# Title

Open Research Data - Tools and a workflow for the modern researcher

# Catalogue Data

## Abstract

Expectations on researchers are increasing. Proposals require data management plans, journals expect data underlying publications to be shared, and the Open Science movement creates pressure to make research reproducible. It can be overwhelming to identify the tools and platforms that are most suitable to satisfy these expectations. In this course, we introduce a set of tools and an opinionated approach commonly used in software development, and increasingly expected to be used in scientific computing. Applying Open Science principles to research products has the potential to make knowledge more equally accessible, partly by publishing without barriers to access and reuse, but also through the use of open-source and free software.

## Objective

By the end of this course, students will be able to:

* write documentation for data using a literate programming tool (R Markdown)
* paraphrase how a file written in a markup language like markdown (.md) differs from a file written in a proprietary file format like .docx
* interact with git version control and GitHub using the RStudio IDE
* archive data using GitHub and Zenodo integration to receive a persistent digital object identifier (DOI)
* add a data publication record to a created ORCiD profile
* apply a workflow that uses R packages devtools, usethis, roxygen2 and pkgdown to document and share data as an R data package
* publish an R data package as a free website using GitHub Pages
* understand how FAIR data principles apply to published data

## Content

This course provides a hands-on opportunity to learn a workflow which uses FAIR data principles to publish data. This workflow uses a set of open source tools and platforms that are accessible to researchers beyond their employment at ETH Zurich. The teaching technique in this course is participatory live coding. Learners are invited to follow along the instructor who writes and narrates code out loud.

Topics covered are:

* Reproducible Research (Open Research, Research Data Management, Version Control, Licensing)
* R Package Development
* Collaboration (git, GitHub)

This course will introduce a range of tools commonly used in software development. None of these tools will be explained in great detail. The goal is to expose learners to these tools rather than develop a complete understand of the underlying concepts. The course therefore requires an openness to foreign concepts.

## Lecture notes

A public website is in preparation.

## Prerequisites / Notice

While this course does not require experience in any programming language, learners that have previously been exposed to data science tools (R, RStudio IDE, Jupyter Notebooks, Python, VS Code) would benefit most from this course.

Prior to the course, learners are expected to have:

* a laptop with internet connection and one of Mozilla Firefox, Safari, Google Chrome, Microsoft Edge (just not the Internet Explorer)
* a registered account on GitHub (https://github.com/)
* a registered ORCID iD (https://orcid.org/)
* tabular data stored in CSV file format that they are willing to publish openly
* read the following paper: https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989
* read the following paper: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005510

The course instructor will contact the registered participants two weeks prior to the course to send a reminder about these prerequisites.

# Performance Assessment

Participants receive 0.5 ECTS for successful participation. This is the equivalent of 15 working hours. The course will be delivered in two modules, each worth 4 hours. The remaining 7 hours will be used for participant preparation prior to the first module and between the first and second module.