

GHE Yearly 2025

openwashdata - past and future

Lars Schöbitz

Global Health Engineering, ETH Zurich

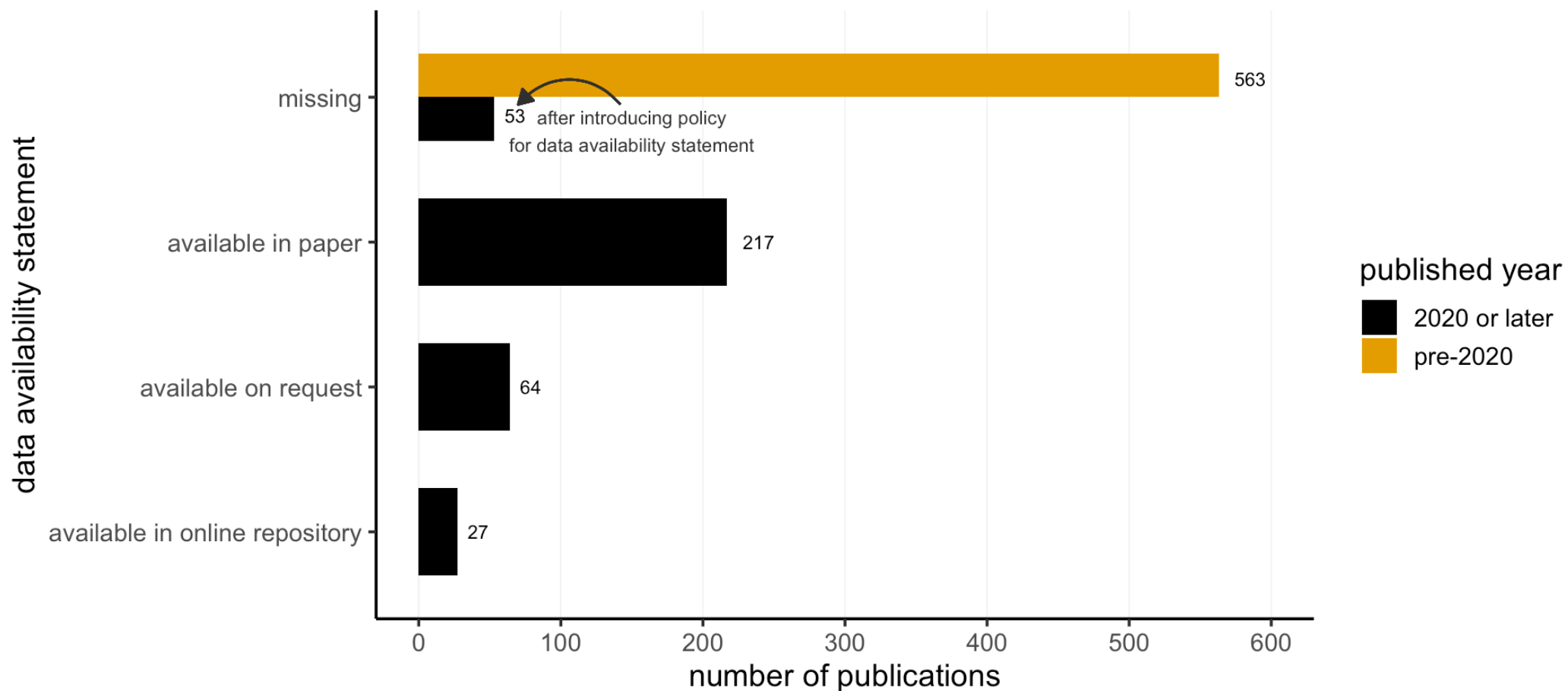
June 2, 2025

The Opportunity

Journal articles

Data Availability Statement

Analysis of 924 articles published in Journal of Water, Sanitation and Hygiene for Development (2011 to 2023)



Journal articles

Supplementary Material

Take-away: Not a single file is in machine-readable, non-proprietary file type format that would qualify for following FAIR principles for data sharing ([Wilkinson et al. 2016](#)).

Good practice: CSV file (comma-separated values), including a data dictionary for all variables/columns in the data

Supplementary Material		
Articles published 2020 or later		
file type	n ¹	%
missing	202	51.4
docx	149	37.9
xlsx	24	6.1
pdf	13	3.3
pptx	4	1.0
png	1	0.3
¹ One article can have multiple files.		

openwashdata community

openwashdata community

- Launched 10 March 2023
- Supported by four projects worth 340'000 CHF (50% in-kind contribution)
- Ends in July 2026
- So many outputs to write about (from 2026 to 2027)

Vision

An active global community that applies FAIR principles ([Wilkinson et al. 2016](#)) to data generated in the greater water, sanitation, and hygiene sector.

Mission

Empower WASH professionals to engage with tools and workflows for open data and code.

VMOST as a method

VMOST

- Vision
- Mission
- Objectives
- Strategy
- Tactics

Objectives (Indicators)

By the end of March 2024

1. Increase the number of datasets published on the website to 20 R data packages.
2. Increase the number of datasets that are donated for publication to 50 datasets.
3. Increase the number of people that have donated, cleaned, and published data independently with support of the openwashdata team to 5.
4. Increase the number of unique visitors to the website to 10 visitors/day.
5. Increase global coverage of visitors to the website to 50% of countries globally.
6. Increase the number of data users who report having used data published through openwashdata community to 2 uses per dataset on average.
7. Increase the number of subscribers to the openwashdata newsletter to 250 subscribers from 50 countries.
8. Increase the number of participants in live coding events to 5 participants on average.

Strategies

... 11

Tactics

... 5

Some Stats

- Number of datasets published: 33 R data packages
- Number of newsletter subscribers: 202 (14 since Jan 2025)
- Number of unique visitors (openwashdata.org): ~5 visitors/day (past 3 months)
- Number of unique visitors (data packages): ~5 visitors/day (past 3 months)

Future

WP2: Governance

- Activity 2.1: Develop a governance structure for a community organization and decision-making processes.
- Activity 2.2: Form a sounding board comprising community members to provide directional feedback.
- Activity 2.3: Create a long-term funding strategy for the openwashdata community.

Open question: Does openwashdata have a long term future?

WP3: Community expansion

- Activity 3.1: Offer advanced data science training and workshops to community members.
- Activity 3.2: Develop a mentorship program to support new members in adopting ORD practices.
- Activity 3.3: Organize community events to foster networking and collaboration.

Priority: Strong focus on WP3 for the remainder of the project.

data science for openwashdata 002

All efforts into the next iteration of the course.

- free, live, online, 10-week programme
- 200 registrations for 2023 iteration
- 100 show-ups
- 20 graduates
- next iteration: from September 2025, sign-up: <https://forms.gle/MP5rNYZagBdfG2ZRA>

ds4owd-001.github.io/website/

ds4owd-002 communication campaign (strategy)

- restart monthly newsletter editions (now)
- publish a blog post on ds4owd-001 (July)
- start publishing a LinkedIn post every Thursday (openwashdata thursday)
- host an information event (late August)
- host a series of workshops for [washr](#) (from November 2025 to March 2026)

ds4owd-002 course preparation

- Platform for access to recordings through authentication
- Prepare quizzes for each module for participants to complete each module
- Write R code for using GitHub API to comment and close issues

Papers from mid-2026

12 months, 4 papers

1. Setting the baseline: FAIR / Open Data practices in the WASH sector
2. Increasing competency: Data from two iterations of data science for openwashdata course
3. Streamling workflows: Developmment of an R package for FAIR data publication ([washr](#) / [fairenough](#))
4. Tracking impact: Analytics from published data packages

Thanks

This project was supported by the [Open Research Data Program of the ETH Board](#).

The slides were created via revealjs and Quarto:
<https://quarto.org/docs/presentations/revealjs/>

You can [view source code of slides on GitHub](#)

Or you can [download slides in PDF format](#)

This material is licensed under [Creative Commons Attribution Share Alike 4.0 International](#).

References

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1).

<https://doi.org/10.1038/sdata.2016.18>.