

A full-semester course on “Version control of code and data”

Lab Meeting of the Max Planck Research Group “NeuroCode” at the Max Planck Institute for Human Development Berlin and Research Group “Mechanisms of Learning and Change” at University of Hamburg

DOI [10.5281/zenodo.8300151](https://doi.org/10.5281/zenodo.8300151) License CC BY 4.0

Talk by Lennart Wittkuhn & Konrad Pagenstedt

2023-08-29

About

-  **WIP:** The presented teaching project is work in progress!
-  **Slides:** Slides are publicly available at lennartwittkuhn.com/ddlitlab-presentation
-  **Software:** Reproducible slides built with [Quarto](#) and deployed to [GitHub Pages](#) using [GitHub Actions](#)
-  **Source:** Source code is publicly available on GitHub at github.com/lennrtwittkuhn/ddlitlab-presentation
-  **License:** Creative Commons Attribution 4.0 International ([CC BY 4.0](#))
-  **Notes:** Please make use of the collaborative notes on [HedgeDoc](#)
-  **Contact:** We are happy for any feedback or suggestions via [email](#) or [GitHub issues](#). Thank you!

Agenda

1. Objectives of this talk (Lennart)
2. Background & history (Lennart)
3. Course overview & resources (Lennart)
4. Introduction to version control (Konrad)
5. Course structure (Lennart)
6. Course contents (Konrad)
7. Course evaluation (Lennart)

Objectives of this talk

Nico: “Before you start, use one slide to clarify your objectives: What’s the intended duration? Are you seeking a rigorous critique of your argument or just some final touch-ups? Do you want feedback on details or more on the overarching theme?”

General questions

- What would you expect from a Git course?
- Is any topic still missing?
- What do you wish you’d known about Git earlier?
- If you use Git, how do you use it in your day-to-day work?
- What are useful exercises for each session?
- Any comments / ideas on the study design of the course evaluation?

Bonus

We would be super happy about feedback on our materials, especially the [Version Control Book](#) and the [course website](#). Thank you! 🙏

Background & History

- October 2022: Lennart realizes that he will have to start teaching at UHH next year 😱
- Lennart thinks: “Can I teach what I am most excited about?” (Quiz: What is Lennart most excited about?) 🤔
- Lennart discovers the [Digital and Data Literacy in Teaching Lab \(DDLitLab\)](#) funding program 😄
- Lennart thinks: “Maybe if I can get some money, I can convince people to let me teach Git.” 💰

Digital and Data Literacy in Teaching Lab

“The Digital and Data Literacy in Teaching Lab (DDLitLab) [...] promotes digital teaching innovation related to data literacy education with teaching projects at all faculties and as part of the [...] Studium Generale.”

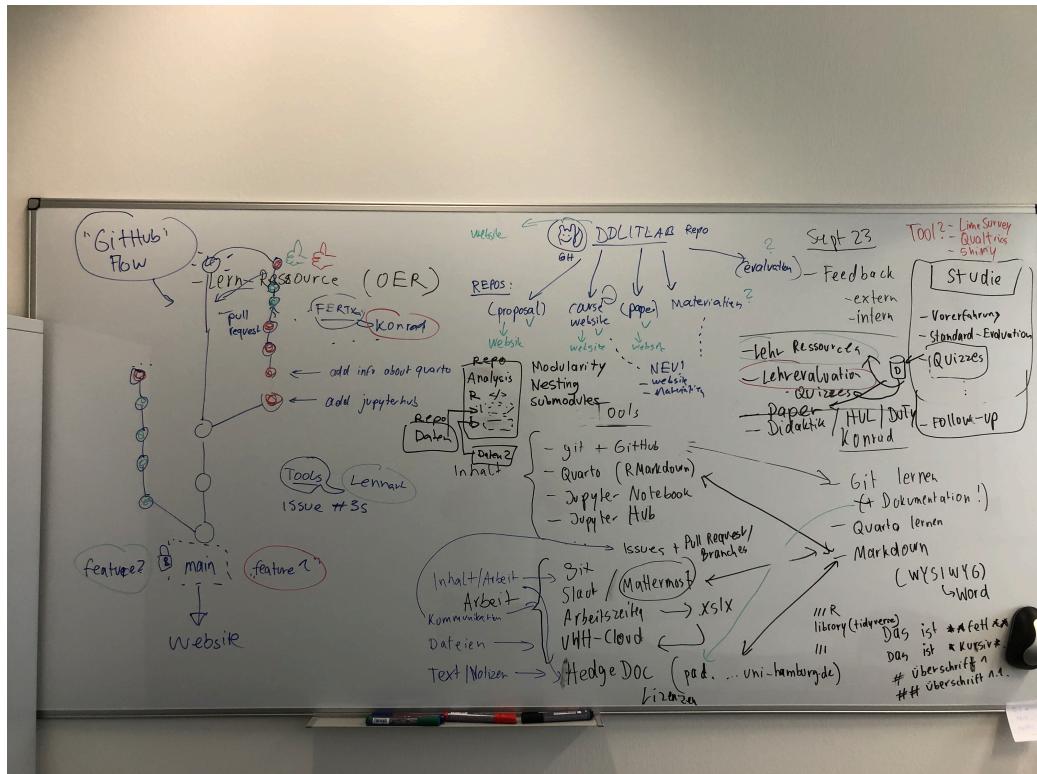
Overview

- ⚒ **Funding:** 50,000 Euro
- 📅 **Period:** 01.04.23 to 31.03.24 (12 months)
- 🌟 **Community-building** across [20 projects](#)

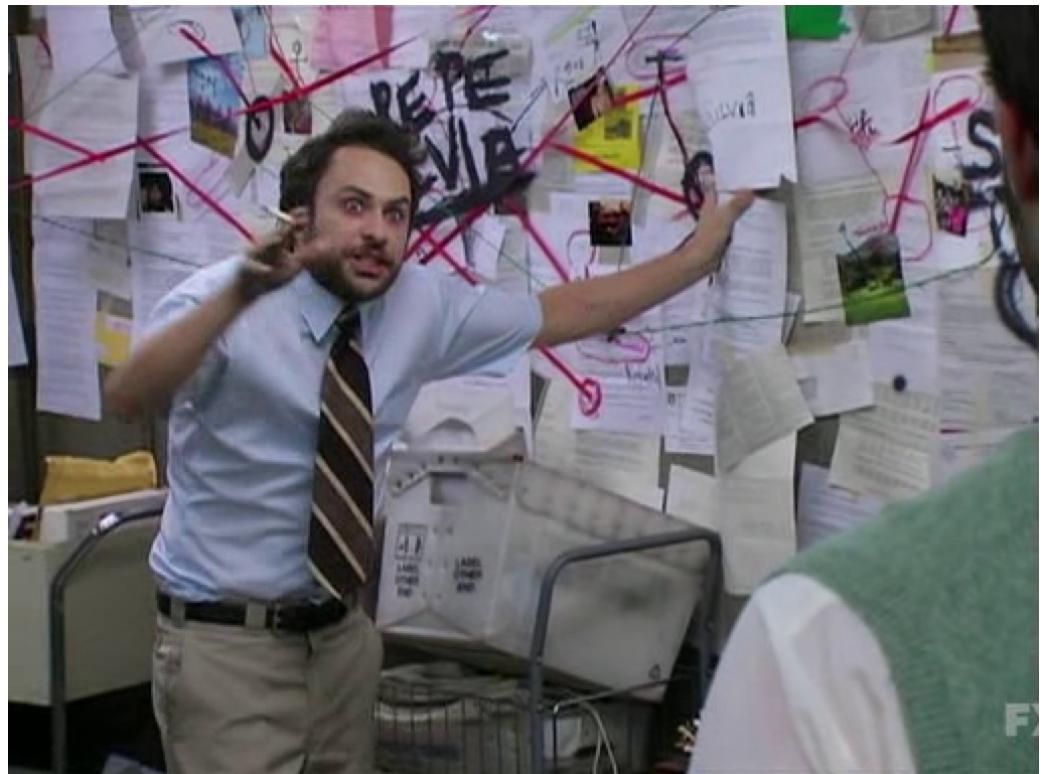
Support

- 🚶 Support by **research** and **teaching assistants**
- 🎤 Support inviting **external speakers**
- 🛠 Support with **technical infrastructure**, didactics, ...

How it started ...



Lennart's whiteboard in the beginning of June 2023



"Pepe Silvia" meme, scene from "It's Always Sunny In Philadelphia"

How it's going ...

 lennartwittkuhn.com/ddlitlab

Resources: Course Website

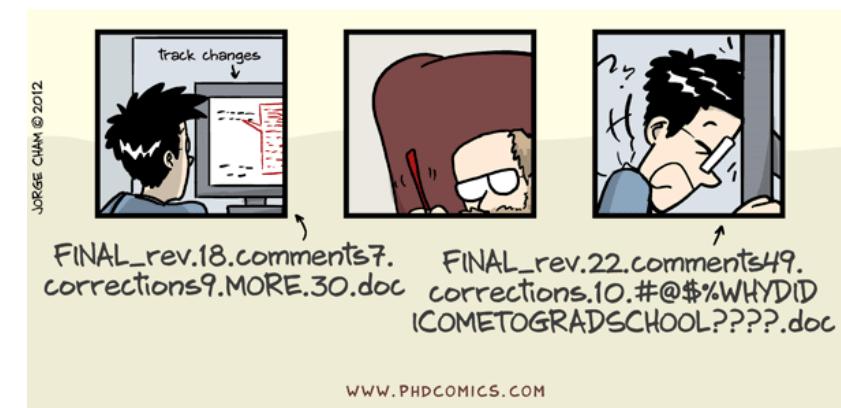
 lennartwittkuhn.com/version-control-course-uhh-ws23

Resources: Version Control Book

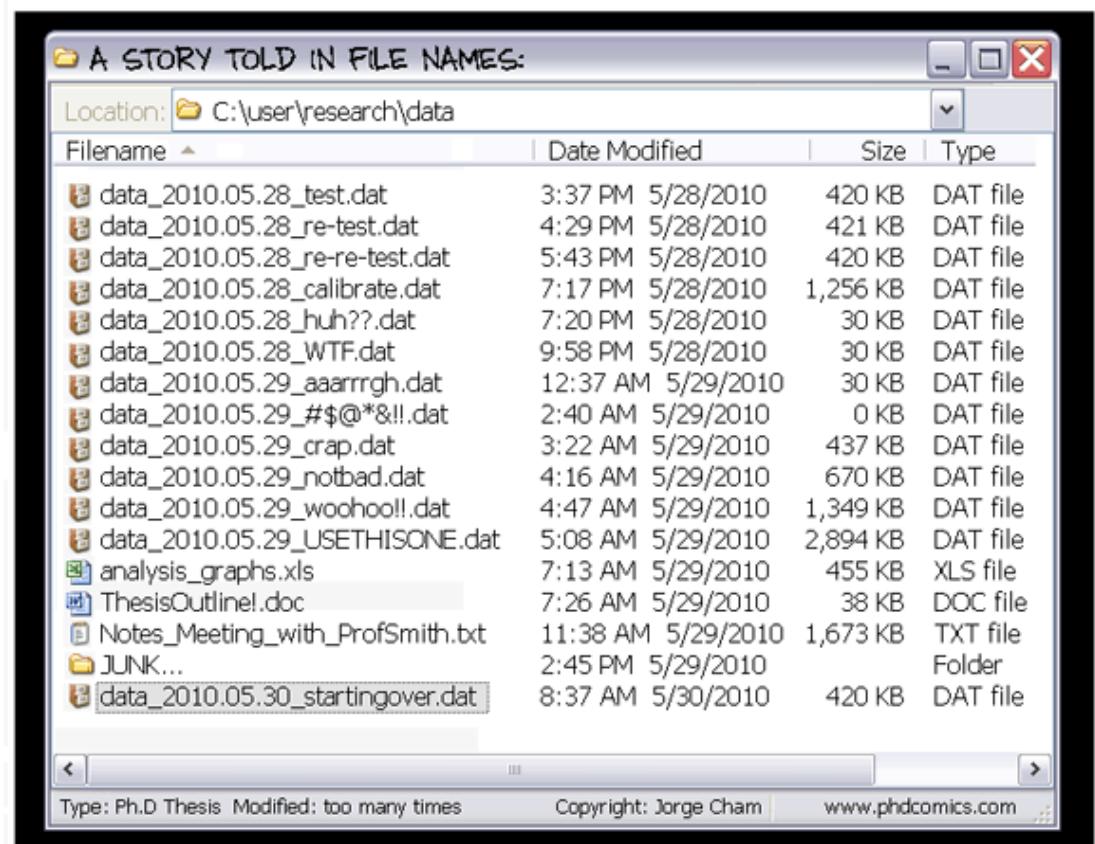
 lennartwittkuhn.com/version-control-book

Why we need version control ...

... for **code** (text files)



... for **data** (binary files)



© Jorge Cham (phdcomics.com)

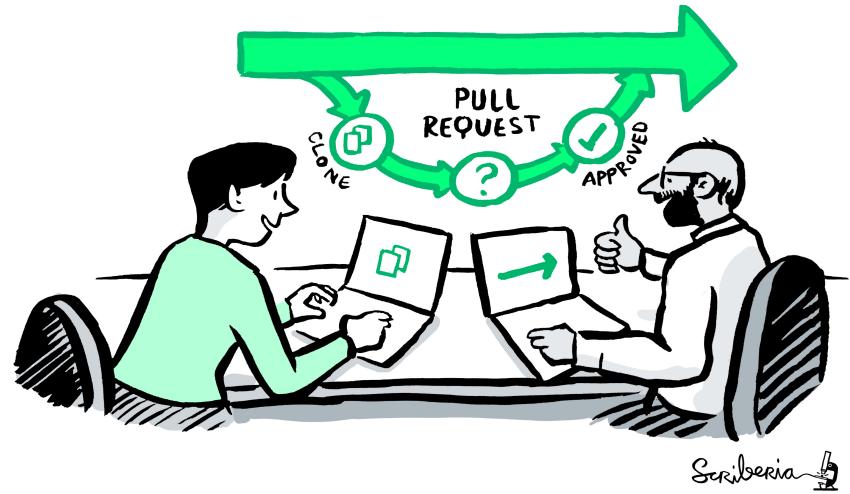
What is version control?

“Version control is a systematic approach to record changes made in a [...] set of files, over time. This allows you and your collaborators to track the history, see what changed, and recall specific versions later [...]” ([Turing Way](#))

- ➡ 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 and go back to any previous state
- ➡ work on parallel “branches” & flexibly “merge” them
- ➡ “push” your repo to a “remote” location & share it
- ➡ 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
- ➡ make your repo public and openly share your work



by Scriberia for The Turing Way community (CC-BY 4.0)



by Scriberia for The Turing Way community (CC-BY 4.0)

What are git and DataLad?



git-scm.com

- most popular version control system
- free, [open-source](#) command-line tool
- graphical user interfaces exist, e.g., [GitKraken](#)
- standard tool for most (all?) software developers
- 100 million [GitHub](#) users ¹



datalad.org

- “git for (large) data”
- free, [open-source](#) command-line tool
- builds on top of [git](#) and [git-annex](#)
- allows to version control arbitrarily large datasets ²
- graphical user interface exists: [DataLad Gooey](#)

Note: We will mainly focus on Git and only refer to DataLad as an outlook.

Course structure

Summary: A hands-on seminar about version control of code and data using Git with curated online materials, interactive discussions, quizzes and exercises, targeted at (aspiring) researchers in Psychology & Neuroscience.

- **Date:** Winter semester 2023/24 (Oct to Feb)
- **Time:** Fridays, 10:15 to 11:45 am (90 minutes)
- **Room:** UHH, VMP 5, Room 4047
- **Program:** MSc Psychology (PO 2014)
- **Events:** 14 seminar sessions
- **Module:** Elective Course
- **Credits:** 4,0
- **Language:** English (German optional)

What does the average seminar session look like?

1. **Content Review (up to 30 minutes):**

Course participants engage with [the online materials](#), supplemented by concise presentations by the instructors. Some course preparation may occur outside of the class.

2. **Interactive Discussions & Quizzes (up to 15 minutes):**

Course participants collectively address any inquiries related to the session's content and online materials. Instructor-led quiz questions may also be interspersed throughout.

3. **Exercises & Implementation (up to 60 minutes):**

Course participants actively delve into hands-on exercises and assignments.

⌚ Learning objectives

-  Benefits of version control using Git
-  Using Git for local version control
-  Using remote Git repositories to collaborate
-  Releasing scientific work using Git repositories

Course content

Sessions

1. Introduction to version control
2. Command line
3. Git Basics
4. Git Branching and Merging
5. Collaborative Workflows with Git (GitHub)
6. Git Workflows
7. Stashing and Rebasing
8. Graphical User Interfaces (GUIs)
9. Project Management and Collaboration Tools
10. Git Tags and Releases
11. Submodules and Advanced Topics
12. DataLad Talk
13. ???
14. Cake? 🍰

Peek in the Book Link

 1 Contents

1 Contents

The estimate of the **reading time** for each chapter is computed by counting t a reading speed of 200 words per minute.



Order By

14 min

August 11, 2023

Introduction

BEGINNER BASICS

Introduction to fundamental concepts of version control.

15 min

August 11, 2023

Command Line

BEGINNER BASICS COMMAND LINE

In this session, we will explore the concept of the command line and discover some commands that work well alongside git.

Course tasks

■ Working on a repository

Example Task - Session 3:

Install Git, Create a Git repository, stage and commit changes to a file and create a .gitignore file.

Example Task - Session 5:

Establish the connection between Git and GitHub, upload your repository to GitHub, and pull and push changes to files.

Example Task - Session 9:

Track a task with an issue, propose a change through a pull requests, and create a README.md files. Choose open-source licenses, set contribution guidelines.

Quizzes

Example quiz questions - Session 2:

Your rights

According to Art. 13(2)(b) of the Ger (GDPR), you have the right to:

- 1. Access (Art. 15 General Data Protection Regulation; §34 Federal Data Protection Act):** You can request information about the data concerning you, including potential recipients, to receive a response within one month of your request.
- 2. Rectification, Erasure, and Right to Object (Art. 16 General Data Protection Regulation; §35 Federal Data Protection Act; BDSG):** You can request correction or deletion of inaccurate or unnecessary data, or object to its processing.

Course evaluation

Background

- ⏱ “Digital Open Science” ([Toelch and Ostwald 2018](#)): Version control among *least* implemented practice.
→ Teaching / learning version control takes time and students should engage in hands-on coding assignments
- ✘ Software Carpentries ([Wilson 2016](#)): Provide opportunities for repeated practice in a controlled environment

Scholarship of Teaching

- 💡 We have a full-semester and dedicate most time to practical implementation!
- ❓ Will this learning setting result in more implementation?

Measurements

- 📈 Record and evaluate responses to quiz questions, implemented in [formR](#) ([Arslan, Walther, and Tata 2019](#))
- 🐣 Assess number of PRs, comments in issues etc. from public repos via GitHub API (cf. [Millman et al. 2018](#))

Open questions and TODOs

- 📈 Details of the study design (follow-up survey after 1 and 2 years?)
- 🚶 Ethics proposal for data collection and publication

Implementation and tools

-  **Impulse lectures & live demonstrations**
-  **Code-along & exercises** (individual and group)
-  **Discussions** on reproducibility, open code & data
-  Fixed computational environments on [JupyterHub](#)
-  Focus on **command-line** interaction
-  Alternative use via **Graphical User Interfaces**
-  **Quizzes** & continuous **evaluation** (in [formR](#))
-  Reuse quiz & evaluation data as **example datasets**
-  **Follow-up** research projects in summer semester
-  Support by **research** and **teaching assistants**
-  Materials shared as **Open Educational Resources**
-  Integration with [GRN](#), [Carpentries Incubator](#), etc.?

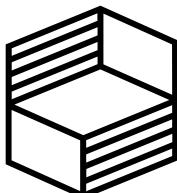
Thank you!

Funding & Support



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Digital and Data Literacy in Teaching Lab (DDLitLab), an initiative by the ISA-Zentrum at University of Hamburg



**Stiftung
Innovation in der
Hochschullehre**

[Stiftung Innovation in der Hochschullehre](#)

People



[Prof. Dr. Nicolas Schuck](#)
(UHH & MPIB)



[Carolin Scharfenberg](#)
(UHH DDLitLab)

Contact

- lennart.wittkuhn@uni-hamburg.de
- lennartwittkuhn.com
- [Twitter](#) [Mastodon](#) [GitHub](#) [LinkedIn](#)

References

- Arslan, Ruben C., Matthias P. Walther, and Cyril S. Tata. 2019. "Formr: A Study Framework Allowing for Automated Feedback Generation and Complex Longitudinal Experience-Sampling Studies Using R." *Behavior Research Methods* 52 (1): 376–87. <https://doi.org/10.3758/s13428-019-01236-y>.
- Millman, K. Jarrod, Matthew Brett, Ross Barnowski, and Jean-Baptiste Poline. 2018. "Teaching Computational Reproducibility for Neuroimaging." *Frontiers in Neuroscience* 12 (October). <https://doi.org/10.3389/fnins.2018.00727>.
- Toelch, Ulf, and Dirk Ostwald. 2018. "Digital Open Science – Teaching Digital Tools for Reproducible and Transparent Research." *PLOS Biology* 16 (7): e2006022. <https://doi.org/10.1371/journal.pbio.2006022>.
- Wilson, Greg. 2016. "Software Carpentry: Lessons Learned." *F1000Research* 3 (January): 62. <https://doi.org/10.12688/f1000research.3-62.v2>.

Footnotes

1. (Source: [Wikipedia](#))

2. see DataLad dataset of 80TB / 15 million files from the Human Connectome Project (see [details](#))