

Open Science and Research Software Engineering

Building Blocks for Quality Research

Dr. Heidi Seibold

Co-executive director Digital Research Academy

21.04.2026, TRR 356 PlantMicrobe Seminar



DIGITAL RESEARCH
ACADEMY

Slides are licensed under CC-BY 4.0
[digital-research.academy](https://creativecommons.org/licenses/by/4.0/)



IS THERE A REPRODUCIBILITY CRISIS?



A computational reproducibility study featuring longitudinal data analyses

Heidi Seibold^{*,1,2,3,4}, Severin Czerny¹, Siona Decke¹, R Steffen Fohr¹, Nico Hahn¹, Rabea Hartmann¹, Christo Dario Lepke¹, Verena Loidl¹, Maximilian Mandl¹, Sara Alexander Piehler¹, Elio Rojas¹, Stefanie Schmid¹, Hann Lennart Schneider¹, Xiao-Yin To¹, Viet Tran¹, Antje V Wagner¹, Maria Waize¹, Hannah Wecker¹, Rui Yang¹,

1 LMU Munich, Germany

2 University of Bielefeld, Germany

3 Helmholtz Zentrum München, Germany

4 LMU Open Science Center

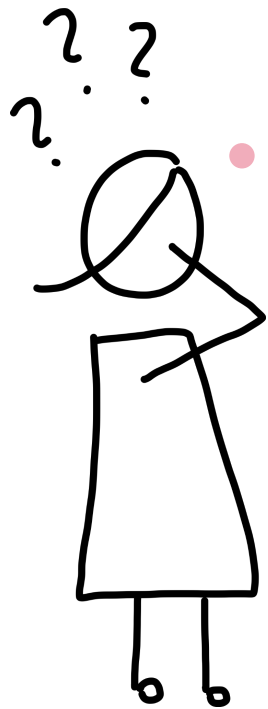
Why Most Published Research Findings Are False

John P. A. Ioannidis

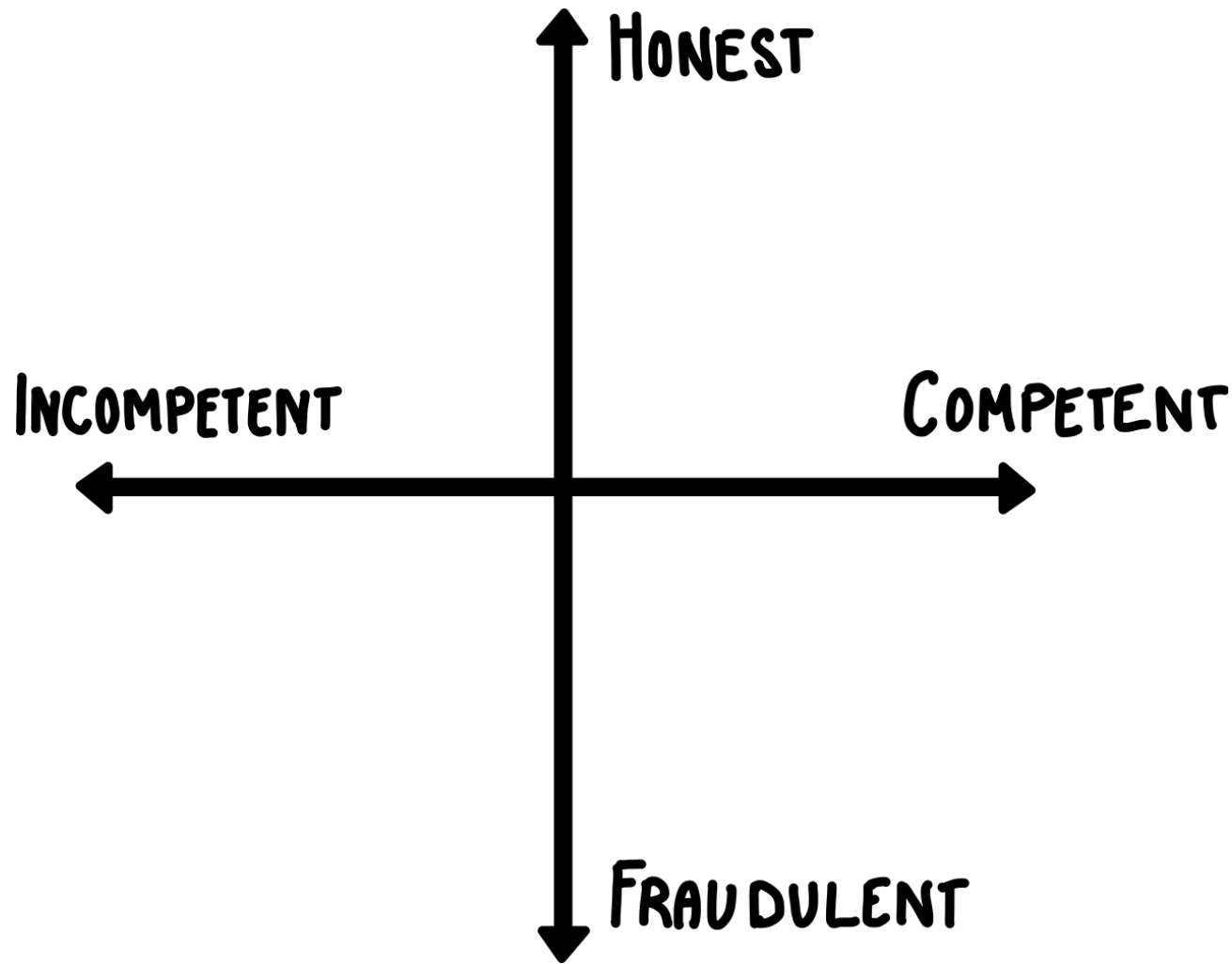
[DOI: 10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)

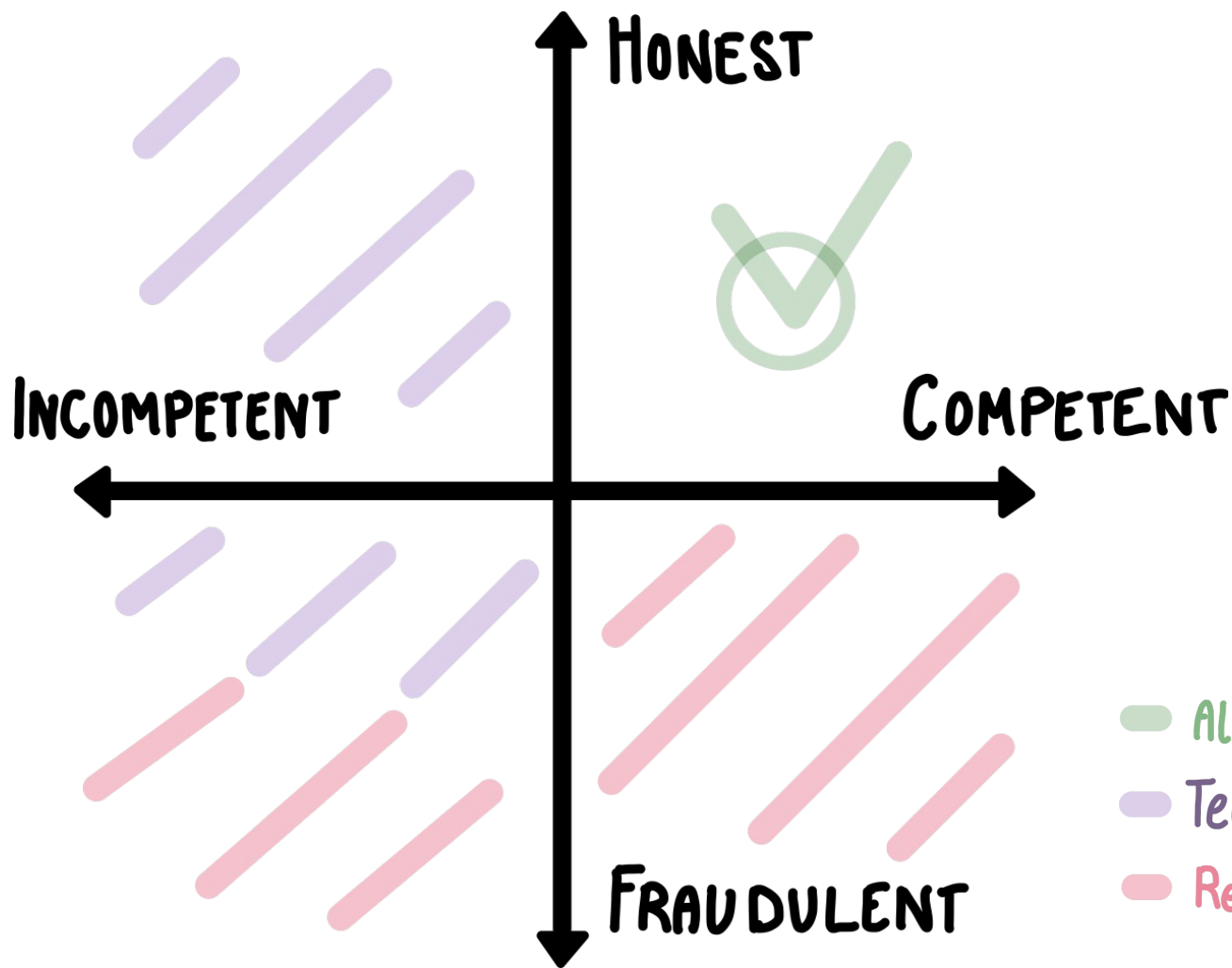
Is this what we want?





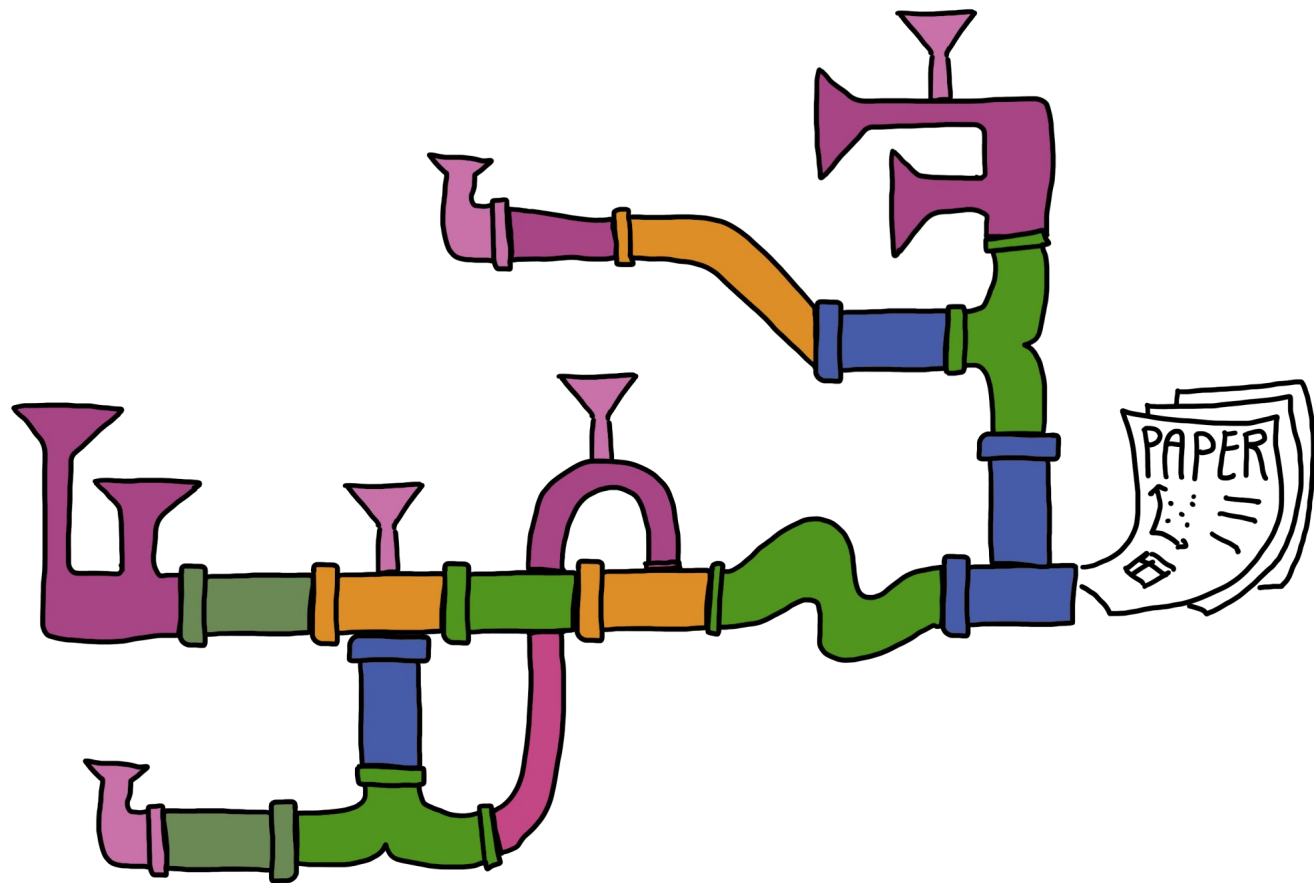
↓ FRAUD





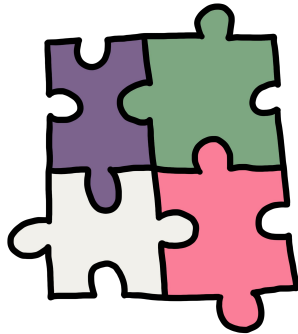
- All good
- Teach
- Regulations / Nothing we can do





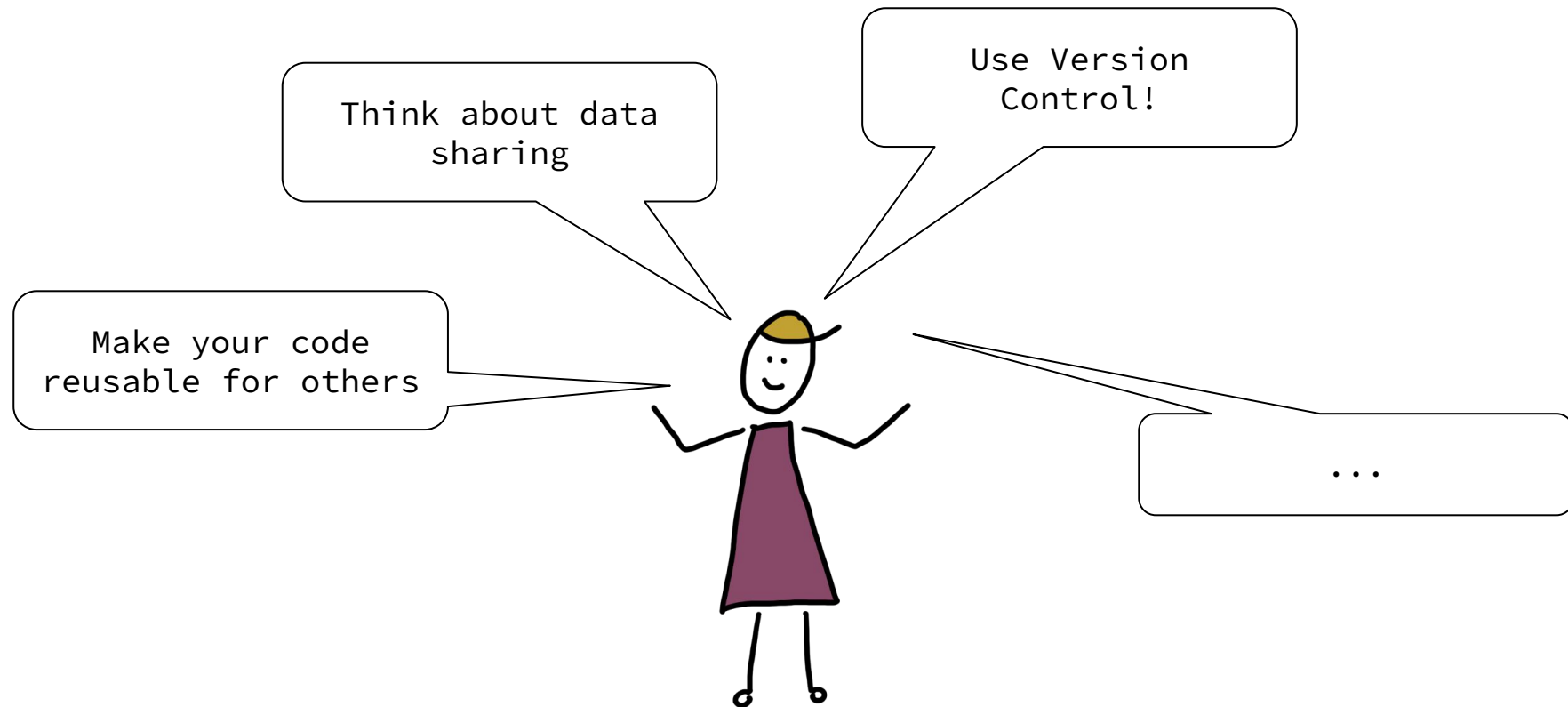
OPEN SCIENCE

as one response to a lack of
quality in research.

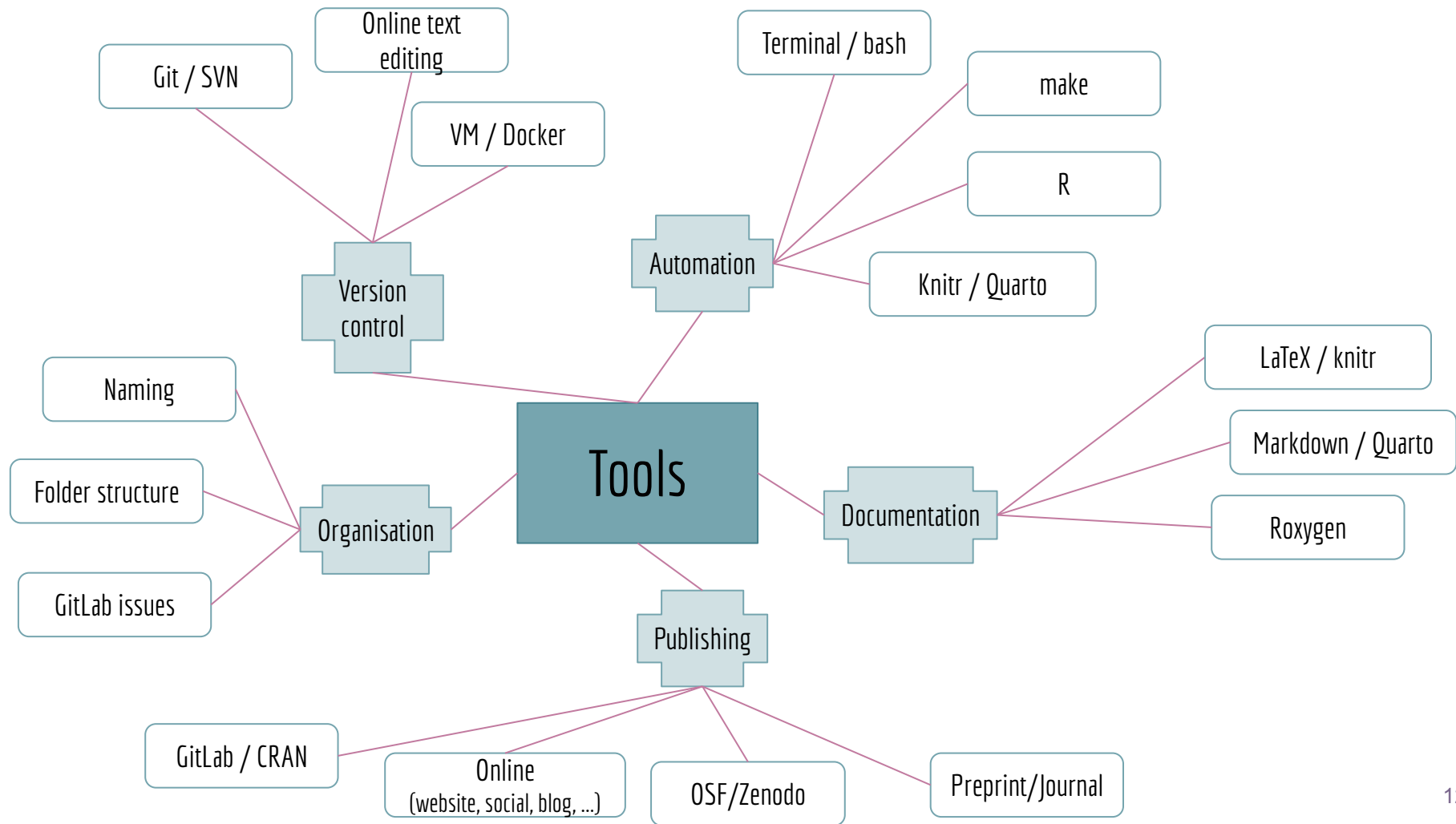


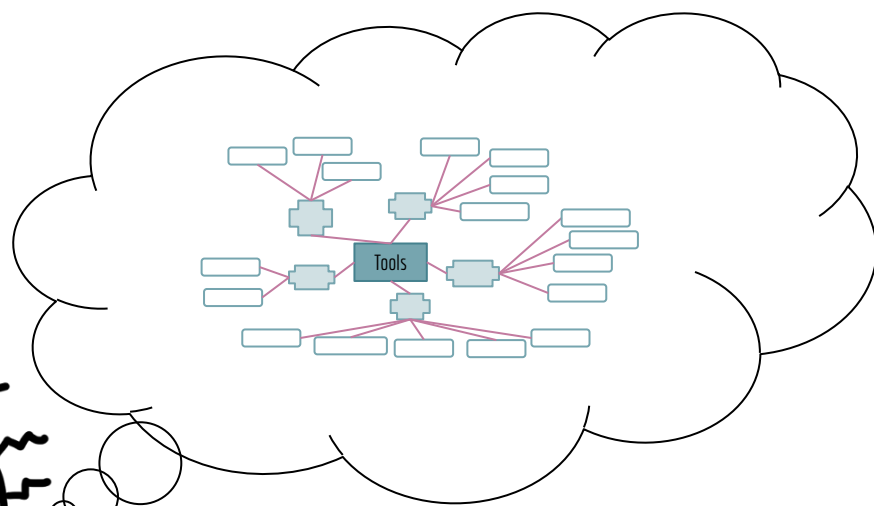
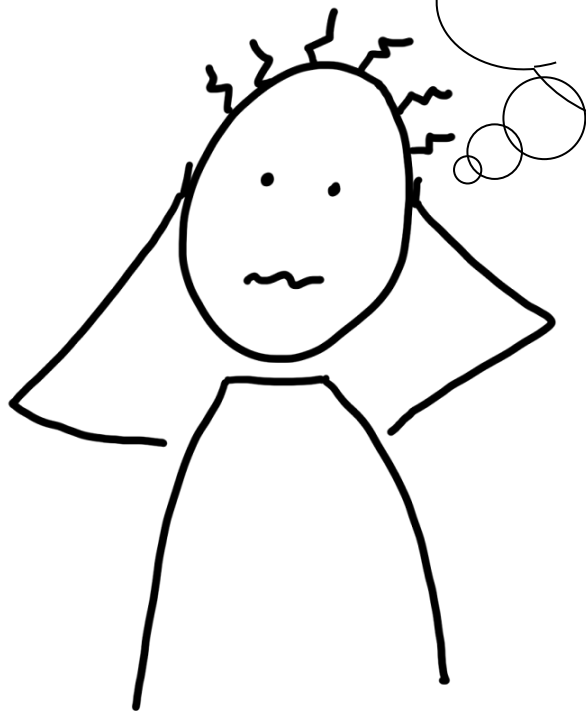
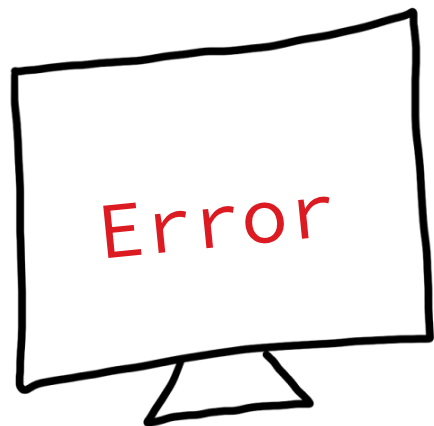
- Collaborative research
- Transparent research process
- Feedback culture
- “Building on the shoulders” of each other
- ...

Me talking about Open Science



And then I show which Tools I use...







Quality
(data-driven) research
requires
Software Skills

PRACTICAL STEPS TOWARDS GOOD SCIENCE/CODE

(a selection)

- **Good organisation**
- **Version control**
- **Code publication**
- Automation
- Environment stabilization
- ...

Get organized!



Good organisation ... starts simple



- Good naming
- Nice file organisation

Good names are important



Photo by [Laura Meinhardt](#)

Naming

NO

- Myabstract.docx
- Joe's Filenames Use Spaces and Punctuation.xlsx
- figure 1.png
- fig 2.png
- JW7d^(2sl@deletethisandyourcareerisoverWx2*.txt

YES

- 2014-06-08_abstract-for-sla.docx
- Joes-filenames-are-getting-better.xlsx
- Fig01_scatterplot-talk-length-vs-interest.png
- Fig02_histogram-talk-attendance.png
- 1986-01-28_raw-data-from-challenger-o-rings.txt

Naming

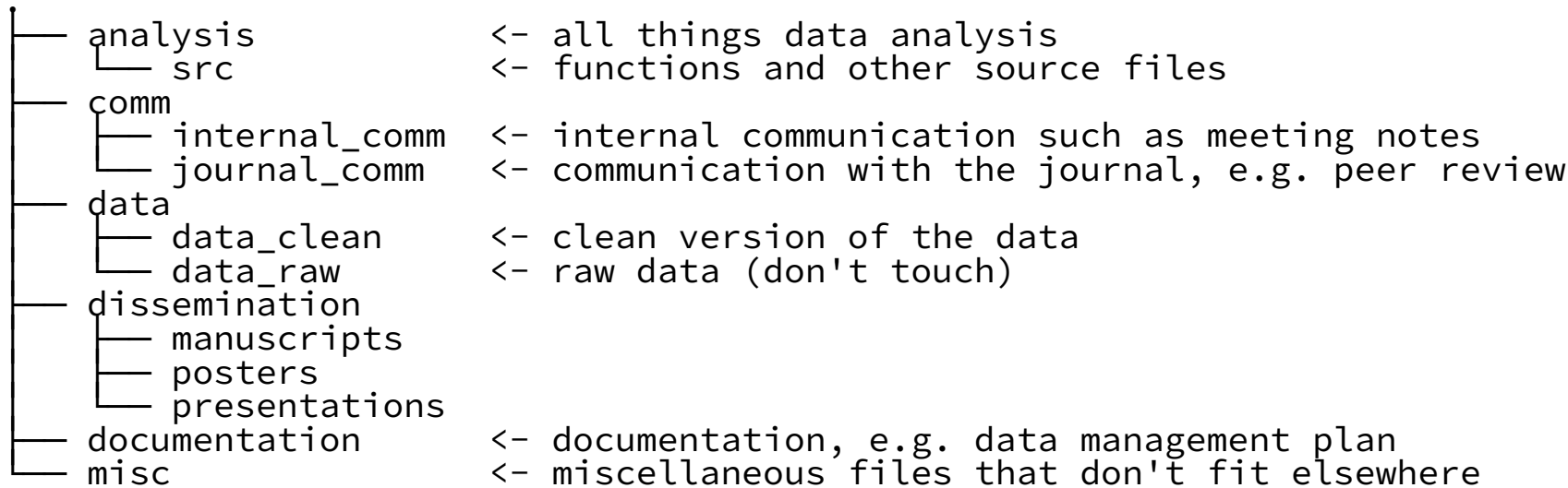
File names should be:

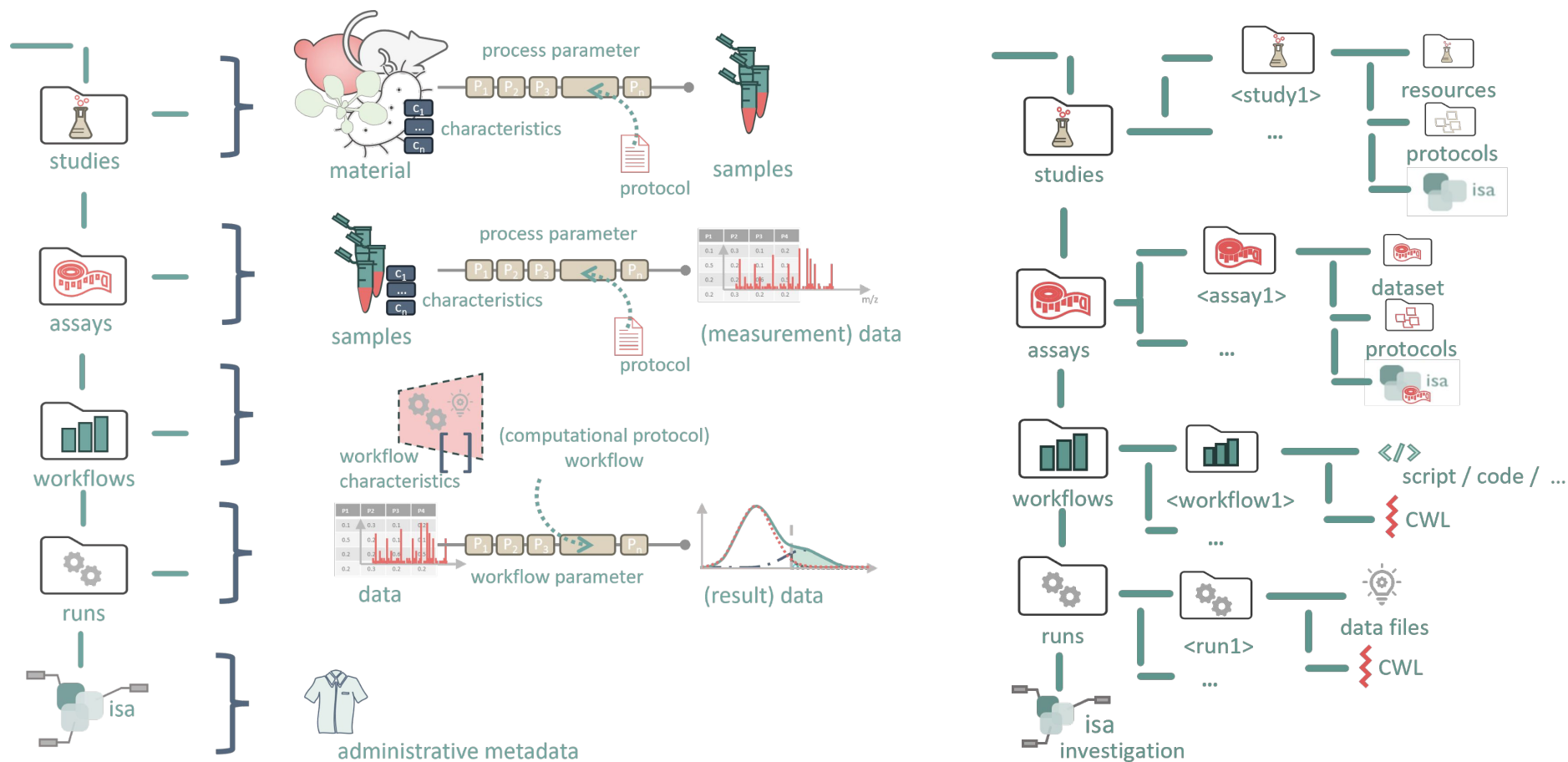
- Machine readable
- Human readable
- Optional: Consistent
- Optional: Play well with default ordering

YES

- 2014-06-08_abstract-for-sla.docx
- Joes-filenames-are-getting-better.xlsx
- Fig01_scatterplot-talk-length-vs-interest.png
- Fig02_histogram-talk-attendance.png
- 1986-01-28_raw-data-from-challenger-o-rings.txt

Organise your files and folders well



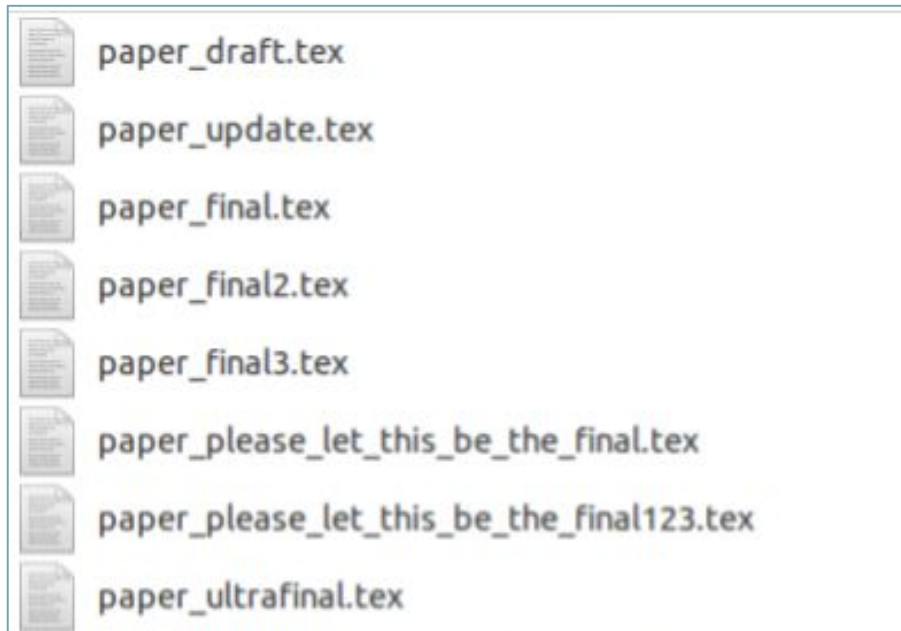


annotated research contexts, see arc-rdm.org/details/organization-principle

Use version control



Version Control in the olden days



Dear colleagues,

attached you find the first public version of the [REDACTED] protocol. Please have a look and do comment. We can also meet to aggregate our reviews.

► 📎 1 attachment: StudyProposa[REDACTED]_Validation_V1_250918docx.docx

Real Version Control

(Including Backup)


Commit **fe6b5538**  authored 10 months ago by  **HeidiSeibold**

[Browse files](#)

[Options](#) ▼

fix typos

 parent [e8a0da8a](#)  master

 No related merge requests found

Showing **1 changed file** ▼ with **3 additions** and **3 deletions**

[Hide whitespace changes](#)

[Inline](#)

[Side-by-side](#)

▼  **2018_swisscore.tex** 



[View file @ fe6b5538](#)

...	...	@@ -56,8 +56,8 @@ In my research I develop algorithms that detect which patient characteristics
56	56	lead to a positive or negative reaction to a therapy. The algorithms can also
57	57	detect and predict which patients are likely to have side effects from a
58	58	therapy. It is important to acknowledge, that \textbf{personalised medicine does not
59		- necessarily mean, that each patient receives a personalised treatment, but it
60		- can also mean that we find that treatments work similarly across patients and
	59	+ necessarily mean, that each patient receives a personalised treatment, but it
	60	+ can also mean that we find that treatments work similarly across patients and

Version Control with Git

*Track **who** made **which** change and **when***

- Track different versions of your code, text, ...
- Description of each version (“commit message”)

Git

push

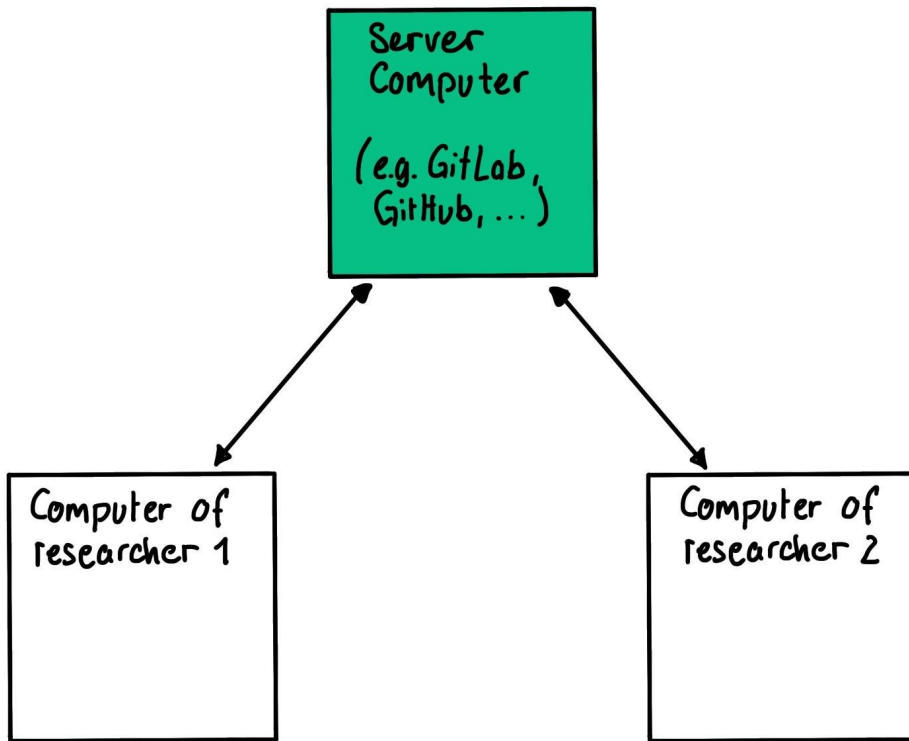


to the server

pull



from the server



Prerequisites for Version Control

- **Text**

- Documents/Papers: LaTeX, Markdown
- Analyses: R, Python

- **Willingness to learn something new: e.g. Git**

Make your work available to others



FAIR

FINDABLE



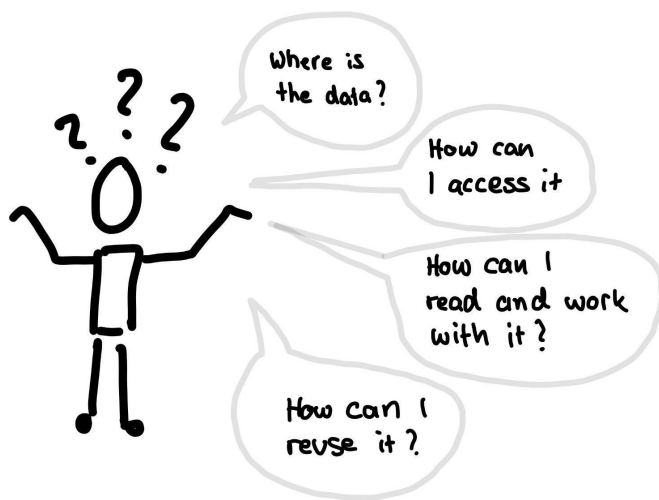
ACCESSIBLE



INTEROPERABLE



REUSABLE

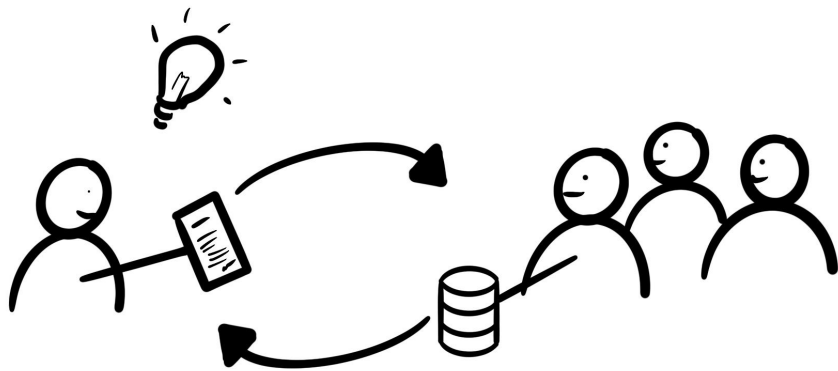
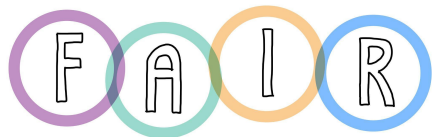


F A I R

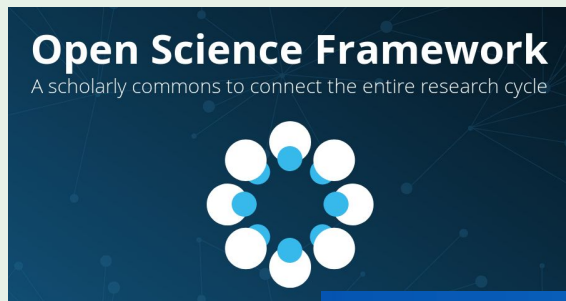
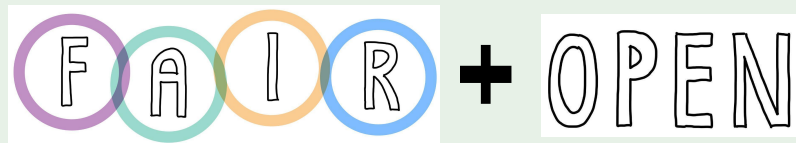
≠

OPEN

SHARING MATERIALS WITH RELEVANT PEOPLE



SHARING MATERIALS ONLINE



SHARING CODE WITH RELEVANT PEOPLE

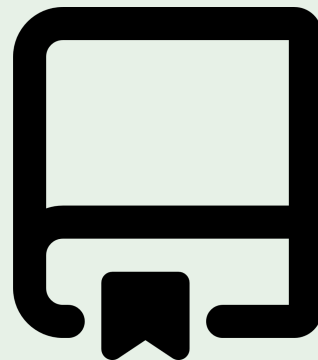
FAIR ?



Private Git repository

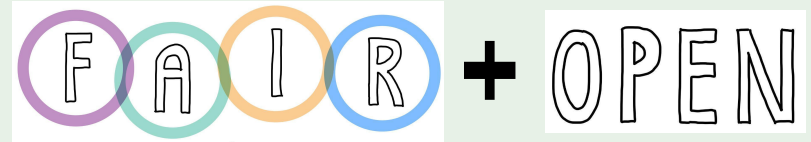
SHARING CODE ONLINE

FAIR + OPEN



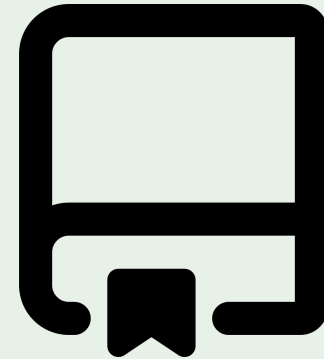
Public Git repository + DOI (Zenodo)

SHARING CODE ONLINE



Try it, it's easier than you'd think!

help.zenodo.org/docs/github



Public Git repository + DOI (Zenodo)

PRACTICAL STEPS TOWARDS GOOD SCIENCE/CODE

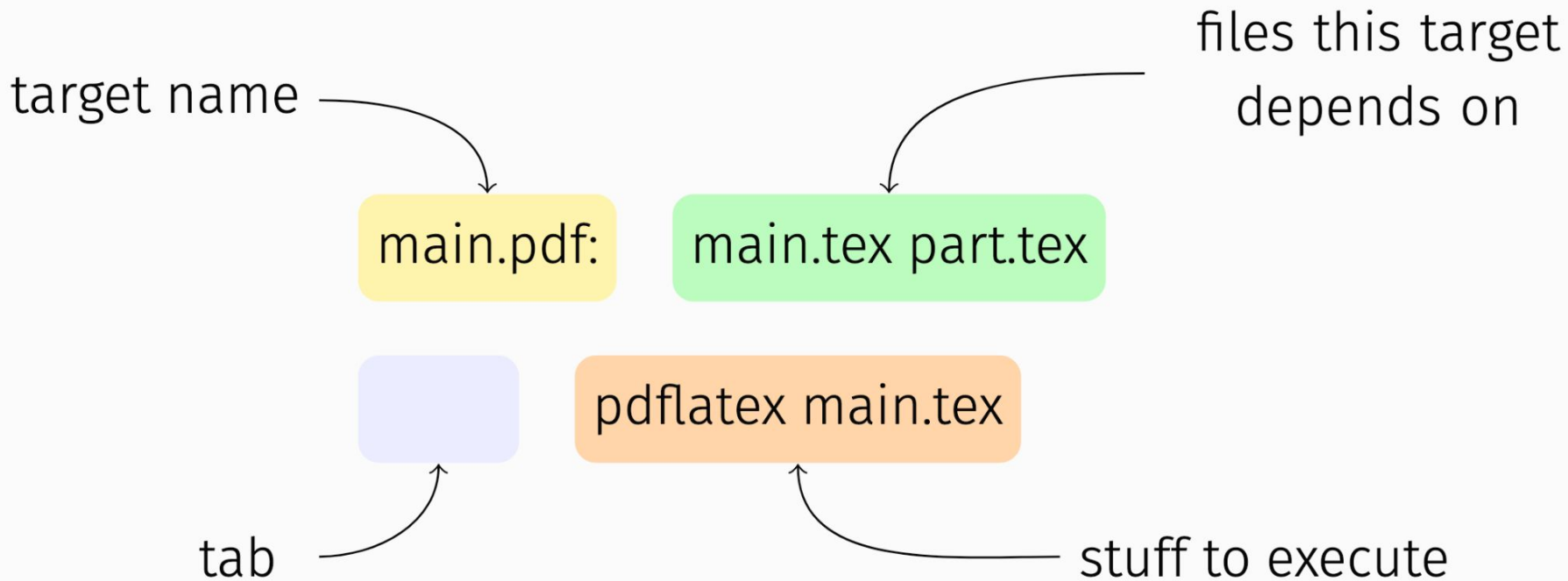
(a selection)

- **Good organisation**
- **Version control**
- **Code publication**
- **Automation**
- **Environment stabilization**
- ...

Automate your work



Make



Make / Makefile

Example: my paper “model4you”

DOI: [10.5334/jors.219](https://doi.org/10.5334/jors.219)

```
man_model4you.tex: man_model4you.Rnw
```

```
Rscript -e "knitr::knit('man_model4you.Rnw')"
```

```
man_model4you.pdf: man_model4you.tex ref_model4you.bib
```

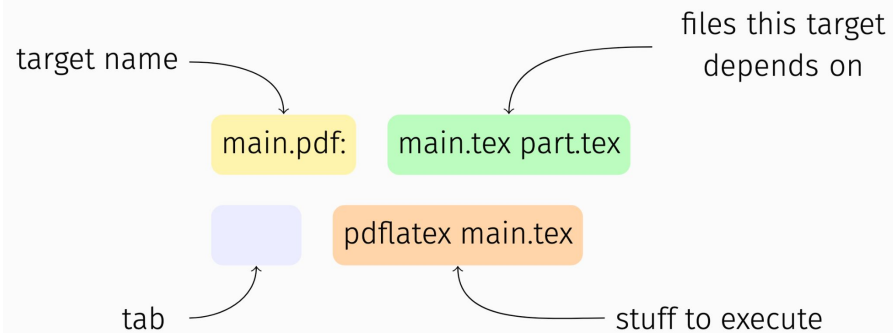
```
pdflatex man_model4you.tex
```

```
bibtex man_model4you
```

```
pdflatex man_model4you.tex
```

```
pdflatex man_model4you.tex
```

```
all: man_model4you.pdf
```



**Stabilize your
computing environment**



Running the same code after a month and getting different results?



Computing Environment?

```
> sessionInfo()
```

```
R version 4.2.2 (2022-10-31)
```

```
Platform: aarch64-apple-darwin21.6.0 (64-bit)
```

```
Running under: macOS Monterey 12.5.1
```

```
Matrix products: default
```

```
LAPACK: /opt/homebrew/Cellar/r/4.2.2/lib/R/lib/libRlapack.dylib
```

```
locale:
```

```
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
attached base packages:
```

```
[1] grid          stats          graphics  grDevices  utils          datasets  methods    base
```

```
other attached packages:
```

```
[1] partykit_1.2-16 mvtnorm_1.1-3  libcoin_1.0-9
```

```
loaded via a namespace (and not attached):
```

```
[1] compiler_4.2.2  Matrix_1.5-1    tools_4.2.2      survival_3.4-0  rpart_4.1.19  
splines_4.2.2
```

```
[7] inum_1.0-4      Formula_1.2-4   lattice_0.20-45
```

Options for stabilizing your computing environment

- 1) Record
- 2) Virtual Machine
- 3) Container
- 4) Other

Record your computing environment

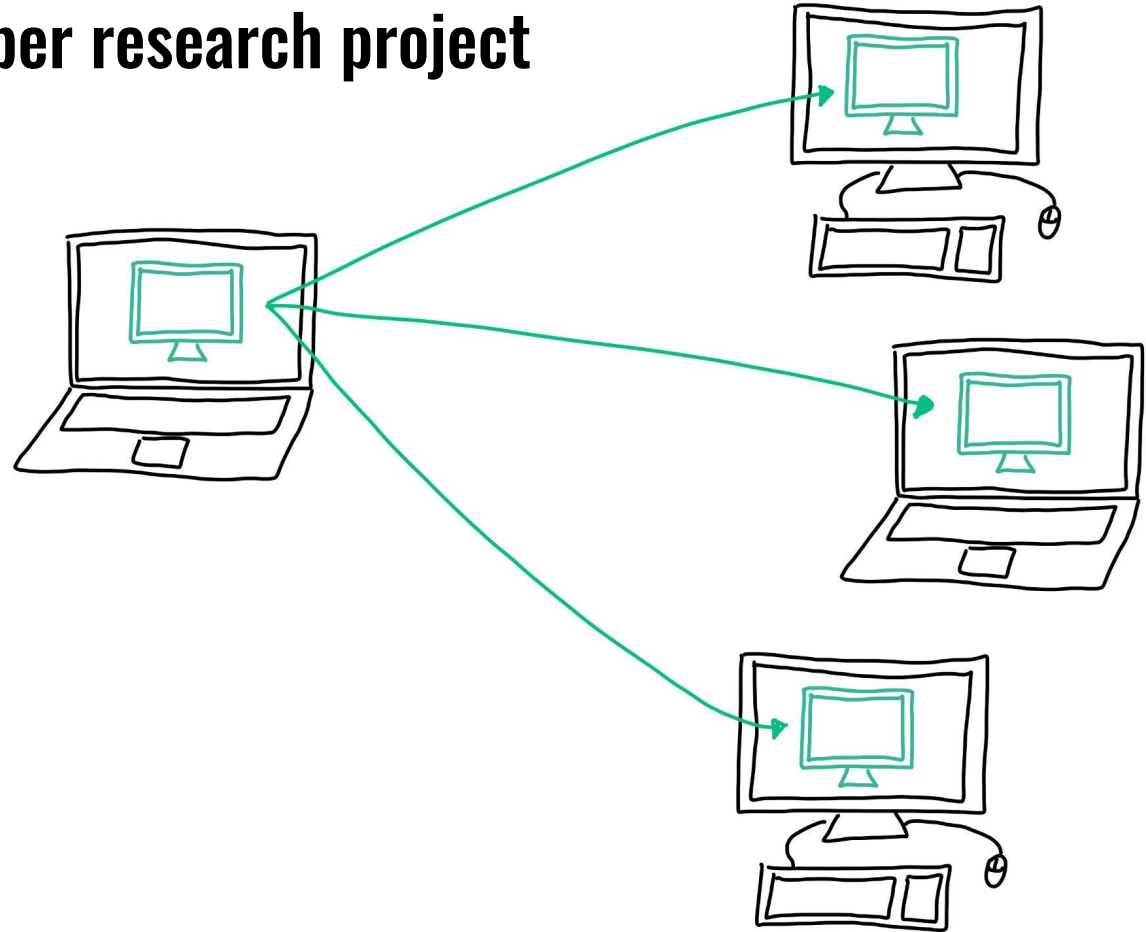
> sessionInfo()   README



Use one virtual machine per research project

**Ask your IT support
to create it for you**

Use one container per research project



Example: Docker

Other

R:

[packrat](#) or [renv](#)



Python:

```
pip install somepackage==1.2.3
```

Conda + mamba

Options for stabilizing your computing environment

- 1) Record
- 2) Virtual Machine
- 3) Container
- 4) Other

WANT MORE?



WANT MORE? → www.osc.lmu.de



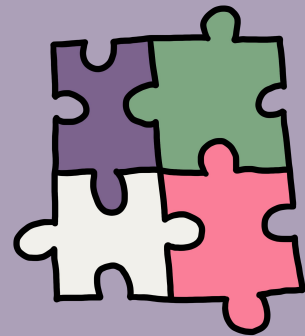


DIGITAL RESEARCH
ACADEMY

Quality Training
for
Quality Research

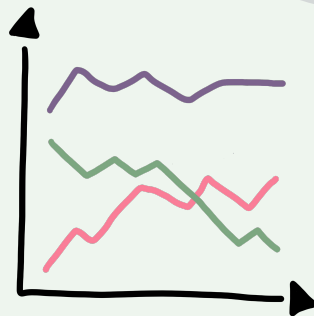
**Improve the quality
of research!**

Open
Science

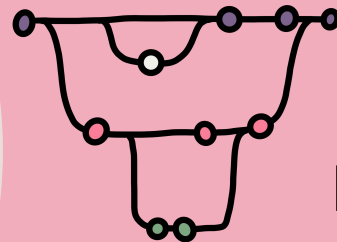


DRA
DIGITAL RESEARCH
ACADEMY

Data
Literacy



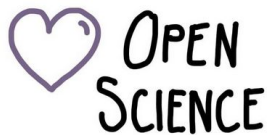
Research
Software
Engineering





Replication Research (R2)

Sent 2 months ago



Open Science Business Models

Sent 3 months ago



Stop celebrating researchers with many publications

Sent 4 months ago



Data sharing (un)culture

Sent 6 months ago



Is Machine Learning and AI solving the problem of p-hacking?

Sent 10 months ago

Check out my newsletter!

heidiseibold.kit.com

