

Abstract for Demonstration at Swiss Research Data Day 2024

Publish with Ease: A Low-Cost, Open-Source Solution for Research Data Sharing

More and more data are being generated, yet public, easy-to-use datasets remain an exception. Too often, research data are collected and cleaned, only to be analyzed *ad hoc* and, if published at all, without documentation that enables reuse. These practices not only make knowledge inaccessible but also lead to redundancies that can be avoided through a more streamlined data publication process.

In this talk, we will introduce the complete data workflow we apply at the Global Health Engineering Group, ETH Zurich, to transform raw data into cleaned datasets that meet FAIR data sharing principles. Our custom R package “washr” helps to create well-documented data, write READMEs and codebooks, and assign digital object identifiers (DOIs), making the final outputs accessible to everyone. This setup allows us to publish large volumes of generated data efficiently without compromising on accessibility or interoperability.

Our aim is to demonstrate a low-cost, low-maintenance approach to data publication, in contrast to many other application areas where large amounts of data can only be accessed and published with sufficient computational power. By being able to run the entire pipeline on an ordinary laptop, using solely open-source software, we aim to provide researchers working with relatively small datasets (under 30MB) with a streamlined, more economical alternative for publishing their data.

Keywords: open source; data publication; FAIRness; rstats

Words: 222 (250)