Version Control of Code and Data

A full-semester course about Git for scientists

Lennart Wittkuhn[1] & Konrad Pagenstedt[1]

[1] Institute of Psychology, University of Hamburg, Germany

Funding & support by:

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DIGITAL AND DATA LITERACY IN TEACHING LAB

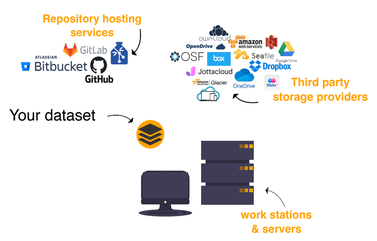
Stiftung Innovation in der Hochschullehre

Science is complex

Scientific building blocks are not static

Version Control

Science is iterative and distributed



Collaboration

"Science as open-source knowledge development"

Web: lennartwittkuhn.com/dditlab-presentation
Contact: lennart.wittkuhn@uni-hamburg.de

DOI 10.25592/uhhfdm.14162

Course 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

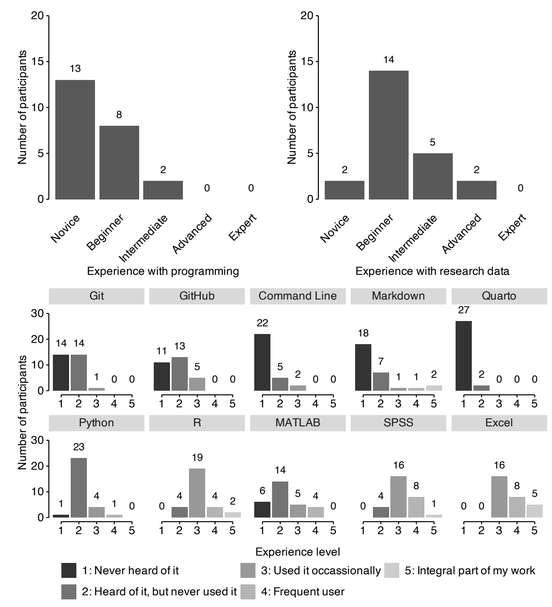
Exercises (~ 60 min): Course participants delve into hands-on exercises

Quizzes (~ 15 min): Supporting quiz questions

- WS + SS 2023/2024
- 14 sessions, 90 min each
- M.Sc. & Ph.D. Psychology
- Elective Module (4 ECTS)
- English
- Web Analytics

Participants

(N = 30 so far)

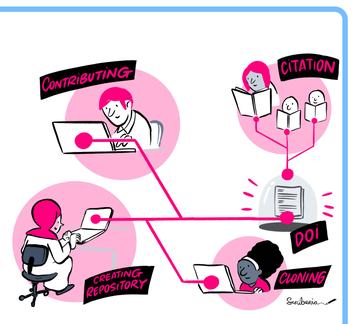


Version Control

- 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



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. Versioning Data

Course Project

- create "recipes" repository
- mimic reproducible paper
- learn same tools that are used for teaching materials

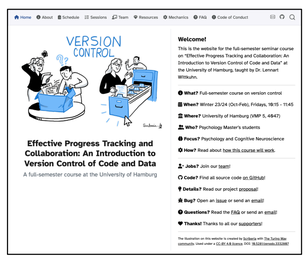
git

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

Open Educational Resources & Project Documentation

Web: lennartwittkuhn.com/{path} **Code:** github.com/lnnrtwttkhn{path}

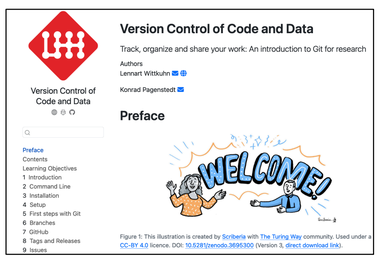
Course Website



{domain}/version-control-course-uhh-ss24

Slides & Schedule
Web: 4.7K views

Online Course Book

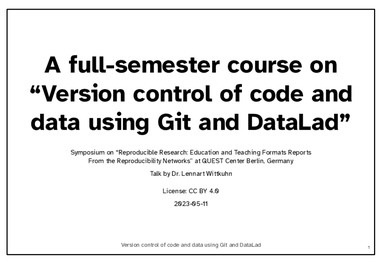


{domain}/version-control-book

DOI 10.25592/uhhfdm.14149

PDF: > 200 pages
Web: 4.5K views

Presentations

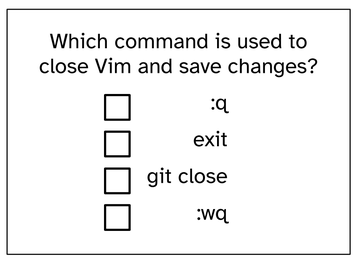


{domain}/ddlitlab-presentation

DOI 10.5281/zenodo.7926692

Talk at QUEST Charité
Talk at Max Planck Group

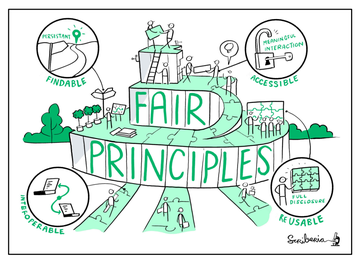
Quizzes



{domain}/version-control-survey

implemented in formR
~ 1 quiz / session

FAIR Teaching



{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)

Course 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**

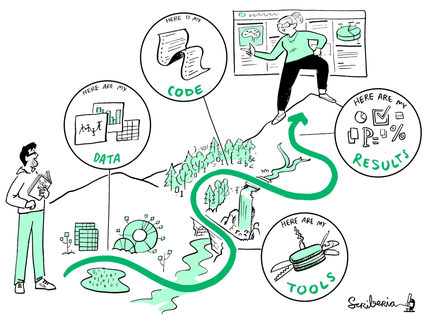
Insights

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 and ongoing work

- Development of "Version Control Book" continues!
- Virtual course with Erasmus University Rotterdam!
- Adapt materials to more formats & target groups!
- More funding by UHH & Claussen Simon Stiftung!

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**)

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

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