

washr an R-package to facilitate data publishing

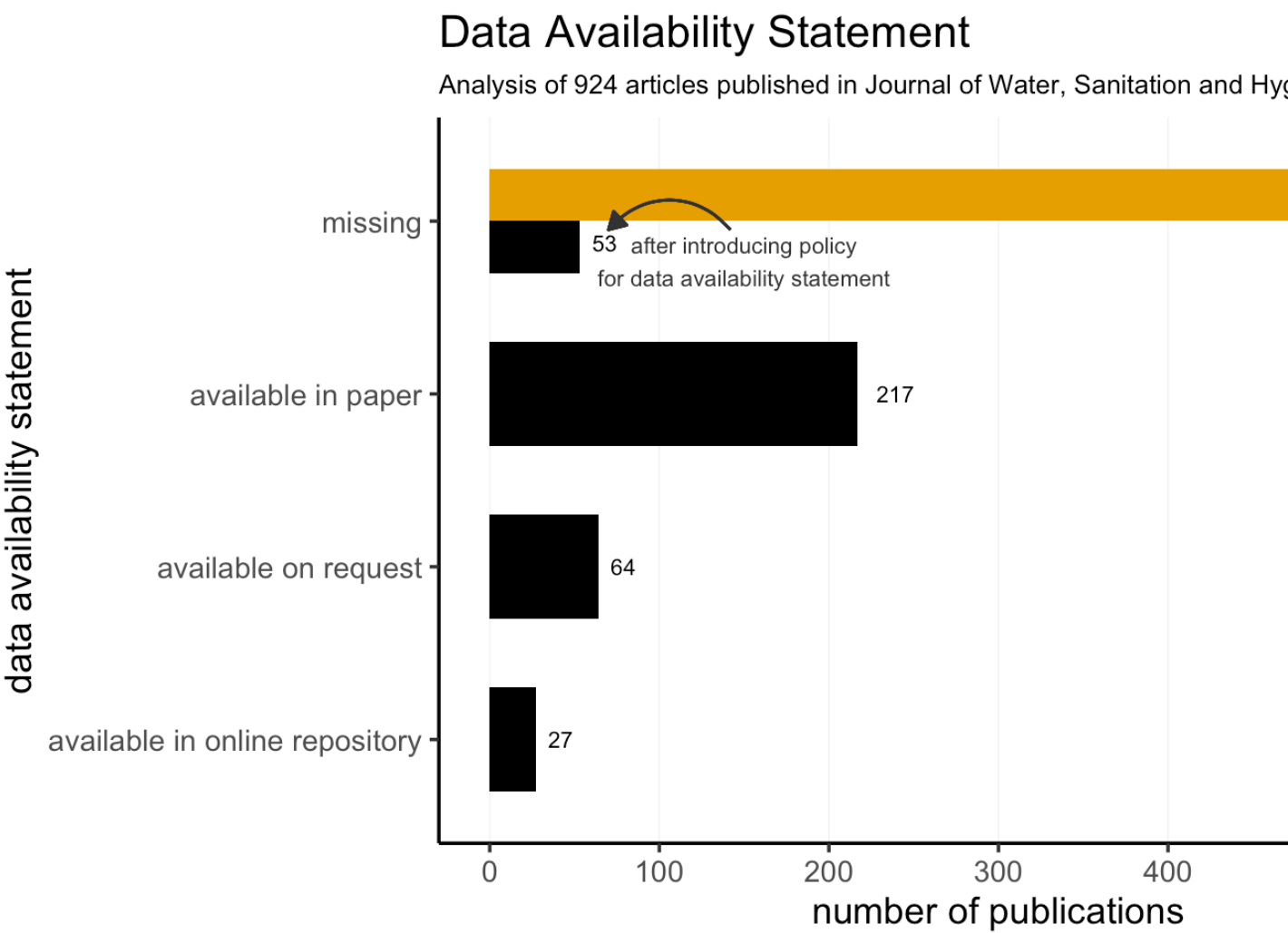
FAIR Data Publishing

Lars Schöbitz

Nov 28, 2024

The Opportunity

Journal articles



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

openwashdata community

Vision

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

Mission

Empower WASH professionals to engage with open data and code.

openwashdata pu

openwashdata.github.io/fsmglobal

fsmglobal 0.0.1 Reference Articles ▼

fsmglobal

This data was first published as part of a journal article by (Greene et al. 2021) and contained the supplemental material as a table in a DOCX file. The following summary table was produced from the data and the code is shown further below.

Demand for faecal sludge emptying services		
summarised for 175 countries		
	population	percent
mechanized	1,030,317,694	25%
no facility	661,998,822	16%

openwashdata academy

- 10-week free data science course to empower professionals to engage with tools and work
- 200 registrations from 46 countries
- 27 datasets submitted as final projects

<input type="checkbox"/>	<input checked="" type="radio"/>	[data] Flood losses and protection measures taken by small businesses	ds4owd	
		#47 opened on May 23 by mianzg		
<input type="checkbox"/>	<input checked="" type="radio"/>	[data] Beneficiaries for different types of WASH technologies from community-based WASH program in Indonesia	ds4owd	tidiness: high
		#46 opened on May 23 by mianzg		
<input type="checkbox"/>	<input checked="" type="radio"/>	[data] Weekly epidemiological reports of cholera cases and death at district level in Malawi	ds4owd	tidiness: mid
		#45 opened on May 23 by mianzg		
<input type="checkbox"/>	<input checked="" type="radio"/>	[data] Laboratory validation of the portable microbial water quality testing kit	ds4owd	tidiness: low
		#44 opened on May 23 by mianzg		
<input type="checkbox"/>	<input checked="" type="radio"/>	[data] Household delivery data on sanitation products from 2019 to 2020 in Cambodia	ds4owd	tidiness: high
		#43 opened on May 23 by mianzg		
<input type="checkbox"/>	<input checked="" type="radio"/>	[data] greenhouse gas data for 22 non-sewered sanitation sites in Canada.	ds4owd	tidiness: high
		#42 opened on May 23 by mianzg		
<input type="checkbox"/>	<input checked="" type="radio"/>	[data] Household Water Insecurity Experiences (HWISE) Scale in Rwanda	ds4owd	tidiness: mid
		#41 opened on May 23 by mianzg		
<input type="checkbox"/>	<input checked="" type="radio"/>	[data] WASH in schools across the various states of Nigeria	ds4owd	tidiness: low
		#40 opened on May 23 by mianzg		

**How can we stream
the data publishing
procedure?**

washr

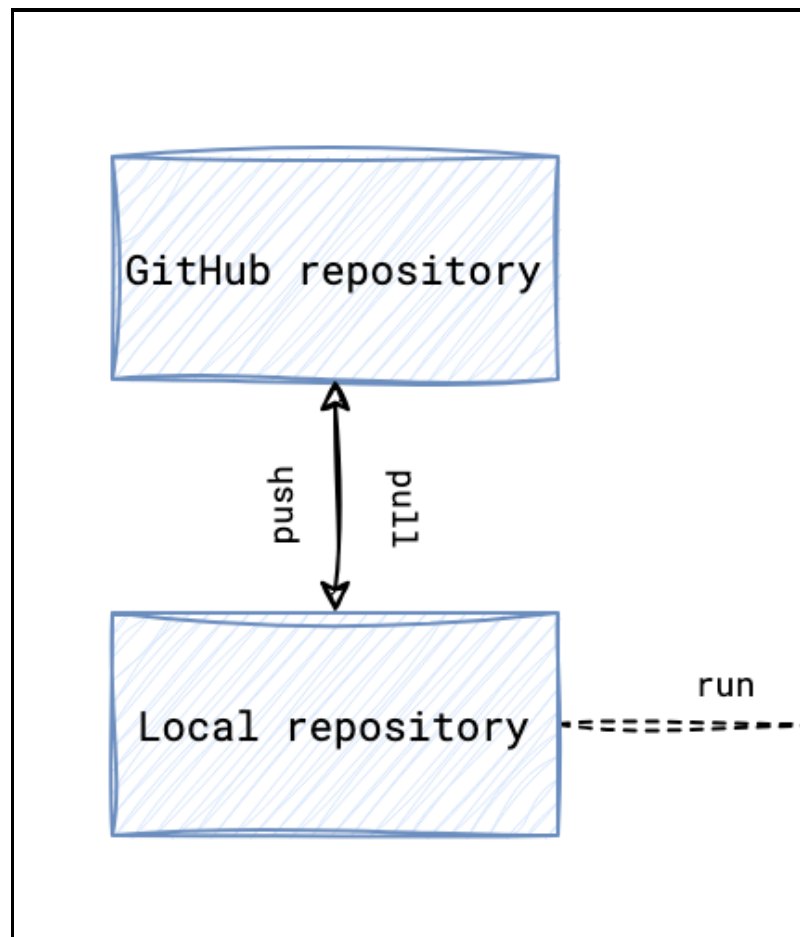
- An R package designed to simplify WASH d
- User-friendly functions to ensure that data principles
- Preparation of a detailed guide and workflow

So far:

- Almost a dozen datasets published
- Requires minimal computational power
- Easily generalizable to benefit the wider com

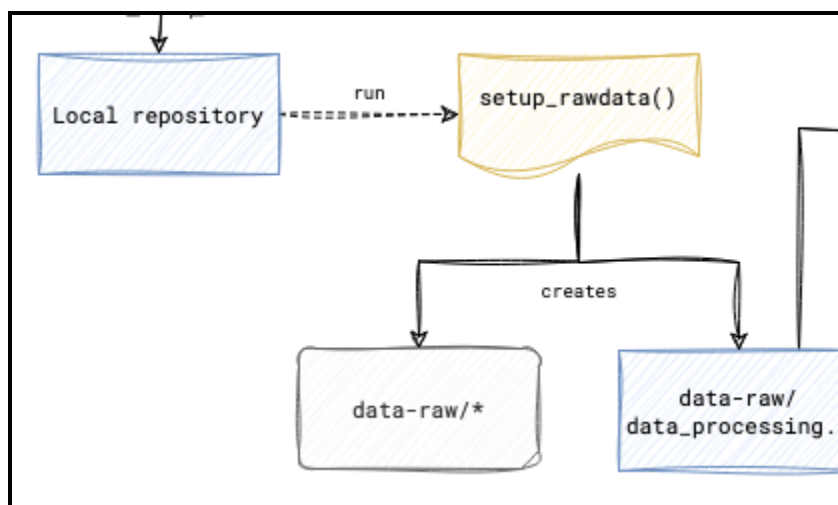
Preparing the data

- Start a local (Posit Cloud for us) version-control connect it to GitHub



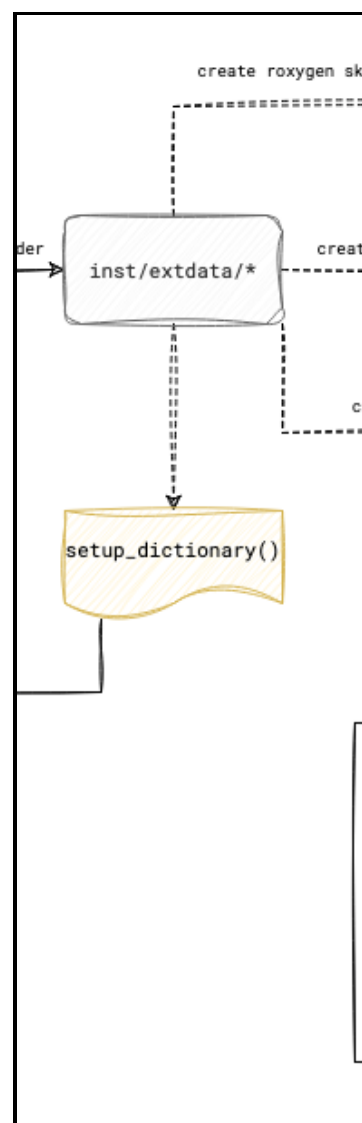
Preparing the data

- `setup_rawdata()`
 - Creates `data-raw` as suggested in `use`
 - Creates `data_processing.R` for data



Documenting the data

- Create roxygen skeletons
- Create README
- Codebook describing each variable
- Website with **pkgdown** R package¹
- Add a license and author(s)



1. See <https://pkdown.r-lib.org/> fairdatapub-washcentre.github.io/website/

Data Publishing G

[global-health-engineering.g](https://global-health-engineering.github.io/ghedatapublishing/)
[ghedatapublishing/](https://global-health-engineering.github.io/ghedatapublishing/)

Data Publishing with **washr**

Welcome

Publishing data can be challenging, especially when adhering to standards of reproducibility. Although technically available, an Excel workbook with multiple tabs, tucked away in an online archive, is far from practical. In other words, it lacks the key principles of being findable, accessible, interoperable, and reproducible — commonly known as FAIR.

This guide aims to provide a detailed walkthrough of how data can be published according to the FAIR principles. It builds on [washr](#), an R package developed for swift data publication. The package emerged from the need to streamline certain steps when publishing the datasets collected during the [openwashdata academy](#).

The guide follows a chronological structure, starting from an empty repository and resulting in the publication of data as a website. [1 Creating a repository](#) introduces version control, guiding readers through setting up both local and remote repositories.

Your turn: Bookmark the guide

1. Navigate to the Data Publishing Guide website engineering.github.io/ghedatapublishing/
2. Bookmark the page and add it to your bookmarks

Thanks!

Links and Downloads

washr source code: <https://github.com/openwashedata/washr>

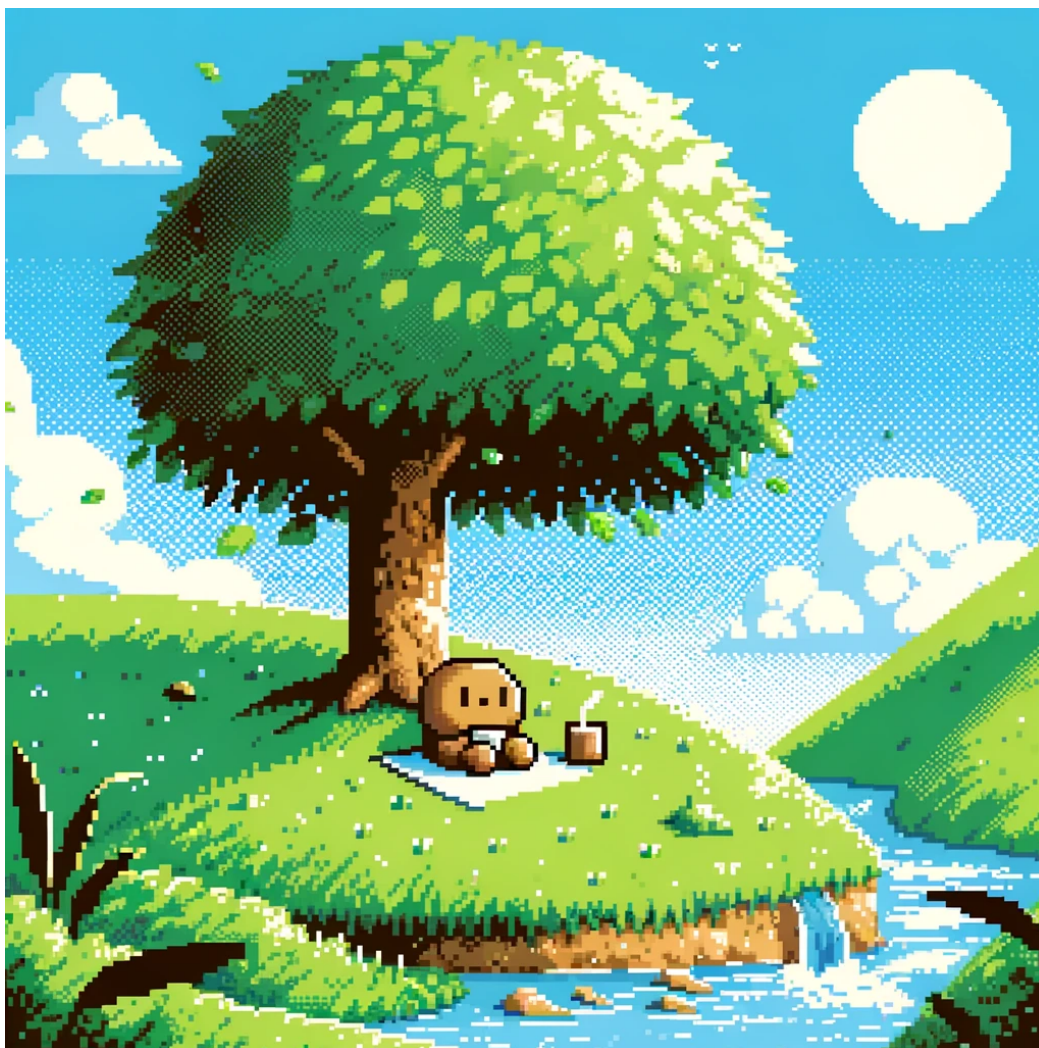
washr guide: <https://global-health-engineering.org/ghedatapublishing>

openwashdata: <https://openwashdata.org>

Download slides as [PDF on GitHub](#)

Take a break

Enjoy your lunch! Let your emails rest in peace



Sign up for the openwashdata r



References

- Greene, Nicola, Sarah Hennessy, Tate W. Rogers, Jocelyn Tsai, and Lars Schöbitz. 2023. "Fsmglobal. Global Faecal Sludge Demand." <https://doi.org/10.5281/zenodo.8208293> .
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Scientific Data Management and Stewardship." *Scientific Data* doi.org/10.1038/sdata.2016.18 .