

Short description of OSW hands-on session

Rostock Open Science Workshop, 17-18 March 2025

Projects' workflow for reproducibility and replicability in Stata (Maarten Buis and Daniel Krähmer)

The end product of a research project is typically an article, manuscript, or some other form of publication. However, in order for that result to be reproducible, it also needs a reproduction package. This is a folder that contains everything someone needs (and you can legally and ethically provide) to reproduce your study.

In the workshop we will approach this from two sides: We will start in the first session from the reproducers' point of view: You get a reproduction package, and you try to make sense of it. We will have different reproduction packages from the same dummy project, all with different problems. The purpose is that we try to establish what makes a good reproduction package, and what are possible pitfalls we want to avoid.

In the second session we will approach this problem from the researchers' point of view: what is a workflow that allows us to do research and have a reproduction package with the characteristics we established in session one come out at the end. The challenge of designing a workflow is that you will be doing that daily for a very long time. It is not enough that this is "the right thing to do". It needs to be comfortable enough, such that we are not tempted to take short-cuts.

Projects' workflow for reproducibility and replicability in R (Jonas Scholey and Egor Kotov)

In this workshop, participants will discover practical strategies for ensuring reproducibility and replicability in their R projects, from robust package management to effective code and data organization. We will introduce tools such as ``renv`` and ``pak`` for maintaining reproducible environments, alongside ``targets`` for workflow management, GitHub for version control, and Zenodo for archiving. We will also touch on containerization options like Docker and Apptainer (Singularity) to support long-term reproducibility.

Replicability of results in the context of private non-sharable data (Lisa DeBruine)

Being able to simulate data allows you to prep analysis scripts for pre-registration, calculate power and sensitivity for analyses that don't have empirical methods, create reproducible examples when your data are too big or confidential to share, enhance your understanding of statistical concepts, and create demo data for teaching and tutorials. This talk will cover the basics of simulation using the R package `{faux}`. We will simulate data with factorial designs by specifying the within and between-subjects factor structure, each cell mean and standard deviation, and correlations between cells where appropriate. We will also discuss other methods for and issues with replicability and non-shareable data.

GitHub (Beginner): Using Git and Github for Open Science & connection with OS repositories for sharing data and codes (OSF and Zenodo)

(Aliakbar Akbaritabar)

Prior knowledge required: None (just being enthusiastic about doing research suffices!)

Technical equipment required: Personal laptop with permission to install software (if you cannot install software, you can follow along on the projector)

ToDo before the workshop: Please download and install Git for Windows (<https://git-scm.com/downloads/win>), Mac (<https://git-scm.com/downloads/mac>) or Linux (<https://git-scm.com/downloads/linux>). Please note: no GUI installation is needed. Only follow the instructions in one of those links, depending on your operating system, and install the Git software. If you were asked for Windows installation to add “context menu shortcut” and “add Git to PATH” please select “yes” for both (selected by default).

Description:

This tutorial assumes no prior knowledge about version control and using Git, GitHub, and similar tools for research. At the end of the tutorial, participants will know what they will gain by using version control and how they can use different version control tools in writing scientific text. The tutorial covers academic writing and analysis scripts, how version control helps, and the benefits of switching to Latex, RMarkdown, and other plain text formats for academic writing. Participants will learn to use Git and GitHub in Terminal (CLI) and popular IDEs such as RStudio and Visual Studio Code. Participants will learn how to deposit replication materials on Open Science Framework (OSF) and Zenodo and share a view-only link for peer review.