# Motivation for Team Award

## Swiss Reproducibility Award 2024

Lars Schöbitz

2024-06-03

This motivation was written in response to the call for nominations for the Swiss Reproducibility Award 2024. The motivation was written by Lars Schöbitz, a team member at the Global Health Engineering group at ETH Zurich.

The result of the review process was announced on 8th May 2024 and this nomination has not been selected as the winner of the category.

## Motivation (word count: 347 words)

We are dedicated to rigorous, transparent, and reproducible research. We aim to make all our research outputs reproducible and foster an Open Science culture. For this purpose, we use public GitHub repositories for version control, collaboration, and integration with Zenodo for long-term archiving and generation of a Digital Object Identifier (DOI).

We categorize our data into three stages:

- Stage 1: unprocessed raw data
- Stage 2: processed analysis-ready data
- Stage 3: data supporting publication results

The raw data's value is relatively low, but the code transforming it into analysis-ready data is invaluable. This code enhances transparency, promotes reusability, and boosts efficiency. It streamlines data cleaning, a crucial but often overlooked task. Moreover, by packaging our code as functions, we enhance efficiency in data cleaning for similar future projects. We ensure a complete audit trail from raw to analysis-ready data by openly sharing both alongside the code within the same repository. In addition, we use a permissive CC-BY license, assign DOIs, and link ORCID IDs to acknowledge all contributors, including those often unrecognized in scientific outputs (e.g. surveyors and lab technicians). Finally, our data is published as an R data package, enhancing accessibility and usability.

In Stage 2, we use the above R data package to prepare a reproducible manuscript. We utilize a literate programming tool to integrate narrative and code. The manuscript is stored and managed in another git repository. It receives a separate DOI, highlighting the differences between data cleaning and curation activities, which is performed by a data steward supporting the group, and the actual work that goes into authoring a manuscript, fine-tuning tables, figures and models, which rests with the scientists.

In Stage 3, we produce CSV files for the data supporting the publication results (e.g. tables, figures, models), allowing others to use our results quickly.

Our team has an Open Science Specialist who mentors our staff and students in applying open science principles, which unifies our team, aligning everyone with our vision and mission. Our team deserves this award to recognize our efforts to make science more rigorous, transparent, and reproducible.