

# Version Control of Code and Data

## A full-semester course about Git for scientists

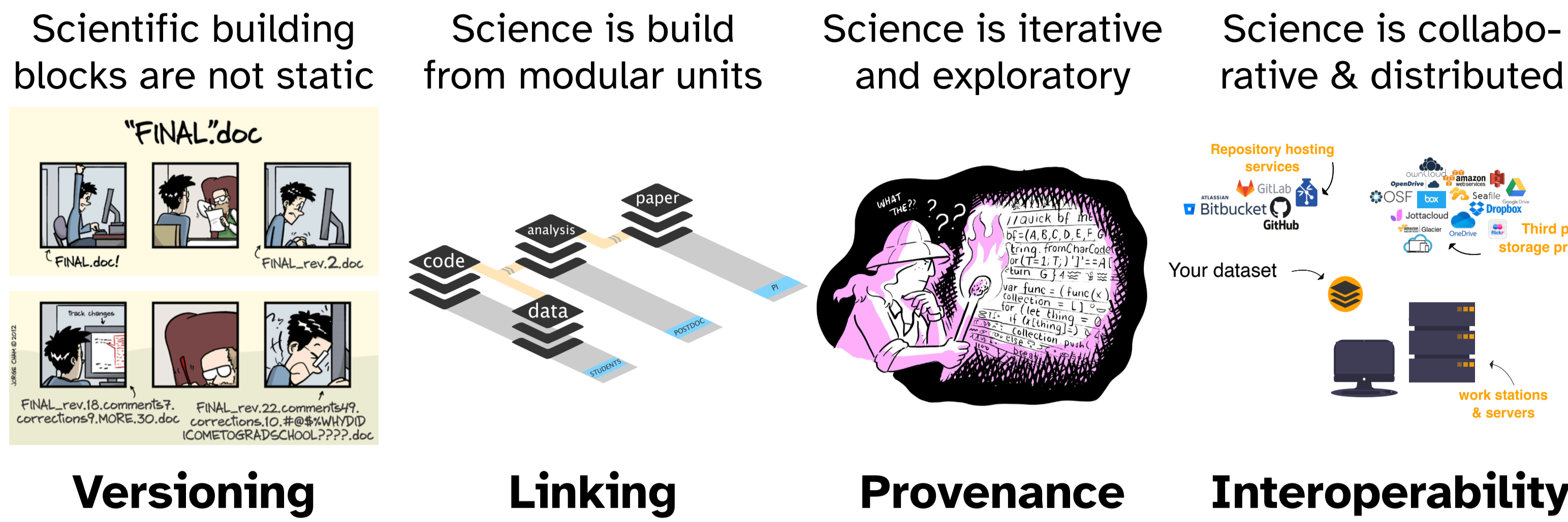
Funding & support by:  
 Universität Hamburg  
 DER FORSCHUNG | DER LEHRE | DER BILDUNG

Lennart Wittkuhn<sup>1</sup> & Konrad Pagenstedt<sup>1</sup>  
<sup>1</sup> Institute of Psychology, University of Hamburg, Germany

DIGITAL AND DATA LITERACY  
 IN TEACHING LAB

Stiftung  
 Innovation in der  
 Hochschullehre

### Background: Science is complex



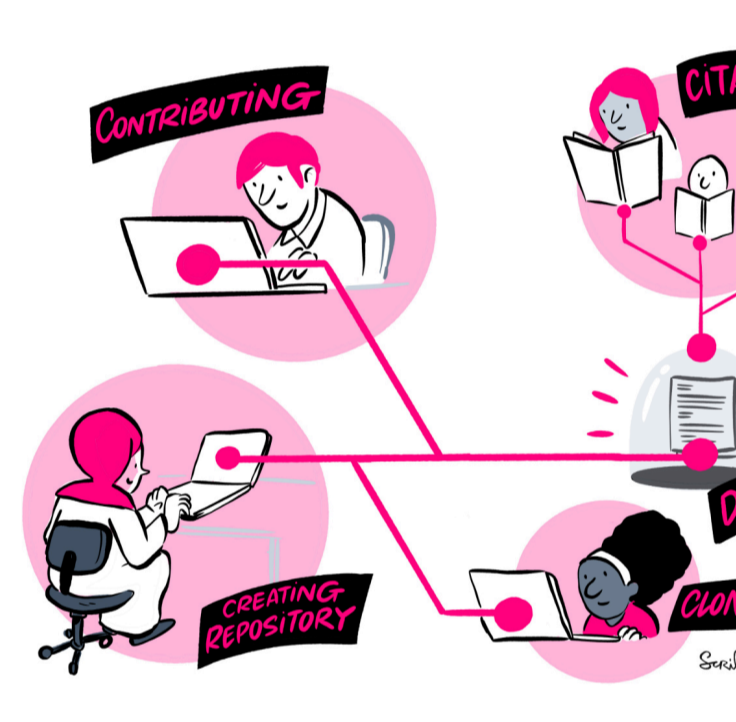
"If everything is relevant, track everything"

### "Science as open-source knowledge development"

- Version Control**
- keep track of changes in a directory (a "repository")
  - take snapshots ("commits") of your repo at any time
  - know the history: what was changed when by whom
  - compare commits or go back to any previous state
  - work on parallel "branches" and flexibly merge them



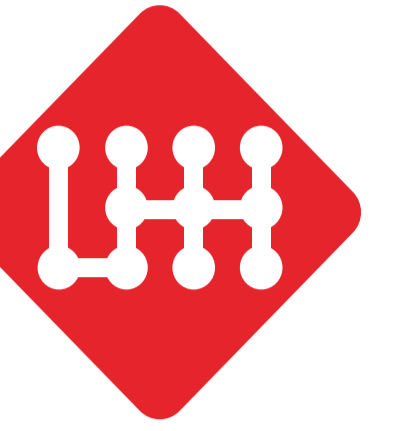
- Collaboration**
- "push" your repo to a "remote" location and share
  - share repos on platforms like GitHub or GitLab
  - work together on the same files at the same time
  - others can read, copy, edit and suggest changes
  - discuss open issues and manage your projects



Web: [lennartwittkuhn.com/dditlab-presentation](http://lennartwittkuhn.com/dditlab-presentation) License: CC BY-SA 4.0 (unless stated otherwise)  
 Contact: [lennart.wittkuhn@uni-hamburg.de](mailto:lennart.wittkuhn@uni-hamburg.de) DOI: 10.25592/uhfdm.14162

### Course v1.0 details

**Summary:** A hands-on course about version control using Git with online materials, exercises & quizzes

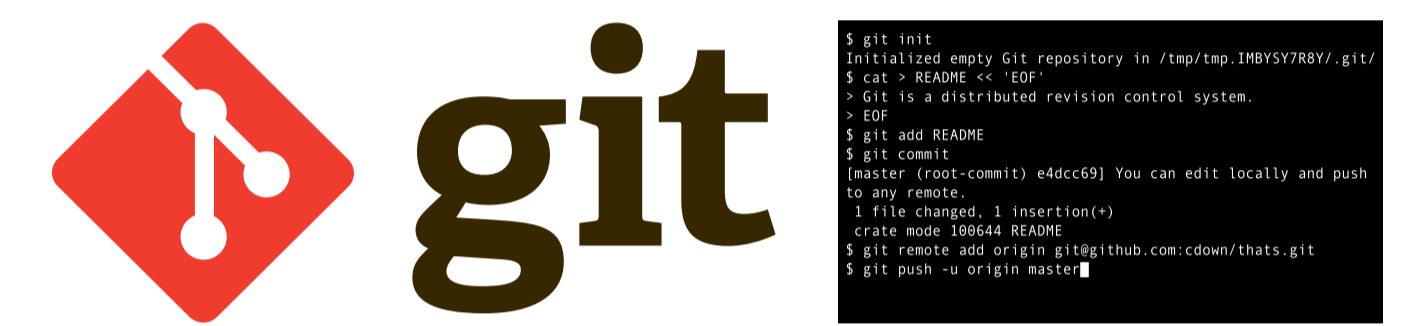


- Content Review** (~ 30 min):  
 Course participants study online teaching materials
- Exercises** (~ 60 min):  
 Course participants delve into hands-on exercises
- Quizzes** (~ 15 min):  
 Supporting quiz questions

- Winter semester 2023/24
- 14 sessions, 90 min each
- M.Sc. & Ph.D. Psychology
- Elective Module (4 ECTS)
- English Web Analytics

### Course Contents

1. Why Version Control?
2. Command Line
3. Installation & Setup
4. Basic Git Commands
5. Quarto Workshop
6. Branching & Merging
7. Basics of GitHub
8. Collaboration in Git
9. Project Management
10. Intro to DataLad



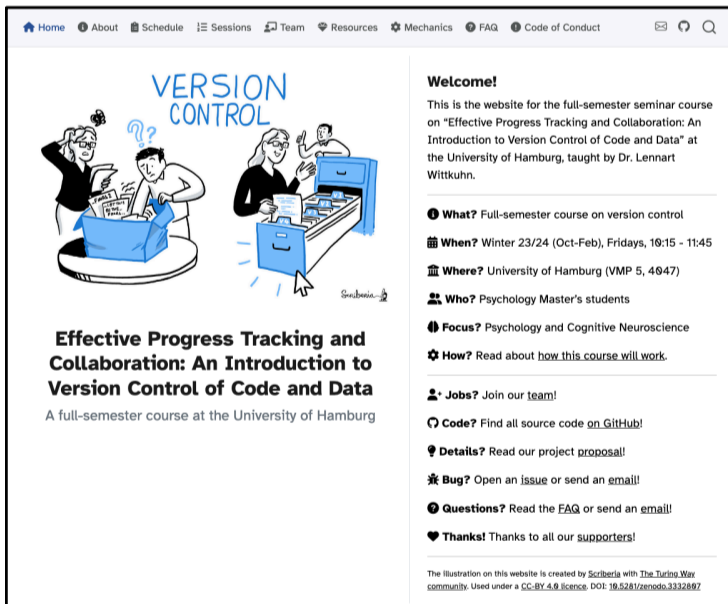
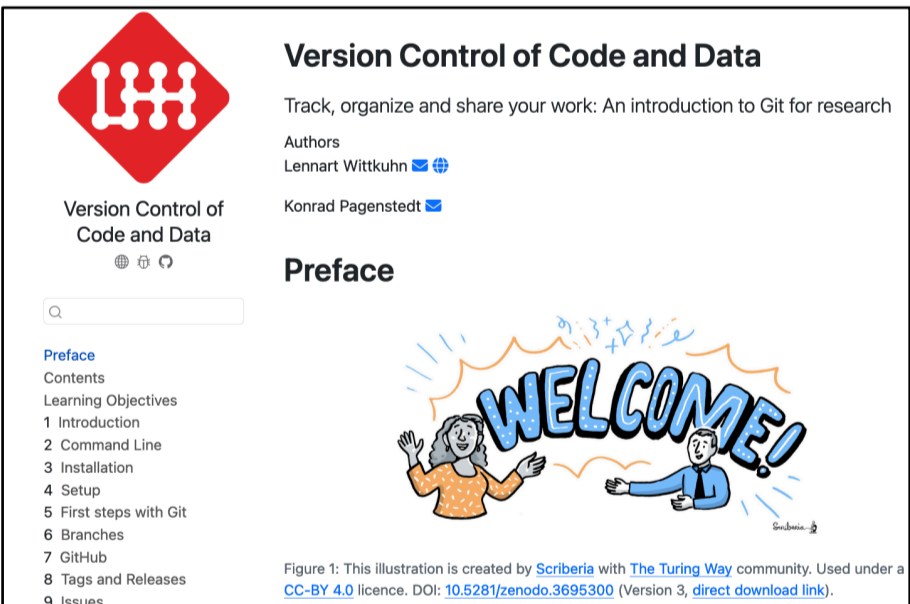
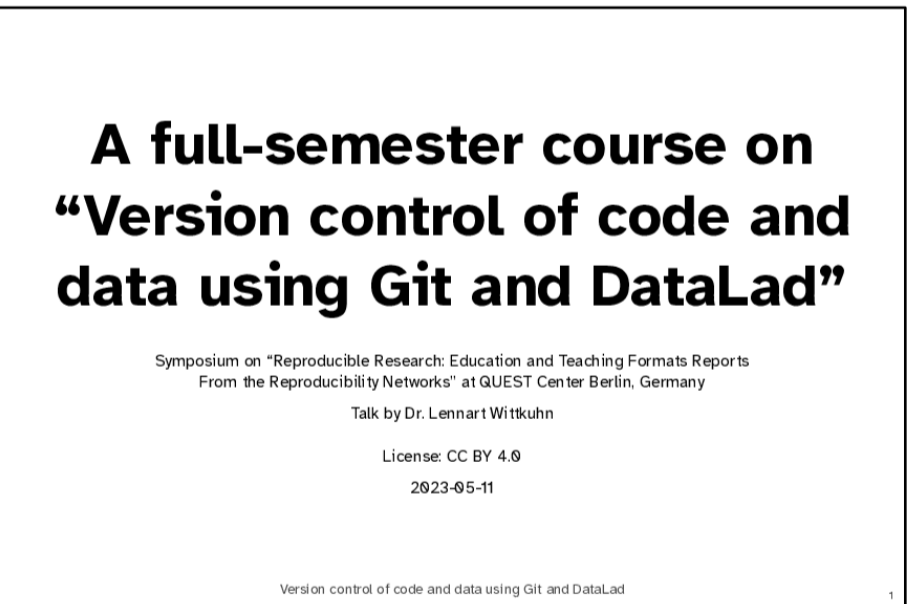
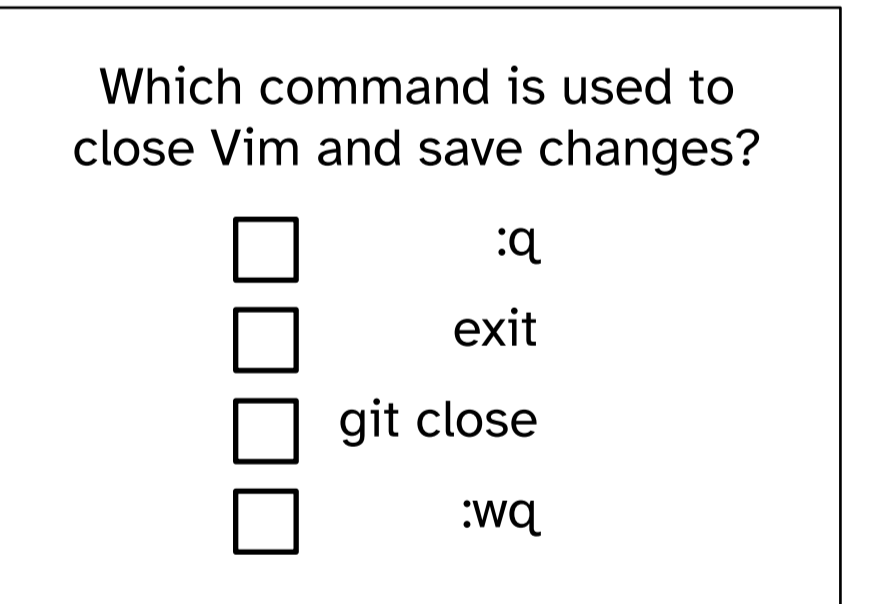
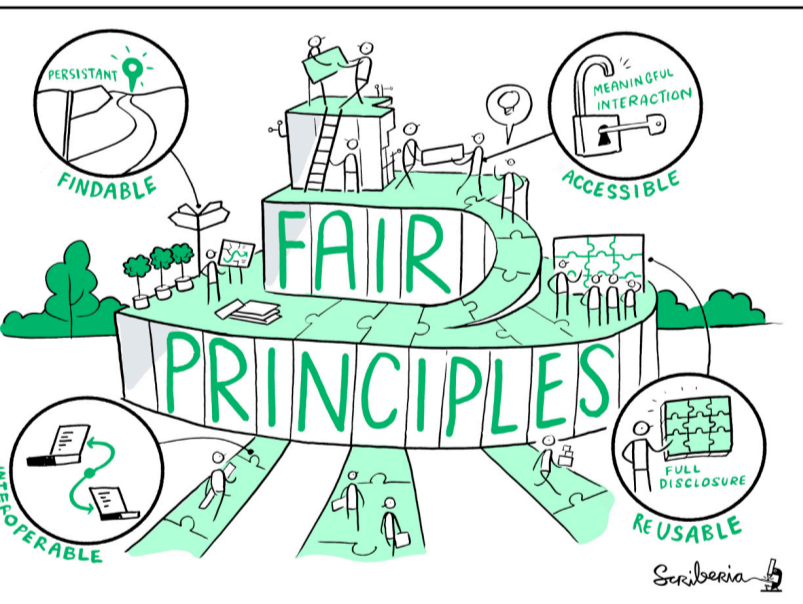
- popular version control system
- free & open-source CLI tool
- software industry standard
- 100 million GitHub users



- "Git for large datasets"
- free & open-source CLI tool
- builds on Git + git-annex

### Open Educational Resources & Project Documentation

Web: [lennartwittkuhn.com/{path}](http://lennartwittkuhn.com/{path}) Code: [github.com/lnnrtwttkhn/{path}](https://github.com/lnnrtwttkhn/{path}) License: CC BY-SA 4.0

Course Website	Online Course Book	Presentations	Quizzes	FAIR Teaching
 Effective Progress Tracking and Collaboration: An Introduction to Version Control of Code and Data	 Version Control of Code and Data	 A full-semester course on "Version control of code and data using Git and DataLad"	 Which command is used to close Vim and save changes?	 FAIR PRINCIPLES
{domain}/version-control-course-uhh-ws23/ Slides & Schedule Web: 2.5K views	{domain}/version-control-book PDF: ~ 180 pages Web: 1.1K views	{domain}/ddlitlab-presentation Talk at QUEST Charité Talk at Max Planck Group	{domain}/version-control-survey implemented in formR ~ 1 quiz / session	{domain}/dra-fair-teaching Talk at Digital Research Academy (DRA)

... and more: DDLitLab Project Website ({domain}/ddlitlab), DDLitLab Proposal built with Jupyter Book ({domain}/ddlitlab-proposal), Slides for 90-min Quarto workshop ({domain}/quarto-workshop), DDLitLab Report ({domain}/ddlitlab-report)

### Evaluation

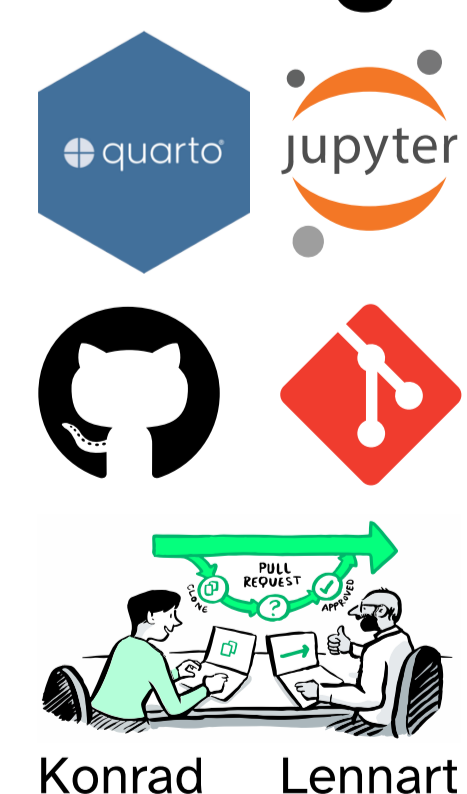
"Learning about version control was **eye-opening** and has left me **eager to implement it** in my academic and research endeavors. The online resources and appropriate pacing of the course allowed for **effective learning and application.**"

"The course is **well structured**, with **clear and understandable** materials provided via the homepage and guide. The practical exercises and **excellent support** from the instructor make learning **engaging and effective.**"

"The course's open-source materials and **emphasis on good scientific practices** beyond version control were particularly appealing. [...] making the course **highly recommended** for anyone interested in the field."

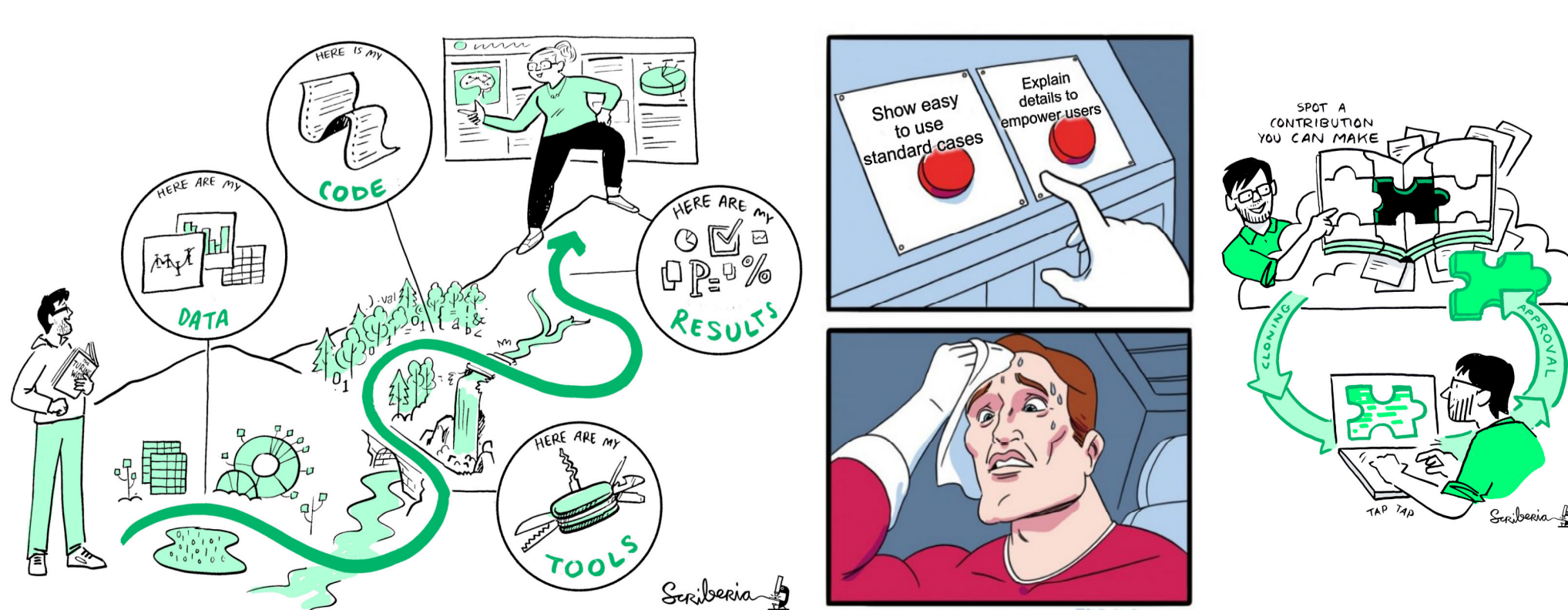
### Open, Reproducible & FAIR Teaching

- Websites, book and slides as **open-source code** build with Quarto (scientific publishing system)
- "Working in the open": All code tracked with Git, openly shared on GitHub, developed collaboratively
- Archiving on Zenodo / UHH FDR (DOIs), **reusable** and **interoperable** data formats, **open licenses**



### Insights

- Content & Course Development**
1. Git is not hard to learn but requires **repeated practice**
  2. Focus on **simple examples** & much **implementation**
  3. Tricky to **connect to discipline-specific** examples
- Admin & Project Management**
1. Challenging to **embed new course** in curriculum
  2. **Collaborative development** of materials is much fun!
  3. Use a **teaching diary** (notes after every session)



### Outlook

- Course v2.0 in summer semester 2024 (n = 21)!
- Development of "Version Control Book" continues!
- Submission special issue of "Neuroinformatics"?
- Virtual course with Erasmus University Rotterdam?

### Acknowledgements

The illustrations on this poster are created by **Scriberia** with **The Turing Way community**. Used under a CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807

"Piled Higher and Deeper" by Jorge Cham ([phdcomics.com](http://phdcomics.com))

**DataLad Handbook** by Adina Wagner et al., (CC-BY-SA 4.0)

Thanks to Prof. Nicolas Schuck, Dr. Adina Wagner, DDLitLab, the students

