



# The Future of Our WASH Data Ecosystem

Hosted by WASHWeb  
April 4<sup>th</sup>, 2024



# Welcome!

Thank you for joining our first ever online event

With speakers from:



# Who are we?

- Dynamic, bottom-up initiative
- Community at the intersection of data & WASH
- Facilitating discussions on:
  1. Social Justice
  2. Using Data Better
  3. Systems Thinking
  4. 4Ws: Who's doing what, where & when?
  5. What are your ideas?

# Our vision



# Agenda

- 01 **openwashdata** openwashdata community & future vision
- 02 **mWater** scalable, functional MIS
- 03 **IRC/WASHNote** AI-based learning companion
- 04 **Akvo/USAID WSSH Data & Analytics** intro to the WSSH D&A program
- 05 **Aquaya** making data discoverable with Project W
- 06 **WASHAI** opportunities in data discovery with LLMs

# Q&A in the chat

- Ask your questions in the chat during the call
- We will select and ask a few questions to our presenters at the end

# Next steps

- Join the WASHWeb community on Matrix to continue the conversation with people you meet here today
- Follow us on LinkedIn for updates @WASHWeb



# Let's get started!

# openwashdata

Hosted by WASHWeb  
April 4<sup>th</sup>, 2024

OpenWASHData

# openwashdata

a community effort to bring open data practices to the WASH  
sector

Lars Schöbitz

Global Health Engineering, ETH Zurich

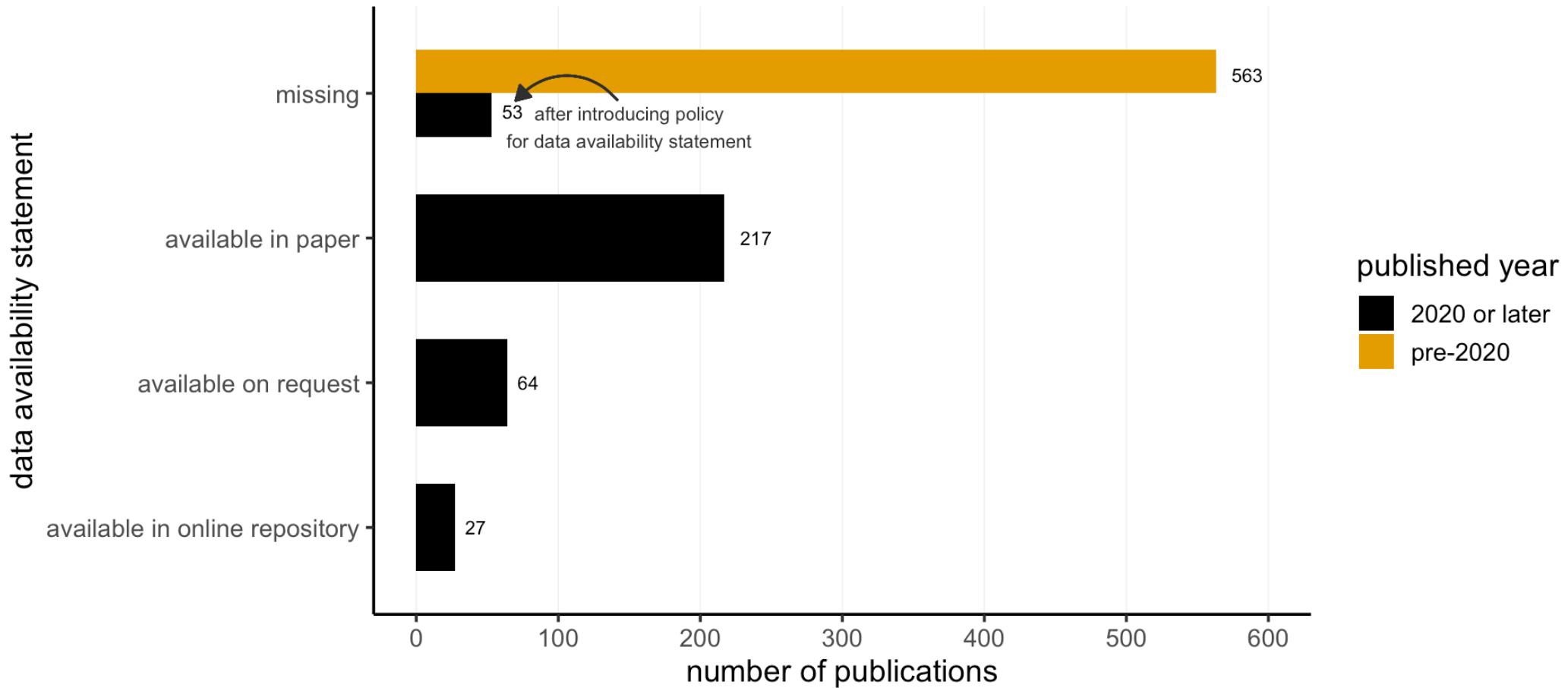
April 4, 2024

# 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

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

**Good practice:** CSV (comma-separated values)

## Supplementary Material

Articles published 2020 or later

file type	n <sup>1</sup>	%
missing	202	51.4
docx	149	37.9
xlsx	24	6.1
pdf	13	3.3
pptx	4	1.0
png	1	0.3

<sup>1</sup> One article can have multiple files.

# PDF reports



## Treatment technologies in practice

On-the-ground experiences of faecal sludge  
and wastewater treatment

**SNV**  **UTS** Institute for Sustainable Futures

# PDF reports

Table 2. Influent and effluent qualities of wastewater treated at Duri Kosambi FSTP plant in 2019, as compared to effluent standards

Parameter	Inlet	Outlet
pH	6, 45-7, 88 pH	7, 12-7, 61 pH
Total suspended solids, TSS	340-8933, 33 mg/L	22, 5-84, 29 mg/L
Biochemical oxygen demand, BOD <sub>5</sub>	106, 38-646, 82 mg/L	2, 76-69, 79 mg/L
Chemical oxygen demand, COD	687, 9-2780, 37 mg/L	41, 25-127, 67 mg/L
Total organic matter, KMnO <sub>4</sub>	108, 04-568, 72 mg/L	54, 21-150, 50 mg/L
Ammonia, NH <sub>3</sub> -N	108, 75-239, 25 mg/L	0, 45-29, 81 mg/L
Methylene blue active surfactant, MBAS	0, 74-2, 69 mg/L	0, 13-0, 78 mg/L

# openwashdata community

# openwashdata community

## 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.

# openwashdata publishing

# [openwashdata.github.io/fsmglobal/](https://openwashdata.github.io/fsmglobal/)

fsmglobal 0.0.1 Reference Articles ▾

Search for

## fsmglobal

This data was first published as part of a journal article by (Greene et al. 2021) and contained in 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
<b>mechanized</b>	1,030,317,694	25%
<b>no facility</b>	661,998,822	16%

### Links

[GitHub repository](#)

### Citation

[Citing fsmglobal](#)

### Developers

So who does the work?

Nicola Greene

Author 

Sarah Hennessy

Author 

Tate W. Rogers

Author 

Jocelyn Tsai

# openwashdata academy

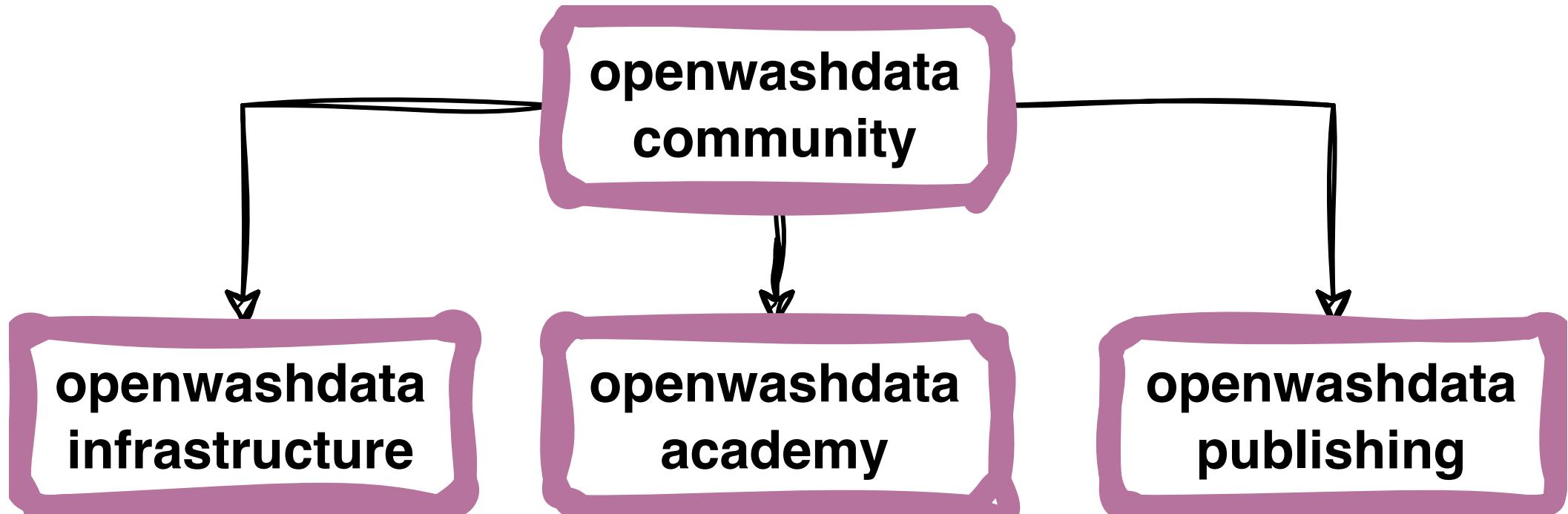
# data science for openwashdata 001

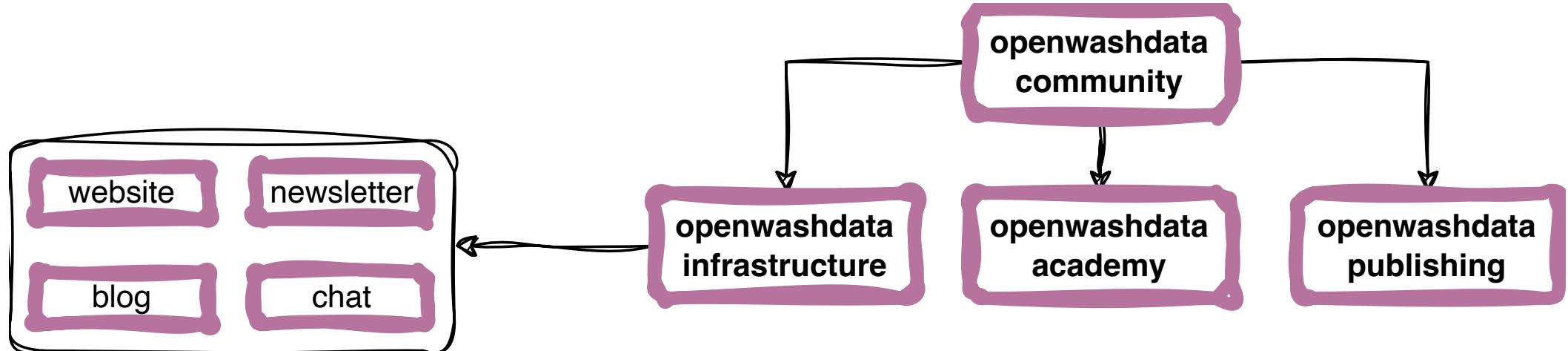
- free, live, online, 10-week programme
- 200 registrations
- 100 show-ups
- 40 graduates
- next iteration: September/October 2024
- sign-up: <https://forms.gle/MP5rNYZagBdfG2ZRA>

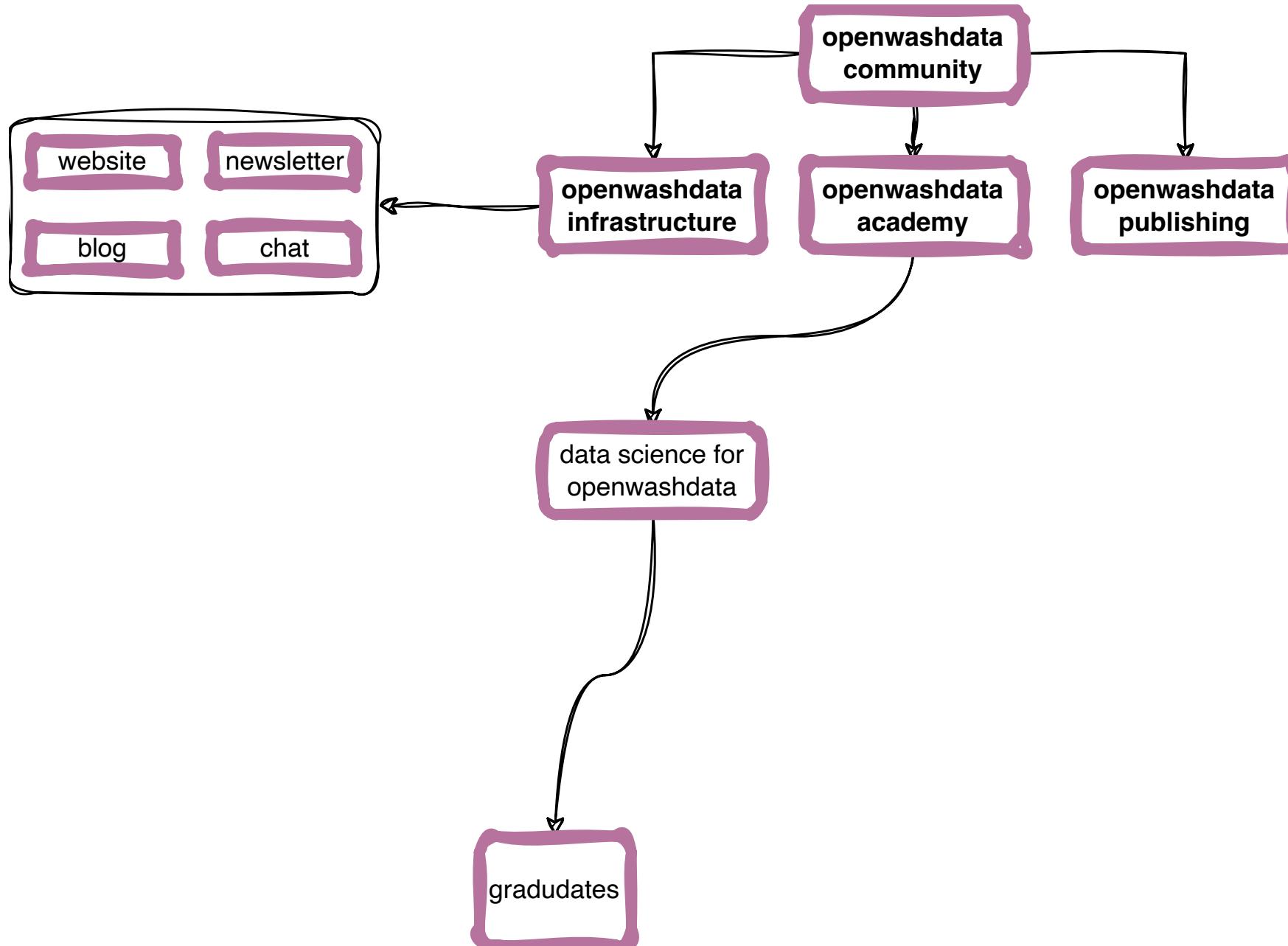
[ds4owd-001.github.io/website/](https://ds4owd-001.github.io/website/)

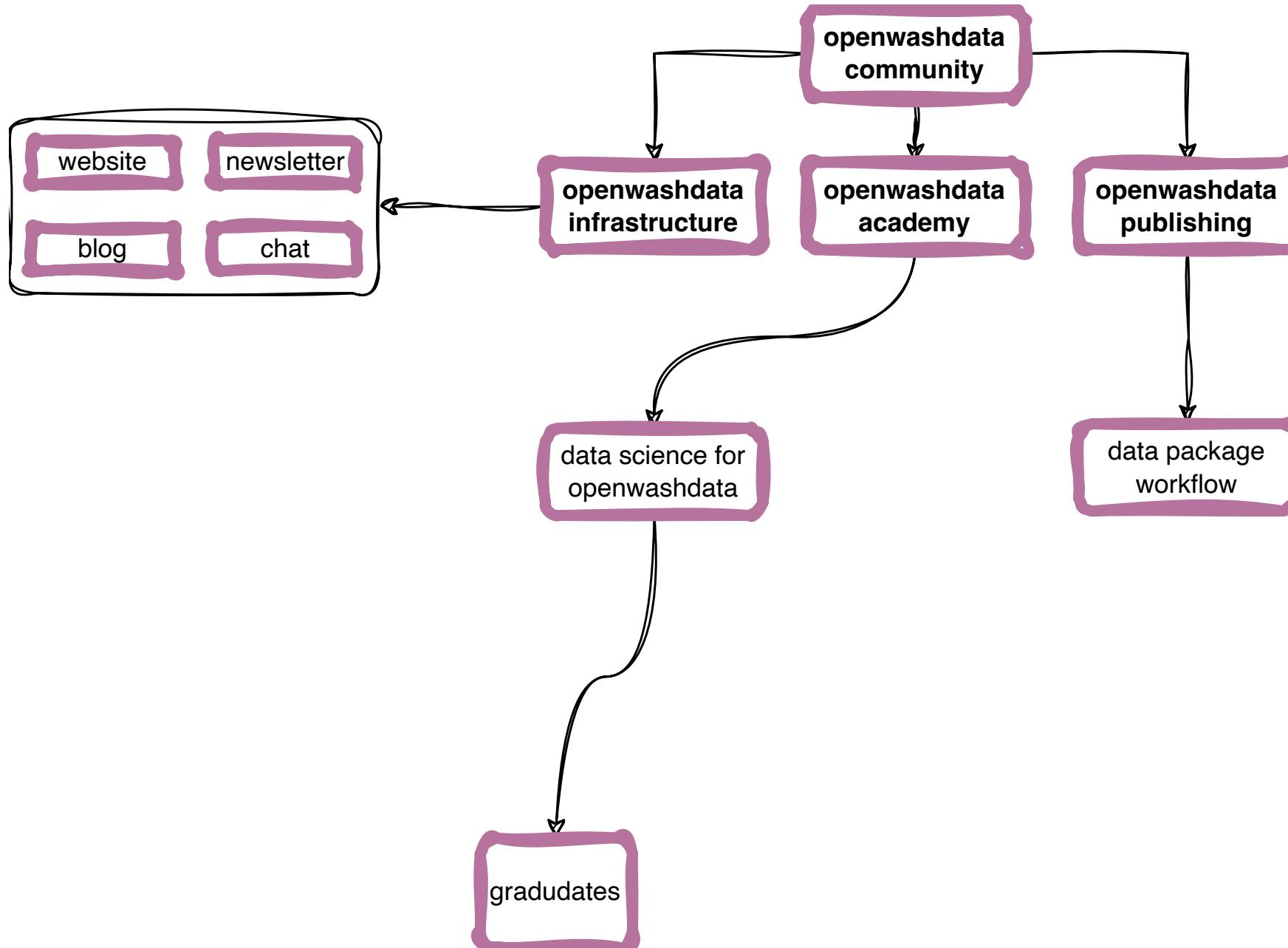
# what's next

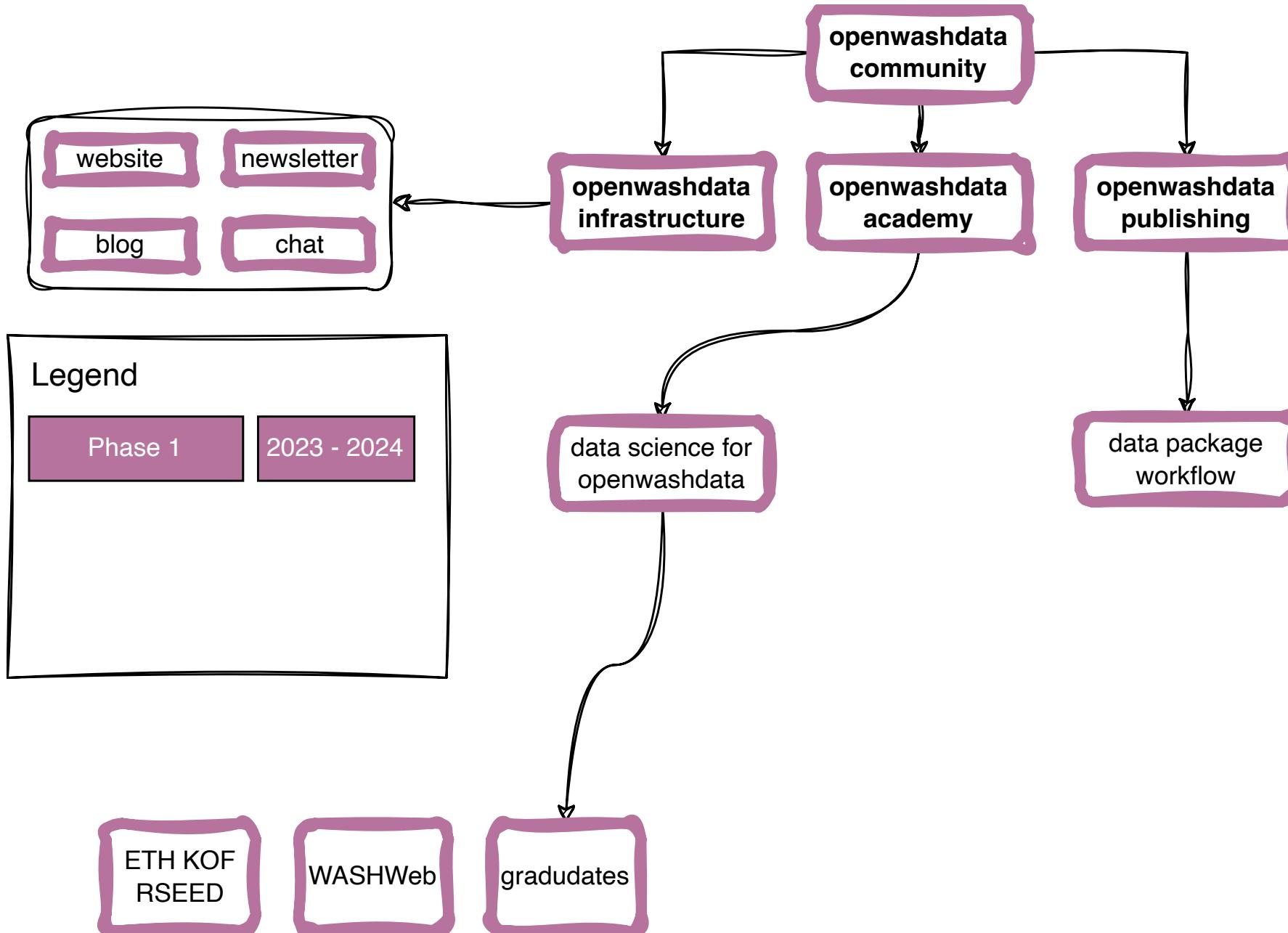
 [openwashdata.org/pages/gallery/slides/](http://openwashdata.org/pages/gallery/slides/)

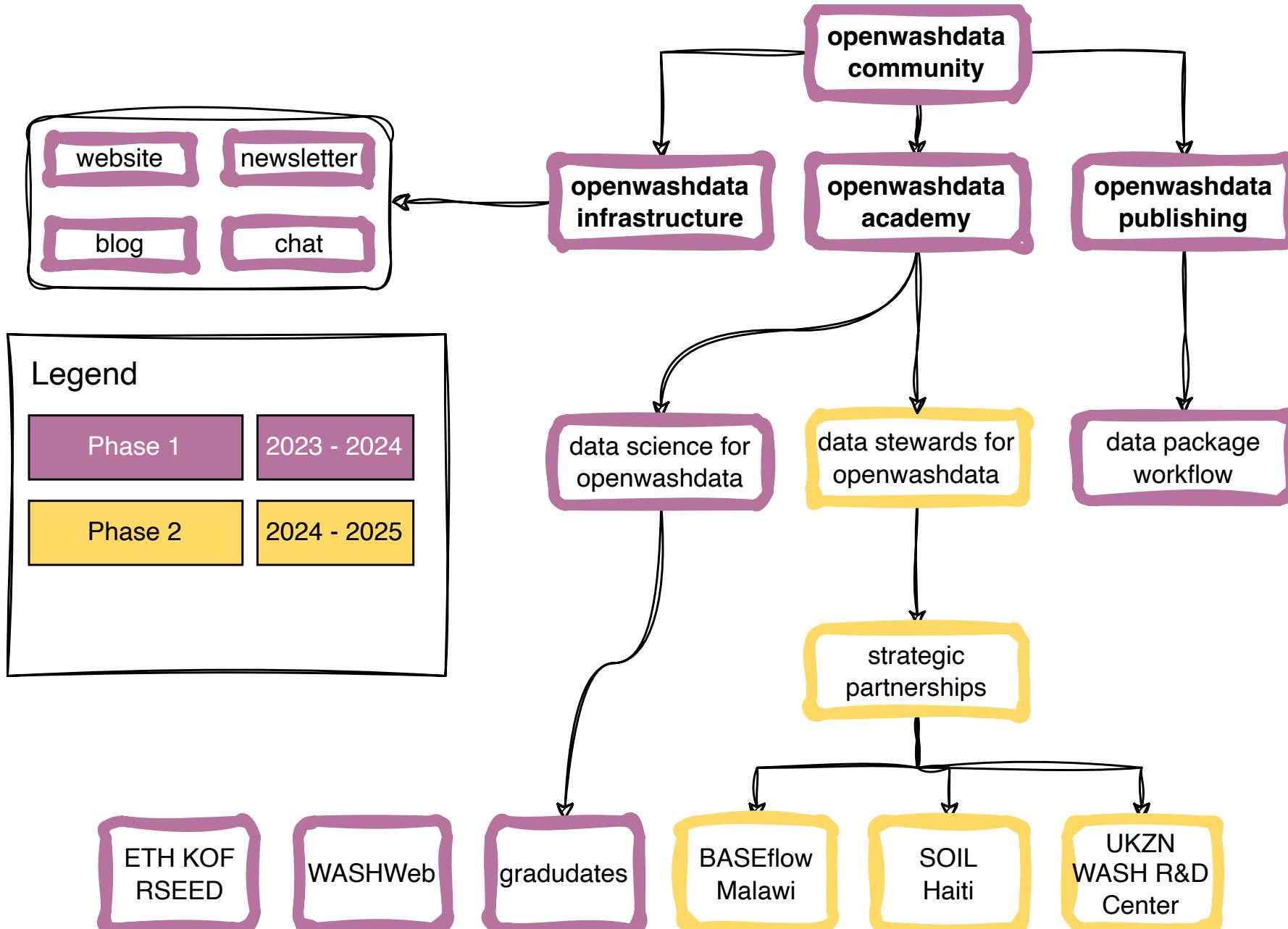


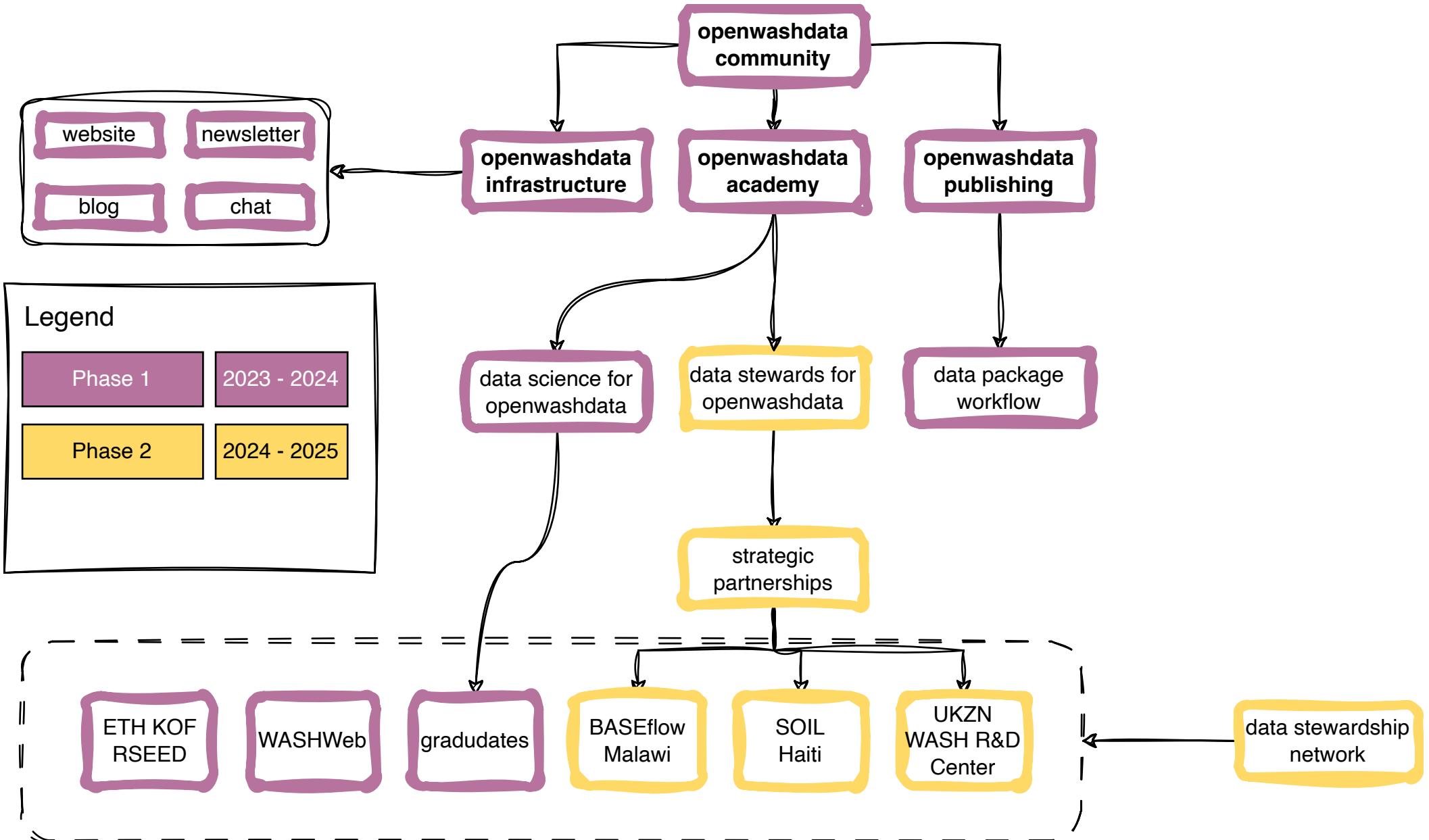


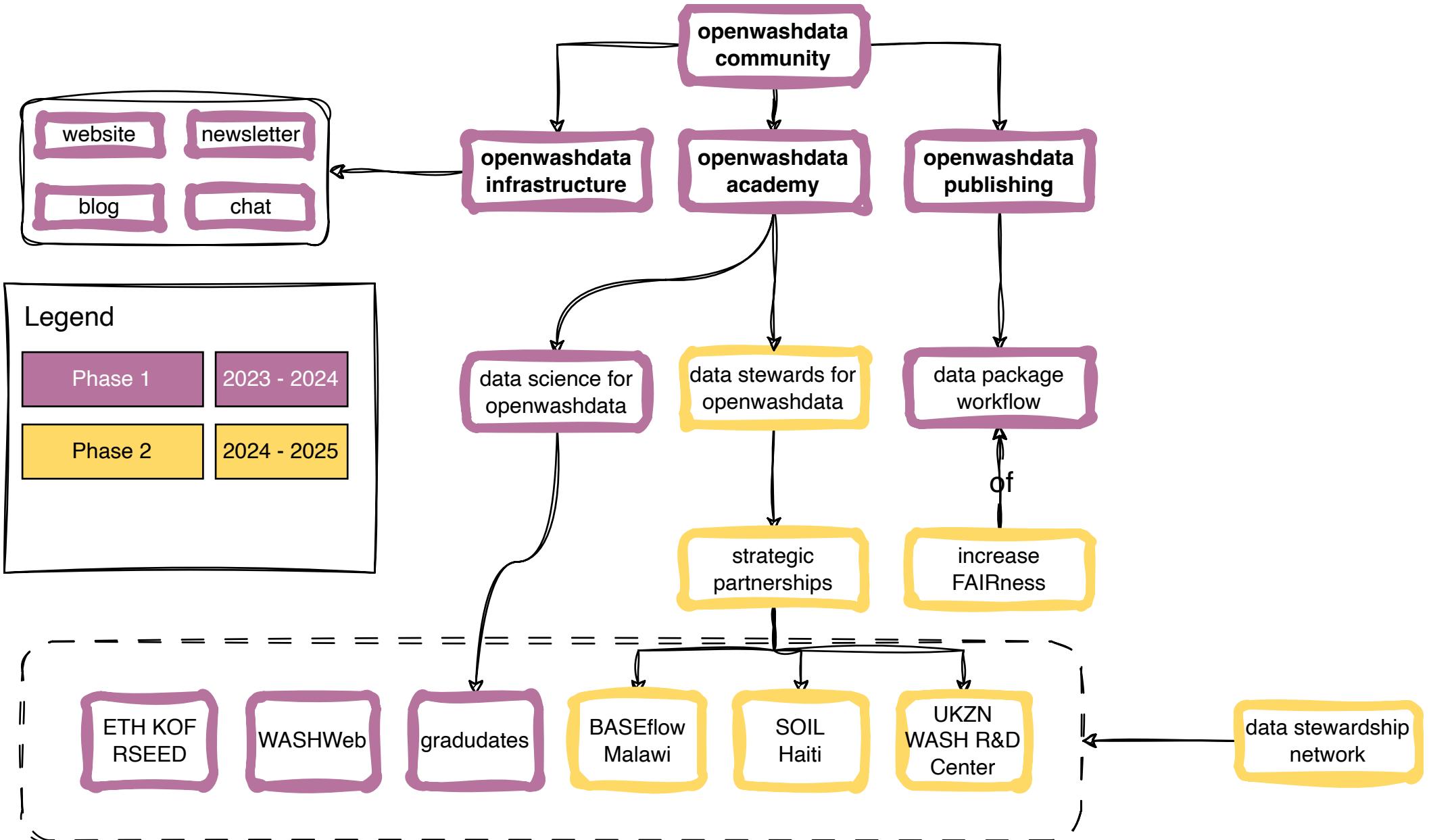


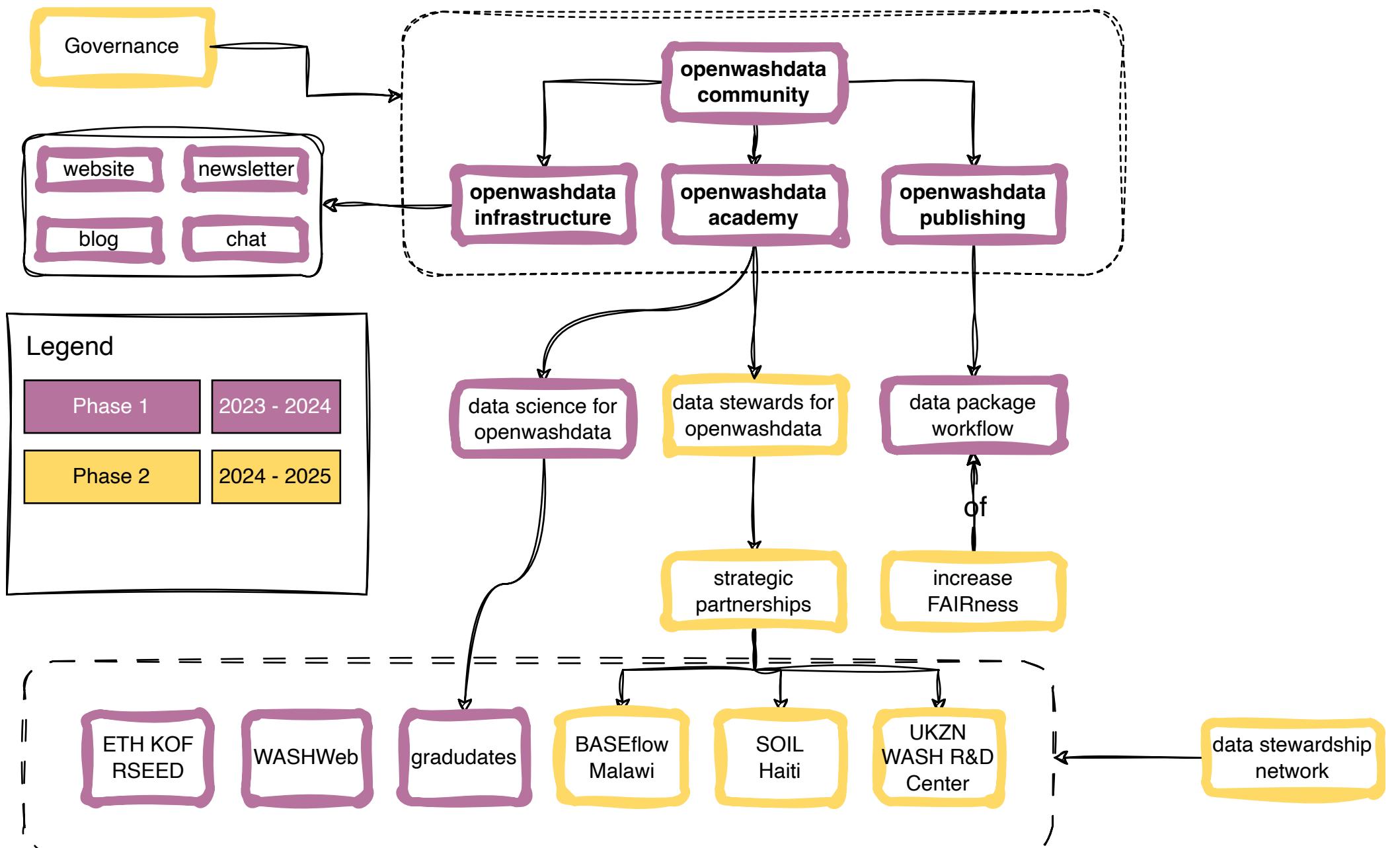


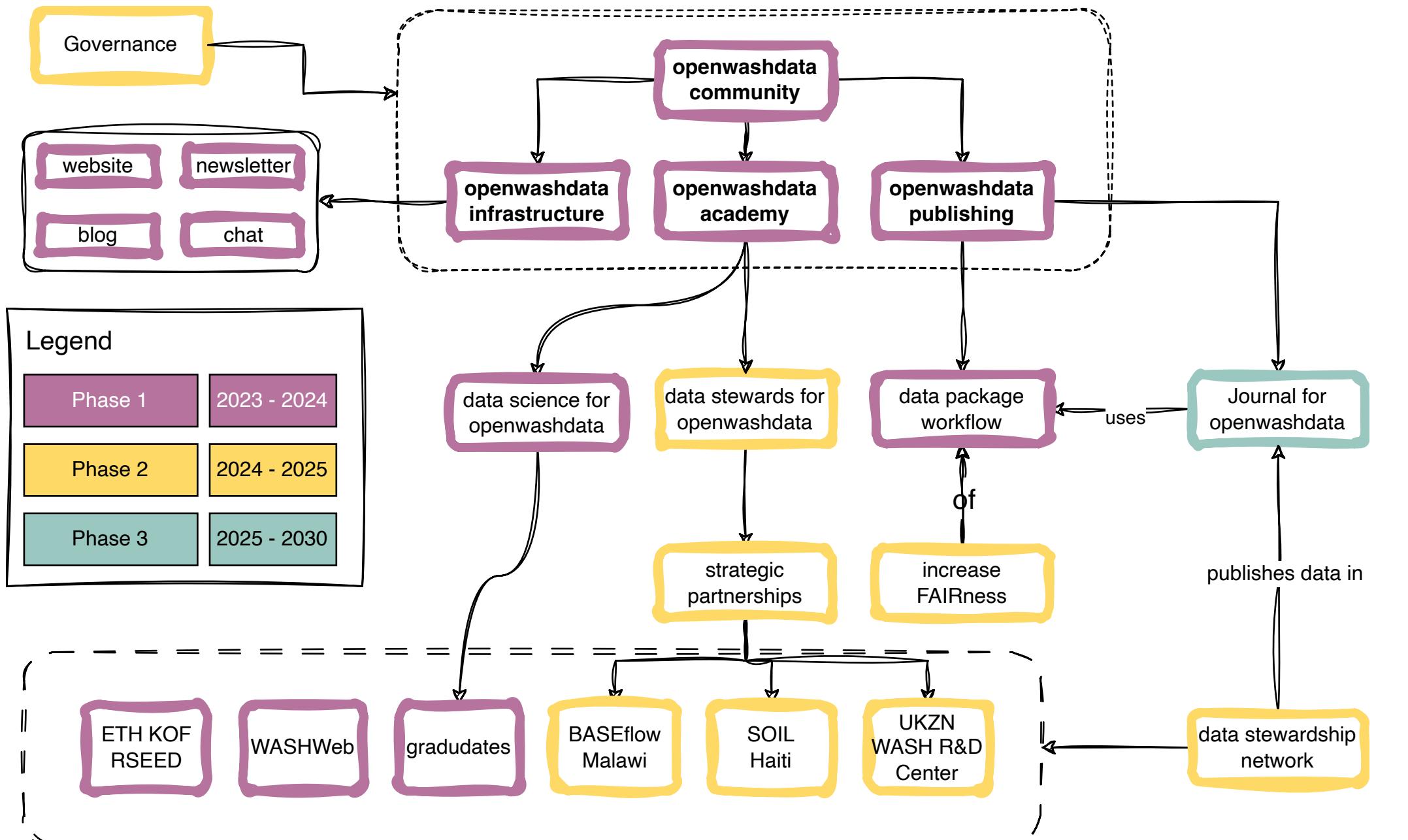












# News

# Sign up to our newsletter



<https://buttondown.email/openwashdata>

Your first name\*

Your location

Your email (you@example.com)

Subscribe

# 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

- Greene, Nicola, Sarah Hennessy, Tate W. Rogers, Jocelyn Tsai, Francis L. de los Reyes III, and Lars Schöbitz. 2023. “Fsmglobal. Global Faecal Sludge Emptying Services Demand.” <https://doi.org/10.5281/zenodo.8208293>.
- Soeters, S, P Mukheibir, and J Willetts. 2021. “Treatment Technologies in Practice: On-the-Ground Experiences of Faecal Sludge and Wastewater Treatment.”
- 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>.

# mWater

Hosted by WASHWeb  
April 4<sup>th</sup>, 2024



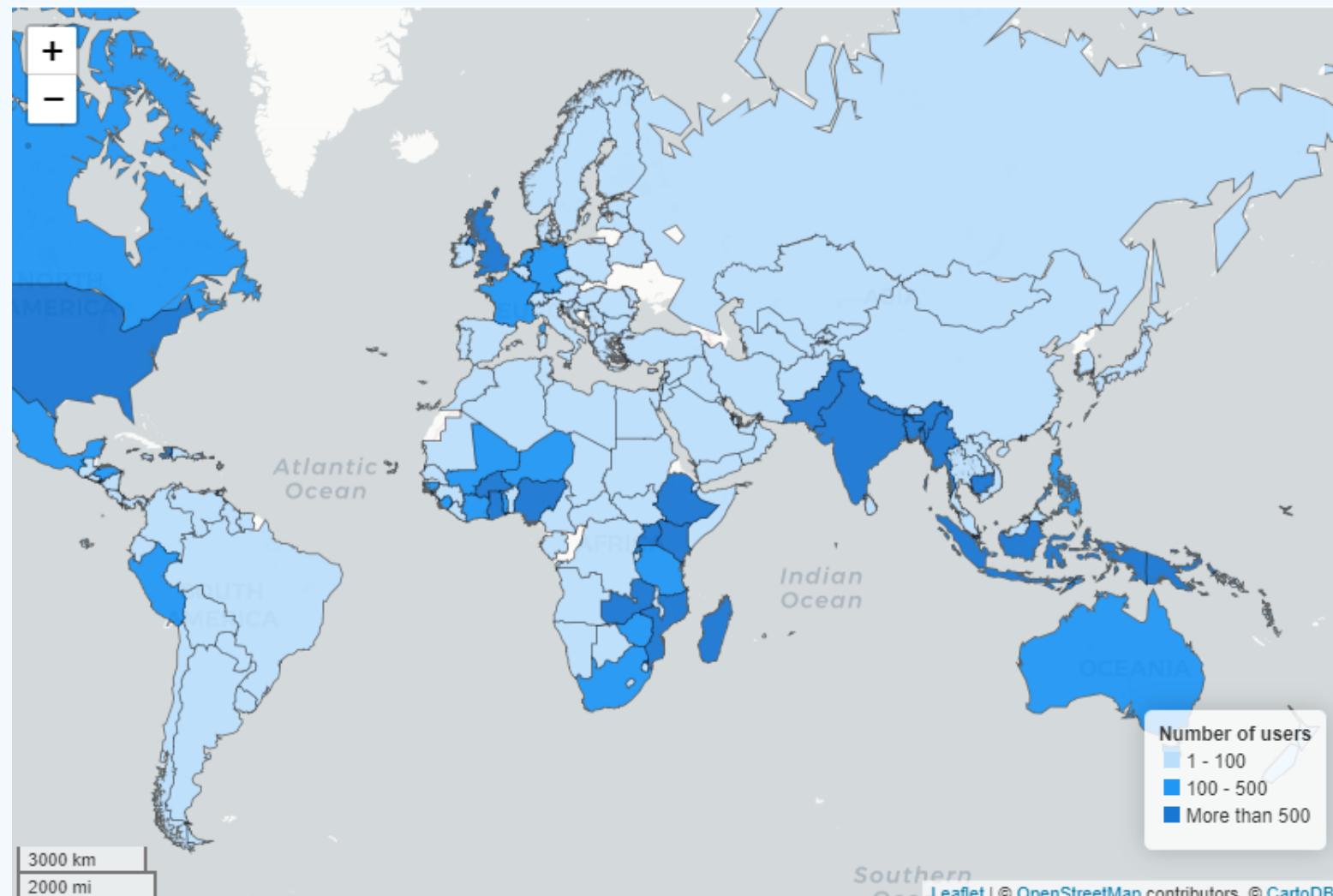
# Data-driven management in mWater

Petri Autio - Head of Product



mWater 2024

- The leading **free** solution for WASH data collection and management
- Collect, manage, analyze, and share data that you control



Users by country

- Empowering front-line workers to make decisions
- Scales from households & water points to water systems, city-wide utilities, and country-level monitoring
- Upward reporting easy, but secondary

## Key stats

- 250,000 users in 200 countries
- 750k surveys per month
- 60k Portal pageviews per month



# MWATER MIS EXAMPLES

## Water utilities

Each water system gets its own homepage in mWater

- 1) Map your assets in mWater
- 2) Immediately get a console of KPIs to see how you're performing now
- 3) Plan ahead to make your utility more efficient and investment-worthy

## National monitoring systems - Example: RUMIS.ug

Different sector actors can focus on their own work, and opt into contributing to national monitoring without duplicating effort

There's always some map...

Example from Taita Taveta  
county in Kenya for USAID STAWI



# UTILITY LEVEL - MWATER ASSET STANDARD

Asset classes:									
System	Facility	Vertical			Horizontal		Natural		
Water system	Water facility	Source	Pump	Tank	Power	Pipe	Canal	Reservoir	River or stream
		Treatment	Meter	Electrical	Valve			Aquifer	Riparian zone
		Hydrant	Junction	Sampling point	Sensor			Infiltration zone	Forest
		Analyzer	Structure	Water point	Other			Wetland	Water-shed

- Standard asset types and data fields allow you to start mapping assets right away
- Use surveys and link them to assets to collect any other data you might need
- Interoperable data prevents vendor lock-in

# UTILITY LEVEL - GLOBAL VIEW OF UTILITIES

mWater Portal    localhost:3002/#/water\_systems

mWater Sites Surveys Visualizations Manage Indicators Help 68 grassick

**Water Assets** - Assets are resources that are owned or controlled by a water system. They can include components such as pipes, pumps, and tanks.

+ Add Water System Search...

Jarry Water System - 26371731	22 Assets
CTE des Cayes - 8729617	4711 Assets
Berase - 4443935	5 Assets
Ayisakro - 6441708	26 Assets
fhghfg - 565916765	

Training woreda/utility - 485764835	11 Assets
Water reserved tank - 9638934	
146566134	
30518049	
321299710	
321299741	
321299758	
Nayu - 607430675	
Anemi Water System - 544286012	
Tnrite Gravity Fed Water System - 544286029	
Mafizi Primary Rain Water System - 479294498	
Gwata Rain Water System - 479294247	
Sangwe Primary Rain Water System - 479294395	
Mitengwe Primary Rain Water System - 479294515	
Kibwemwenda Primary Rain Water System - 479294412	
Wnaluhifera - 12566292	

My Organizations Systems All Systems

Microsoft Bing

Copyright © 2022 Microsoft and its suppliers.

# UTILITY LEVEL - WATER SYSTEM PAGE

mWater    Sites ▾    Surveys    Visualizations ▾    Manage ▾    Indicators    Help    50    Petri ▾

← Water System - PDAM Salatiga

**Water System**

Water system managed by PDAM Salatiga (private ⓘ)

**Location**  
Location: BLOTONGAN, SIDOREJO, Salatiga, Jawa Tengah, Indonesia  
Urban or rural classification: Urban  
Unique ID: 604040718

**Photos:**  

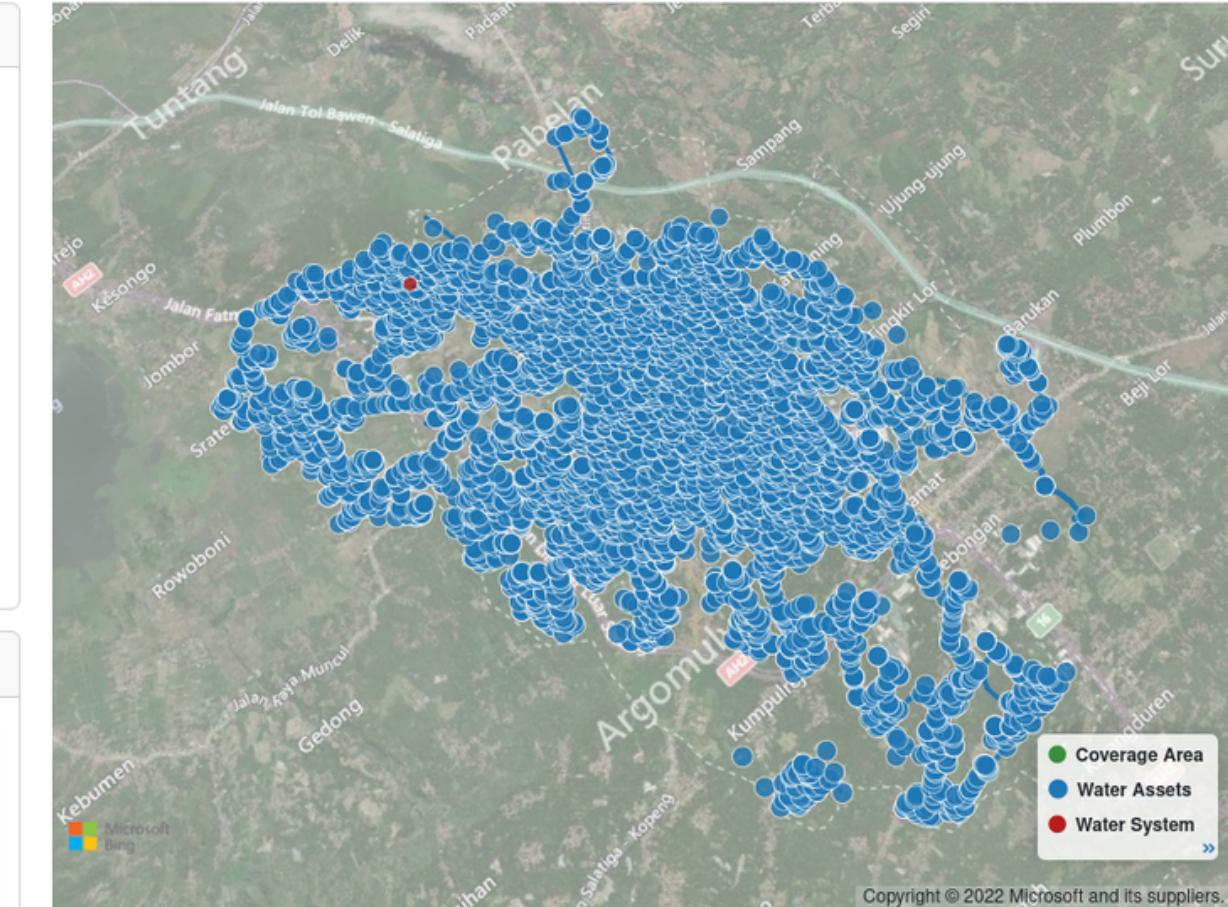

**Attributes**  
Population served: 115990

**Assets**    Go to Asset Registry

Asset Type	Count
Connection	37269
Hydrant	8
Junction	4749
Meter	65
Other vertical	4
Pipe	2788
Pump	1
Source	46
Subsystem	7
Tank	17
Valve	373
Water facility or building	1

**Edit Water System**

**Configure Page**



Copyright © 2022 Microsoft and its suppliers.

Salatiga water utility we are supporting as part of USAID Tangguh

mWater 2024

# UTILITY LEVEL - KPI CONSOLE

“In our evaluations reporting goes from something that takes a month to something that takes a moment.”

mWater Sites Surveys Visualizations Manage Indicators Help Petri

← Water System - PDAM Salatiga

Water System Asset Overview Asset Registry Customers Meters Accounting Issues

Water system summary Asset details Maintenance Financials Asset register System map Create As Console Fullscreen Print Refresh Export as PDF

### Water System - PDAM Salatiga - Asset details

This page displays detailed information for key asset types of the system. Additional assets type can be added to the page as needed.

Contents Sources Pumps Tanks Pipes Power Treatments Water points

#### Sources (46 total)

Name	Unique ID	Source type	Source meter installed	Source pressure (bar)	Nominal flow (m³/s)	Status	Condition	Model	Serial number	Installation date	Rehabilitation date	Last service date	Photo
Ngemplak 2	604771278	Borehole								Jan 1, 2016			
Kalilanang	604771429	Spring											
Warak	604771481	Borehole											
Tegalsari	604771883	Borehole											
Kaligethek	604770576	Spring											
Noborejo	604770583	Borehole											
Ngronggo	604770590	Borehole											
Randuacir	604770600	Borehole											
Kecandran	604770617	Borehole											
Sukowati	604770624	Borehole											
Mata Air	604770631	Spring											
Senjoyo													

Jan 1, 2007

#### Pumps (1 total)

Sources by type

Type	Count
Intake	1
Spring	10
Borehole	34
Dug well	0
Rainwater catchment	0
Other	1

Configure Page

- Asset details
- Maintenance data
- Financial information
- Complete asset register
- Full-page map

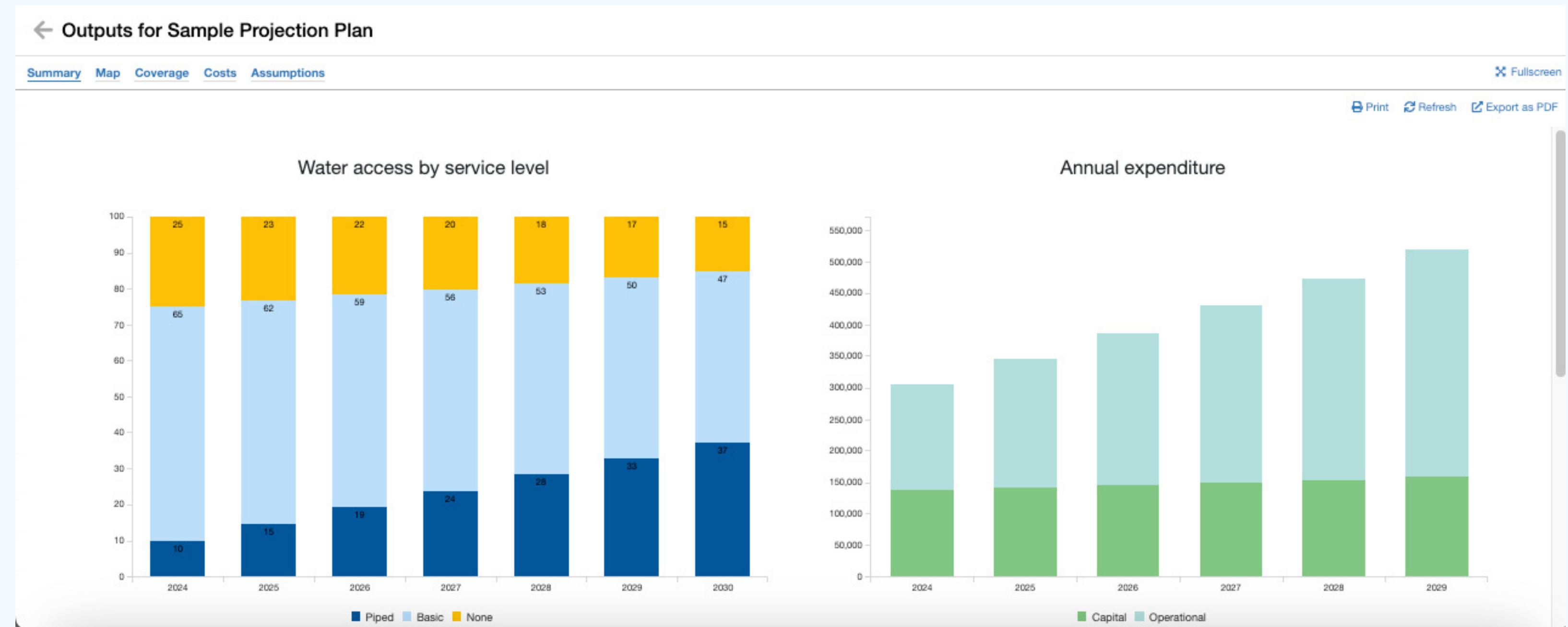
Reporting to national standards can take a long time.

We can save a lot of time on that.

E.g. in Kenya water scheme operators can rapidly get to the point where they report to WASREB

# UTILITY LEVEL - MODULES

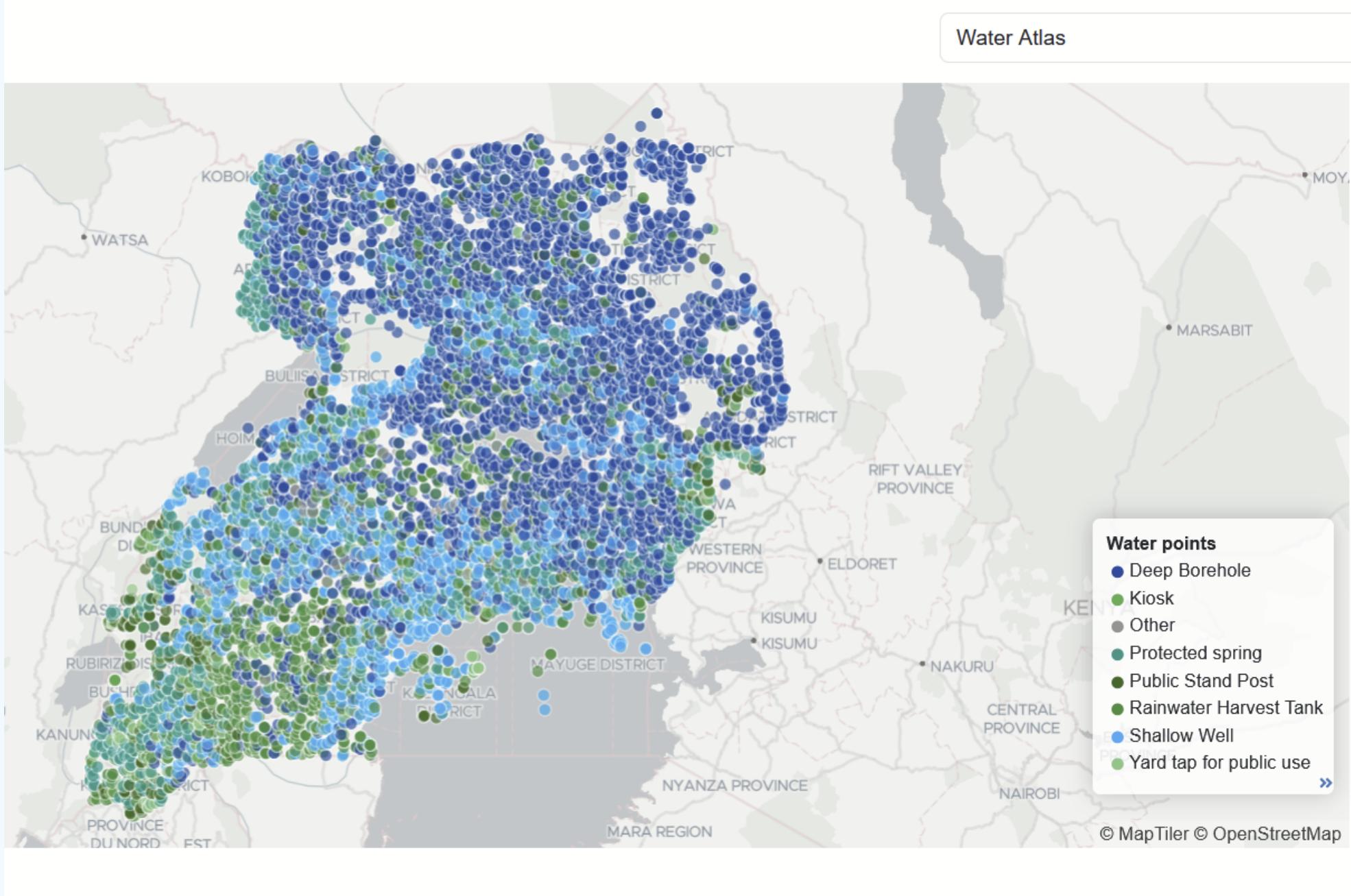
Once your data is in, you'll be able to get to data-driven decisions using modules we are releasing.  
The planning tool will help utilities justify their investment-worthiness.



Also customer management, meter reading, accounting, issue management coming up.

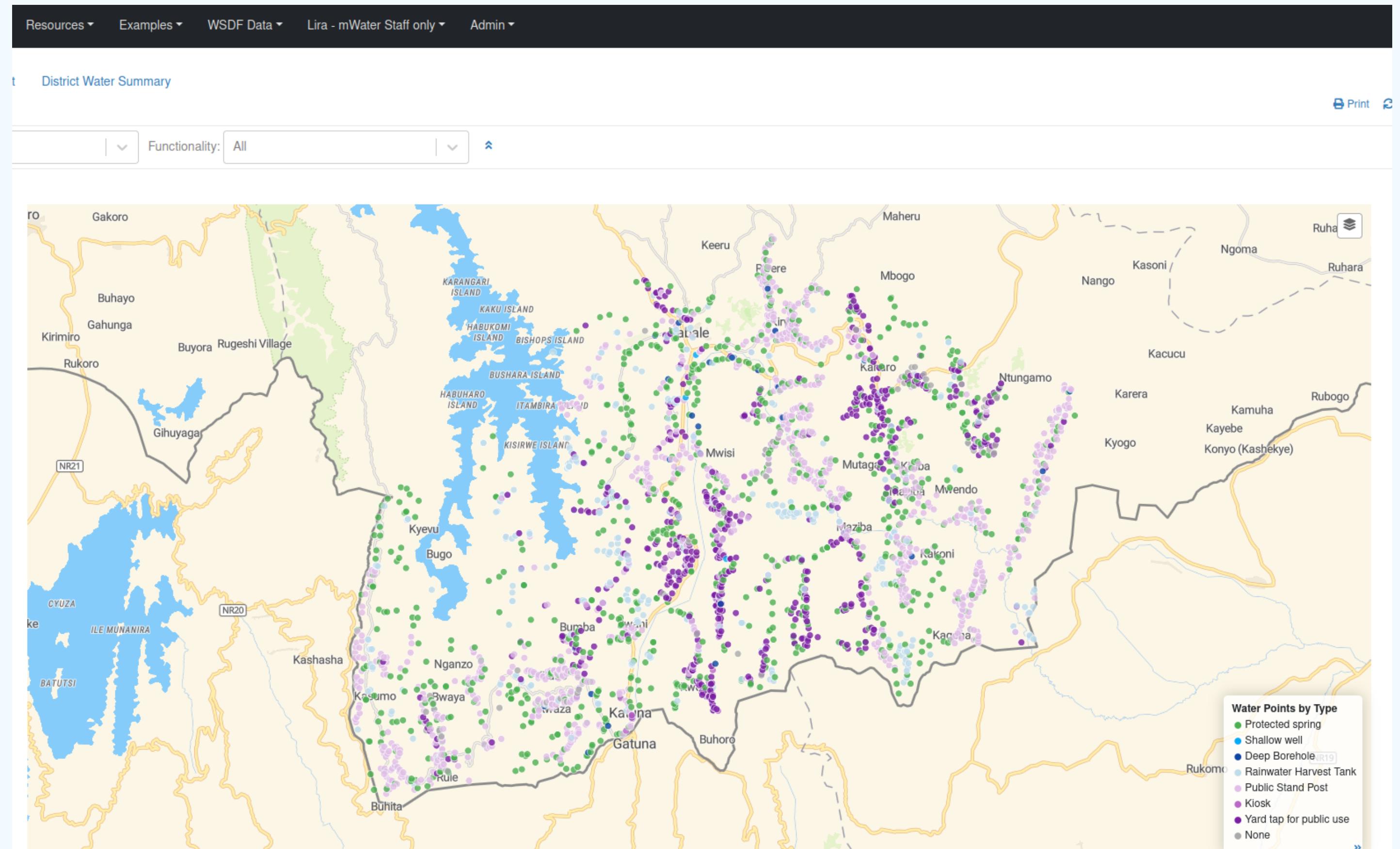
# RUMIS.UG - COUNTRY LEVEL MONITORING

Uganda Real Time Water Supply Atlas in mWater



- Imported Water Atlas
- District level updates
- Organizations can use mWater for their own work, complete the nationally mandated forms and opt into the reporting platform

# RUMIS.UG - DISTRICT LEVEL

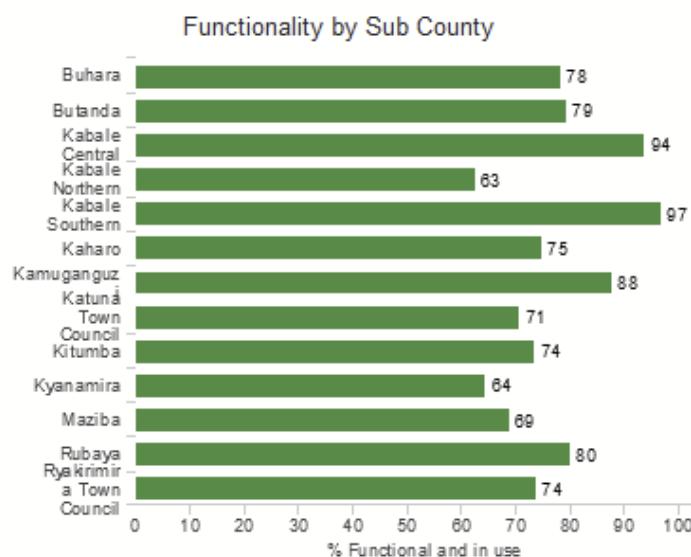
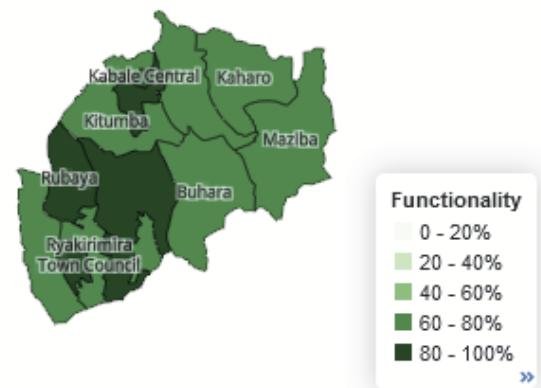


# RUMIS.UG - DISTRICT LEVEL

## KABALE District

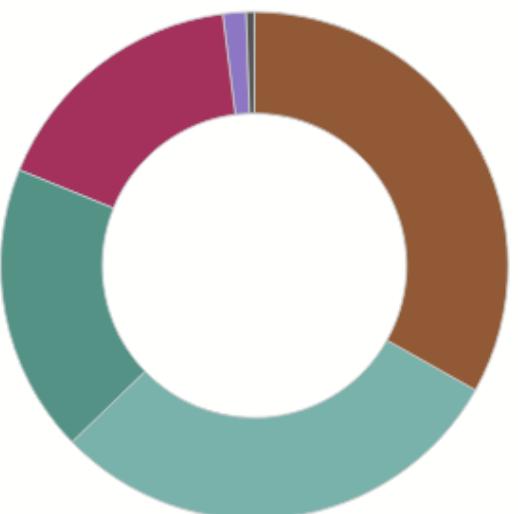
Has a total of 3,589 water points.

### Functionality of Point Sources

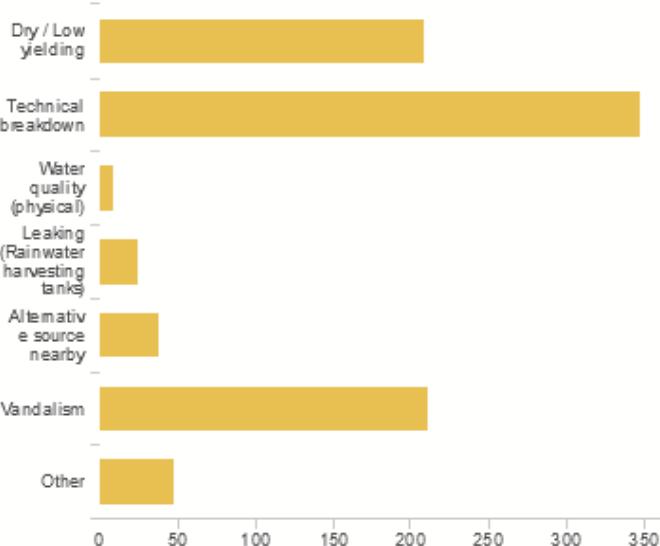


Note: Functionality calculations exclude water points marked as "Recommended for Decommissioning."

### Point sources by technology



### Reasons for Non-Functionality



## Real-time situational awareness:

- Functionality
- Technology types
- Funding sources
- Types of management

# EXAMPLES OF MWATER BASED MIS FROM DIFFERENT COUNTRIES:

**Haiti**  
National water and sanitation monitoring platform

Système d'Information sur l'Eau Potable et l'Assainissement (SIEPA)

DINEPA

Taux d'accès moyen à l'eau national\*: 55%

- Taux d'accès des Communes: 68%
- Urbaines: 68%
- Taux d'accès des Communes Rurales: 48%

Taux d'accès par département

Département	Taux d'accès (%)
Artibonite	51%
Centre	67%
Grande-Anse	30%
Nippes	53%
Nord	71%
Nord-Est	71%
Nord-Ouest	45%
Ouest	56%
Sud	54%
Sud-Est	38%

50 km

**Malawi**  
National water asset inventory

District Improved water point

- Functional
- Partially functional but in need of repair
- Not functional

**Ethiopia**  
Regional Water Asset Management System

Latest functional status by zone

Zone	Total systems	% Functional	% with data	Data age (days)
1 Awsi Rasu	148	77%	61%	989
2 Kilbatti Rasu	45	87%	87%	968
3 Gabt				
4 Fanti				

Click on a row in the table to view the systems in that zone

View definitions

Zone Total systems % Functional % with data Data age (days)

**Guinea Bissau**  
National water point monitoring system

Guinea-Bissau National Water Point Monitoring Dash...

unicef

Guinea-Bissau Water Point Monitoring Overview

Total Water Points: 12,420

Water points visited: 6,858

Functionality of visited water points: 75%

Functionality of visited water points

50 km

**Papua New Guinea**  
National WASH MIS / RapidWASH (with WaterAid)

RapidWASH Baseline PNG

Overview Projects About English John Feighery Settings

Indicators Communities Households Data Approval

Select a community: All Show history Print

Water

Household water quality - Does the household drink from an improved water source and perceive it to be of good quality?

Community water quality - Does the community have access to an improved drinking water source that is clean, well maintained, and meets WHO guidelines for E. coli?

Water Quantity - Does the water source used by a household produce a sufficient quantity of water (measured in liters per day) to meet all the water needs of the household?

Water access - Water is available within the home or yard/plot; or if not, does it not take a long time to collect?

Water reliability - Water source is functional with minimal downtime over the past year; or if non-functional, a repair plan is in place

Water availability - Does the water point provide water all day, at least once a day, or discontinuously?

Water finance - Are water user fees or tariffs collected and managed as agreed?

Payment or tariffs Local n/a

**Indonesia**  
Water utility index (USAID IUWASH Tangguh)

IUWASH PLUS PDAM Index

Overall Score - Skor keseluruhan rata-rata

Avg. total score split by area - rata-rata skor keseluruhan dibagi berdasarkan area

Avg. total score split by area, proportional - sebanding

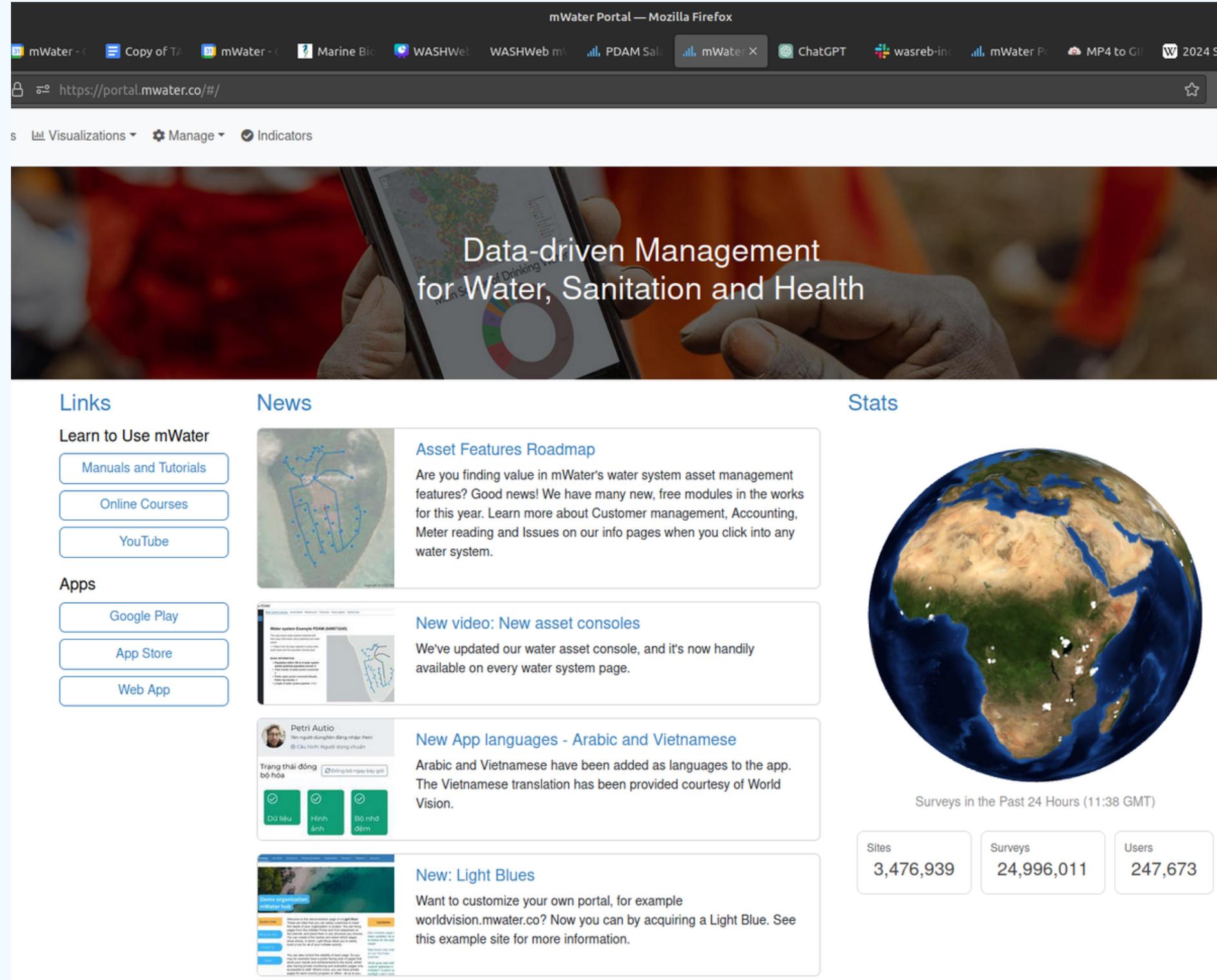
Shows the average for multiple institutions. Quickfilter to show the overall score for a specific institution.

Shows the average for multiple institutions. Quickfilter to show the overall score for a specific institution.

Always sums to 100. Shows the relative change. Selalu berjumlah 100. Menunjukkan perubahan relatif.

	Overall Score	Financial (Max. 28)	Service (Max. 8)	Operational (Max. 28)	HR (Max. 12)	Admin (Max. 18)	Untreated water (Max. 6)	
2016	Kota Magelang	57	22	4	10	6	13	2
	Kota Salatiga	67.5	16	8	19.5	8	13.5	2.5

# SUMMARY



The screenshot shows the mWater Portal homepage. At the top, there's a banner with the text "Data-driven Management for Water, Sanitation and Health". Below the banner, there are three main sections: "Links", "News", and "Stats". The "Links" section includes links to "Learn to Use mWater" (Manuals and Tutorials, Online Courses, YouTube), "Apps" (Google Play, App Store, Web App), and "Petri Auto" (Arabic and Vietnamese app interface). The "News" section features articles about "Asset Features Roadmap", "New video: New asset consoles", "New App languages - Arabic and Vietnamese", and "New: Light Blues". The "Stats" section displays a globe and data points: "Surveys in the Past 24 Hours (11:38 GMT)" (3,476,939 Sites, 24,996,011 Surveys, 247,673 Users).

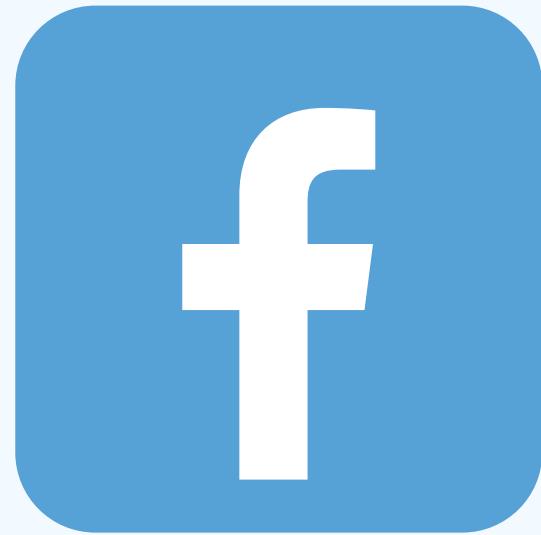
Focus on empowering front-line workers

Ease-of-use leads to rapid achievement of situational awareness

Also facilitates reporting to national standards and other stakeholders

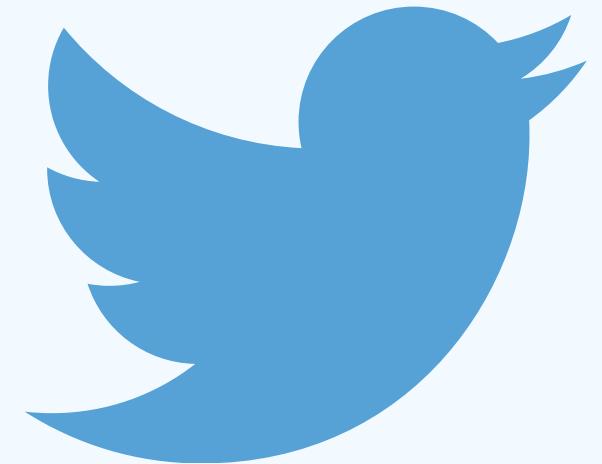
Leads to active management information systems instead of inert data repositories

# Connect with us



**FACEBOOK**

<https://www.facebook.com/mwaterco/>



**TWITTER**

<https://twitter.com/mwaterco>



**LINKEDIN**

<https://linkedin.com/company/mwater>



# IRC/WASHNote

Hosted by WASHWeb  
April 4<sup>th</sup>, 2024



**WASH SYSTEMS**  
academy

# WASH Systems Academy

## Learning Companion

WASHWeb webinar

4 April 2024

By Jeske Verhoeven and Nick Dickinson

**WASH  
NOTE**

**IRC**

Supporting water sanitation  
and hygiene services for life

Why a Learning Companion?

What are we doing?

- Telegram integration
- Curated knowledge base
- Knowledge graphs

Possible linking to WASH sector





**The learning platform for water, sanitation and  
hygiene systems strengthening**

**Launched in September 2019**

**8000+**

total  
registered  
users

**125**

countries

From

**95%**

of users  
recommend our  
courses

# 90%

of users found opportunities to use course  
content in their work within 6 months

# Why a learning companion?



Registration support and troubleshooting



Help learners identify and access courses that meet their needs



Provide insight on content in courses and excerpts



Provide relevant sector information not contained in the course such as country/user context specific information from the knowledge base or tips on how to find more information

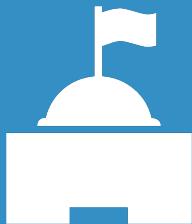


Support learners to brainstorm about how to critically apply their new knowledge in their own WASH system



8 and 16 hours  
basic course and building block courses

# Looking into use cases for 3 key users:



**Government staff**  
(26% of users)  
Key audience for  
systems strengthening.



**Women**  
(27% of users)  
The project aims to  
increase female  
participation and  
completion.



**Young professionals**  
(62% of users)  
Our largest audience and  
early adopters.  
Often need  
contextualisation.

## Testing the learning companion with different modalities:

### MASSIVE OPEN ONLINE COURSE (MOOC)

- Open access
- 4-16 hour MOOC
- Free
- Self-paced
- Self-guided
- Support via email

### MOOC WITH STRUCTURED SUPPORT

- Closed group
- 4-16 hour MOOC
- Live webinars/  
virtual workshops
- Adapted content
- Group work
- WhatsApp group

### BLENDED LEARNING

- Closed group
- 4-16 hour MOOC
- 1-5-day face-to-face  
training workshop
- Adapted content
- Group work
- WhatsApp group

 ...24 LC Learning Companion bot

the regulator. Who are the service authorities in Rwanda? 11:39 ✓

In Rwanda, the service authorities responsible for Water, Sanitation, and Hygiene (WASH) include the Rwanda Environment Management Authority (REMA) and the Rwandan Bureau of Standards. REMA is pivotal in regulating water quality and managing environmental impacts, while the Rwandan Bureau of Standards is instrumental in developing standards for industrial and wastewater management. These organizations play a key role in ensuring the effective regulation and oversight of WASH services in the country. 11:39

What is then the role of municipalities and districts? 11:42 ✓

In Rwanda, municipalities and districts have a significant role in the management and implementation of Water, Sanitation, and Hygiene (WASH) initiatives. Their responsibilities include:

1. **Designing Resilient Water Infrastructure**: They work in partnership with the Ministry of Infrastructure and other stakeholders to design water infrastructure that is resilient and sustainable. This collaborative effort is crucial for developing effective Water Resource Management (WRM) Plans.
2. **Developing Water Resource Management Plans**: These plans are developed to provide sustainable water services by quantifying available water resources, assessing demand, and

Rwanda - Zotero

File Edit View Tools Help

All Fields & Tags

Title	Creator	Year
National Sanitation Policy Implementation Strategy	Ministry of Infrast...	
All Systems Connect : WASH Learning Theme 2 - Ma...		
Water For People market system development update	Sugden et al.	2020
Climate change, water resources and WASH systems	Butterworth et al.	2021
Everyone Forever : Water For People's WASH system...	Water for People	2021
With 5 years of change, looking 10 years into the fut...	Smits	2022
All Systems Connect : CONNECT Finance presentatio...	IRC Team	2023
All Systems Connect: WASH Learning Theme 4 - Gov...	IRC Team	2023
Payment modalities for low-income households	Smits and Mekonta	2023

Info Notes Tags Related

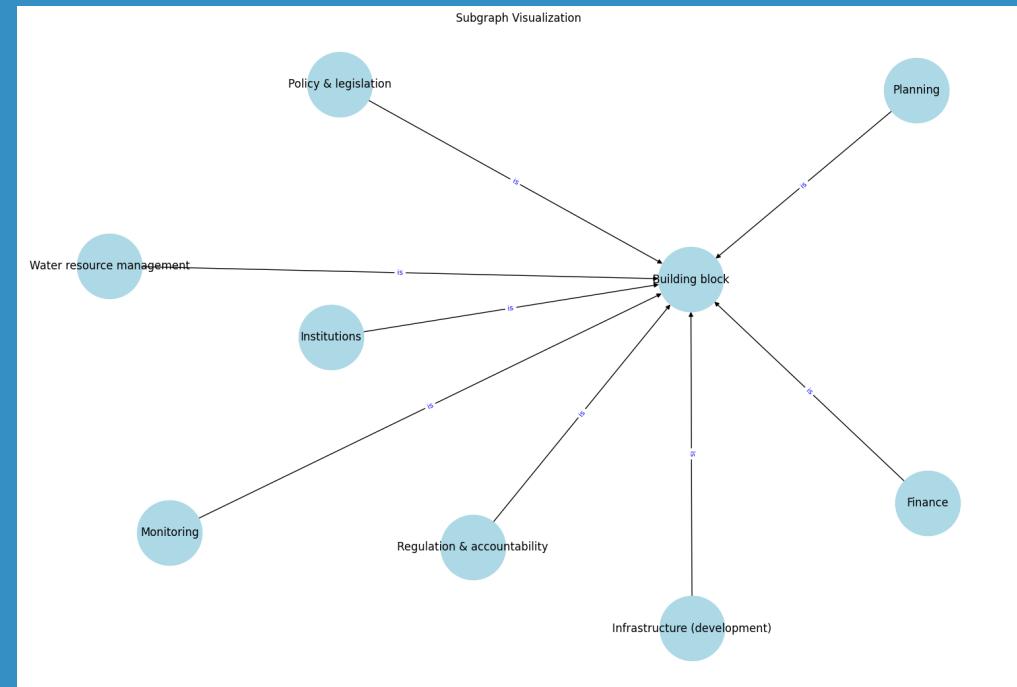
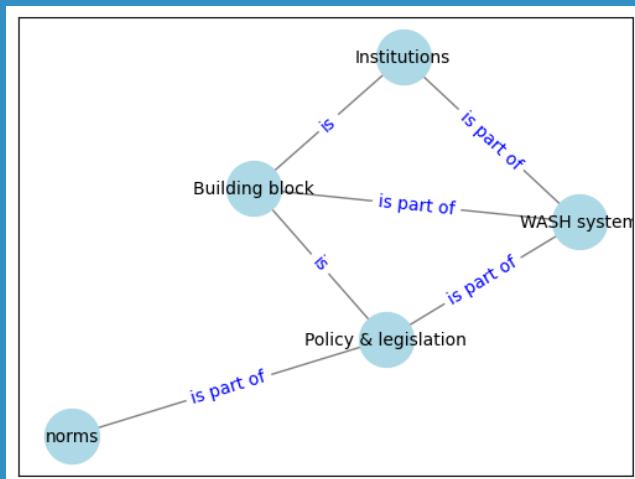
Item Type Report  
 Title Payment modalities for low-income households  
 Author Smits, Stef  
 Author Mekonta, Lemessa  
 (...) Abstract This study aims to generate insights into the...

Report Number  
 Report Type Research Report  
 Series Title  
 Place  
 Institution  
 Date 2023  
 Pages  
 Language  
 Short Title  
 URL <https://www ircwash org/resources/payment-modalities-for-low-income-households>

# A curated knowledge base

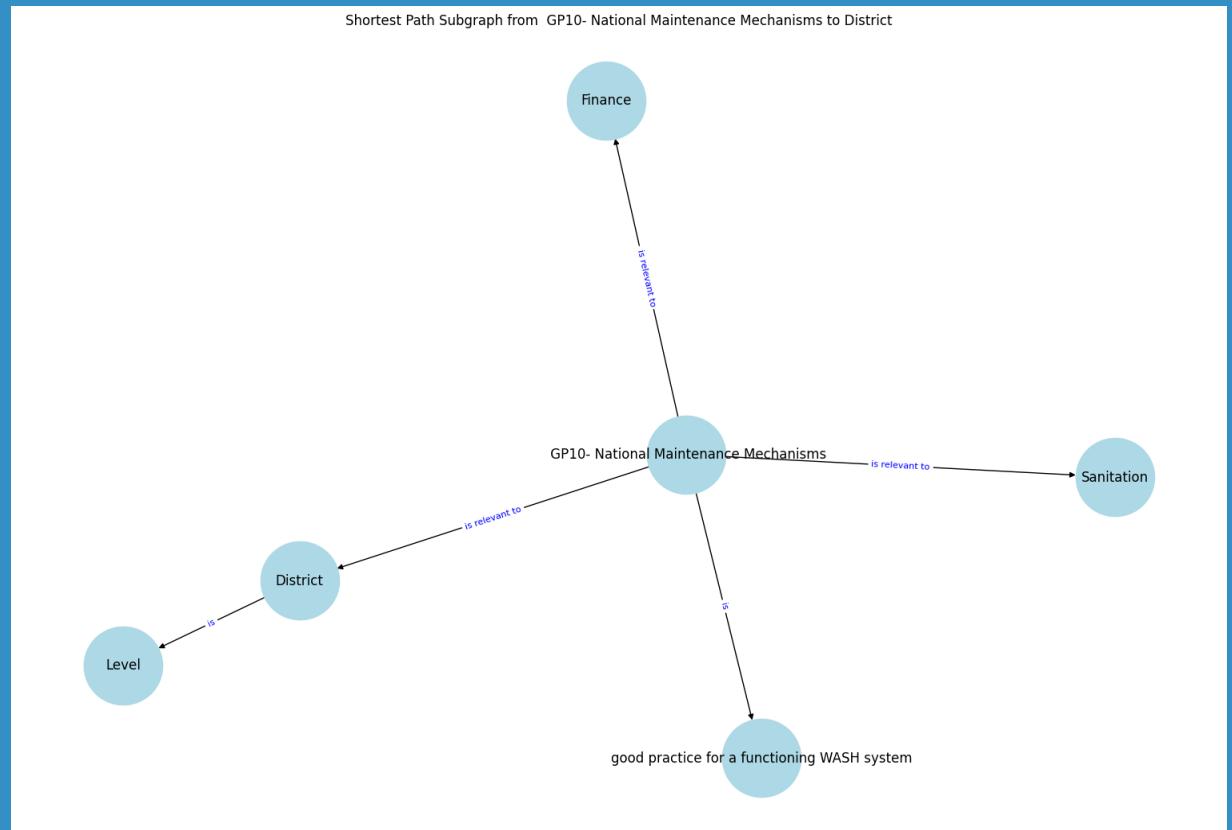
# IRC Building Blocks as a Knowledge Graph

- WASH system  
Modeling the language and ideas of the Academy
- Learning Companion ‘speaks’ the language of the Academy



# IRC Building Blocks as a Knowledge Graph

- Discover contextual information  
Good practices under the Finance Building block
- “I’d like to know about the sanitation finance in Ethiopia.”  
→ Only documents addressing sanitation and the Finance Building Block and related concepts (taxes, tariffs, transfers)



# Where are we now?

- Prototype is live and testing internally
- Curated knowledge base is live
  - Core methodology and course content
  - Selected top 10 documents for Rwanda and Ethiopia to contextualize responses
  - Full academic citations
- Knowledge graph loaded in a database; to be integrated in next round of testing
- Each new course will have a Learning Companion tailored to its content.

## Reflections for the future of the WASH data ecosystem

- AI and knowledge graphs have a role in making data on WASH systems accessible: WASHNote's goal
- Use-cases and specific geography improve AI responses
- Curation of information is invaluable
- Ground-truthing data and validation are critical for GenAI
  - A community, like with Wikidata or Wikipedia, is needed
  - User-testing/evaluation results to be communicated



# WASH SYSTEMS

## academy

[training@ircwash.org](mailto:training@ircwash.org)  
[washsystemsacademy.org](http://washsystemsacademy.org)

# USAID WSSH D&A

Hosted by WASHWeb  
April 4<sup>th</sup>, 2024

WASHWeb



# WATER SECURITY, SANITATION, AND HYGIENE DATA & ANALYTICS (WSSH D&A)

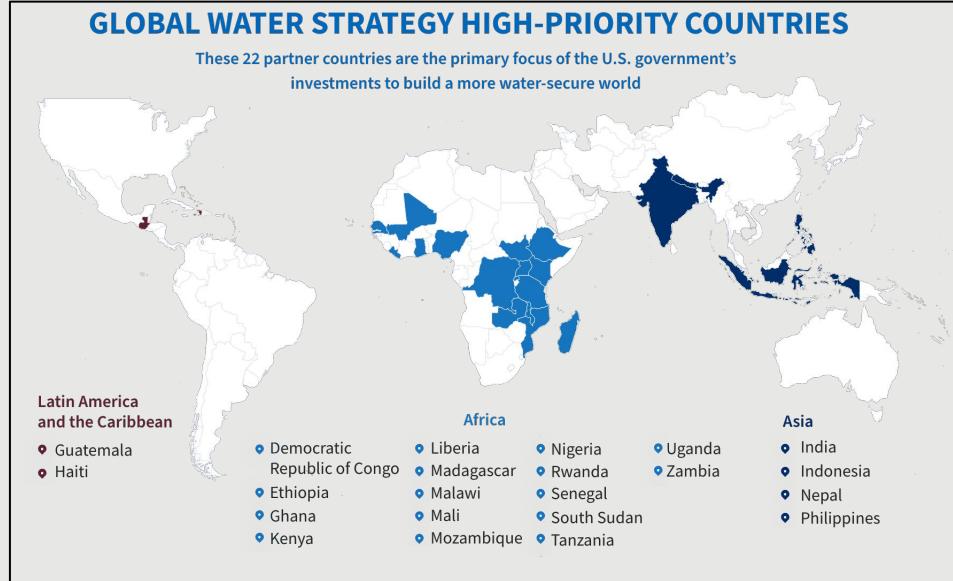
washweb webinar  
April 4, 2024

# WATER SECURITY, SANITATION AND HYGIENE DATA & ANALYTICS (WSSH D&A) ACTIVITY (2023-2028)

## Objectives:

- Strengthening the WSSH Data Landscape
- Supporting Data-Informed Decisions
- Monitoring WSSH Acceleration Resulting from USAID's System Strengthening Work
- Providing Tailored, Evidence-Informed Decision Support

## Implemented by:



## With consortium members:



## Learn more:

<https://www.globalwaters.org/HowWeWork/Activities/water-security-sanitation-and-hygiene-data-analytics>

## Harnessing the Opportunity

### Component 1: Strengthening the WSSH Data Landscape

Convene WSSH Data Expert Advisory Panel  
Conduct research on forward-looking WASH data questions

Build momentum around WASH data through sector engagement

Invest in public goods that support use of WASH data

### Component 2: Supporting Data-Informed Decisions

Support Missions and Implementers to use data-driven insights to improve programming

### Component 3: Monitor WSSH Acceleration Resulting from USAID's System Strengthening Work

Monitor indirect impact of USAID's work

### Component 4 Providing Tailored Data-Informed Decision Support

Support Missions with on demand WASH data analysis

Long Term Progress (Sector)

Near Term Impact (USAID)

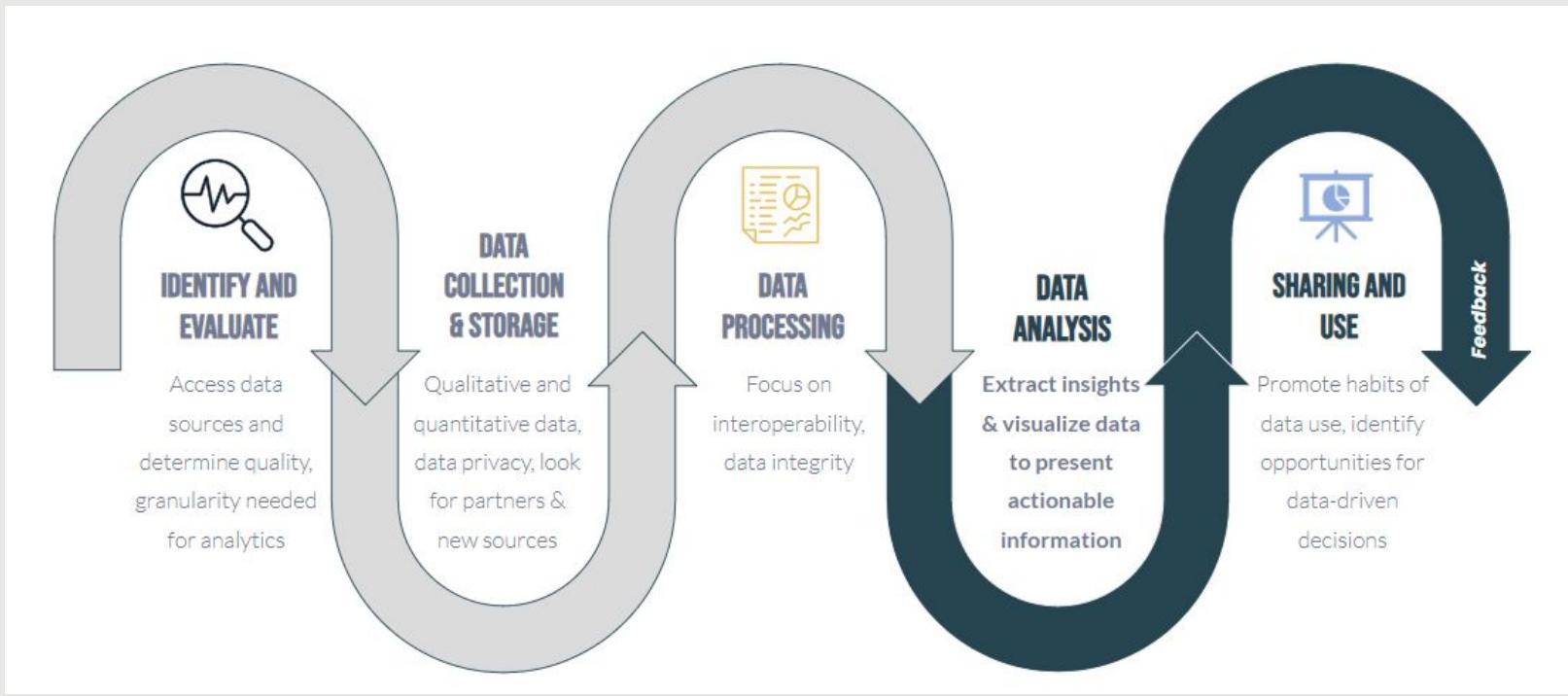
Investments must support the sector for longer term progress while meeting urgent USAID needs

# WSSH D&A OVERARCHING PRINCIPLES

- Ensuring equity throughout the data lifecycle
- Responsible use of data
- Alignment with the Principles for Digital Development
  - Special focus on understanding the existing ecosystem
- Focusing on support rather than data



# WSSH D&A - Data Value Chain



## **Strengthening WSSH Data Landscape - Key Activities**

- Annual USAID Research Briefs on a priority research topic
- Increasing the “drumbeat” on uses of data in the WASH and WRM sectors through increased communications channels and support for conference events
- Quarterly webinars to highlight existing tools which can support USAID and implementing partner staff improve programming
- Identifying existing data efforts where USAID can contribute (data and/or via grant resources)
- Annual virtual convenings on Evidence Informed Decisions in WSSH (starting in Year 2)

# Supporting Data Informed Decisions

## Envisioned Decision Support Tools

PLANNED WSSH DECISION SUPPORT TOOLS	
Decision Support Tool	Use Case
WASH Scenario Forecast Exploration	A country WASH lead would like to understand the possible impact of accelerating or decelerating access to WASH services on access rates as well as broader development results, such as health, prosperity, and stability.
Country Context Benchmarking Dashboard	A country WASH Lead would like to understand how the country is progressing towards each of the Intermediate Results and Strategic Objectives of the Global Water Strategy, and understand if the country is relatively ahead of its peers or has a relatively significant need for investment.
District Level Prioritization Tool	A country WASH Lead would like to identify optimal locations for implementation for specific WASH activities.

For more information, please contact:

Katy Sill, Ph.D., Chief of Party, WSSH Data & Analytics  
[ksill@seguraconsulting.net](mailto:ksill@seguraconsulting.net)

---

Thank you!



# Aquaya

Hosted by WASHWeb  
April 4<sup>th</sup>, 2024

WASHWeb.org

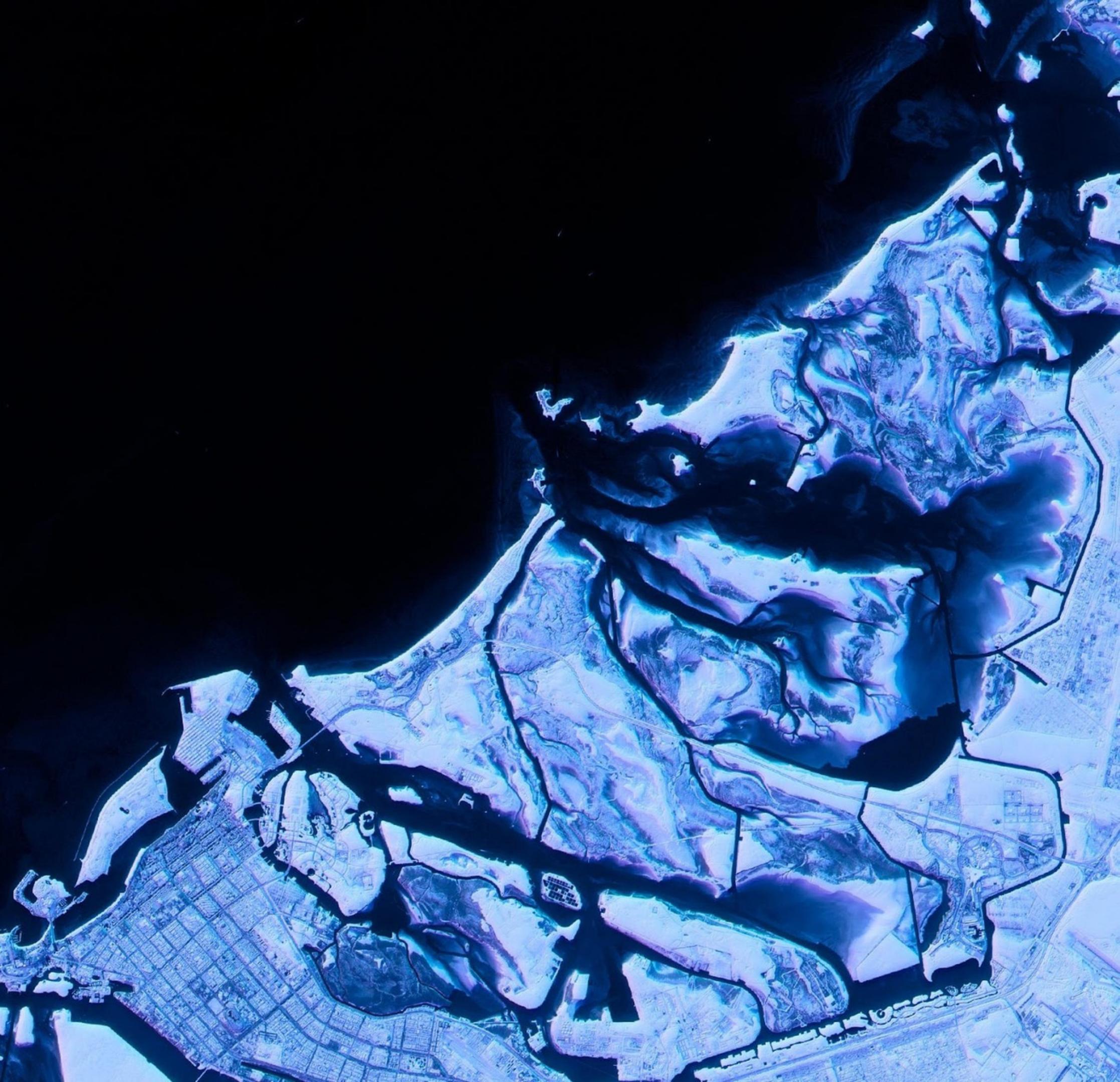
# PROJECT W

---

*The challenge, the gap,  
and the future*

WASHWeb Innovation  
Showcase

April 2024  
Duncan Penfold-Brown  
[duncan@aquaya.org](mailto:duncan@aquaya.org)





# Who are we?

## A brief intro to Aquaya

**44**

Peer-  
Reviewed  
Publications

**7**

Countries  
with Full-Time  
Staff

**24**

Countries  
Worked In



# The Problem



How can we **find** the data we need to make more evidence-informed decisions?



**Data is discoverable** and **usable**

How can we be confident we've covered **everything** that's out there?



**Full coverage** of the data ecosystem

How much is it going to **cost**? How **long** is it going to take?

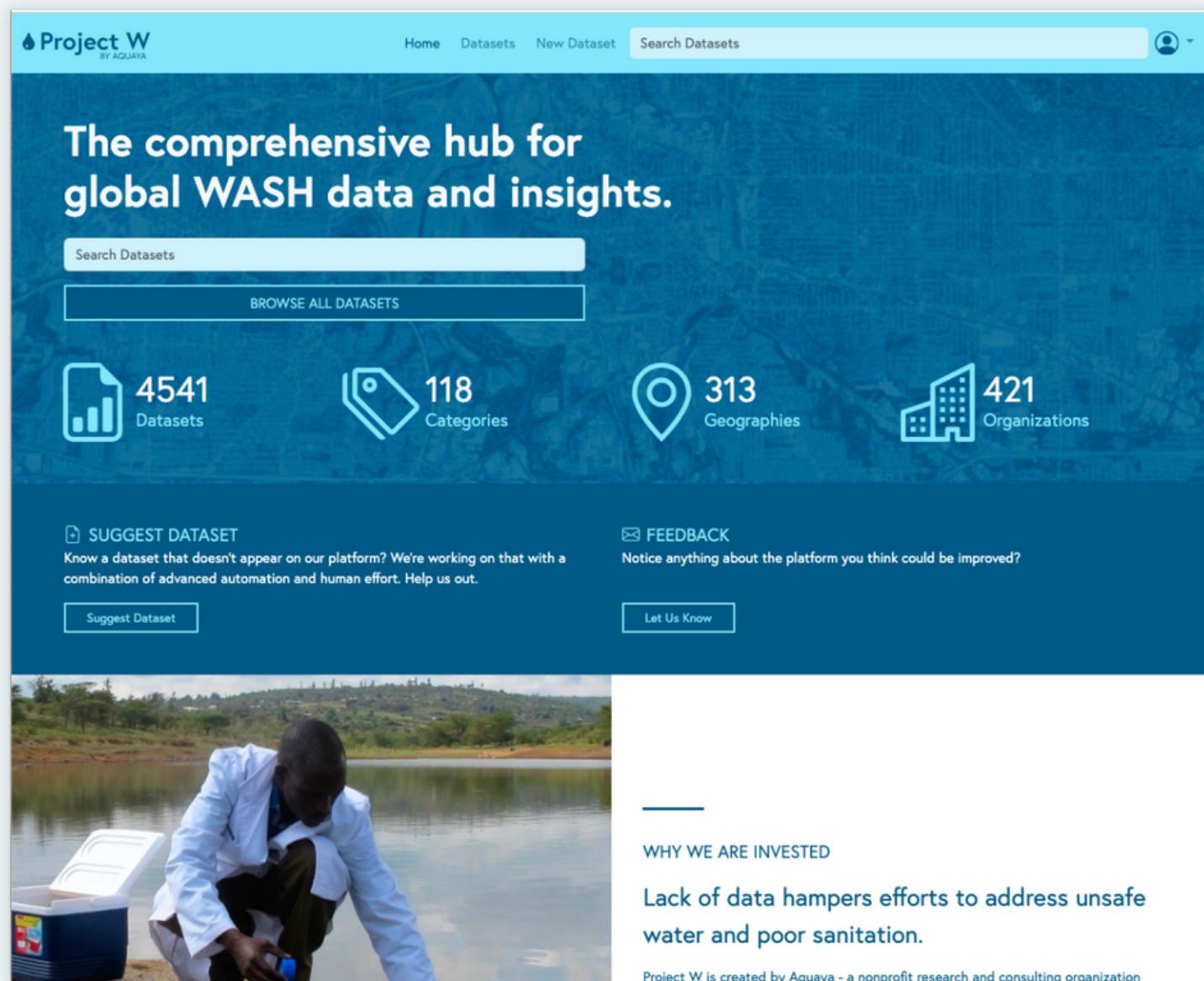


**Efficient access** and organization



# Project W

## Directory, Platform, and Service



## Who is it for?



Governments



Funders



Implementers



People

Project W is for the people who use data for WASH in their day-to-day, month-to-month, year-to-year.



# Project W

## A directory for all WASH data

The comprehensive hub for global WASH data and insights.

Ghana

- Demographic and Health Survey (DHS 2022)
- Ghana Flooding Nov. 2023 - Volta Region map of hum...
- World Vision Ghana Annual Report 2022
- Ghana: Inequalities in Food Insecurity Report (Dec 2023)
- Ghana: World Food Programme Annual Country Repor...
- All Datasets with keyword **Ghana** (98)

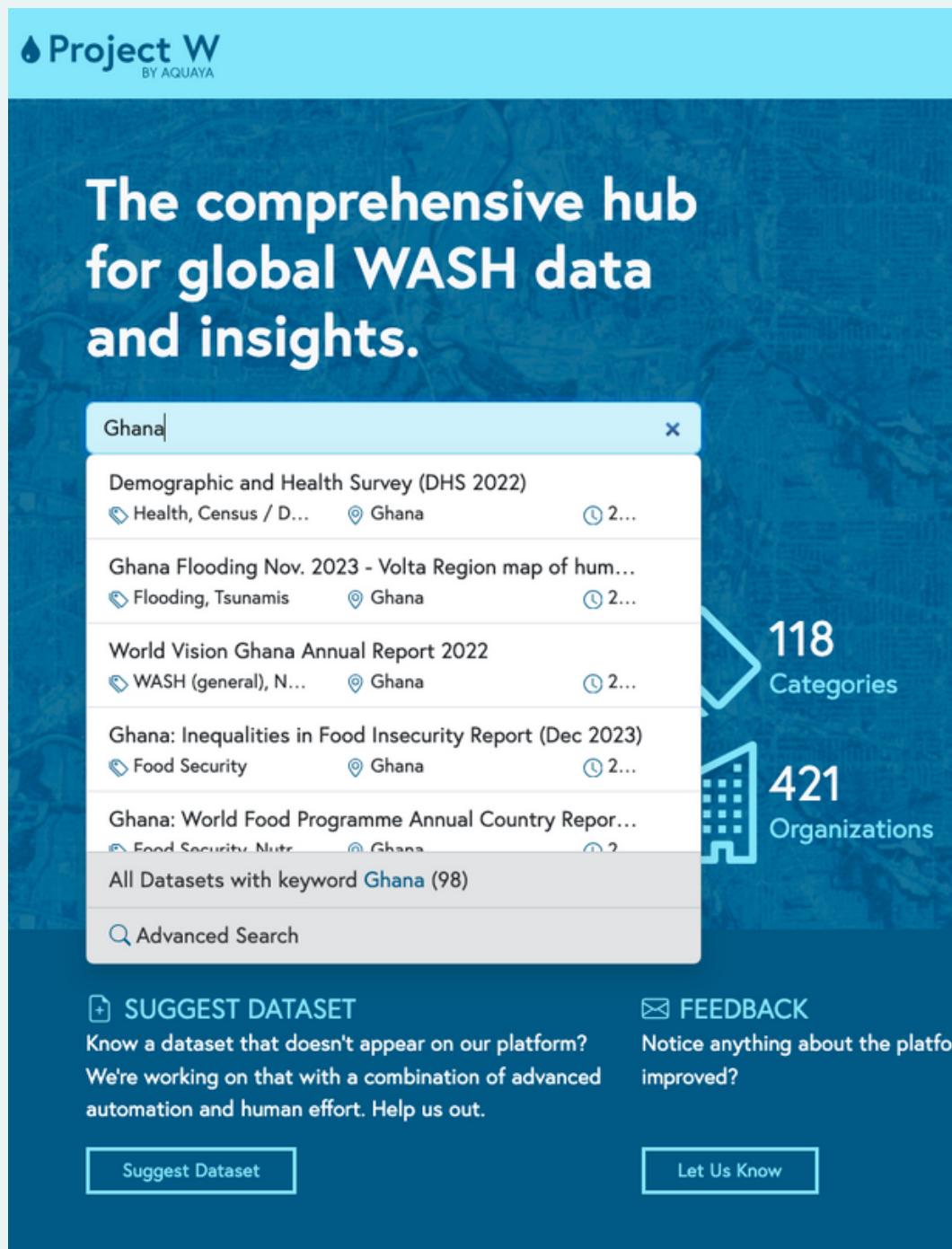
Advanced Search

**SUGGEST DATASET**  
Know a dataset that doesn't appear on our platform? We're working on that with a combination of advanced automation and human effort. Help us out.

**FEEDBACK**  
Notice anything about the platform you think could be improved?

Suggest Dataset Let Us Know

118 Categories  
421 Organizations



Categories

- E Coli
  - Health (99)
    - Waterborne Diseases (9)
      - E. coli (E coli), Salmonella**

Geographies

  - Search geographies
  - West Africa (9)
    - Benin (7)
    - Burkina Faso (43)
    - Cabo Verde (0)
    - Côte d'Ivoire (1)
    - Gambia (2)
    - Ghana (91)**
    - Guinea (4)

Latest Data Update

Select data ages

Data Format

Select dataformats

Spatial Resolution

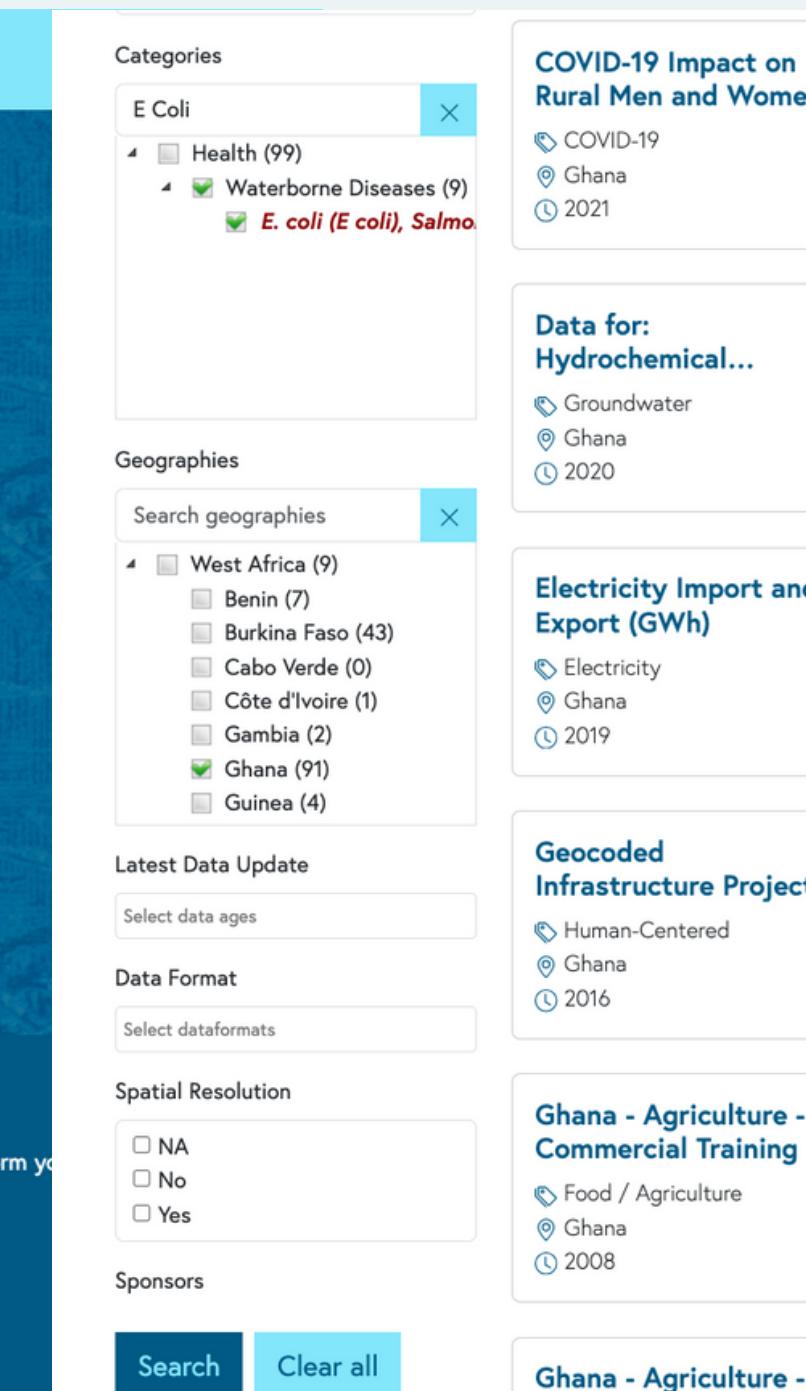
NA

No

Yes

Sponsors

Search Clear all



COVID-19 Impact on Rural Men and Women...

COVID-19  
Ghana  
2021

Data for: Hydrochemical...

Groundwater  
Ghana  
2020

Electricity Import and Export (GWh)

Electricity  
Ghana  
2019

Geocoded Infrastructure Project...

Human-Centered  
Ghana  
2016

Ghana - Agriculture - Commercial Training

Food / Agriculture  
Ghana  
2008

Ghana - Agriculture - Community

Ghana - Community

Dataset Objective

To visually represent humanitarian activities at the regional level in Ghana on November 11, 2023, through national administrative layers.

Latest Data Update 2023

Update Frequency Not Specified

DRAINEAGE NETWORK AND INFRASTRUCTURE, FONGORO DISTRICT,...

Civil Structures (Schools, Hospitals, Transportat...  
Chad  
2008

Major Airfields in Kenya

Civil Structures (Schools, Hospitals, Transportat...  
Kenya  
2008

Rwanda Road Network - Start Points

Civil Structures (Schools, Hospitals, Transportat...  
Rwanda  
2008

School Damages - Hurricane Matthew

Civil Structures (Schools, Hospitals, Transportat...  
Haiti  
2008

Subset of OpenStreetMap Roads, v1 (2014)

Civil Structures (Schools, Hospitals, Transportat...  
Global  
2008

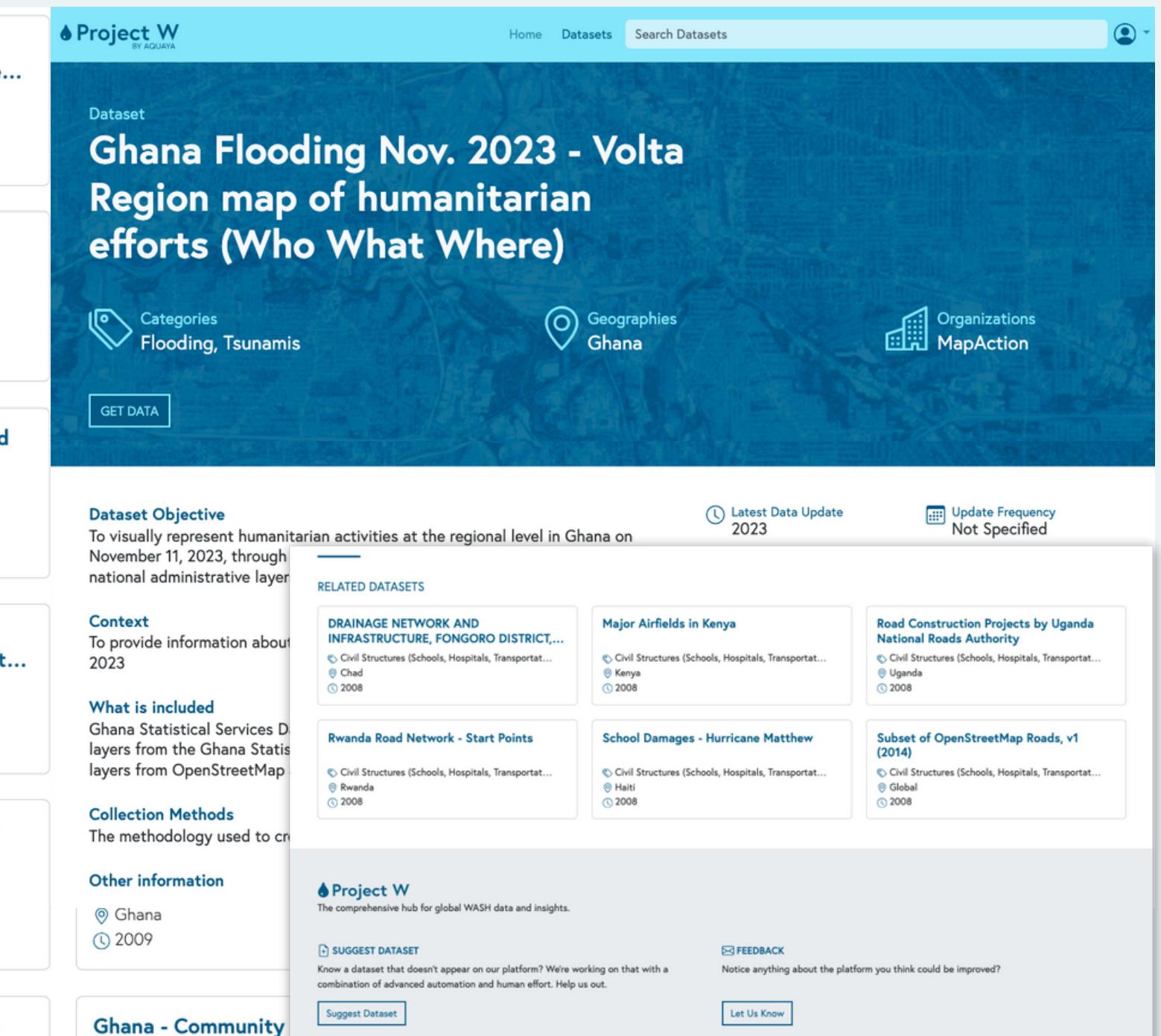
Project W

The comprehensive hub for global WASH data and insights.

**SUGGEST DATASET**  
Know a dataset that doesn't appear on our platform? We're working on that with a combination of advanced automation and human effort. Help us out.

**FEEDBACK**  
Notice anything about the platform you think could be improved?

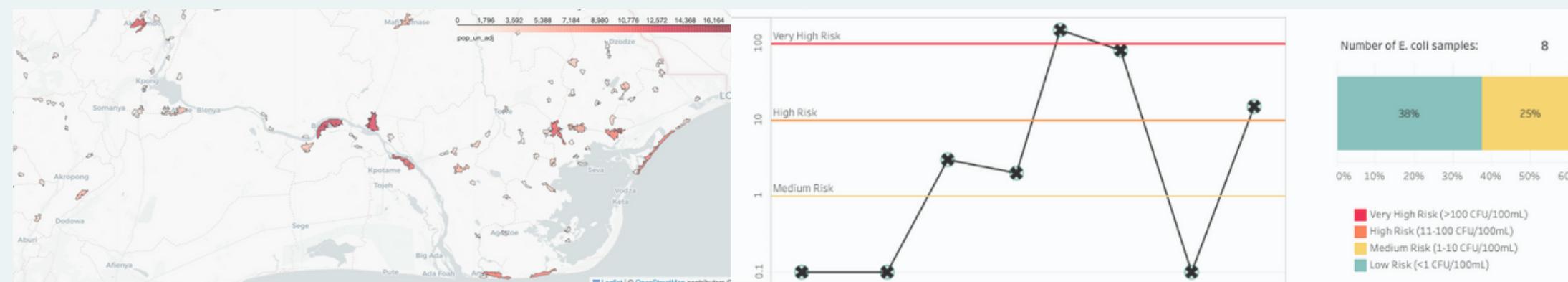
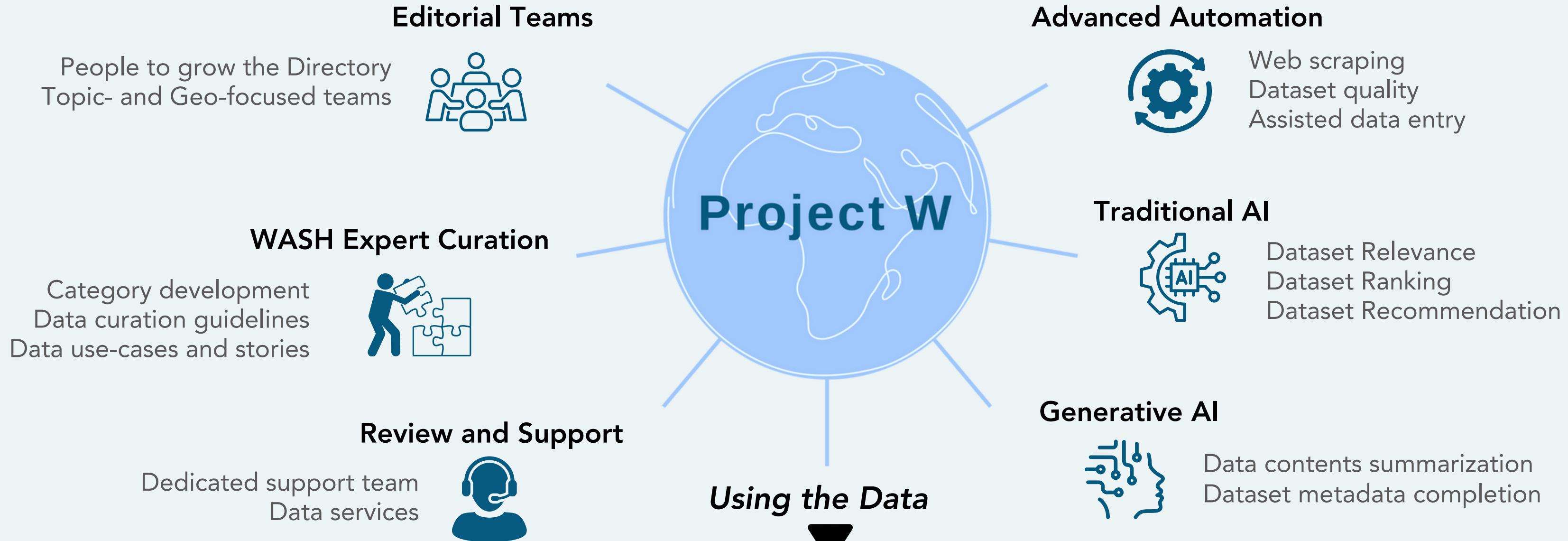
Suggest Dataset Let Us Know



# Where We're Going

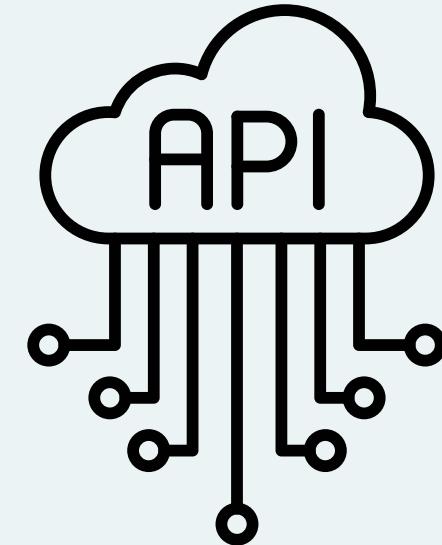
People

Technology

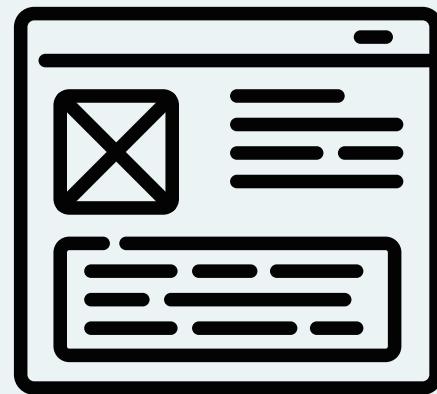


# Collaboration

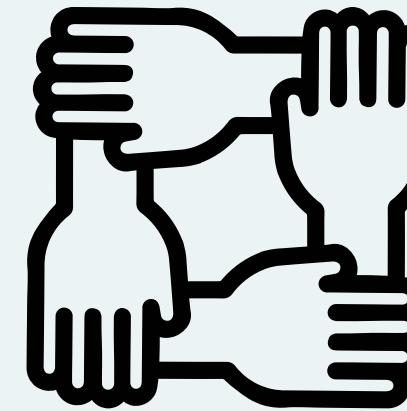
*Accessibility, Standards, Communication*



- Making data available
- Making services available



- WASH metadata specification



- Talking about it; working together

*projectw@aquaya.org  
duncan@aquaya.org*

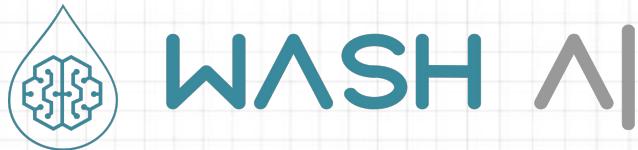


# WASHAI

Hosted by WASHWeb  
April 4<sup>th</sup>, 2024



# The Future of the WASH Data Ecosystem: AI-powered Data Discovery



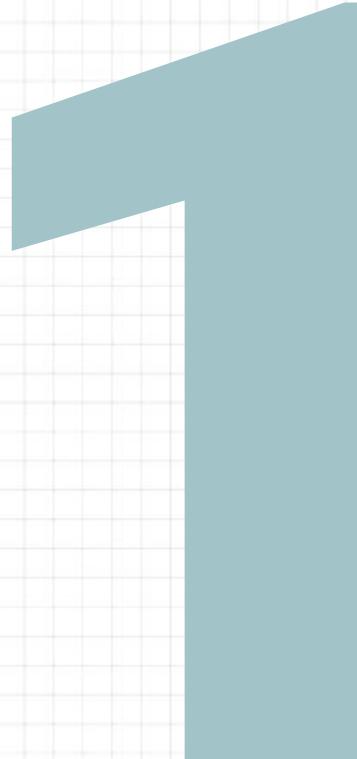
# WASH Data Discovery

1 What are LLMs, **WASH AI**,... and Data discovery? 

2 How can we use **AI to help** with Data Discovery? 

3 What next? 





# **Intro to LLMs, WASH AI and Data Discovery**

1940s-50s: Cybernetics & Connectionism - Early neural network theories

1956: Dartmouth Workshop - "Artificial Intelligence" coined

1960s: Symbolic AI & Expert Systems - Rule-based approaches

1980s-90s: "AI Winter" - Reduced funding & interest in AI

*More classic ML*

1986: Backpropagation - Efficient way to train neural networks

2000s: Statistical Machine Learning - Empirical ML methods

*+ Random forests*

**2010s: Deep Learning** - Multilayer neural nets learn representations

*+ GBMs*

2014: Sequence-to-Sequence Models - Neural nets for sequential data

**2017: Transformers** - Self-attention mechanism for handling long-range dependencies

2018: BERT - Bidirectional transformer for language understanding

2019: GPT-2 - Transformer language model with 1.5B parameters

2020: GPT-3 - Transformer with 175B parameters & few-shot learning

2022: InstructGPT - Aligning language models with instructions

2022: PaLM - 540B parameter model with breakthrough capabilities

**2022: ChatGPT** - Public facing, popularization of Large Language Models

**2023: Claude** - Large context window (200K tokens ~500 pages of text)

# Large Language Models.. with “language” data

Summarization

*“Summarize this meeting I fortunately missed”*

Simplification

*“Explain this paper on chemical analysis in simpler terms”*

Key takeaway extraction

*“Extract the key takeaways from this case study”*

Focused extraction and summarization

*“Extract and summarize the key takeaways from this report focusing on the practical applications of x in context y”*





How it works

# Democratizing WASH Knowledge

The days of google'ing for Water, Sanitation & Hygiene (WASH) knowledge is over.

Delivering interactive context-specific insights in your language, WASH AI informs your decision-making with its advanced AI capabilities, and helps you understand WASH complexities using simple language and references to resources.

Join Waitlist

Learn more →

The screenshot shows a web browser window titled "WASH AI" with the URL "app.washai.org". The interface includes a search bar with the placeholder "Maria" and a question input field containing "How frequently should I test turbidity for an open well?". An AI-generated response is displayed in a box, starting with "For an open well water supply that is community managed, be conducted at a minimum frequency of once per month [1] frequency guidance applies to monitoring of microbial indic [1]. .... [2]". Below this is a "References" section with two numbered items: "1. WHO Guidelines, Page 72" and "2. WHO Guidelines, Page 68". At the bottom are buttons for "Give feedback" and "Start new chat".

WASH AI

app.washai.org

Maria

How frequently should I test turbidity for an open well?

AI

For an open well water supply that is community managed, be conducted at a minimum frequency of once per month [1] frequency guidance applies to monitoring of microbial indic [1]. .... [2]

References

- WHO Guidelines, Page 72
- WHO Guidelines, Page 68

Give feedback

Start new chat

# What is WASH AI?

Suite of AI-powered services and products  
to support better access to WASH information  
and informed decision making for WASH practitioners

...that integrates with **existing channels, websites**

& **knowledge bases**

You are  
here

Web-based  
Knowledge Chat



WhatsApp



Learn



TL;DR



Data



Pulse



Translate



# What is “Data Discovery”

1. Search for the data(sets)
2. Find the relevant ones
3. Clean the data
4. Transform the data
5. Interpret the data



# Current practices, challenges & limitations

Data is hidden in databases,  
csv files, appendices, tables in  
longs reports

Which datasets are  
being used?

How do I read this data?

How useful was it?

I am not a data scientist!  
Can you transform this data  
so it makes sense to me

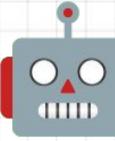
Here is my  
experience

For the data that  
is “open” its  
still not used

I have all this data, how can  
I make it useful to others?





How can we use  
AI  to help  
with Data  
Discovery?

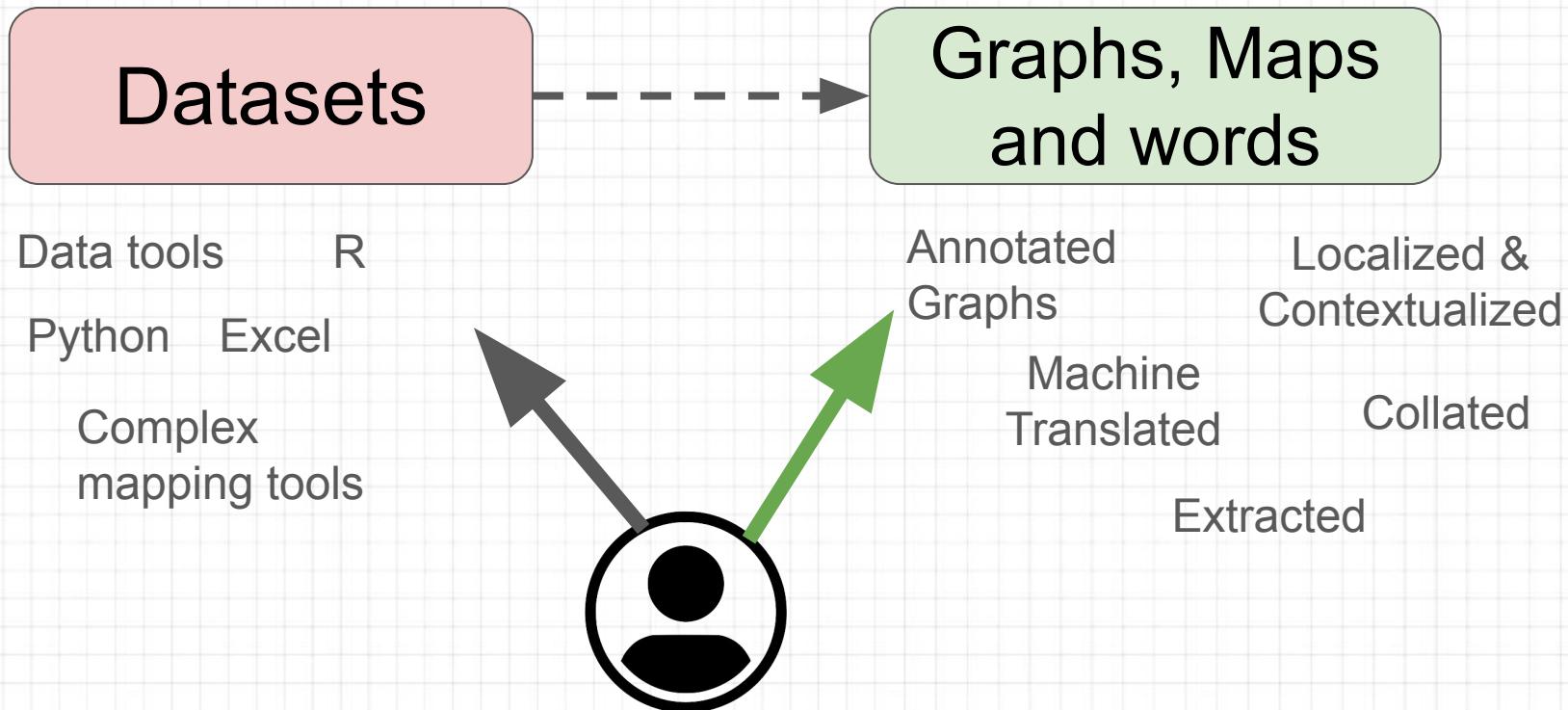
# What is “Data Discovery”

1. Search for the data(sets)
2. Find the relevant ones
3. Clean the data
4. Extract the data
5. Transform the data
6. Present the data



# Information & Knowledge: *Transformation*

Accessibility of data: get it out of the dataframe!



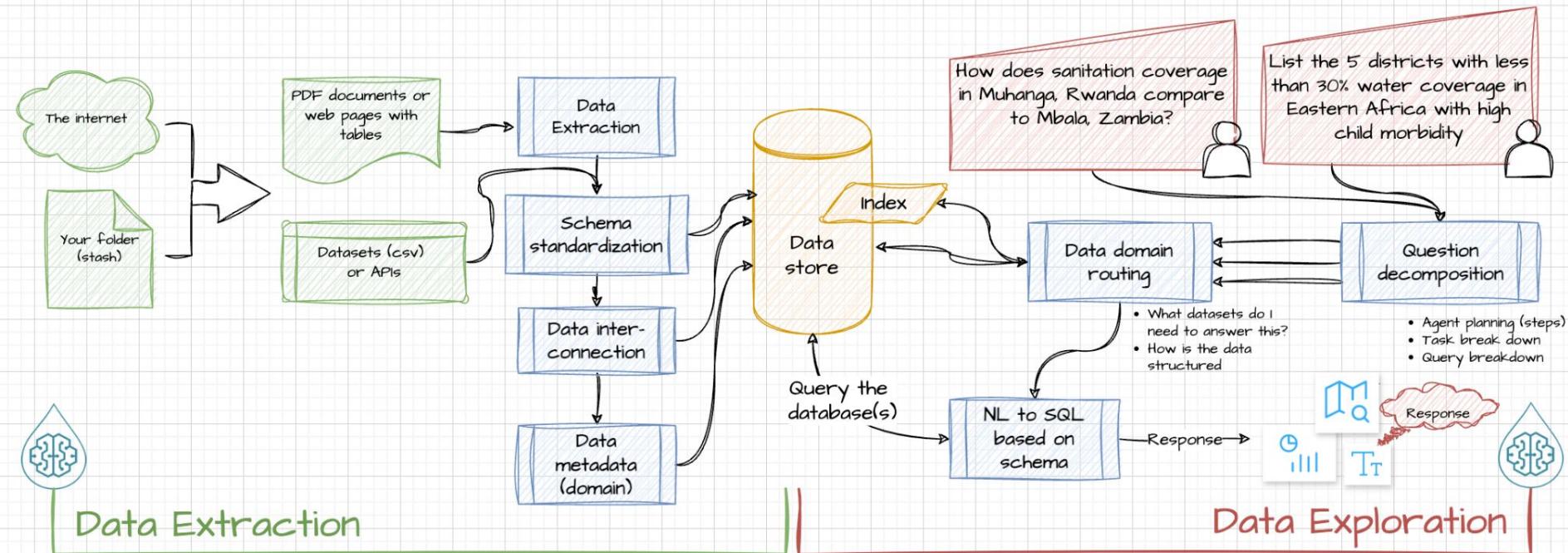


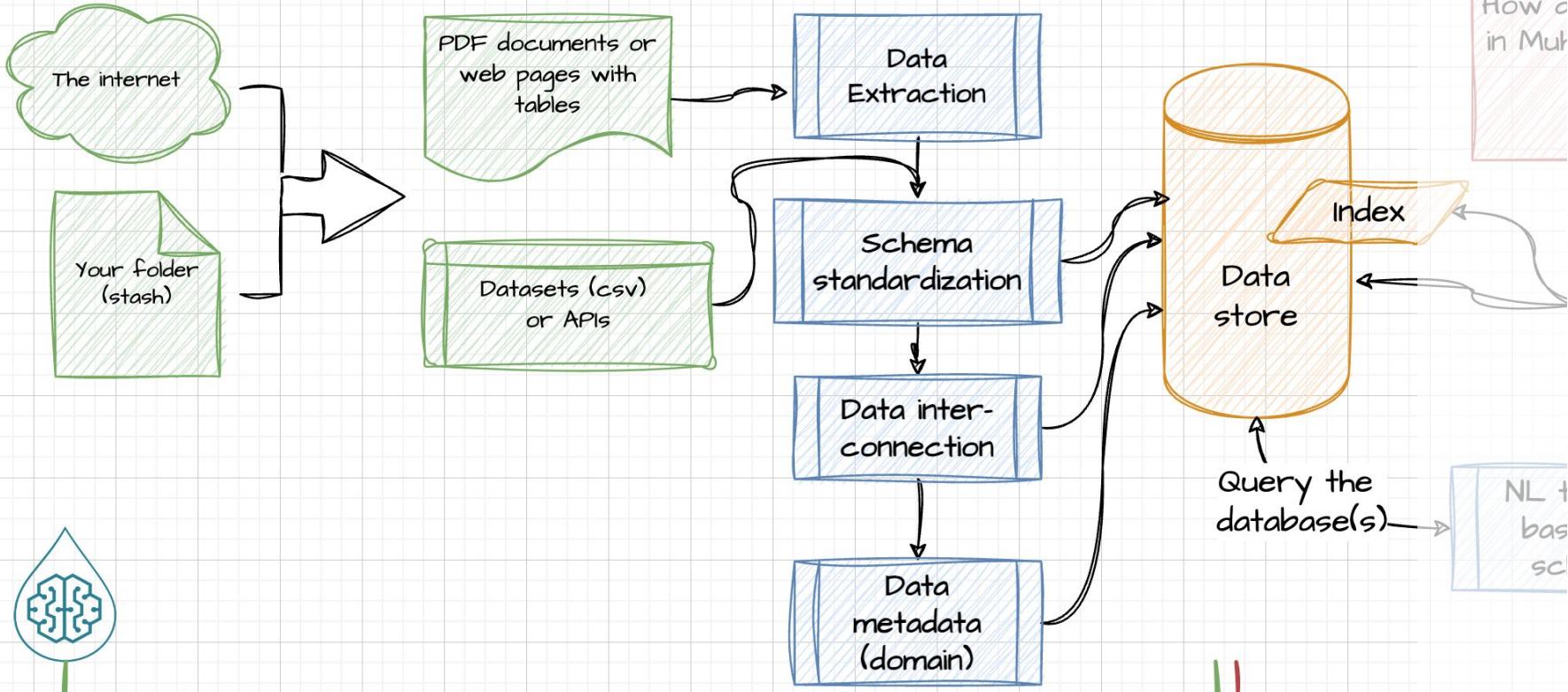
How to get the most  
relevant data in the hands  
of those who need it, now,  
in way they can understand  
and interact with ?



# Data Extraction & Discovery

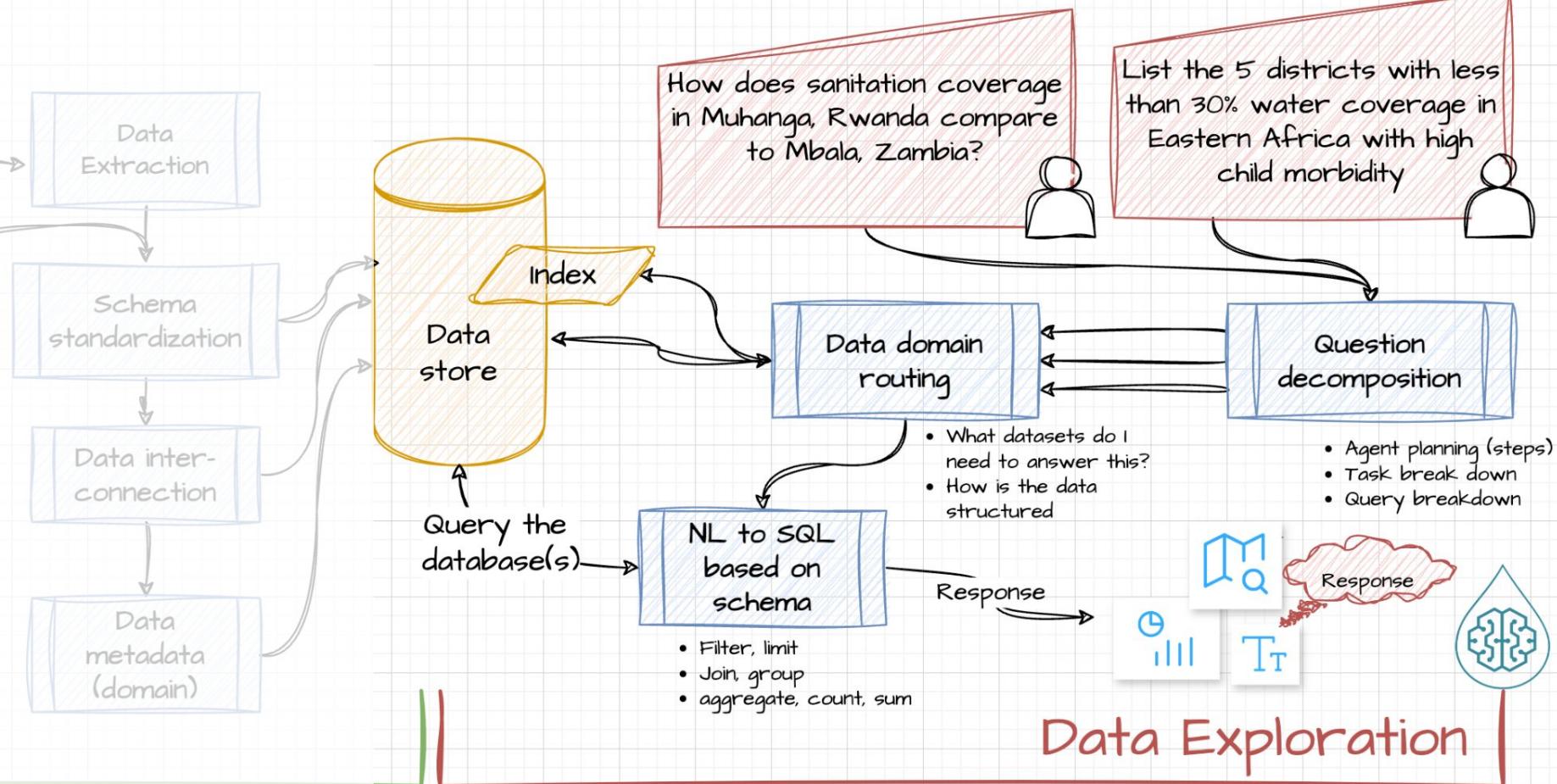
v0.1: Proof of concept





How c  
in Mul

NL +  
bas  
sc



# Outcomes & Opportunities



- AI can **simplify complex data** for everyday use
- **Better data access** nurtures informed decision-making
- AI enhances **data transparency**
- **Data literacy** broadens through user-friendly AI tools
- AI-driven **insights** level the playing field



# Challenges and limitations 😞

- Availability of valuable datasets (“openness”)
- Complex and very niche datasets
- Quality of datasets  ⇒ 
- Complex questions/queries



3

**What next?**

# Q&A – selections from the chat

# Next steps

- Poll
- Join the WASHWeb community on Matrix
- Continue the conversation with people you meet here today
- Follow us on LinkedIn for updates @WASHWeb