

openwashdata Data Package Hackathon

Mian Zhong

Global Health Engineering, ETH Zurich

June 21, 2024



Collaborate on this Deploy Preview

Log in to the Netlify Drawer →



Welcome!



Data Package - Raw Data

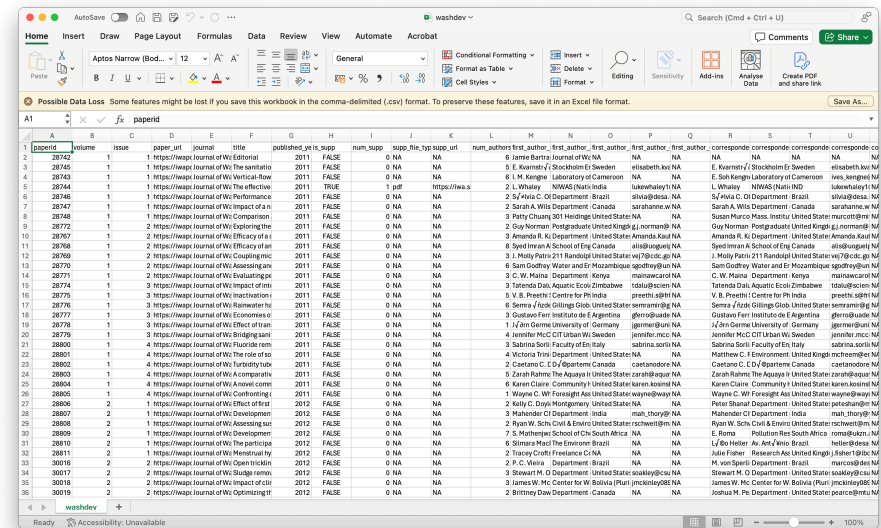
Conceptually

- Raw Dataset(s)

ID	Date	Country
1	21.06.2024	U.S.A.
2	June 22 2024	United States
...

Technically

- Spreadsheets, PDF reports, Hand-written documents



Data Package - Tidy Data

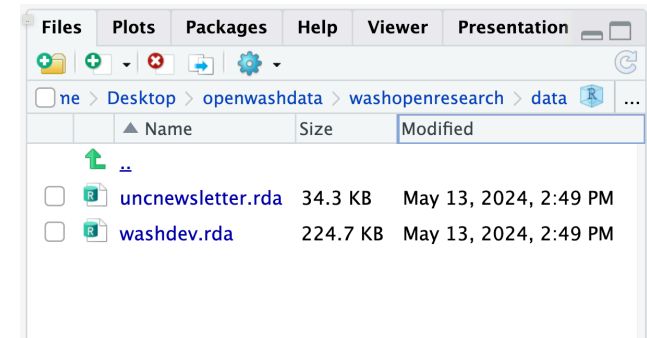
Conceptually

- Tidy/Processed Dataset(s)

id	date	country
1	2024-06-21	United States
2	2024-06-22	United States
...

Technically

- .rda Data Objects



Data Package - Dictionary

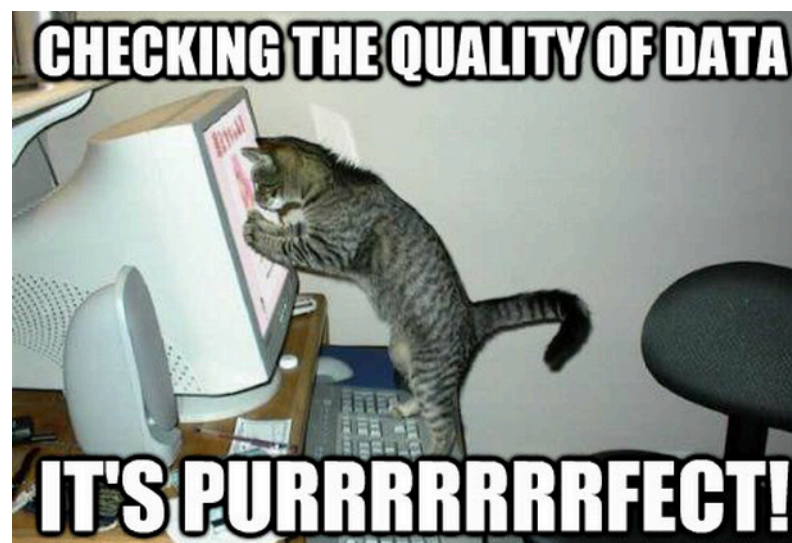
Conceptually

- Dictionary

variable	description
id	The unique ID
date	Date in the format YYYY-MM-DD
country	Full name of the country

Technically

- Dictionary csvfile
- R Roxygen documentation



Data Package - Summary

Conceptually

- Raw Dataset(s)
- Tidy/Processed Dataset(s)
- Dictionary

Technically

- Spreadsheets, PDF reports, Hand-written documents
- .rda Data Objects
- Dictionary csvfile
- R Roxygen documentation
- R package metadata
 - DESCRIPTION
 - CITATION

openwashdata Workflow

- Initialize package repository
- Create dataset
- Document dataset
- Communicate dataset

openwashdata Workflow

- Initialize package repository
 - create GitHub repo
 - create local repo
 - sync local with GitHub
- Create dataset(s)
 - Import raw data
 - Clean data to tidy version
 - Export tidy data
 - Write data dictionary
- Document dataset(s)
 - write package metadata (e.g. authors)
 - document dataset in R
- Communicate Dataset(s)
 - README
 - Website
 - Tutorial/Example

washr Overview

A set of automation tools to set up an openwashdata data package in a consistent structure. It helps to reduce manual entries on tidy data exporting, README writing, pkgdown configuration and content, and etc.

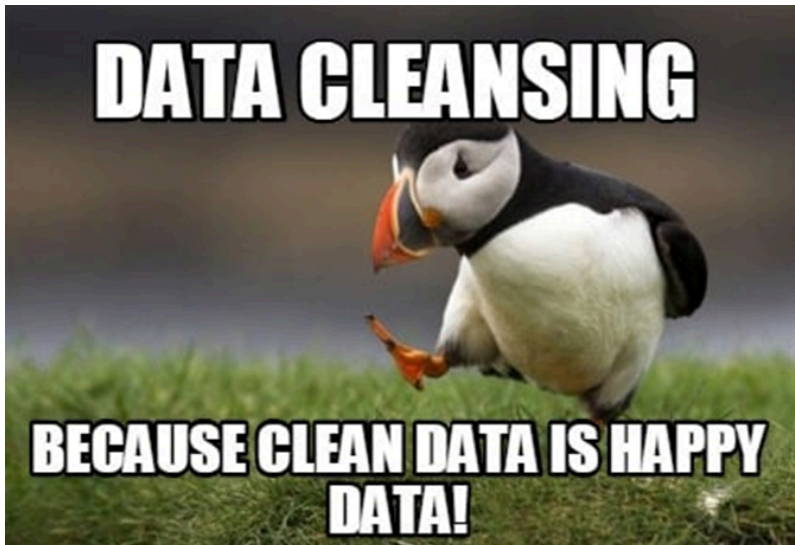
Create Dataset

1. Initialize raw data directory

```
1 setup_rawdata()
```

2. Import raw data into `data-raw`

3. Go to `data-raw/data-processing.R` and happy data cleaning



Document Dataset

1. Update Package Metadata

```
1 update_description()  
2 # add author information(All authors)  
3 usethis::use_author(given = "Mian", family = "Zhong",  
4                     role = c("aut", "cre"), email = "mzhong@ethz.ch",  
5                     comment = c(ORCID = "0009-0009-4546-7890"))  
6 usethis::use_author(given = "Lars", family = "Schöbitz",  
7                     email = "lschoebitz@ethz.ch",  
8                     comment = c(ORCID = "0000-0003-2196-5441"))
```

2. Create a dictionary for the dataset

```
1 setup_dictionary()
```

Document Dataset

3. Go to `data-raw/dictionary.csv`, and fill in the descriptions of the variables
4. Convert the dictionary csvfile to R documentation

```
1 setup_roxygen()
```

Go to `R/` and fill in the title and description for the dataset.

DOCUMENT ALL THE DATASETS



Communicate Dataset

1. Create README and a stand-alone webpage

```
1 setup_readme()
2 # Go to README.Rmd and complete it
3 build_readme()
4 setup_website(has_example = TRUE)
```

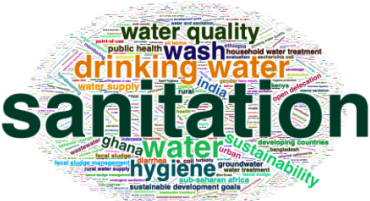
washopenresearch 0.0.1
Reference
Articles
Search for

washopenresearch

https://www.icloud.com

The goal of washopenresearch is to provide an overview of open research data related to Water Sanitation and Hygiene (WASH). The current version contains two datasets from the following sources:

- washdev : Open access journal [Journal of Water, Sanitation and Hygiene for Development](#)
- uncnwslatter : Research section of the newsletter [North Carolina Water News](#)



Installation

You can install the development version of washopenresearch from [GitHub](#) with:

```
# install.packages("devtools")
devtools::install_github("openwashdata/washopenresearch")
```

Links

[Browse source code](#)
[Report a bug](#)
[GitHub repository](#)

Citation

[Citing washopenresearch](#)

Developers

So *who* does the work?

Mian Zhong
Author, maintainer

Ludwig Luz
Author

Lars Schöbitz
Author

Thanks all!

Dev status

License [CC BY 4.0](#)
R-CMD-check [passing](#)
DOI [10.5281/zenodo.11185700](#)

Funding

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

openwashdata.org/pages/gallery/slides/