presentation abstract

# Format

Live or pre-recorded presentation (15-20 min) followed by Q&A (10-15 min).

# Title

The data and code the behind the JMP WASH estimates

# Abstract

**Why is the topic relevant to the current trends in the WASH sector?**

The WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (the JMP) was established in 1990. Since then, a database of nearly 20’000 data points has been established for sanitation alone.

The data is collected from various sources, including censuses, household surveys, administrative data, and others (e.g. research). Data structures between these sources are highly heterogenous, and a significant amount of effort is invested into extracting data from each source and bringing it into a defined format. Anyone can access the result of this work through the so-called “country files”, which are MS Excel workbooks containing each country’s raw data.

While the JMP uses this input data specifically to produce indicator estimates for SDG indicators 6.1 and 6.2, there is great potential for unforeseen use cases (e.g. research, teaching, joining with other data, etc.). These could be enabled by making the data readily accessible in a form that follows FAIR data principles.

Availability of data is one of the key constraints within the WASH sector, and from observing current trends, it becomes a key focus area for organisations (e.g. UNICEF Game Plan to Reach Safely Managed Sanitation; WHO monitoring of the safe management of on-site sanitation (SMOSS) services).

We have extracted the data contained in the JMP MS Excel country files and produced a documented data package that can be readily accessed using the R statistical computing software. In our first use case, we reproduce indicator estimates for SDG 6.2 using R. The JMP uses linear regression models to produce these estimates for reasons including the often small number of data points per indicator, reliability over time, and comparability between countries. In our research, we have experimented with alternative statistical models that could be applied to some countries.

In this session, we will introduce the “jmpraw” R data package and show a live coding demonstration of how it can be used to reproduce indicator estimates. We will present key results from our research on applying alternative non-linear models to produce indicator estimates.

This session highlights how making data accessible in a format that follows FAIR data principles can positively benefit other people’s work. It supports the mission of the Colorado WASH Symposium by applying tools that are open and free for anyone to use.