

“Track, organize and share your work: An introduction to Git for research”
Concept paper for the full-semester course on version control with Git, prepared for the
Teaching Award 2024 at the Institute of Psychology, University of Hamburg
Dr. Lennart Wittkuhn, 2024-11-15

i This concept paper summarizes content, structure, and didactic methods of the full-semester course “Track, organize and share your work: An introduction to Git for research” by [Dr. Lennart Wittkuhn](#). An online version is available at <https://lennartwittkuhn.com/version-control-course-concept>. A PDF version is archived at [doi: 10.25592/uhhfdm.16783](https://doi.org/10.25592/uhhfdm.16783). It was prepared for the Teaching Award 2024 at the Institute of Psychology, University of Hamburg. More details about the course can also be found in this [project description](#), this [project report](#), and a [20-minute video interview](#). Links to external resources are highlighted in blue.

Concept of the course

Content In today’s digital age, most scientific insights in quantitative research are gained from the analysis of data using programming code. Despite the importance of data and code, researchers often deal with digital objects on their computers in a chaotic manner. Questions such as “Which version did I use?” or “What did I change?” are common, compromising the reproducibility and trustworthiness of science and leading to frustration and loss of time. The good news is that scientists can learn about practices and tools from other disciplines, particularly software engineering, that have professionalized collaborative work on digital objects like programming code. At the core of this toolkit is the free and open-source software [Git](#), a so-called “version control system” that allows managing the evolution of digital objects on computers and sharing them via platforms like [GitHub](#) or [GitLab](#). In this hands-on seminar with practical exercises, course participants learn how to track, organize and share their work using Git.

Structure and didactic methods The full-semester seminar consists of up to 14 sessions, with 90 minutes per session. Every course features a dedicated course website (see [WS 23](#), [SS 24](#), [WS 24](#)), serving as a centralized platform for all course-related details. This includes the course schedule, course information, course mechanics, and session-specific materials. Each session page provides a transparent overview of the procedure, clearly defined learning objectives, exercises with example solutions, and accompanying slides. The sessions follow a consistent and structured format as outlined in the following:

During the *Reading* phase (ca. 30 minutes), participants explore chapters from the “[Version Control Book](#)”. This free, open-source online textbook, specially created for the course, encompasses over 200 pages of in-depth course content. It also serves as a standalone open educational resource, supporting self-paced learning beyond the classroom and making knowledge about version control accessible to anyone.

During the *Exercises* phase (ca. 50 minutes), the course emphasizes hands-on implementation. Participants collaborate on coding assignments inspired by “[pair programming](#)” ([Williams, 2001](#)), working in small groups to encourage peer exchange and mutual support. Over the course duration, participants develop a [version-controlled online recipes book](#). While this project may appear trivial or unscientific, it effectively teaches critical skills in file management and the development of computational notebooks in a fun and engaging way. These skills are directly transferable to academic tasks, such as thesis projects.

In addition, *Quizzes* (ca. 10 minutes) are offered regularly to help students assess their progress toward the learning objectives. A [pre-course survey](#) collects information on prior experience with relevant tools, expectations, concerns, and motivations. All surveys are built using the open-source survey framework [formr](#) ([Arslan et al., 2019](#)), an open-source alternative to proprietary tools like [Qualtrics](#), with [survey code](#) openly shared. Survey data is analyzed to identify common errors or misunderstandings, which are addressed in subsequent sessions and guide further development of the course materials.

The course also integrates several additional components alongside this core structure. Each session includes clear [learning objectives](#). [Exercises](#) are divided into mandatory and optional tasks, with example solutions and a clean project folder available for download. [Cheatsheets](#) provide a quick reference for commands introduced in each session. Additionally, we use [Plausible](#), an open-source, GDPR-compliant web analytics tool, to track resource use and analyze online interactions on all course-related websites.

Reference to teaching award criteria

Student centering With most class time dedicated to hands-on exercises, active participation is key to the course. Students can approach tasks in their preferred learning style and allocate time for reading, exercises, and quizzes as needed. This flexibility is balanced by clear task goals and learning objectives, providing reassuring structure and guidance. All materials (course website, online textbook) are openly accessible, supporting asynchronous self-study. Students can review content, complete assignments, and take quizzes at their own pace, even outside class hours. Developing a version-controlled online recipes book as a computational notebook combines skill-building with a fun project, boosting motivation and practicing skills that can transfer beyond the course.

Connection of scientific and practical relevance The seminar bridges theory and practice through hands-on training in version control with Git, a tool of growing importance in academic research. Students also learn theoretical concepts related to reproducibility and best scientific practices for managing programming code and research data. These skills are increasingly vital in psychological research and highly applicable across various career fields where data-driven decision-making is essential.

Topicality and degree of innovation Against the backdrop of the reproducibility crisis, particularly in psychological research (e.g., [Crüwell et al., 2023](#); [Hardwicke et al., 2021](#); [Obels et al., 2020](#)), this seminar addresses a pressing need by introducing students to transparent, reproducible workflows - a cornerstone of open science. The seminar's innovation lies in both its subject matter and execution. All teaching materials are open-source educational resources, created using innovative tools and by the same principles taught in the course. Adopting a "working in the open" approach, the seminar enables transparency and collaboration throughout, from initial conception to implementation in programming code. Recognizing these efforts, the project received funding to generalize and promote this [approach to FAIR and reproducible teaching](#).

Promoting research-based learning and critical thinking The seminar concept draws on the Scholarship of Teaching approach ([Hutchings & Shulman, 1999](#)), which emphasizes reflecting on teaching practices from a research perspective. A comprehensive questionnaire at the start of the course captures students' prior knowledge, motivation, concerns, and expectations. Key concepts are assessed repeatedly to monitor understanding. Systematic versioning of all resources allows to identify connections between students' knowledge and the evolution of materials, even retrospectively. Combined with web analytics, this enables data-driven improvements to refine resources and teaching strategies. The seminar encourages students to critically assess current scientific practices, focusing on transparency, effectiveness, and collaboration.

Feedback culture Students can influence the seminar's progression through questions and feedback at any time. Additionally, the course teaches the tools (primarily [Git](#), [GitHub](#), and [Quarto](#)) used to create the course materials. As all materials are open-source, students are encouraged to contribute directly to the resources. For example, students [provided feedback via GitHub](#), and one student directly [fixed a typo in the Version Control Book](#), demonstrating the principles taught in the course.

Promotion of equality We value the social elements of the learning experience, fostering a welcoming environment, which is also emphasized in a [code of conduct](#). Collaborative learning is encouraged through exercises, and to accommodate students' workloads, course content is designed to be largely covered during class time. All coursework is clearly outlined, establishing fairness and transparency while making it easy to catch up on missed sessions. [Quarto](#)'s modern technical framework also includes accessibility features, such as alternative text for images and hyperlegible font, to support individuals with visual impairments. The tools taught in the course are free, open-source, and compatible with all major operating systems, avoiding reliance on proprietary, costly software.

Internal Differentiation Differentiation is achieved by allowing students to study materials at their own pace. They can pair with peers at varying learning stages, creating diverse learning interactions. The concept of pair programming is also implemented, encouraging students to support each other in completing tasks. Advanced students can share their knowledge, reinforcing their understanding while taking on a teaching role. Tasks are divided into mandatory and optional components. Mandatory tasks are designed for all students, while optional exercises offer deeper engagement with the content. The chapters of the Version Control Book are structured to be completed during class, but also offer expandable sections and cross-references providing additional background for deeper exploration of the topics.

Interdisciplinarity The course is interdisciplinary, bringing together students with diverse interests in psychological research, as revealed by our pre-course survey. Despite the varied research focuses, the core skills taught - version control, reproducibility, and collaboration - are universally applicable across disciplines. The course provides a common framework for managing research data, sharing code, and enabling transparency and effective collaboration in scientific practices. These skills are valuable not only in Psychology but also in any research field or industry where data-driven decision-making and collaboration are key.

Internationalization All course materials (website, book, slides) are available in English. With support from the University of Hamburg, the [third version of the course](#) is being offered as part of an international teaching collaboration between the University of Hamburg and Erasmus University Rotterdam. This partnership provides students from both institutions the opportunity for both knowledge and cultural exchange. Website usage data of the Version Control Book shows over 1,000 unique users from more than 30 countries.

Media use While the course materials focus on *digital media*, they provide a wide range of resources. These include course websites with slides, the comprehensive Version Control Book, and online quizzes. All materials are fully open-source and rely on open-source tools, ensuring high potential for interoperability and reusability - key elements of the FAIR principles ([Wilkinson et al., 2016](#)). Major releases are tagged, annotated with metadata, and archived in repositories like [Zenodo](#), ensuring long-term accessibility. This course thereby exemplifies an approach to FAIR and collaborative development of educational materials (cf. [Garcia et al., 2020](#)), opening new avenues for participatory and adaptable teaching methods.

Appendix

Acknowledgements This project was funded by the [Digital and Data Literacy in Teaching Lab \(DDLitLab\)](#), an initiative by the [ISA-Zentrum](#) at the [University of Hamburg](#). The Digital and Data Literacy in Teaching Lab program is in turn funded by the [Stiftung Innovation in der Hochschullehre](#). The project was also funded by the Federal Ministry of Education and Research (BMBF) and the Free and Hanseatic City of Hamburg under the Excellence Strategy of the Federal Government and the Länder as part of the [International Teaching and Learning Formats \(ILLF@home\)](#) program at [University of Hamburg](#).

Teaching Award Criteria Student centering (e.g. promote active participation and engagement, encourage self-study, motivate, enable participation), teaching skills (e.g. understandable communication, commitment, professional competence), connection of scientific and practical relevance, topicality and degree of innovation, research-based learning, stimulate critical thinking, feedback culture, promotion of equality, internal differentiation, interdisciplinarity, internationalization, media use, evaluation results ([Source](#)).

References

- Arslan, R. C., Walther, M. P., & Tata, C. S. (2019). formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*, 52(1), 376–387. <https://doi.org/10.3758/s13428-019-01236-y>
- Crüwell, S., Apthorp, D., Baker, B. J., Colling, L., Elson, M., Geiger, S. J., Lobentanzer, S., Monéger, J., Patterson, A., Schwarzkopf, D. S., Zaneva, M., & Brown, N. J. L. (2023). What's in a Badge? A Computational Reproducibility Investigation of the Open Data Badge Policy in One Issue of *Psychological Science*. *Psychological Science*, 34(4), 512–522. <https://doi.org/10.1177/09567976221140828>
- Garcia, L., Batut, B., Burke, M. L., Kuzak, M., Psomopoulos, F., Arcila, R., Attwood, T. K., Beard, N., Carvalho-Silva, D., Dimopoulos, A. C., Angel, V. D. del, Dumontier, M., Gurwitz, K. T., Krause, R., McQuilton, P., Le Pera, L., Morgan, S. L., Rauste, P., Via, A., ... Palagi, P. M. (2020). Ten simple rules for making training materials FAIR. *PLOS Computational Biology*, 16(5), e1007854. <https://doi.org/10.1371/journal.pcbi.1007854>
- Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., deMayo, B. E., Long, B., Yoon, E. J., & Frank, M. C. (2021). Analytic reproducibility in articles receiving open data badges at the journal *Psychological Science*: An observational study. *Royal Society Open Science*, 8(1). <https://doi.org/10.1098/rsos.201494>
- Hutchings, P., & Shulman, L. S. (1999). The Scholarship of Teaching: *New Elaborations, New Developments*. *Change: The Magazine of Higher Learning*, 31(5), 10–15. <https://doi.org/10.1080/00091389909604218>
- Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology. *Advances in Methods and Practices in Psychological Science*, 3(2), 229–237. <https://doi.org/10.1177/2515245920918872>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1). <https://doi.org/10.1038/sdata.2016.18>
- Williams, L. (2001). Integrating pair programming into a software development process. *Proceedings 14th Conference on Software Engineering Education and Training*. 'In Search of a Software Engineering Profession' (Cat. No.PR01059). <https://doi.org/10.1109/csee.2001.913816>