WASHCollab

Proposal for the 7th call for Contribute projects by the ETH Board

Table of contents

# 1. Abstract

TODO: Get instructions from submission portal

# 2. Keywords

# 3. Proposal full title

Enhancing Global WASH Data Accessibility through Collaborative Initiatives (washcollab)

# 4. Background and motiviation

## 4.1 Introduction to the partnership

WASHWeb and openwashdata were connected through one of the lead statisticians of the Joint Monitoring Programme (JMP), which is the custodian of global data on Water Supply, Sanitation and Hygiene (WASH) hosted by the World Health Organization (WHO) and UNICEF. Sharing common goals for an improved WASH data ecosystem, WASHWeb and openwashdata formed a partnership. Both teams met in-person at ETH Zurich in order to collaboratively submit this funding proposal.

## 4.2 WASHWeb

WASHWeb is a practice-oriented community for fostering advancements in water, sanitation, and hygiene (WASH) data. Through four thematic working groups, citizens, NGOs, governments, coders, researchers and publishers enrich WASH data discoverability, usability, and representation. WASHWeb’s platform exists to accelerate the collective improvement of the WASH data ecosystem towards the achievement of universal access to water, sanitation and hygiene. WASHWeb has authored the ORD WHO/UNICEF JMP R package [1].

## 4.3 openwashdata

openwashdata is a vibrant, global community that applies FAIR (Findable, Accessible, Interoperable, Reusable) principles to data generated in the WASH sector. The community focuses on data generated by researchers without making this a requirement, but due to the nature of openwashdata being placed within the Global Health Engineering group at ETH Zurich. More generally, the mission of openwashdata is to empower WASH professionals to engage with tools and workflows for open data and code, which is realised through education by the openwashdata academy providing training in the required competencies.

## 4.4 Common goals

WASHWeb and openwashdata share a collective vision of enhancing the accessibility, usability, and representation of WASH data globally. By leveraging their respective strengths in WASH practice and research, the partnership aims to maintain existing open-source tools for analysing key WASH data, develop new tools in collaboration with data providers, and disseminate its ongoing work through webinars and conference events. The collaboration strives to democratise access to critical information by involving smaller organisations and communities in the pursuit of improved WASH outcomes.

## 4.5 Problem statement

Despite the wealth of information in the water, sanitation, and hygiene (WASH) sector, data remains siloed, challenging to access, and underutilised. Existing datasets from authoritative sources such as the World Health Organization (WHO) and UNICEF’s Joint Monitoring Programme (JMP), World Bank, UN-Water Global Analysis and Assessment of Sanitation and Drinking-Water (GLAAS), and the Organisation for Economic Co-operation and Development (OECD) are not easily reusable for research. Furthermore, the scattered and disconnected nature of WASH data ecosystems impedes accountability and hampers collaborative efforts. WASHWeb and openwashdata aim to address these challenges by fostering a participatory approach, engaging diverse stakeholders, and promoting ethical open data and open source principles to make crucial information more discoverable, representative, and actionable on a global scale.

# 5. ORD project plan

## 5.1 WP1: Maintain

### 5.1.1 Goal

Work package 1: “Maintain” involves maintaining open source tools for researchers and product developers who analyse indicators of Sustainable Development Goal 6: “safe water and sanitation for all”. Specifically, WP1 will focus on maintaining the Joint Monitoring Programme (JMP) dataset for wider, more accessible use in the WASH community. Thanks to the funding the Global Health Engineering group received from the third call for Contribute projects by the ETH Board, we discovered that WASHWeb was actively working on tools we intended to develop. Instead of replicating efforts and creating inefficiencies, we invested in building a partnership between Global Health Engineering and WASHWeb. As researchers with experience in research data management, we now actively contribute to developing the tools to ensure compliance with FAIR data sharing practices [2]. To work productively, we request the support of WASHWeb developer Nick Dickinson to maintain the technical parts of the package for increased usability. This will enable us as researchers to use the data more efficiently for our purposes (e.g. scientific blogs, teaching material, scientific publications).

### 5.1.2 Activities

1. Update the WHO/UNICEF JMP R data package to represent the latest figures published, add functions for automatic updates and testing.
2. Liaise with JMP on issues identified in their data files to improve the (automated) data pipeline and create greater awareness.
3. Provide helper functions to shape data for better useability by researchers.
4. Update ORD documentation to provide examples on how to load and transform the datasets for analysis.

### 5.1.3 Aims addressed

Specifically, WP1 addresses two aims of the Contribute projects: We will be **contributing** by extending the ORD package to be able to shape the data into (long) formats for better usability by researchers. We are **curating** by updating and maintaining the data package.

The JMP dataset includes WASH indicators for all countries and therefore holds a very central role in many WASH analyses. This work is critical to improve both accessibility and (re)usability. We provide a community service by coordinating on the **development** of a flexible **ORD standard** for this specific dataset.

## 5.2 WP2: Extend

### 5.2.1 Goal

Work Package 2: “Extend” builds on the work from WP1 by engaging with other authoritative data providers of economic data on water and sanitation in order to produce a new R dataset package to support practitioners and researchers with their analyses. Reliable economic data, including disbursements and expenditure are critical for research on WASH investments. Together with the data from WP1, researchers will be able to compare investments and economic conditions against outcomes (actual service levels). Organisation identifiers for financiers, such as bilateral donors, government bodies, and their receiving organisations are lacking in these formal datasets. This work package will add identifiers so that metadata on these organisations, such as the type of organisation and their geographic location, can be used to answer more targeted research questions on investments in WASH.

Targeted data providers are:

* UN-Water Global Analysis and Assessment of Sanitation and Drinking-water (GLAAS)\*
* Organisation for Economic Co-operation and Development Creditor Reporting System (OECD-CRS)
* Sanitation and Water for All (SWA)
* International Accountability and Transparency Initiative (IATI)

The openwashdata and WASHWeb teams have worked with and are in contact with all of these organisations. The final dataset selection will be based on the level of collaboration of each data provider with our teams and overall technical feasibility. As a fallback, the public and freely available GLAAS data will be used. The technical development lead for the R data packages that are products of this WP2 will be with Nick Dickinson from WASHWeb, while the Global Health Engineering group will contribute by providing input on data sharing following FAIR data sharing practices.

### 5.2.2 Activities

1. Engage with authoritative data providers to determine feasibility and level of collaboration to produce an R package for WASH economic data.
2. Develop and publish an R package dataset for at least one of the identified authoritative data providers.
3. Add organisation identifiers in the dataset in collaboration with the data provider and a public registry (https://org-id.guide/ and/or Wikidata). With identifiers, researchers can leverage metadata on the type of investors for analysis.
4. Provide documentation on how to use the economic data.

### 5.2.3 Aims addressed

WP2 addresses the first two aims of the Contribute call by **contributing** new open source tools based on publicly available information from the most relevant international organisations, to support wider research capabilities and analyses in WASH research and practice. Additionally, this data will be **curated** to ensure the data is better organised with metadata to be able to classify the economic data for more refined analysis.

## 5.3 WP3: Disseminate

### 5.3.1 Goal

For wide scale use of the open data packages for WASH indicators, it is key that we promote and disseminate our work to a wider audience. In Work Package 3: “Disseminate”, we will host a webinar on innovations in WASH open data practices and use this opportunity to highlight ongoing progress under WP1 and WP2. Furthermore, the WASHWeb and openwashdata partnership will host a session at Stockholm World Water Week (SWWW) in August 2024 on the WASH data ecosystem, in collaboration with cross-sectoral partners representing research (University of KwaZulu-Natal, ETH Zurich), government (Government of South Africa), NGOs (IRCWASH, BASEflow Malawi, DigDeep), the UN (WHO-GLAAS), and the private sector (WASHNote). This session was accepted through a competitive application process for the largest, most prestigious water conference on Thursday, 11th April 2024. Finally, we will engage interested individuals and organisations to join discussion groups online, through Matrix.

### 5.3.2 Activities

1. Host a webinar on innovations in WASH open data practices and highlight ongoing work under WP1 and WP2.
2. Host a session at Stockholm World Water Week “Unlocking the hidden potential of open WASH data”.
3. Invite WASHWeb and openwashdata communities and others to join discussion groups for all thematic areas (1. Using Data Better, 2. Accountability, 3. Systems Strengthening, 4. Social Justice & Privacy) and facilitate ongoing conversations (live and in-chat).

### 5.3.3 Aims addressed

WP3 addresses the third aim of the Contribute call. We will **provide community services** through an online webinar (exp. 300 participants) to disseminate information about our ongoing work and the other work packages (WP1 and WP2), and we will host a workshop session at the largest annual water conference in the world (exp. 100 participants in the session, 15,000 attendees to the conference). Finally, these events will also engage the authoritative data providers to strengthen their link to ORD communities.

# 6. Impact

* How sustainable is the proposed project inside the ETH Domain?

This Contribute proposal is our next step towards the long-term sustainability of the openwashdata community within the ETH Domain. By collaborating with WASHWeb, an external ETH Domain actor, we will be able to promote them as an important, centralised partner who could also collaborate with other institutions within the ETH Domain (e.g., the Swiss Federal Institute of Aquatic Science and Technology (Eawag)). Therefore, the Global Health Engineering group can continue to invest in developing researcher-focused workflows in the ETH Domain and see these projects promoted through external actors who are experienced and respected professionals in the WASH sector—sustainability increases as we tap into sectors outside the research communities but benefit from them in return.

* To what extent may an existing or a newly formed community (be able to) engage with the ORD practice(s) built-up during the project?

As this Contribute proposal engages with two communities and the activities all revolve around engagement with community members, we expect significant take up in the ORD practices built up during the project. We will see an increased use of highly valuable data in research, practice, policy-making, and educational material. Due to our collaborative approach and the established openwashdata community, we can promote the data package through our numerous (ETH and external) educational activities. The communities will be able to identify several points of entry to engage with ORD practices as we grow our suite of available data and training.

# 7. Work Packages and milestones

The following Table 2 shows a basic gantt chart against the four work packages, including program activities. Column “Lead” abbreviations: ND (Nick Dickinson), SA (Scientific Assistant), ML (Merel Laauwen)

Any publications derived from this program will be published as open access material, following ORD practices and Open Science standards for computational reproducibility and sharing of data and code under FAIR principles.

Table 2: https://docs.google.com/spreadsheets/d/1xMlMepWoq4lTRRP89x8A1A-1Bkqkn7hjslvpJ5XPXU8/edit#gid=0

# 8. Resources (including project costs)

Table 4: https://docs.google.com/spreadsheets/d/1p77mMyCWZ3r8N73ZjYB\_u038cwsdsMWs6rMjDVwBW\_c/edit#gid=0

# 9. Bibliography

1. Nick D. WASHNote/jmpwashdata. 2024 Apr;

2. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, ’t Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. [The FAIR Guiding Principles for scientific data management and stewardship](https://doi.org/10.1038/sdata.2016.18). Scientific Data. 2016 Mar;3(1):160018.