openwashdata conf 2025

where do we head?

Lars Schöbitz

Global Health Engineering, ETH Zurich

July 15, 2025



openwashdata conf 2025



Programme

Time	Title	Speaker
09:00 - 09:30	Arrival at Villa Hatt. Welcome coffee, bread rolls, croissants (Gipfeli).	-
09:30 - 10:00	Warm-up activity	<u>Lars</u>
10:00 - 10:45	openwashdata community - past and future	<u>Lars</u>
10:45 - 11:00	Break	-
11:00 - 12:00	Data stewardship at BASEflow Malawi	<u>Emmanuel</u>
12:00 - 13:00	Lunch break	-
13:00 - 13:30	Warm-up - Positive Gossip	-
13:30 - 14:00	Governance Part 1	<u>Lars</u>
14:00 - 14:15	Break	-
14:15 - 15:00	Governance Part 2	Lars
15:00 - 15:30	Hackathon Ideation	_
15:30 - 15:45	Break with small cakes	<u>Lars</u>



Session Roadmap (45 min)

- Lack The Opportunity (~10 min) Why open data matters in WASH
- **openwashdata Journey** (~10 min) Our story and framework
- **Q** Activity 1 (5 min) Your strategies & tactics
- **Future Plans** (~10 min) WP2/WP3 & ds4owd-002
- Funding & Collaboration (~5 min)
- Activity 2 (5 min) Building together
- **Wrap-up** (~2 min)

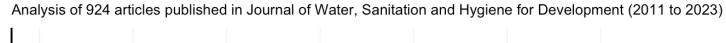


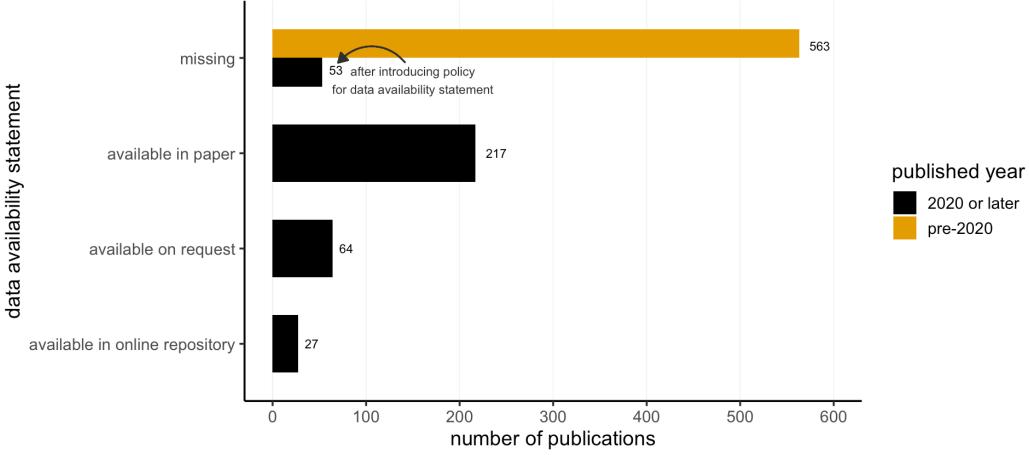
The Opportunity



Journal articles

Data Availability Statement







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 [@wilkinson2016].

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	n¹	%
missing	202	51.4
docx	149	37.9
xlsx	24	6.1
pdf	13	3.3
pptx	4	1.0
png	1	0.3

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)





An active global community that applies FAIR principles [@wilkinson2016] to data generated in the greater water, sanitation, and hygiene sector.





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



VMOST as a method



VMOST

- **O** Vision
- Mission
- **£** Objectives
- İr Strategy
- X 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



- Develop and maintain a data warehouse on the openwashdata website that provides an overview of published datasets.
- Develop a guide as a companion to workshops, live coding events, etc. that documents how to participate in the community and publish data following FAIR principles.
- Build a cohort of students, scientists, practitioners, and civil servants, that are comfortable
 using R statistical software for exploratory data analysis and Git version control and GitHub
 for communication and collaboration.
- Prepare all communication material for openwashdata using Quarto publishing framework¹ and R statistical software.
- Provide tools and resources to promote the use of open data in the WASH sector
- Publish workshops as open educational material.
- Introduce people to the concept of "open by default", as well as the use of open source software, the concept of open science, and benefits of open government (data).
- Build material always in mind with learner personas3 that were defined for the community.
- Communication material does not refer to openwashdata as a project, but as a community.
- Design a common corporate identity using defined color palettes, fonts, etc.
- Ensure that material developed for openwashdata follows best practices for accessibility.



^{1.} https://quarto.org/

Tactics

- Provide a 10-week online workshop for a selected group of participants to share tools and workflows that support publishing of open data following FAIR principles.
- Publish monthly blog posts on the openwashdata website about selected open datasets, community stories, workflows, insights into community management, use cases, etc.
- Publish monthly issues of the openwashdata newsletter.
- Host quarterly community meetups with invited speakers that share stories from their organisations related to data management, data analysis, open data, etc.
- Visualize and disseminate published open data using interactive dashboards, maps, articles, etc.



Get Ready to Share!

In a moment, I'll pause to hear your ideas

- Think about strategies & tactics for your context
- Consider your (local) challenges & opportunities
- What partnerships could accelerate progress?



Activity: (5 minutes)

In our shared Google Doc, please add:

What other strategies or tactics could help build an open data culture in WASH?

Think about:

- Your (local) context
- Partnerships you could leverage
- X Tools or resources you need



Future



WP2: Governance (what today is all about)

- 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: What is the medium-term future of openwashdata and what does it look like?



WP3: Community expansion (What I need your support for)

- 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, 9 module programme (goal: new AI module)
- 200 registrations for 2023 iteration (goal 2025: 500)
- 100 show-ups (goal 2025: 250)
- 20 participants completed capstone project (goal: 100)
- 5 participants published data packages (goal: 50)
- next iteration: from 11th September 2025, sign-up link: https://ee-eu.kobotoolbox.org/x/7V3qeDYD

ds4owd-002.github.io/website/



Course Overview

Thank you for your interest in this course. Your course instructors: <u>Lars</u>
<u>Schöbitz</u> & <u>Mian Zhong</u> & <u>Sophia Skorik</u> are looking forward to meet you.

We will meet on Zoom for 10 modules over 17 weeks (see Course Calendar below) at the following times:

- Start: 31st October 2023 2 pm to 4:30 pm CET
- End: 20th February 2024 2 pm to 4:30 pm CET

We will use Posit Cloud infrastructure, so you do not need to install any

and the second s



ds4owd-002 communication campaign (strategy from early June)

- restart monthly newsletter editions (now)
- publish a blog post on ds4owd-001 (July) (Thanks, Adriana)
- start publishing a LinkedIn post every Thursday (openwashdata thursday)
- host an information event (late August)
- host a series of workshops for washr / fairenough (from October 2025 to March 2026)



ds4owd-002 course preparation

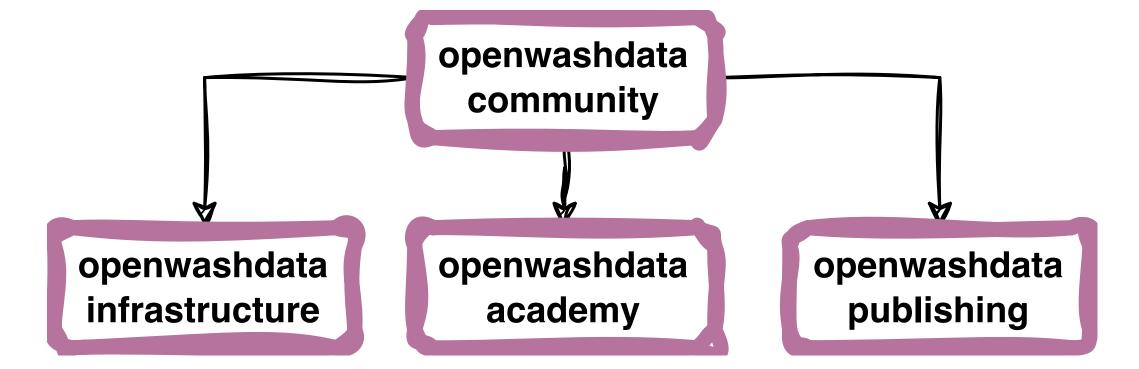
- Platform for access to recordings through authentication (Zoom with registration for each session and detailed usage statistics)
- Prepare quizzes for each module for participants to complete each module (with feedback on whether participants watched recording or joined live module)
- Share information about course through personal channels, newsletters, etc.
- Write templates / R function / Claude Code Slash commands for reviewing homework assignments, capstone projects, data packages, etc.
- Establish mentorship programme



Granogram (Phase 3?)

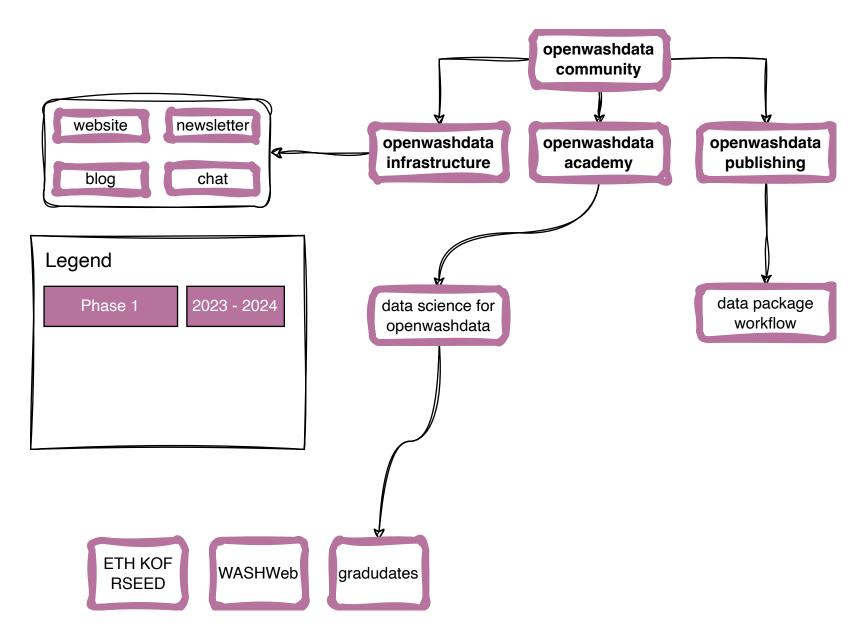


Starting Point



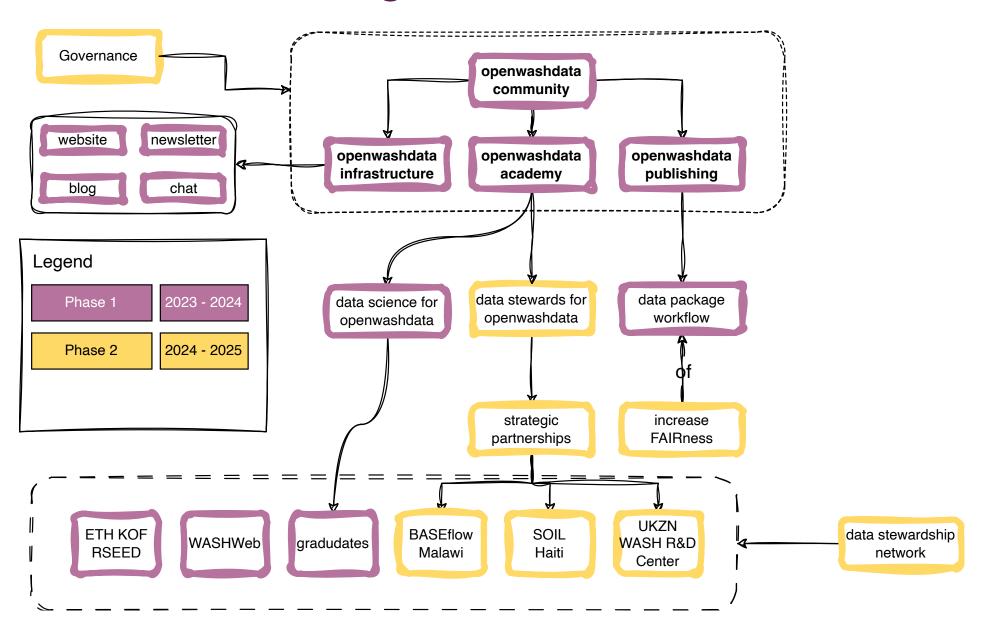


Mid-development



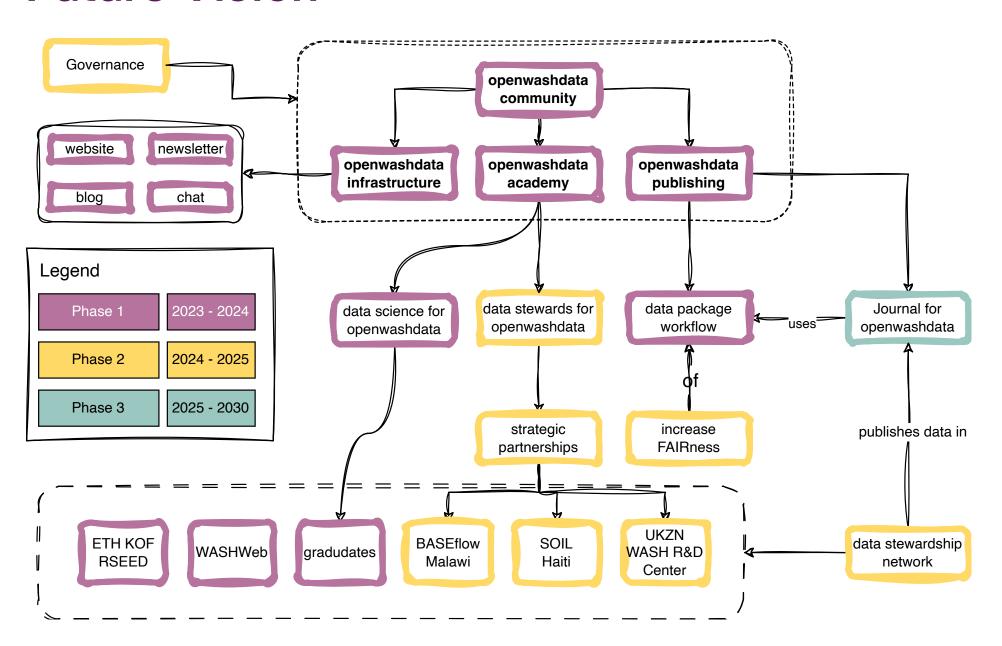


Community Integration





Future Vision





Funding opportunities



On the list

- 2025-09-01: <u>ETH Open Research Data Program</u> Creating and publishing open educational resources on research data management. CHF 10,000 CHF 25,000. (10 months)
- 2025-09-30: <u>Swiss Leading House Africa</u> Call for Research Partnership Grant. CHF 15,000 CHF 40,000 (3 to 12 months).
- 2025-10-31: Congressi Stefano Franscini. 2nd open(wash)data / conference in July 2027.
- TBD: <u>ETH Foundation</u>



Activity 2 Coming Up

Prepare to share your knowledge about:





What opportunities exist in your Who would you partner with? region/sector?



Activity: Building Together (5 minutes)

In our shared Google Doc, please share:

- 1. What other funding opportunities exist for open data initiatives?
- 2. Who would you want to collaborate with on an open(wash)data project?



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. Streamlining workflows: Development of an R package for FAIR data publication (washr / fairenough)
- 4. Tracking impact: Analytics from published data packages
- 5. What else?



Thanks 🌻

This project was supported by the <u>Open Research Data Program</u> of the <u>ETH Board</u>.

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 <u>download slides in PDF format</u>

This material is licensed under <u>Creative Commons Attribution</u> <u>Share Alike 4.0 International</u>.

