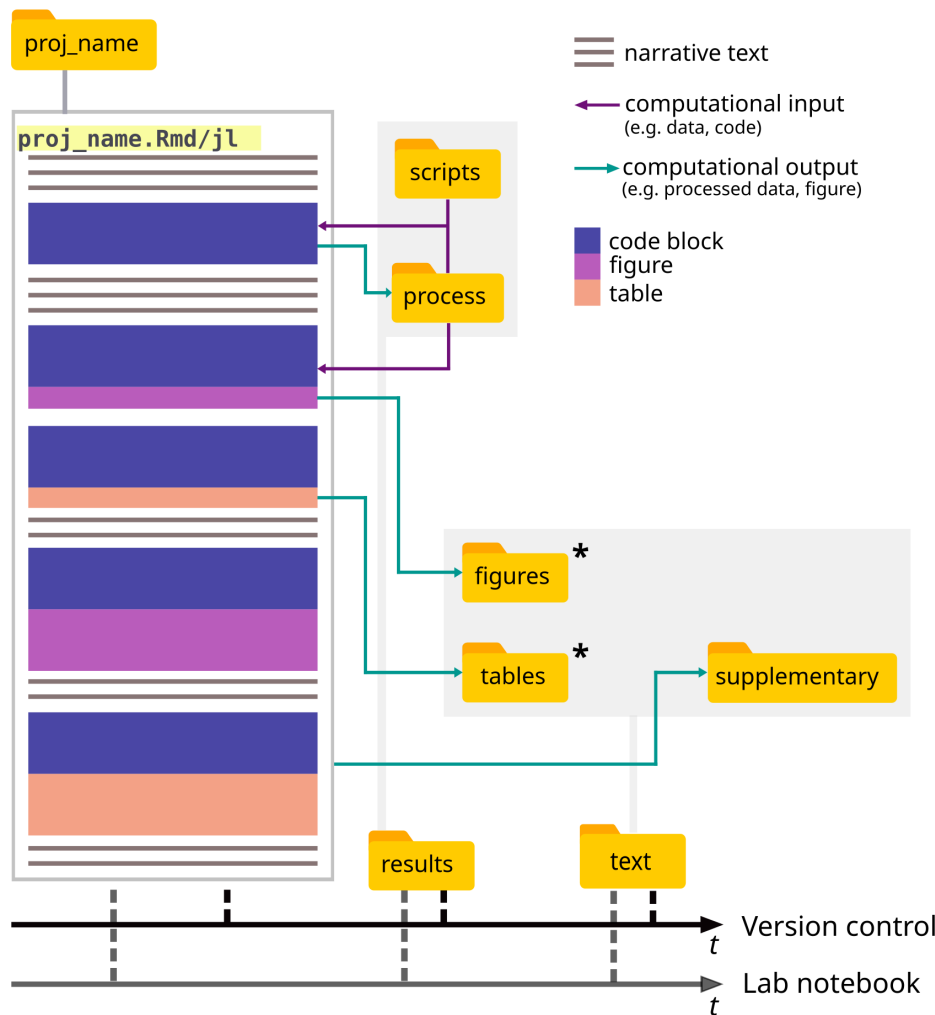
One notebook to rule them all



One notebook to rule them all



One notebook to rule them all



One notebook to rule them all

Examples
of
Markdown
syntax

Short
paths to
relevant
directories

Suggestion
of use

```

RStudio Source Editor
test.Rmd
1-
2 title: "Your title here"
3 output:
4   html_document:
5     code_folding: hide
6     toc: true
7     pdf_document: default
8-
9
10 <!-- ----- VERY MINIMAL INTRO -----
11
12 This is your notebook. When filling it, you should obey the
13 Text written directly in the file like this one you are reading
14 this file (click the "Knit" button in the upper-left part
15
16 <!-- Text written like this will not be included in HTML
17
18 This line includes `inline code`. Inline code is only formatted
19
20 Blocks of code (also called 'chunks') are marked like this:
21 ```{r}
22 print("This is an example of a code block")
23 ```
24
25 To execute the code in the chunk and have its results appear
26
27 To insert add a chunk, click the 'Insert' button on the toolbar
28
29 It is also possible to include hyperlinks: [link to further details]
30
31 very important, check how to adjust how much of your code and
32
33 <!-- ----- END OF VERY MINIMAL INTRO -----
34
35 ```{r set-up}
36 # Create objects to provide easy access to the folders related to the project.
37 # The code in this chunk is not relevant for the reader, and thus is not included in the knitted version (therefore, 'include = FALSE').
38 # This is also useful if you are not using the folder structure suggested by this kit, and want to preserve your privacy.
39 # Feel free to edit this path as you wish.
40
41 data_dir <- file.path("results", "processed") ## Do NOT play with stuff in raw_data. That is your back-up. Work only on 'processed'.
42 scripts_dir <- file.path("results", "scripts")
43 suppl_dir <- file.path("results", "supplementary")
44 semiprods_dir <- file.path("results", "semi_products")
45 figures_dir <- file.path("text", "figures")
46 tables_dir <- file.path("text", "tables")
47
48
49 # Read data
50 ```{r}
51 # e.g.
52 # data <- read.csv(file.path(data_dir, "your_data.csv")) ## replace with the name of the file you want to read in
53 ```

```

Your title here

- Read data
- Data analysis
 - Figure 1
 - Figure S1

This is your notebook. When filling it, you should obey the Markdown syntax. Text written directly in the file like this one you are reading, with no special markings, is what we call "narrative", and will appear without any special formatting when you generate a .html or .pdf version of this file (click the "Knit" button in the upper-left part of this panel to do it).

This line includes `inline code`. Inline code is only formatted differently from narrative text, it is not executed.

Blocks of code (also called 'chunks') are marked like this:

```
## [1] "This is an example of a code block"
```

To execute the code in the chunk and have its results appear below it, click the green "play" ("Run") button within the chunk or place your cursor inside the chunk and press `Ctrl+Shift+Enter`.

To insert add a chunk, click the `Insert` button on the toolbar (upper-right of this panel) or press `Ctrl+Alt+I`.

It is also possible to include hyperlinks: [link to further details on the Markdown syntax used in R notebooks](#).

Very important, check how to adjust how much of your code and outputs you want to show in your notebook when you convert it to .html or .pdf [here](#).

```

# Create objects to provide easy access to the folders related to the project.
# The code in this chunk is not relevant for the reader, and thus is not included in the knitted version (therefore, 'include = FALSE').
# This is also useful if you are not using the folder structure suggested by this kit, and want to preserve your privacy.
# Feel free to edit this path as you wish.

data_dir <- file.path("results", "processed") ## Do NOT play with stuff in raw_data. That is your back-up. Work only on 'processed'.
scripts_dir <- file.path("results", "scripts")
suppl_dir <- file.path("results", "supplementary")
semiprods_dir <- file.path("results", "semi_products")
figures_dir <- file.path("text", "figures")
tables_dir <- file.path("text", "tables")

```

Read data

```

# e.g.
# data <- read.csv(file.path(data_dir, "your_data.csv")) ## replace with the name of the file you want to read in

```

A) What you edit

B) What the reader sees

Live Demo



Try it out!

Kits:

https://github.com/FellowsFreiesWissen/computational_notebooks

<http://tinyurl.com/notebooks2023>



Data:

<https://idata.idiv.de/ddm/Data/ShowData/3519?version=0>

<http://tinyurl.com/dataECEM2023>



Next steps

- Electronic lab notebooks
- Writing manuscripts in Rmarkdown or Quarto
- R packages to build workflow:
 - workflowr (Blischak et al. 2019)
 - targets (Landau 2021)
- Organize your work as an R package or research compendium
- renv, Binder, Docker image/containers
- Emacs' Org mode: fully customizable digital lab notebook



Thank You!

Prof. Juliano Sarmento Cabral & Dr. Cedric Scherer
Center for Computational and Theoretical Biology, Uni Wuerzburg

Wikimedia Deutschland & Freies Wissen Program