Learn about tools for **reproducibility**

Agata Bochynska, PhD
Digital Scholarship Center
University of Oslo Library

agata.bochynska@ub.uio.no

UNIVERSITY
OF OSLO

Bochynska, A (2026). *Learn about tools for reproducibility*

CC BY

Roadmap

- Some definitions
- Open research and reproducibility
- Reproducible data acquisition, processing, analyses and reports/publications (with some useful tools)
- Take-aways

		DATA	
		Same	Different
ANALYSIS	Same	**Reproduced**	**Replicated**
	Different	**Robust**	**Generalized**

Based on: https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html

Reproduced

results are consistent when following the same method and analysis steps with the **same input data**

https://www.nap.edu/catalog/25303

Quantitative studies

computational reproducibility

Qualitative studies

process transparency

Quantitative studies

computational reproducibility

“obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis”

https://www.nap.edu/catalog/25303

Quantitative studies
computational reproducibility

Re-running analyses/code with the same data

Qualitative studies
process transparency

Arriving at similar (consistent) interpretation by following the same analysis process

Qualitative studies
process transparency

Following the step-by-step reasoning and interpretation process

https://blog.ukdataservice.ac.uk/show-me-the-data/

Reproducibility is strongly associated with **transparency**

“**Open Science** has the potential of making the scientific process more **transparent**, **inclusive** and **democratic**. It is (…) a true game changer in bridging the science, technology and innovation gaps and fulfilling the **human right to science**.”



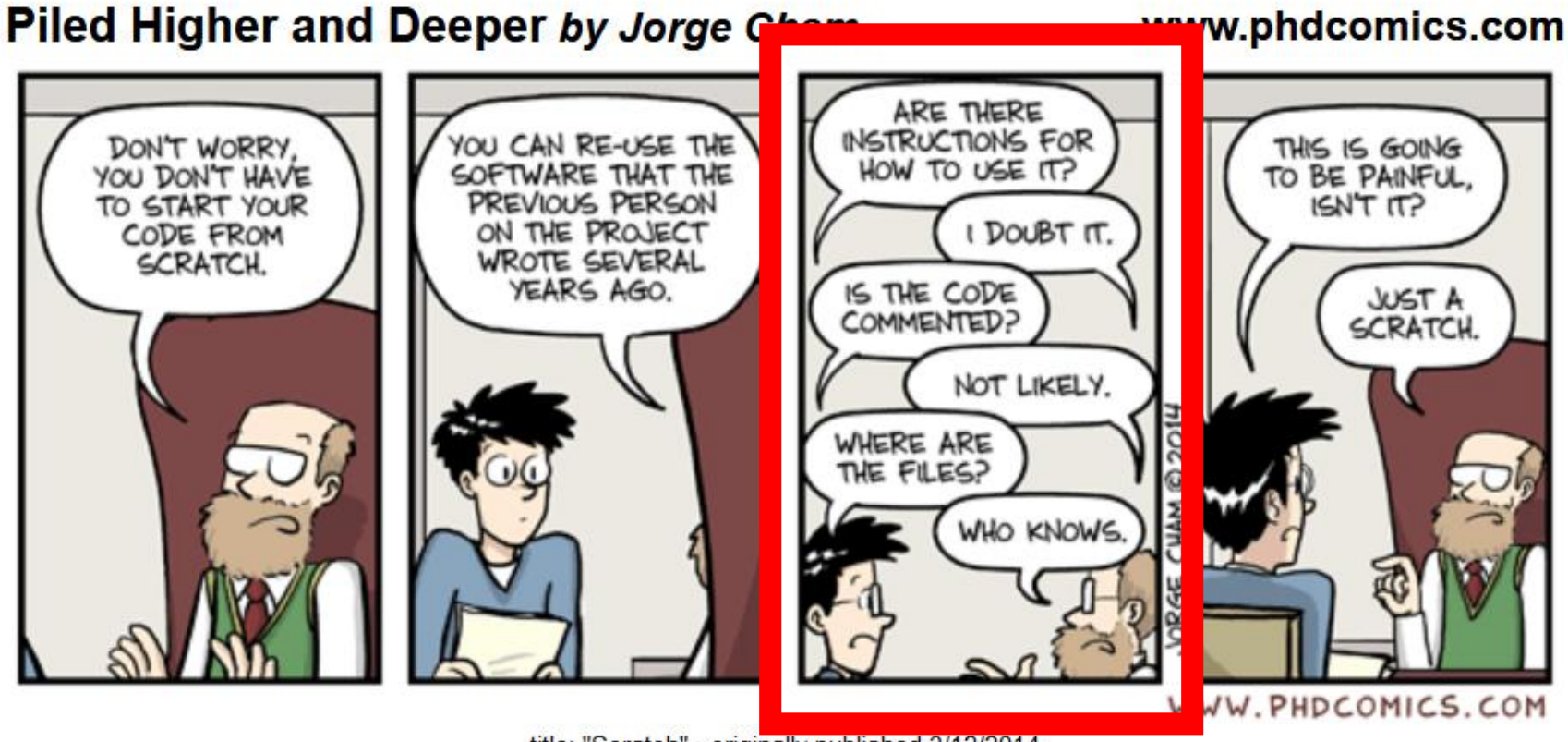
UNESCO Recommendation on Open Science

https://youtu.be/I3Wkvx_ZaFo
https://www.unesco.org/en/natural-sciences/open-science

Reproducible does not (have to) mean fully **open**

As **open** as possible,
as **closed** as necessary

Open but not usable



title: "Scratch" - originally published 3/12/2014

Reasons for irreproducibility:

- Unavailablility of materials, data and/or analyses
- Poor data management
- Unclear analysis specification
- Lack of documentation
- Errors in reporting numbers
- Lack of quality checking procedures
- Insufficient peer review

Reproducible research workflows

Data acquisition and processing
Data analyses
Data reports (manuscripts)

Data acquisition and processing

Data organization
Data documentation
Version control

Organized data

```
project_name/
├── README.md             # overview of the project
├── data/                 # data files used in the project
│   ├── README.md         # describes where data came from
│   └── sub-folder/       # may contain subdirectories
├── processed_data/       # intermediate files from the analysis
├── manuscript/           # manuscript describing the results
├── results/              # results of the analysis (data, tables, figures)
├── src/                  # contains all code in the project
│   ├── LICENSE           # license for your code
│   ├── requirements.txt  # software requirements and dependencies
│   └── ...
└── doc/                  # documentation for your project
    ├── index.rst
    └── ...
```

Research project with a proper data file structure. Image taken from CodeRefinery, Lesson on Reproducible Research. Shared under CC-BY 4.0.

https://coderefinery.github.io/reproducible-research/organizing-projects/

Versioned data

Versioning refers to saving **new copies** of your files when you make changes so that later you can go back and **retrieve** specific **versions** of your files

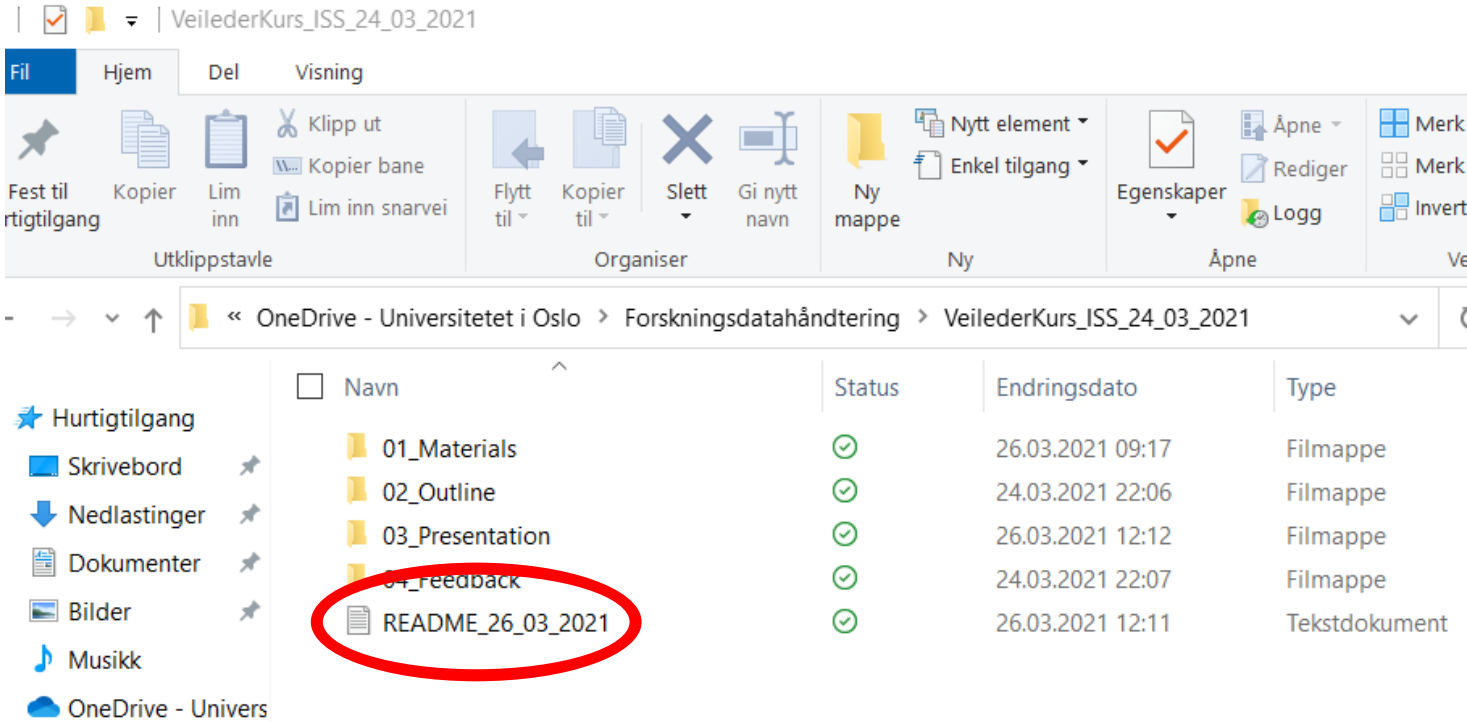
DataFileName**_v1_0** = original document

DataFileName**_v1_1** = original document with minor revisions

DataFileName**_v2_0** = document with substantial revisions

Documented data: README-files

- The first file to open
- Map for navigating and exploring files and their content
- One README.txt file per folder



Documented data: CODEBOOK

- Explains all variables and their codes in the dataset
- **It typically contains:**

❑ variable names, variable labels, variable codes, variable formats, missing data (in quantitative research)

❑ codes, code definitions, examples of what to include with a given code (in qualitative research)

- Can be also called **Data Dictionary**

What additional information do we need here?

	A	B	C	D	E	F	G
1	subjid	sex	gender	recruitment	condition	group	age
2	10001	1		2	9999	9999	272
3	10002	1	1	1	2	2	187
4	10003	1		1	9999	9999	291
5	10004	1	1	1	1	1	300
6	10005	1	1	1	1	1	288
7	10006	1		6	0	9999	254
8	10007	1	1	4	1	2	233
9	10008	0	2	9999	0	0	266
10	10009	1	1	2	1	1	195
11	10010	1	2	2	1	1	207
12	10011	1	1	1	1	2	234
13	10012	1		2	1	2	231
14	10013	1	1	0	1	2	257
15	10014	0	2	0	0	0	309
16	10015	0	2	1	2	2	254
17	10016	0	2	5	0	9999	161
18	10017	1		6	0	9999	
19	10018	1	1	1	1	2	314
20	10019	1	1	5	0	0	188
21	10020	1		0	1	9999	
22	10021	1	1	4	9999	9999	
23	10022	1		1	1	2	294
24	10023	1	1	0	9999	9999	
25	10024	1	1	1	1	1	291
26	10025	1	1	2	1	1	214
27	10026	1	1	1	2	2	332
28	10027	1	1	0	1	2	348
29	10028	1	1	1	1	2	324
30	10029	0	2	9999	0	0	293
31	10030	1		0	1	2	365

Sex: biological sex self-reported by the participant (0 – Female, 1 – Male)

Gender: self-reported gender (0 – Female, 1 – Male, 2 – Other)

Age: age in months at the date of test (automatically calculated in the system as the difference between the date of test and the date of birth)

	A	B	C	D	E	F	G
1	subjid	sex	gender	recruitment	condition	group	age
2	10001	1		2	9999	9999	272
3	10002	1	1	1	2	2	187
4	10003	1		1	9999	9999	291
5	10004	1	1	1	1	1	300
6	10005	1	1	1	1	1	288
7	10006	1		6	0	9999	254
8	10007	1	1	4	1	2	233
9	10008	0	2	9999	0	0	266
10	10009	1	1	2	1	1	195
11	10010	1	2	2	1	1	207
12	10011	1	1	1	1	2	234
13	10012	1		2	1	2	231
14	10013	1	1	0	1	2	257
15	10014	0	2	0	0	0	309
16	10015	0	2	1	2	2	254
17	10016	0	2	5	0	9999	161
18	10017	1		6	0	9999	
19	10018	1	1	1	1	2	314
20	10019	1	1	5	0	0	188
21	10020	1		0	1	9999	
22	10021	1	1	4	9999	9999	
23	10022	1		1	1	2	294
24	10023	1	1	0	9999	9999	
25	10024	1	1	1	1	1	291
26	10025	1	1	2	1	1	214
27	10026	1	1	1	2	2	332
28	10027	1	1	0	1	2	348
29	10028	1	1	1	1	2	324
30	10029	0	2	9999	0	0	293
31	10030	1		0	1	2	365

Tools that help: Templates

- Cornell University template and guide to README.txt-files:

  https://data.research.cornell.edu/content/readme

- README.txt-files from DataverseNO:

  General template

  Example for social sciences

  Example for life sciences

Tools that help: Nettskjema codebook

View | Form builder | **Codebook** | Settings | Collect responses | See results

Codebook

Mapping questions and alternatives to variables is necessary if the results of a survey are going to be processed in an external analysis tool (e.g. SPSS, STATA or R).

Read more about, and get an introduction to the codebook in Nettskjema

The codebook is not activated for this form

Activate codebook

https://www.uio.no/english/services/it/research/sensitive-data/help/create-nettskjema.html
https://www.uio.no/tjenester/it/adm-app/nettskjema/hjelp/kodebok.html

Tools that help: ELN

Electronic Lab Notebooks

help document research, experiments and procedures performed in laboratories



https://www.uio.no/english/services/it/adm-services/elabjournal/

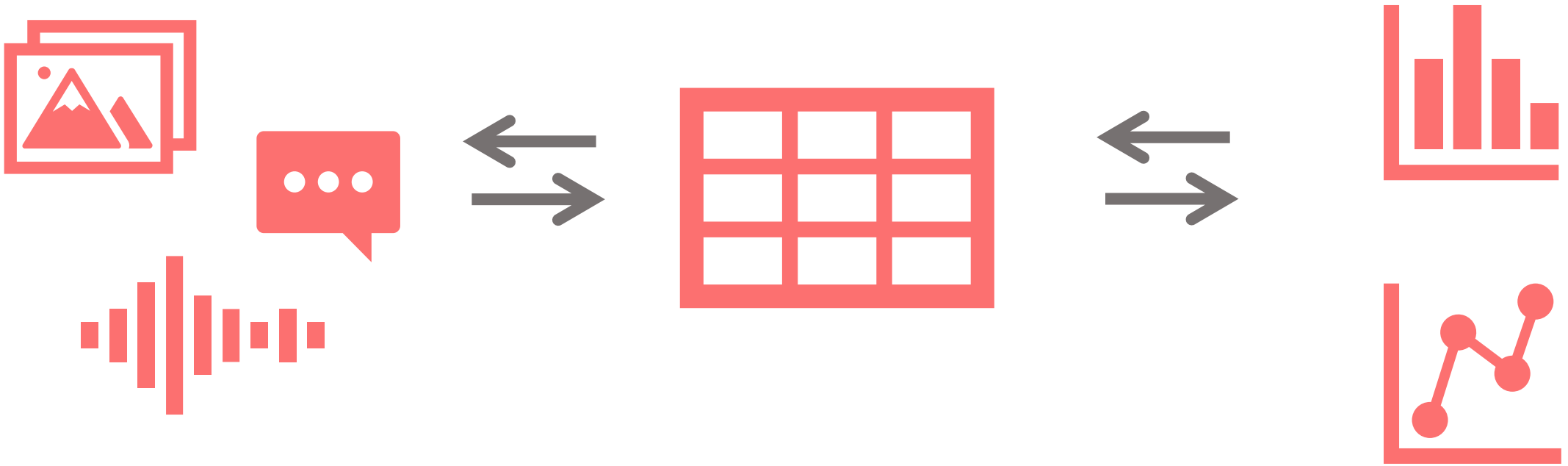
Data analysis

Step-by-step documentation
Version control
Cloud computing and/or containers

Raw data

Processed data

Analyzed data



Qualitative studies
process transparency

Following the step-by-step reasoning and interpretation process

https://blog.ukdataservice.ac.uk/show-me-the-data/

Tools that help: Annotations

Use annotations to comment on selected parts of a source or node

Like scribbled notes in the margin, annotations let you record comments, reminders or observations about specific content in a source or node.

Annotated content is highlighted in blue and the text of the annotation is displayed in the **Annotations** tab at the bottom of the window.

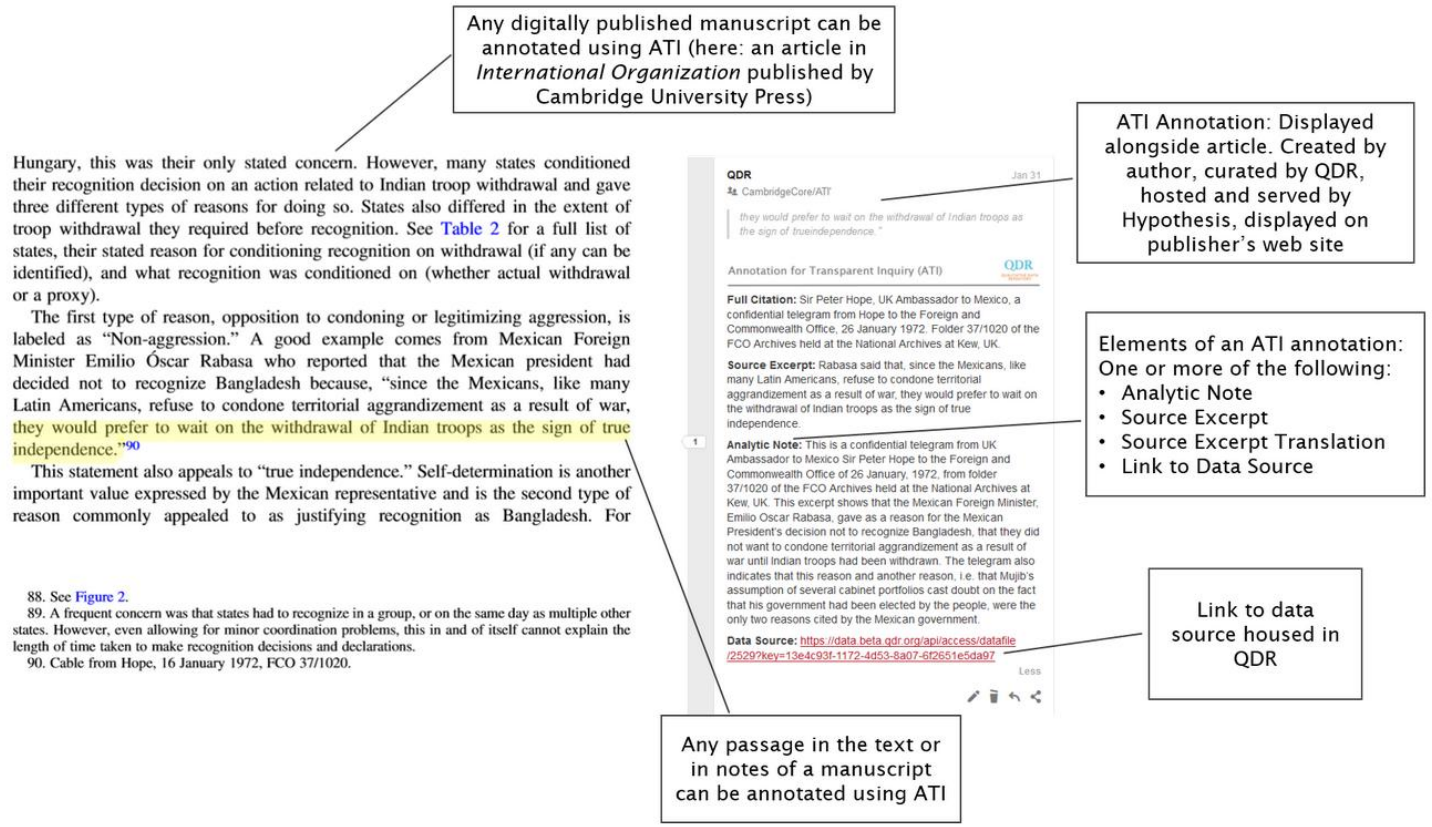


https://help-nv11.qsrinternational.com/desktop/concepts/about_annotations.htm

Annotation for Transparent Inquiry (ATI) at a Glance

- Annotation for Transparent Inquiry (ATI)
  - ATI Models
  - ATI Instructions
  - Why ATI?
  - Empowering Openness in Law-Related Research: A Pilot
- Working with Sensitive Research Data (WSRD)
- Publications

Annotation for Transparent Inquiry (ATI) facilitates transparency in qualitative research by allowing scholars to "annotate" specific passages in an article. Annotations amplify the text and, when possible, include a link to one or more data sources underlying a claim; data sources are housed in a repository.



https://qdr.syr.edu/ati

SYMPOSIUM

Active Citation: A Precondition for Replicable Qualitative Research

Andrew Moravcsik, *Princeton University*

https://www.princeton.edu/~amoravcs/library/ps.pdf

Quantitative studies
computational reproducibility

Re-running analyses/code with the same data

Tools that help: Analysis via code





Tools that help: Version control

- Git: Free and open source version control system



- GitHub: is an internet hosting service for software development and version control using Git



https://git-scm.com/
https://github.com/
https://youtu.be/gY2JwRfin1M

Tools that help: Shared notebooks

jupyter

Try Install Get Involved Documentation News Governance Security About



Free software, open standards, and web services for interactive computing across all programming languages

https://jupyter.org/

Tools that help: Quality check

Language	Static code analysis tool
Python	Pylint, prospector
R	lintr
Shell/Bash	shellcheck

Language	Formatter Tool
Python	Black, yapf
R	formatR
Shell/Bash	ShellIndent
HTML	Tidy

https://the-turing-way.netlify.app/reproducible-research/code-quality.html

Tools that help: Code review



Catching bugs. *The Turing Way* project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807.

https://the-turing-way.netlify.app/reproducible-research/testing.html
https://the-turing-way.netlify.app/reproducible-research/reviewing.html

Tools that help: Checklists

The **TADA**-Wiki: Best practice for sharing analytical code
https://eivimeycook.github.io/TADA-Wiki

T – Transferable
A – Available
D – Documented
A – Annotated



Tools that help: Checklists

Code (TADA) | Data

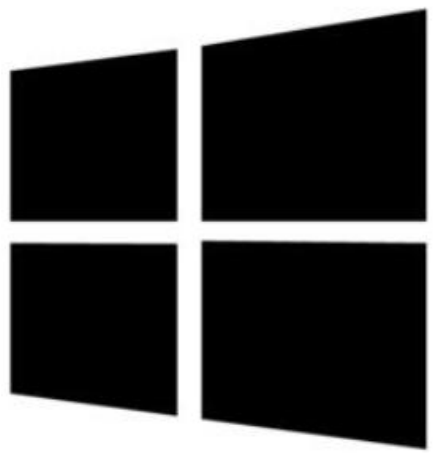
Condensed from the TADA guidelines.

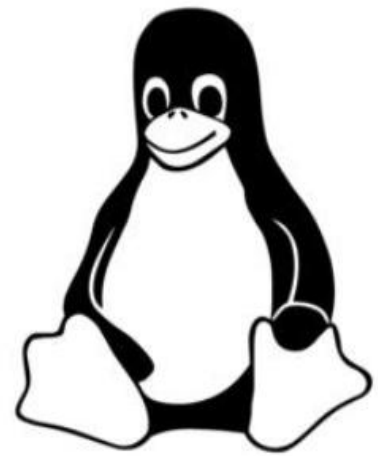
Transferable — full chapter

- [ ] Every script is saved with a standard code extension (`.R`, `.py`, `.cpp`, ...), not inside a `.docx` or `.pdf`
- [ ] No `setwd()` (R) or `os.chdir()` (Python) appears anywhere in the code
- [ ] No absolute paths (`/Users/you/…`, `C:/Users/you/…`) remain; paths are relative (written from the project folder, e.g. `data/raw/file.csv`)
- [ ] Data, code and materials all live in **one** project directory
- [ ] The project has been opened from a clean start (restart R or Python with nothing already loaded) and still runs



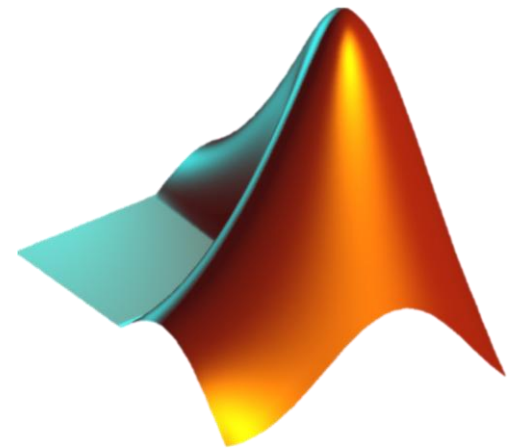
However…







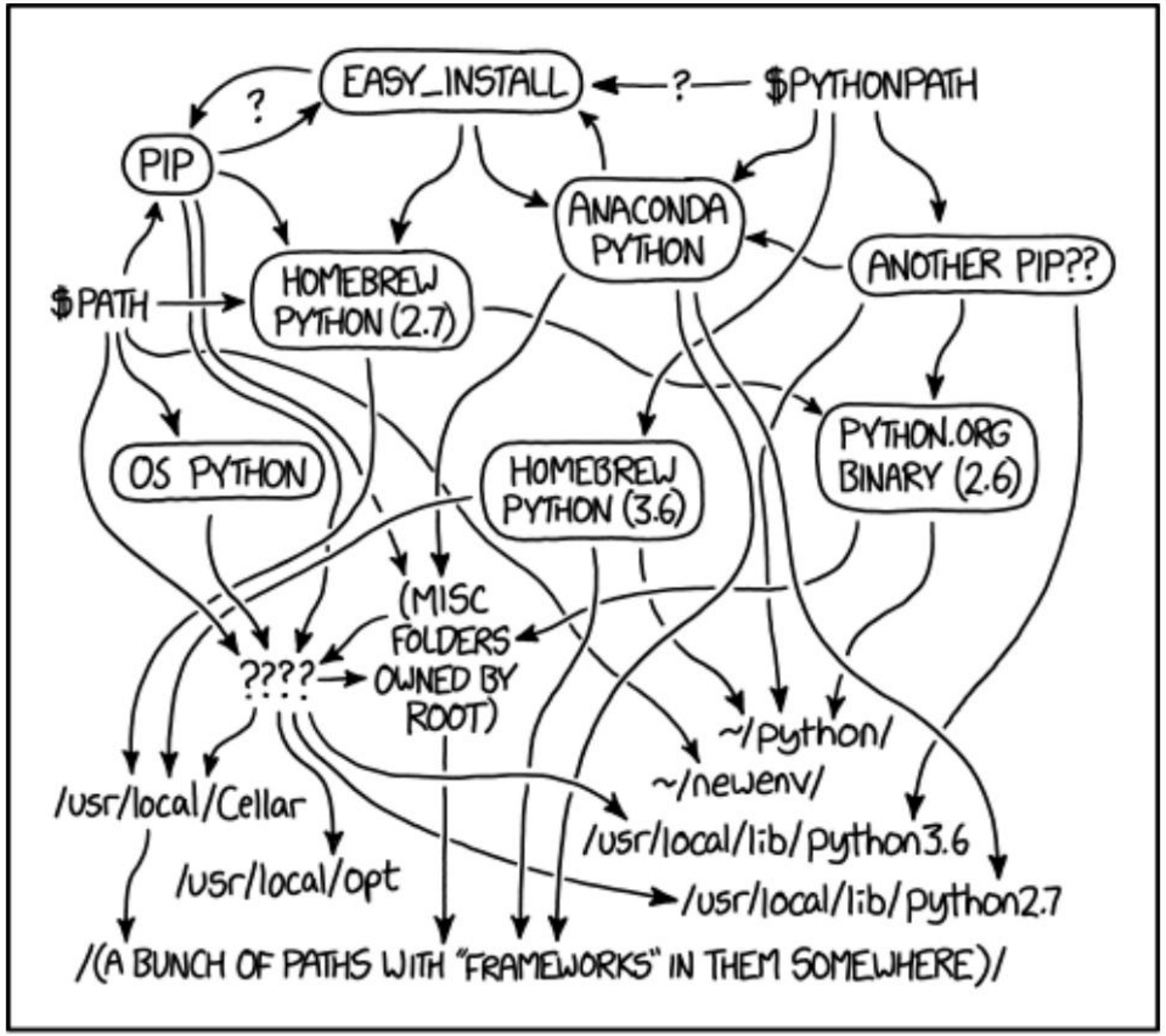








https://jiayiwu.me/blog/2021/05/13/how-to-make-a-r-package-in-rstudio.html
https://nazca-design.org/wp-content/uploads/2017/04/python-modules.png



MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

https://xkcd.com/

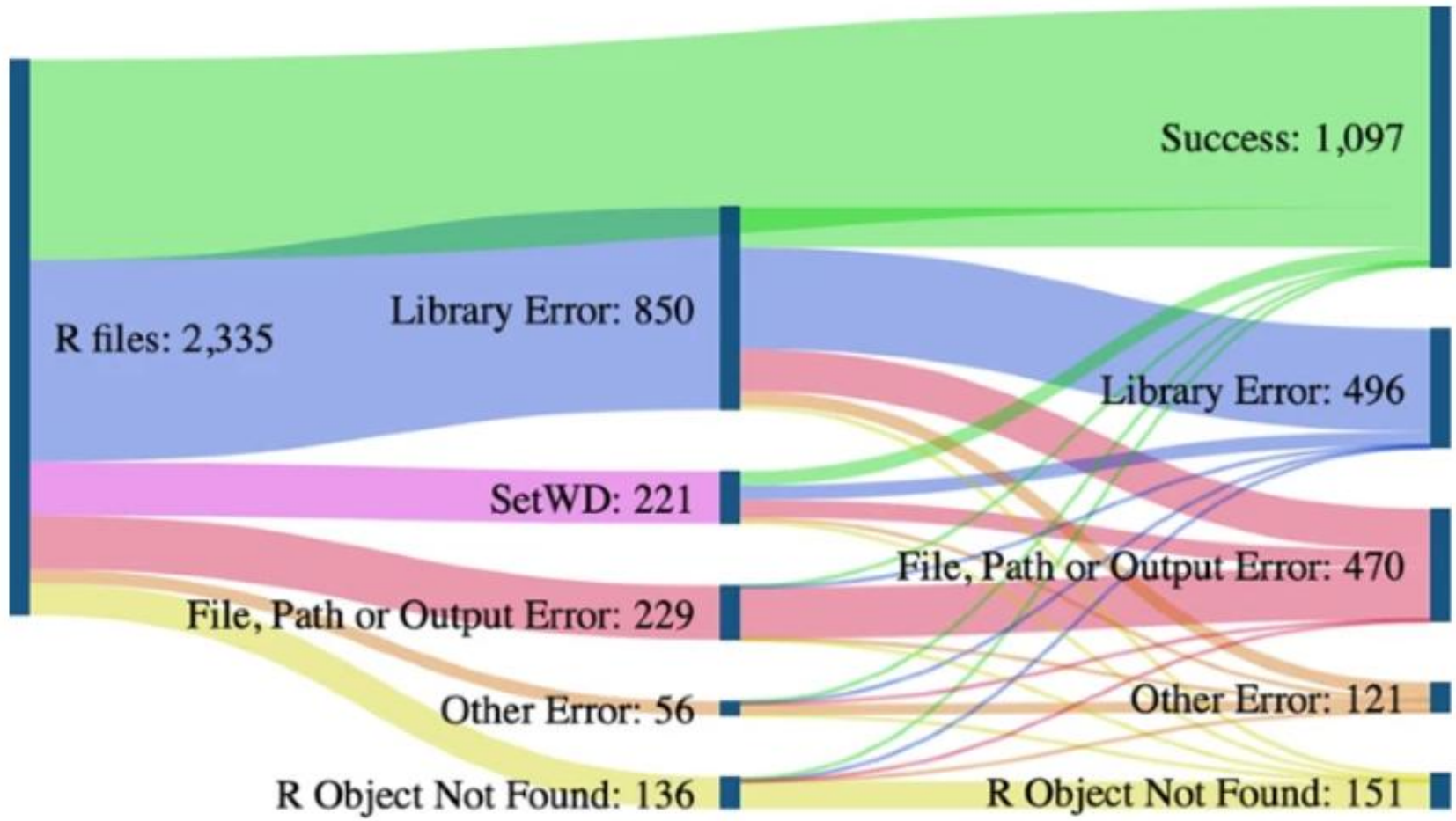
Analysis | Open Access | Published: 21 February 2022

A large-scale study on research code quality and execution

Ana Trisovic, Matthew K. Lau, Thomas Pasquier & Mercè Crosas

Scientific Data **9**, Article number: 60 (2022)

19k Accesses | **7** Citations | **399** Altmetric



https://www.nature.com/articles/s41597-022-01143-6

Tools that help: Containers

What is reproduced? \ Interaction style	Graphical	Command line
Software & versions		
Entire system		

https://the-turing-way.netlify.app/reproducible-research/renv/renv-options.html

Data reports (manuscripts)

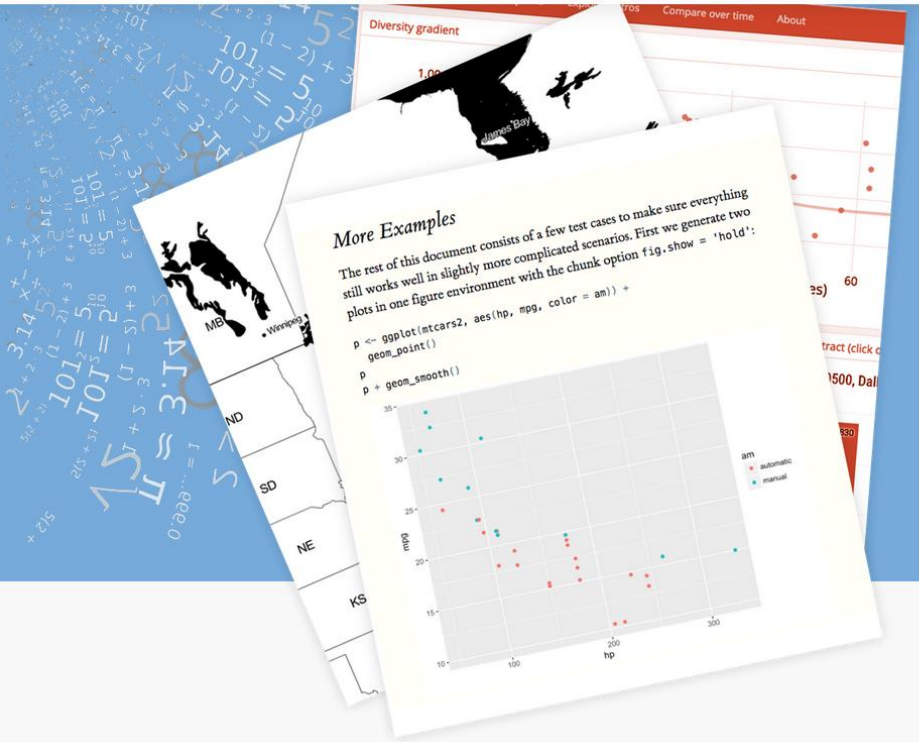
Linked tables and analyses
Version control
Collaboration

Tools that help: R Markdown

R Markdown

from RStudio

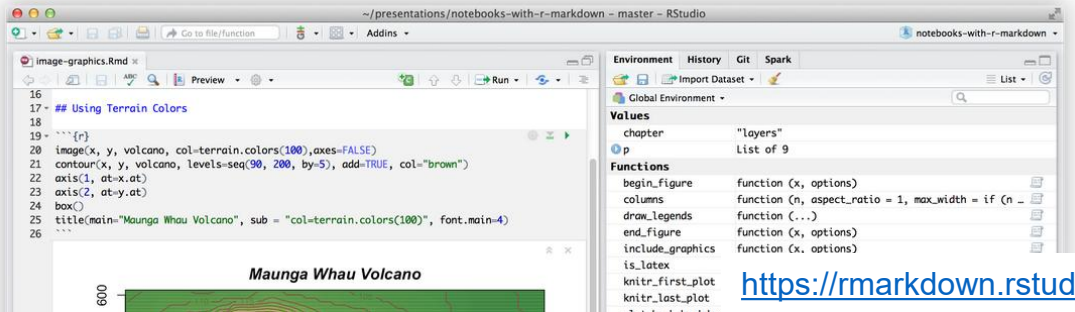
Get Started Gallery Formats Articles Book References



Analyze. Share. Reproduce.

Your data tells a story. Tell it with R Markdown. Turn your analyses into high quality documents, reports, presentations and dashboards.

R Markdown documents are fully reproducible. Use a productive notebook interface to weave together narrative text and code to produce



https://rmarkdown.rstudio.com/

Tools that help: Quarto



https://quarto.org/

Other tools

- Overleaf (collaborative LaTeX editor)
- HackMD (a realtime web-based collaborative Markdown editor)
- Manuscripts.io (a collaborative authoring tool that support scientific content and reproducibility)
- Rrtools (instructions, templates, and functions for making a basic compendium suitable for writing a reproducible journal article or report with R)
- Jupyter Notebooks (can be used for supplementary material with journal articles.

https://coderefinery.github.io/reproducible-research/organizing-projects/

Reproducible research workflows

Data acquisition and processing
Data analyses
Data reports (manuscripts)

UNIVERSITY
OF OSLO

Digital Scholarship Centre

ReproducibiliTea

Hands-on ReproHack

Learn how to reproduce research!



November 19, 13:15 - 16:00
Blindern, Georg Sverdrups hus
In-person only

<<< **INFO & REGISTRATION**

More resources:

The Turing Way: Guide for Reproducible Research
https://the-turing-way.netlify.app/reproducible-research/reproducible-research.html

CodeRefinery: Reproducible Research
https://coderefinery.github.io/reproducible-research/motivation/



UNIVERSITETET
I OSLO

ReproducibiliTea

Journal Club

JOIN IN AND DISCUSS WITH FELLOW STUDENTS AND RESEARCHERS

OPEN RESEARCH, REPRODUCIBILITY and RESEARCH IMPROVEMENT



Join us

Everyone is welcome to join us - whether you are an enthusiast of open and reproducible research, a skeptic, or a cautious explorer. Currently, all meetings are hybrid with the possibility of joining on-site at Blindern or via Zoom. Grab a cup of tea (coffee?) and join us!

Subscribe to our mailing list →

https://www.ub.uio.no/english/libraries/dsc/open-repro-research/reproducibilitea/



Welcome to Norway's Reproducibility Network!

Norwegian Reproducibility Network (NORRN) is a peer-led consortium located in Norway. It follows an organisational format adopted internationally as nationwide "Reproducibility Networks". NORRN collaborates with the other Reproducibility Networks whilst remaining a unique and independent community.

https://www.norrn.no/

UNIVERSITY
OF OSLO

DSC

Digital Scholarship Centre

at the University of Oslo Library

Courses, guidance, networks, and resources

- Research Data Management
- Digital tools and methods
- Science visualization
- Open research



Det senteret for digitalforskerstøttes nyhetsbrev,
en del av Universitetsbiblioteket i Oslo

The Digital Scholarship Centre's Newsletter,
part of the University of Oslo Library

DSC NEWS

Senter for digitalforskerstøtte
Digital Scholarship Centre

https://sympa.uio.no/ub.uio.no/subscribe/dsc-news/subscribe

Thank you!

- Be **transparent** about the full research workflow: questions, methods, data, step-by-step procedures and analyses
- Make sure you have good **documentation** for all outputs and all stages of your research process
- Keep track of **versions** and do a solid **quality check** of your methods, data and analyses
- **Verify** your own work: try to reproduce your own results and/or have others do it
- Make your methods, data and/or analyses **open** (if you can)

Agata Bochynska, PhD
Digital Scholarship Center
University of Oslo Library

agata.bochynska@ub.uio.no

UNIVERSITY
OF OSLO

CC BY