Open Science

Webpage for ghe.ethz.ch

**Version history**

* draft outline completed by Lars
* draft outline reviewed by Mian and Liz
* draft narrative completed by Lars on 2023-09-19

# Open Science (Overview page)

We are committed to Open Science. It is a movement aimed at making scientific research more accessible, inclusive, and transparent. Our vision is that all our research output is reproducible. To ensure this, it is our mission to foster an Open Science culture and empower students, researchers, and staff with tools and workflows that enable open data and open code.

Our tactics achieve this by:

* holding a 3-hour Quarterly team event, which discusses and integrates elements of Open Science
* teaching courses at ETH that build competencies in research design, research data management, scientific communication and computational tools (R, Quarto, Git, GitHub)
* contributing to networks, communities, and working groups for Open Science and reproducibility
* writing grant applications that target programmes focused on Open Science

We have developed a data publishing workflow, which support ideas around next generation metrics. Raw, but analysis-ready data is documented and shared prior to the preparation of scientific articles. Where applicable, data from BSc and MSc projects is published. Every data publication is considered a standalone research output with its unique DOI (digital object identifier) and all contributors, including research support staff (e.g. enumerators, laboratory technicians) are added to publications with their unique ORCID iD.

The diagram below highlights the elements and tools of our workflow:

**Communciation**

Within in our group we avoid Email by all means. For management and communication on tasks, we use the GitHub issue tracker and Asana task management tool. For instant communication, we use the Matrix chat protocol in combination with Element. For anyone that is interested to connect with us on Open Science, we provide an open room on Element that can be accessed and joined at: https://matrix.to/#/#ghe-open:staffchat.ethz.ch

## Projects (sub-pagge)

### openwashdata

Add Mian’s content.

### openjmp

Open JMP - unlocking the potential of global indicator data

* Project start: 2023-01-05
* Project duration: 12 months
* Project lead: Global Health Engineering
* Funding: Open Research Data Program of the ETH Board
* Funding: 30,000 CHF

Decades of manual data structuring have resulted in the most comprehensive and internationally-comparable information on Water, Sanitation, and Hygiene (WASH) coverage. The databases are maintained by the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP) and data are shared in spreadhsheet-based proprietary software. This under utilizes the potential those data could have for other purposes than the national, regional and global monitoring of progress in WASH mandated to the JMP.

The project has three Work Packages, which can be summarised as follows.

I have analyzed the attached project proposal in Markdown format[1]. Here is a concise three-bullet summary of each work package:

#### WP1: Document and publish R data package jmpinput

* Goal: Make JMP input data readily accessible for researchers and practitioners.
* Main output: An R data package with detailed documentation, metadata, codebooks, and a public website.
* Activities: Write documentation, prepare use cases as vignettes, publish data and documentation as an R package, and create a public website.

#### WP2: Prepare and document R software Package jmpmodel

* Goal: Develop open-source software for transparent documentation of modeling activities for primary indicators.
* Main output: An R software package with complete documentation and a public website.
* Activities: Collaborate with JMP team to replicate production of basic indicators, prepare R functions for data manipulation, modeling, and visualization, write documentation, and publish R functions as a software package.

#### WP3: Dissemination

* Goal: Seek active collaboration on package development to increase potential usage and long-term maintenance of data and software package.
* Main output: Increased community engagement and dissemination of developed products.
* Activities: Continuously update JMP team on progress, host two public participatory live coding online workshops to showcase developed R packages from WP1 and WP2.

Read the full funded proposal, including Project Plan and Work Packages here: https://github.com/openwashdata/proposal-eth-ord-scheme-contribute/blob/main/proposal-eth-ord-scheme-contribute.md

### Data Stewardship Network

## Courses (sub-pagge)

### rbtl

Research Beyond the Lab: Open Science and Research Methods for a Global Engineer

Offered for the first time in 2022, we will teach a revised version of this course in spring 2024. We designed this course to teach learners how to conduct a research project out of the lab, and apply Open Science principles to their data analysis projects.

Over the course of the semester, students will develop a research project and learn the necessary quantitative methods required to collect data from people. We will use tidyverse R packages to work with data, git and GitHub as tools for version control and collaboration, and Quarto publishing system for scientific writing. By the end of the course, students will have a complete overview of how a typical research project is designed, implemented, analysed, and communicated.

All materials are published as [Open Educational Resources with permissve licenses through GitHub](https://github.com/rbtl-fs22). For students, these course materials are available as a public website: https://rbtl-fs22.github.io/website/.

### Open Science Tools

Open Science Tools - authoring and publishing workflows for collaborative scientific writing

This course offers researchers a hands-on opportunity to learn open source tools and platforms that are accessible beyond institutional employment. The participatory live coding teaching technique involves an instructor writing and narrating code as they teach, and inviting learners to write and execute the same code.

Topics covered include:

* Scientific communication.
* Reproducible research (open research, version control, licensing).
* Collaboration through Git and GitHub.

During the course, participants will learn how to use the Quarto file format to prepare various scholarly documents, including slides, scientific articles, books, and websites. Additionally, we will demonstrate how GitHub can serve as both a project management tool and a publishing platform using GitHub Pages.

While the course will introduce commonly used tools in software development, they will not be explained in great detail. Instead, the goal is to expose learners to the tools and concepts, requiring openness to new ideas.

Course Catalogue: https://www.vorlesungen.ethz.ch/Vorlesungsverzeichnis/lerneinheit.view?lerneinheitId=174400&semkez=2023W&ansicht=LEHRVERANSTALTUNGEN&lang=en Website: https://ost-hs23.github.io/website/ GitHub organisation: https://github.com/ost-hs23/

## Fellowship (sub-pagge)

In 2022, we started offering an Open Science fellowship. The goal of the fellowship is to support students in applying Open Science principles to their research. Students are invited to join our group at ETH for a duration of five months between July and October. The fellowship is offered as part of the[ETH project mobility students programme](https://ethz.ch/en/studies/non-degree-courses/project.html).

The conditions for application are:

* student from a university in Africa
* enrolled in MSc. programme
* data for thesis has already been collected

We support the fellowship with 20’000 CHF which covers:

* flights
* accomodation in Zurich
* a monthly stipend of 1750 CHF
* administrative costs (first monthly ZVV pass, registration)

Conditions for application are:

* A max 400 word summary covering MSc. thesis objectives, collected data formats, and experience using computational tools (R, Git, GitHub)

## Strategy & Planning (sub-pagge)

Developing an Open Science culture within a research group requires addressing and dispelling common misconceptions about Open Science. It also requires opportunities for learning tools and methods to apply Open Science principles. We have incorporated three regular events into our research group’s organizational structure and culture:

1. Weekly 1-hour event: Weekly event with a shared Google Doc as a notebook. Prior to the event, everyone adds their notes to the Google Doc in four categoires (Doing this week, Done last week, Support needed, Other announcements). This events purpose is to keep everyone in the group aligned updated, but also an opportunity to discuss Open Science elements of individual projects.
2. Quarterly 4-hour event: Our Quarterly event allows for work and sharing of group-wide projects. We also invite external speakesr to help us learn more about working together efficiently. Every event has at least one element that addresses Open Science principles.
3. Yearly 2-day event: We organize an annual 2-day retreat where the research group comes together for an extended period to discuss their work, share best practices, and develop strategies for incorporating Open Science principles into our research.

When fostering an Open Science culture, we are confronted with misconceptions around publishing data and code openly with permissive licenses. Some common misconceptions we are tackling in our strategic meetings are:

**Misconception 1: Publishing my data does not benefit anyone**

Counterargument: Publishing research data can benefit both the researcher and the scientific community. Sharing data increases the visibility of the research, potentially leading to more citations and recognition for the researcher. It promotes collaboration and enables other researchers to build upon the existing work, fostering innovation and scientific progress while avoiding ineffiencies.

**Misconception 2: Others may criticize my code**

Counterargument: While it is true that open-source code may be subject to criticism, this can be a positive aspect of Open Science. Constructive criticism can help identify potential issues, improve the code, and contribute to the overall quality of the project.

**Misconception 3: Publishing my content under CC-BY-NC will prevent people from exploiting my content commercially.**

Counterargument: While the CC-BY-NC (Creative Commons Attribution-NonCommercial) license does restrict commercial use of the content, it may also inadvertently limit the potential impact and reach of your work.

By using a CC-BY-NC license, you prevent the following use cases:

* Commercial use in academic research: Researchers working in collaboration with commercial entities or receiving funding from commercial sources may be unable to use your content, limiting the potential for interdisciplinary research and collaboration.
* Commercial use in educational materials: Publishers of textbooks, online courses, and other educational materials that are sold for profit may be unable to include your content, reducing its potential reach and impact on students and educators.
* Commercial use in software development: Companies developing software or applications that incorporate your content may be unable to do so under a CC-BY-NC license, limiting the potential for innovation and the development of new tools and technologies.
* Commercial use in creative works: Artists, writers, and other creators who wish to incorporate your content into their commercial works may be unable to do so, limiting the potential for your work to inspire and influence others.

## Data Management Strategy (sub-pagge)

We are actively working on a coherent Data Management Strategy for our group. At this poing, we prepared a strategy for BSc and MSc projects. This so-called [GHE students Wiki is publicly available](https://unlimited.ethz.ch/display/ghestudents/GHE+students+Wiki) and published using the Atlassian Confluence tool, for which ETH covers a license agreement.