

# openwashdata

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

Lars Schöbitz

Global Health Engineering, ETH Zurich

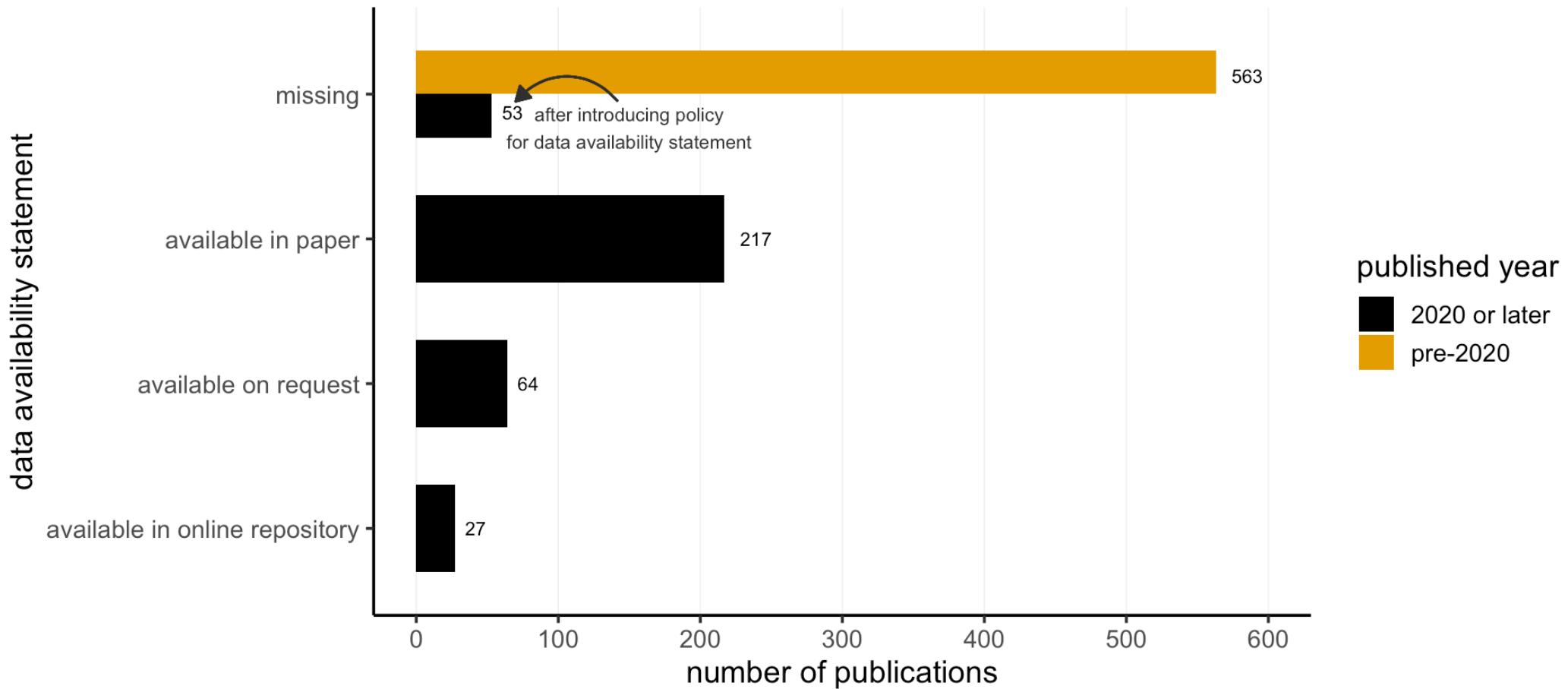
September 12, 2024

# 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

## 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

---

### 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%

# 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/)

## Course Overview

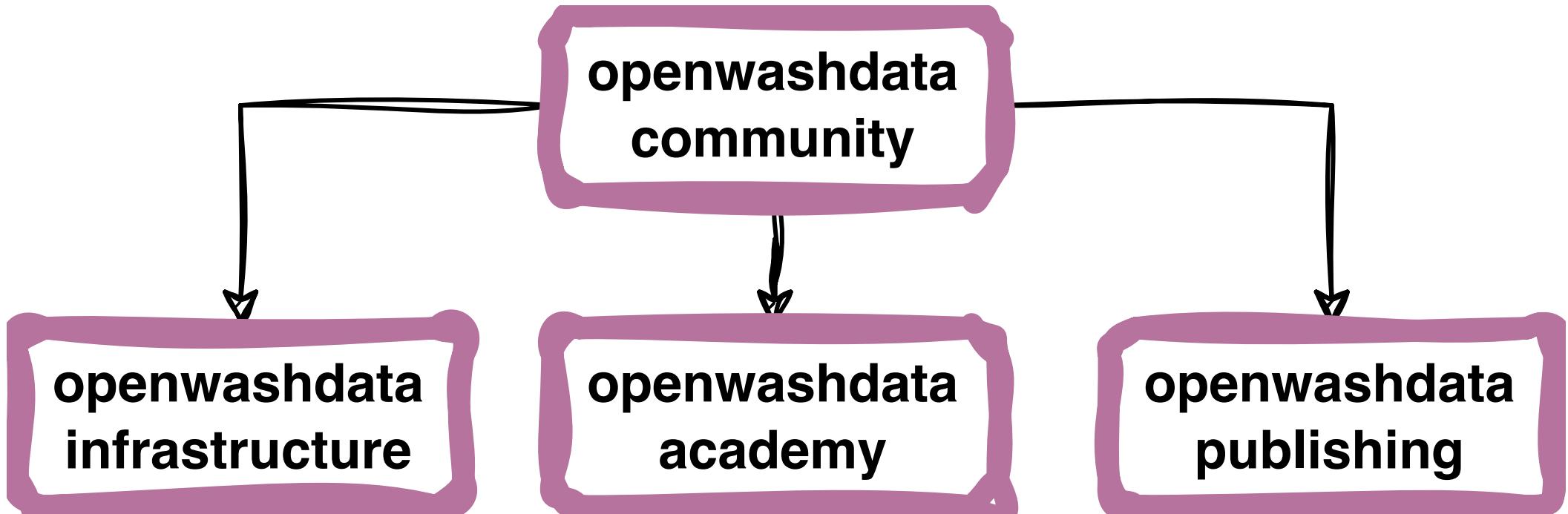
Thank you for your interest in this course. Your course instructors: [Lars Schöbitz](#) & [Mian Zhong](#) & [Sophia Skorik](#) are looking forward to meet you.

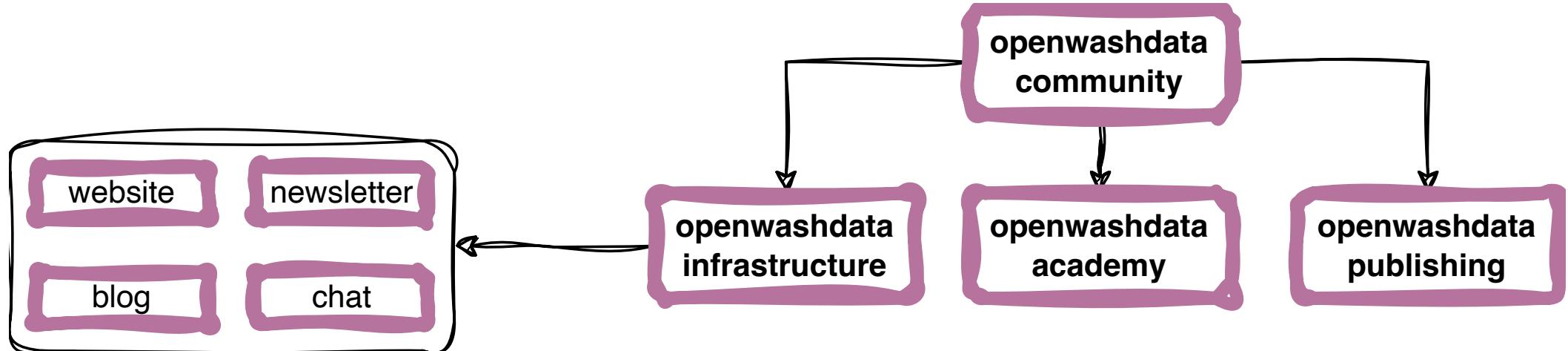
We will meet on Zoom for 10 modules over 17 weeks (see Course Calendar below) at the following times:

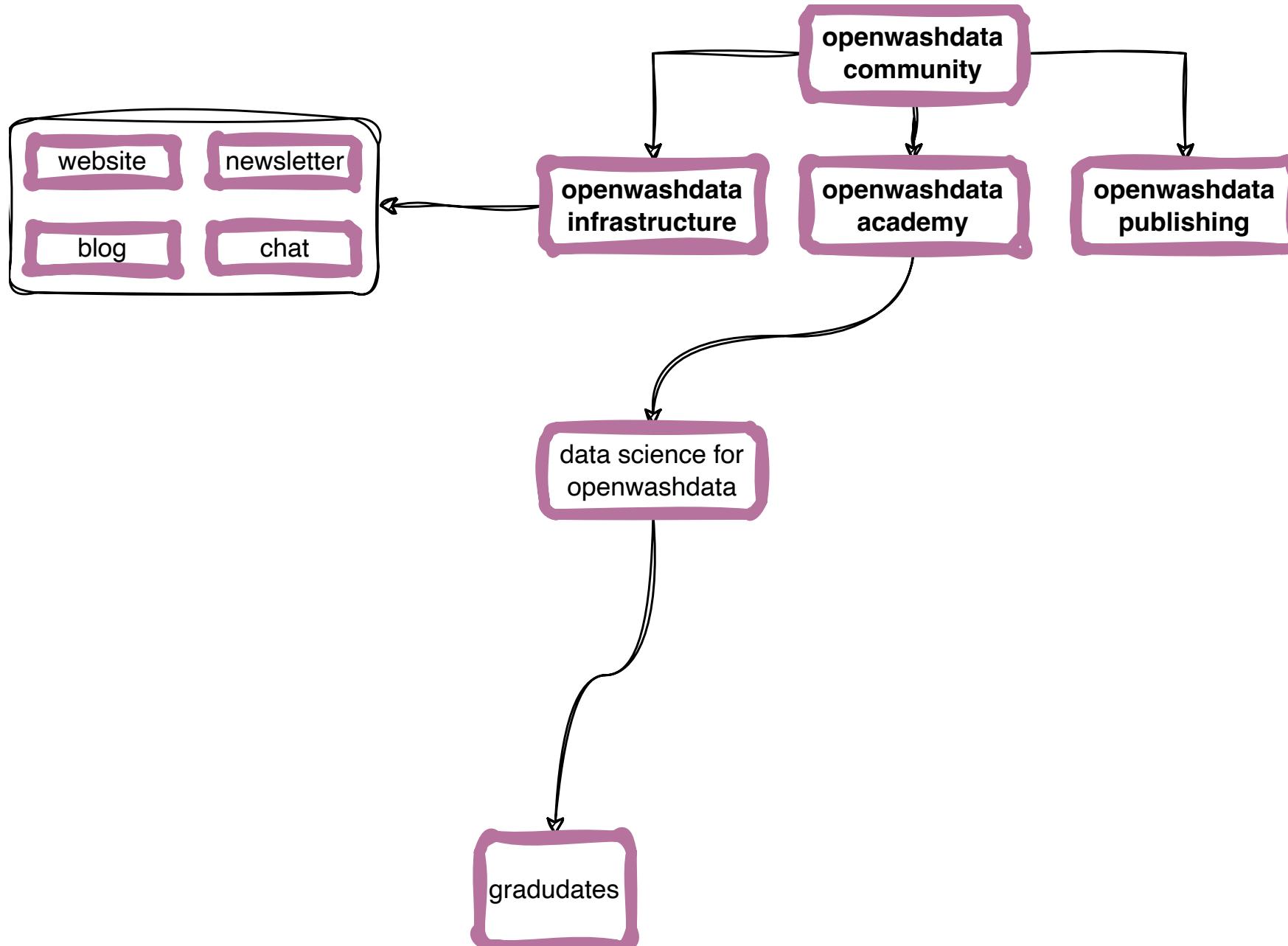
- **Start:** 31st October 2023 - 2 pm to 4:30 pm CET
- **End:** 20th February 2024 - 2 pm to 4:30 pm CET

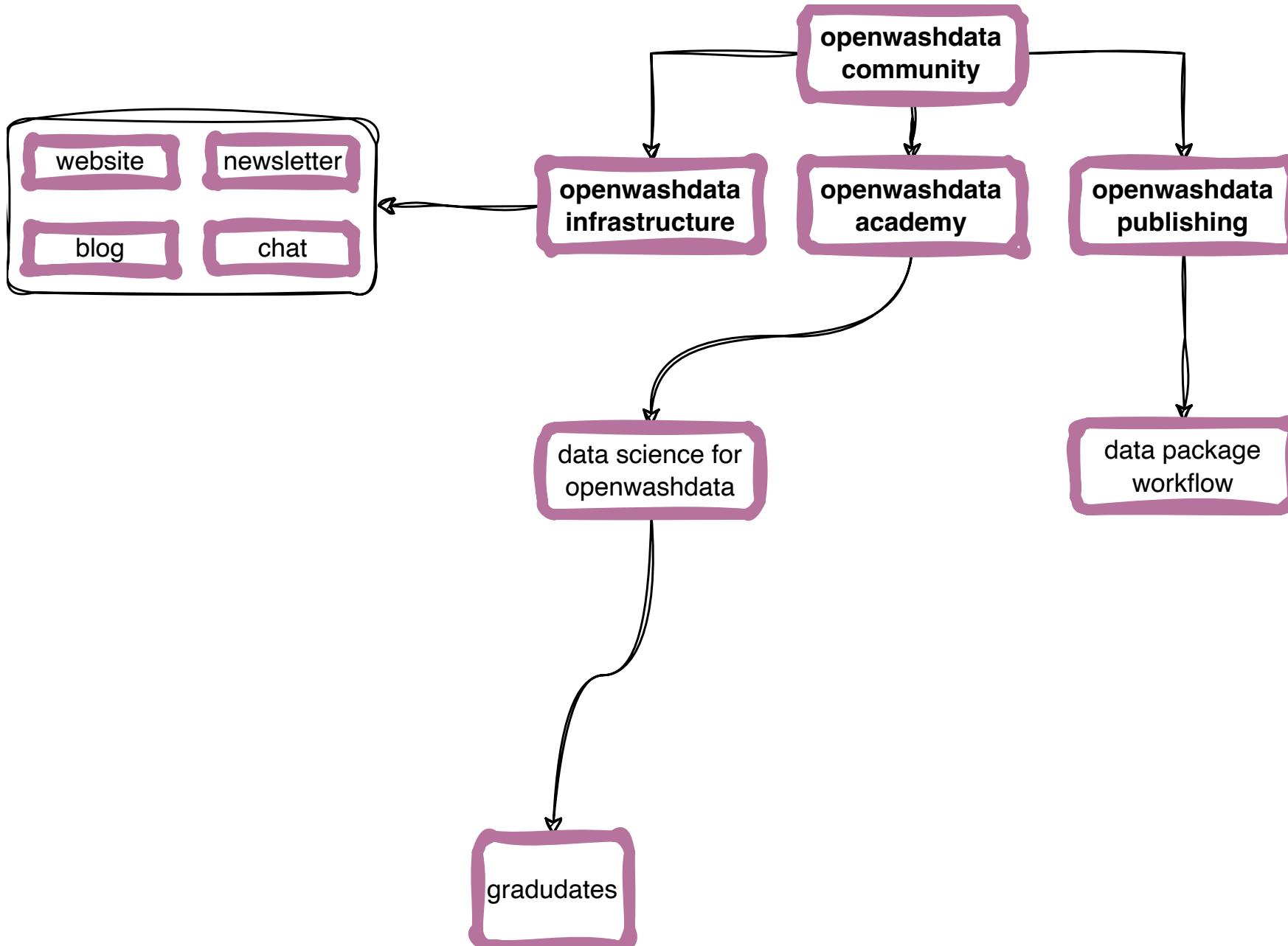
We will use Posit Cloud infrastructure, so you do not need to install any

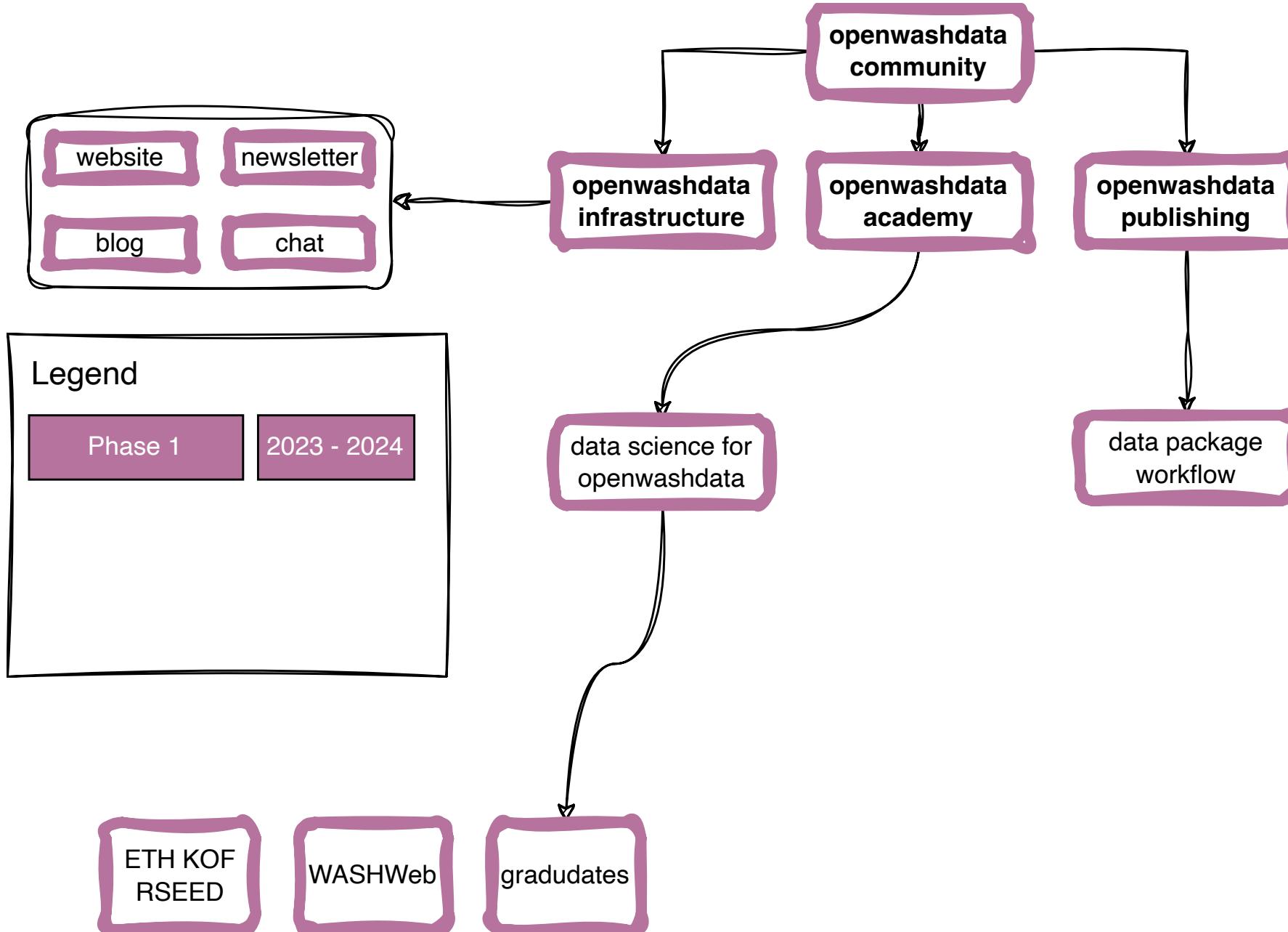
# what's coming

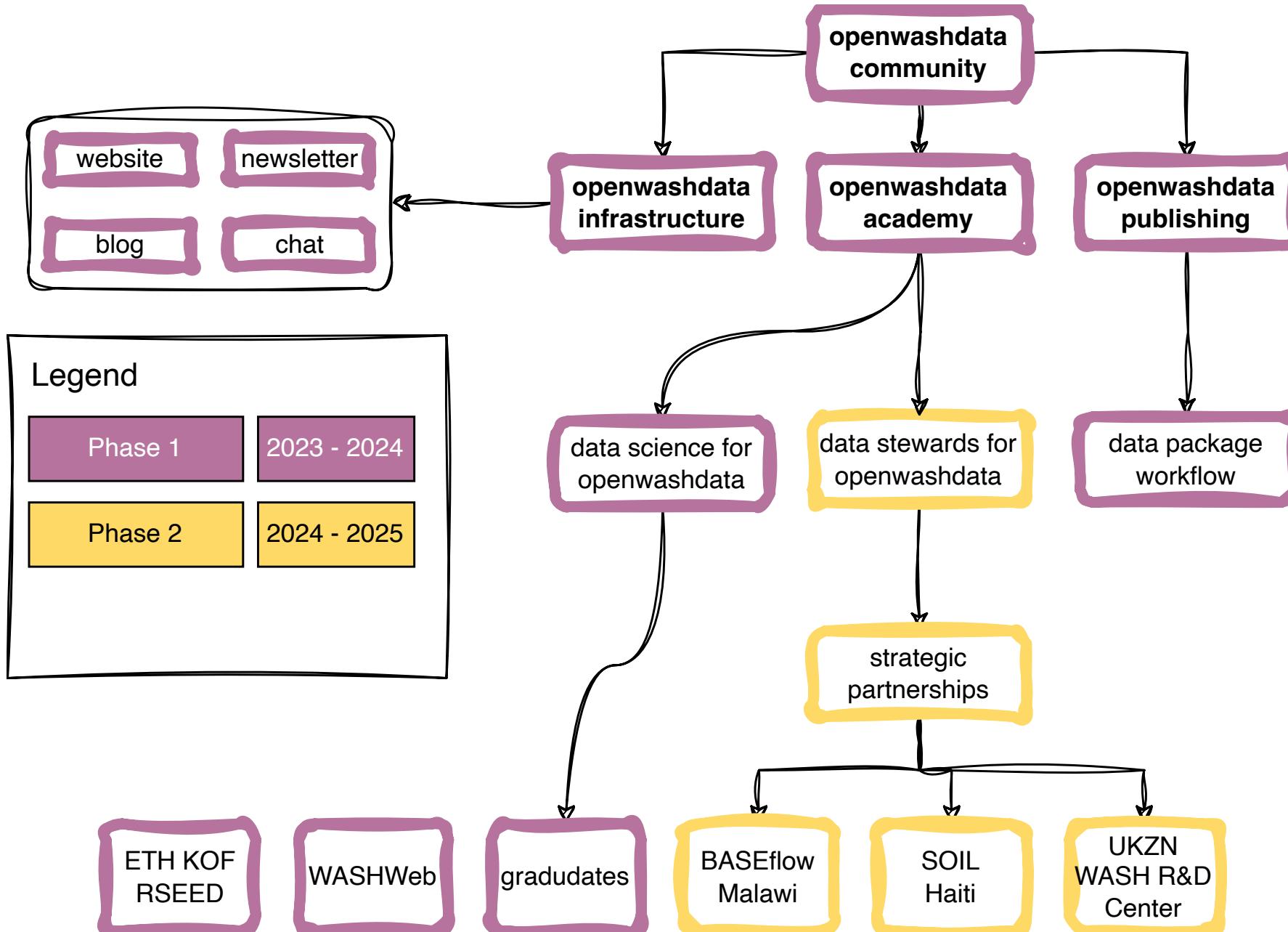


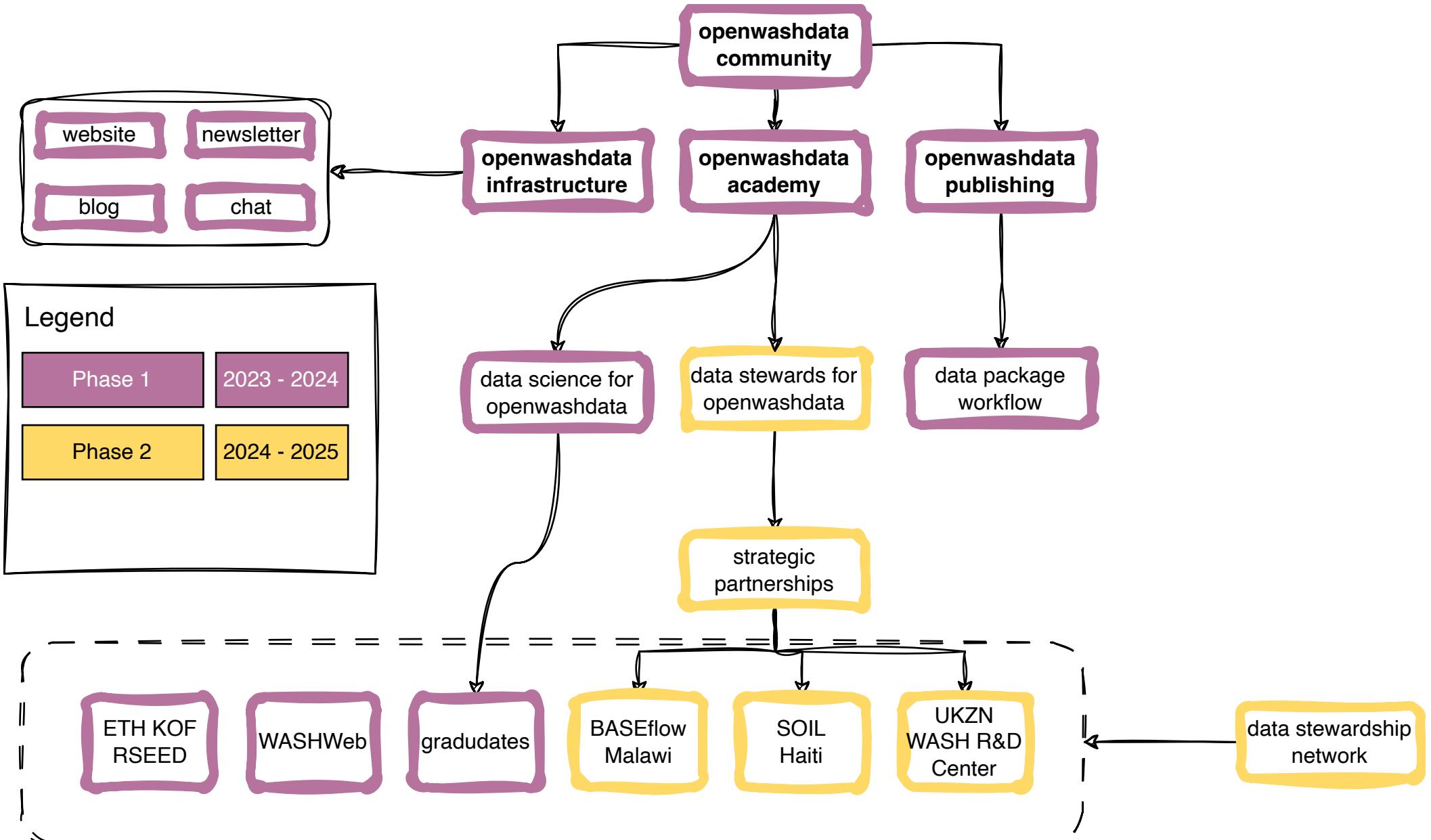


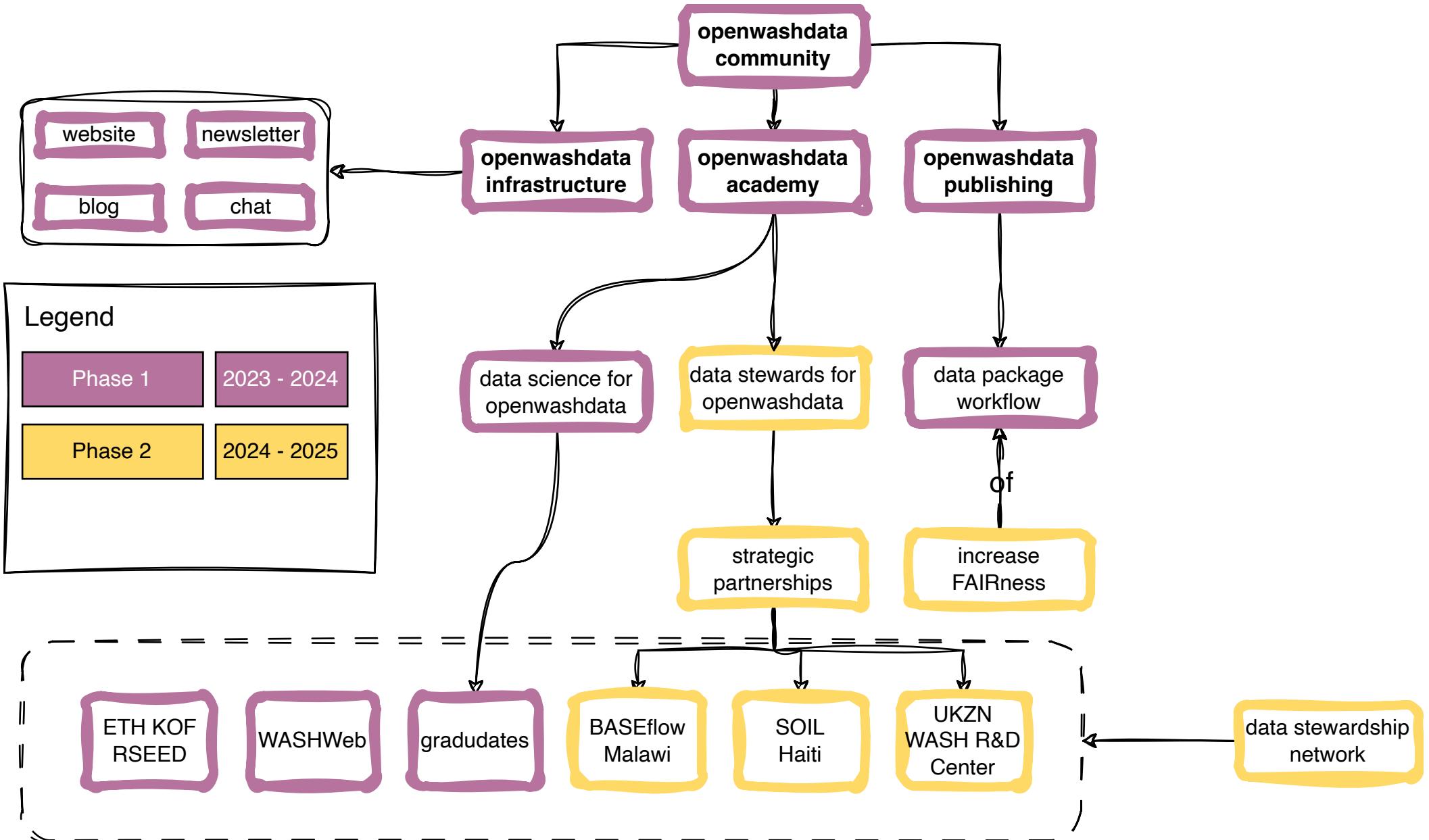


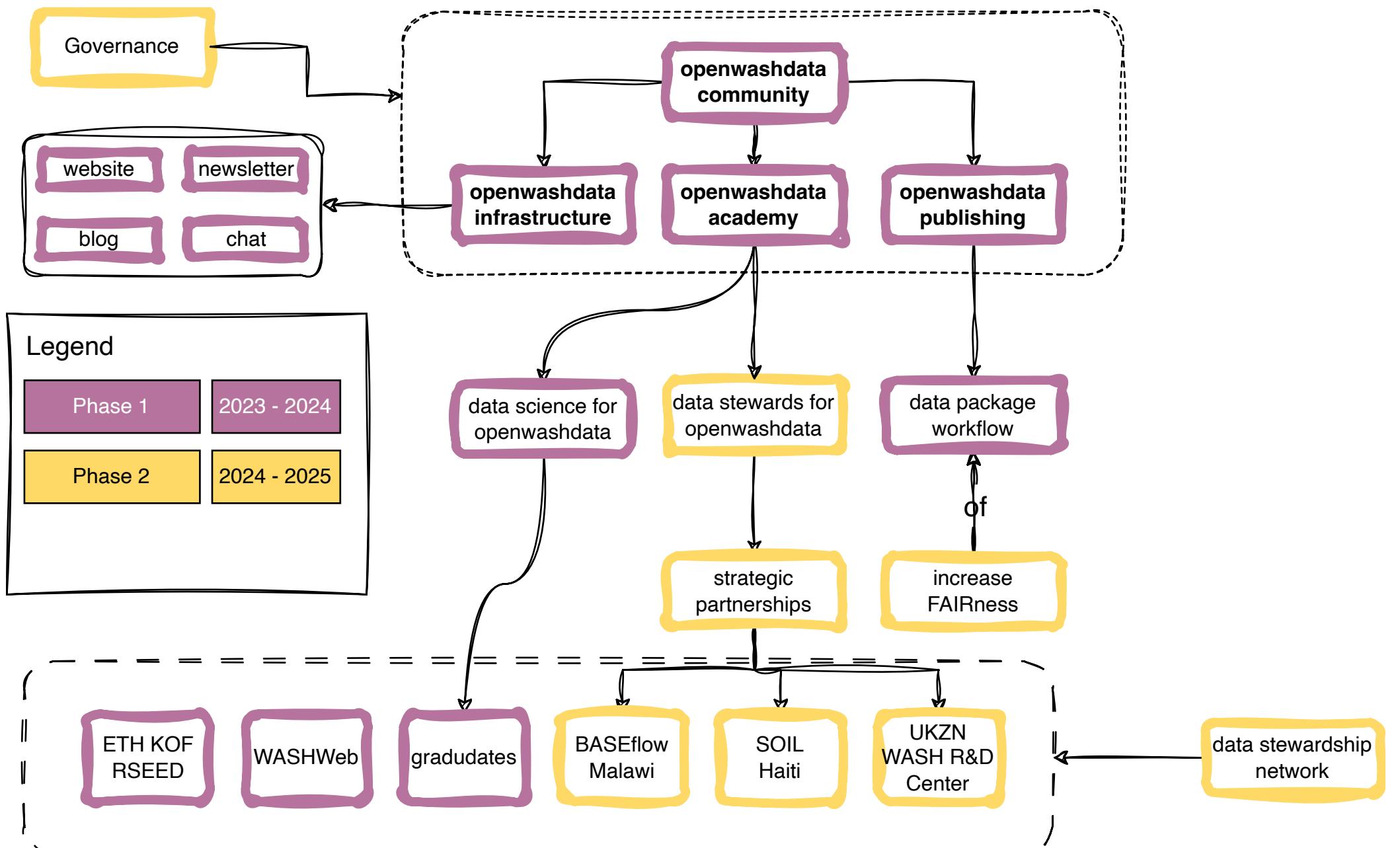


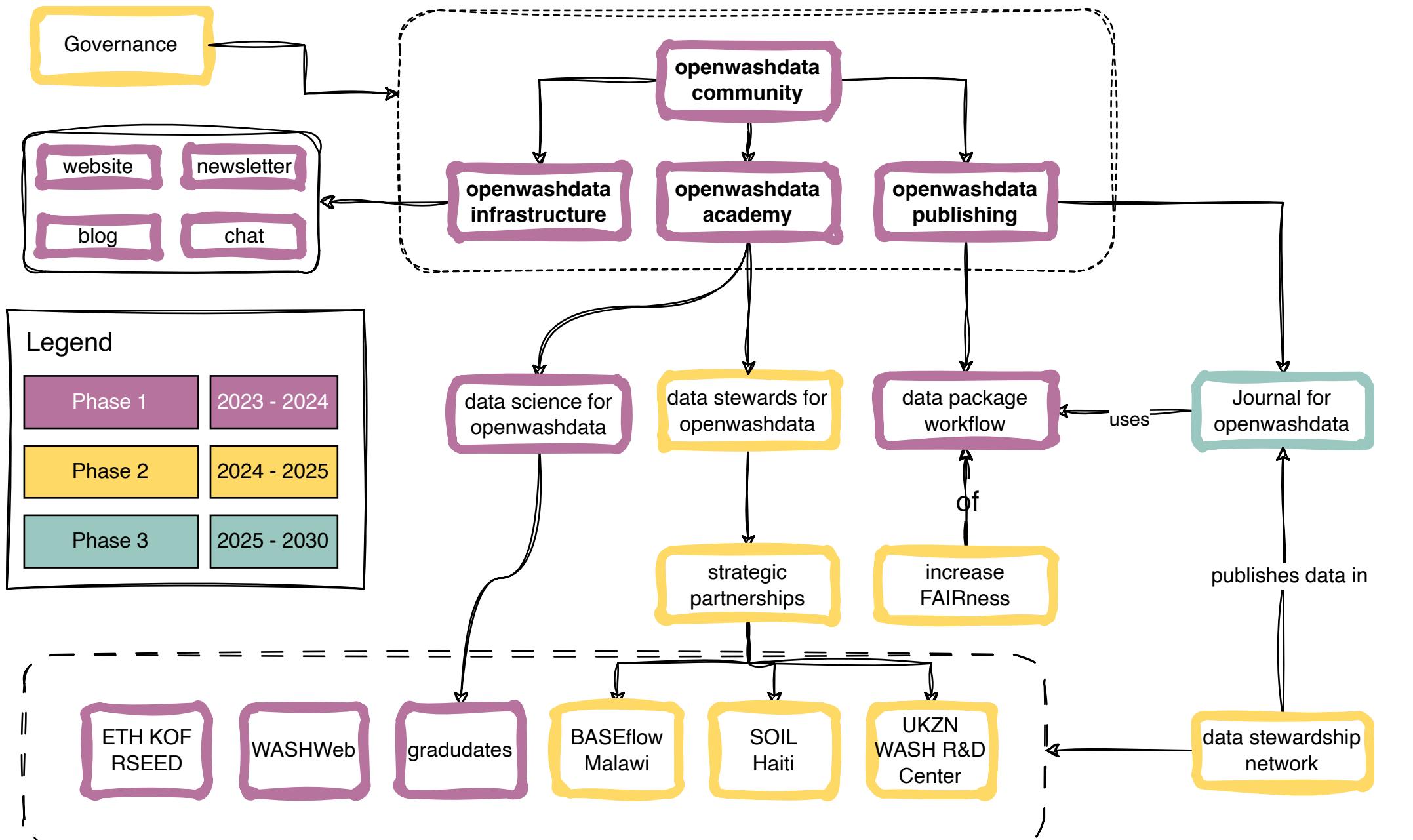












# Data stewardship (openwashdata phase II)

✋ hands up ✋

Who has an ORCID iD?



Who has published a scientific article in a journal?

# Meet a data steward

## I have:

- 10+ years work experience (5 in research)
- empathy, compassion, patience, persistance
- an affinity for IT
- teaching experience

## I don't have:

- a doctoral degree
- a qualification in computer science
- a qualification in statistics
- a lot of time

# Your turn: Think & Note

For 1 minute, think about these two questions and take some notes for later:

1. How should I be rewarded scientifically?
2. Which career paths are there for data stewards?

# Research Data Management

# Three terms for three stages

term	explanation	file format
unprocessed	data that is not processed and remains in its original form and file type	often XLSX, also CSV and others
raw data		

# Three terms for three stages

term	explanation	file format
unprocessed raw data	data that is not processed and remains in its original form and file type	often XLSX, also CSV and others
processed analysis-ready data	data that is processed to prepare for an analysis and is exported in its new form as a new file	CSV, R data package

# Three terms for three stages

term	explanation	file format
unprocessed raw data	data that is not processed and remains in its original form and file type	often XLSX, also CSV and others
processed analysis-ready data	data that is processed to prepare for an analysis and is exported in its new form as a new file	CSV, R data package
final data underlying a publication	data that is the result of an analysis (e.g descriptive statistics or data visualization) and shown in a publication, but then also exported in its new form as a new file	CSV

# Data Management Strategy

research  
questions

experimental  
design

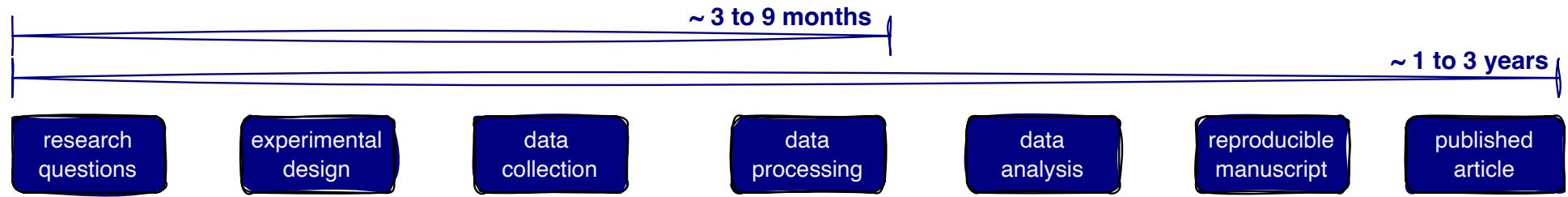
data  
collection

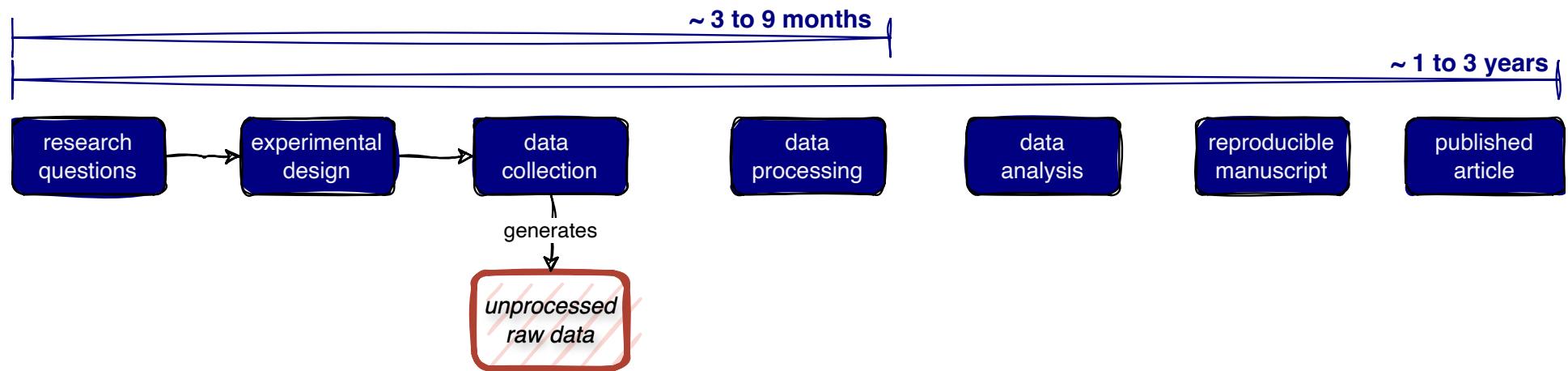
data  
processing

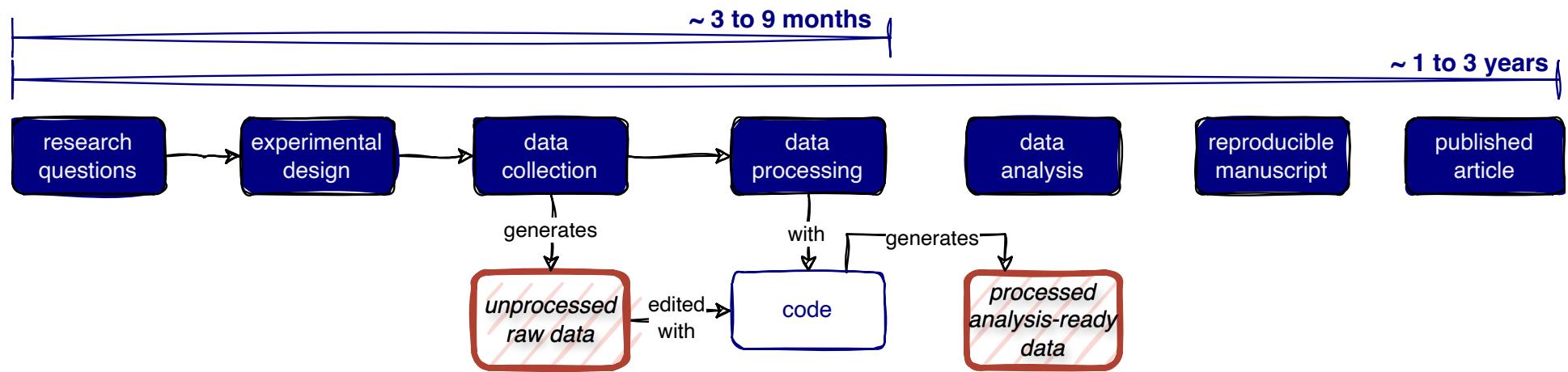
data  
analysis

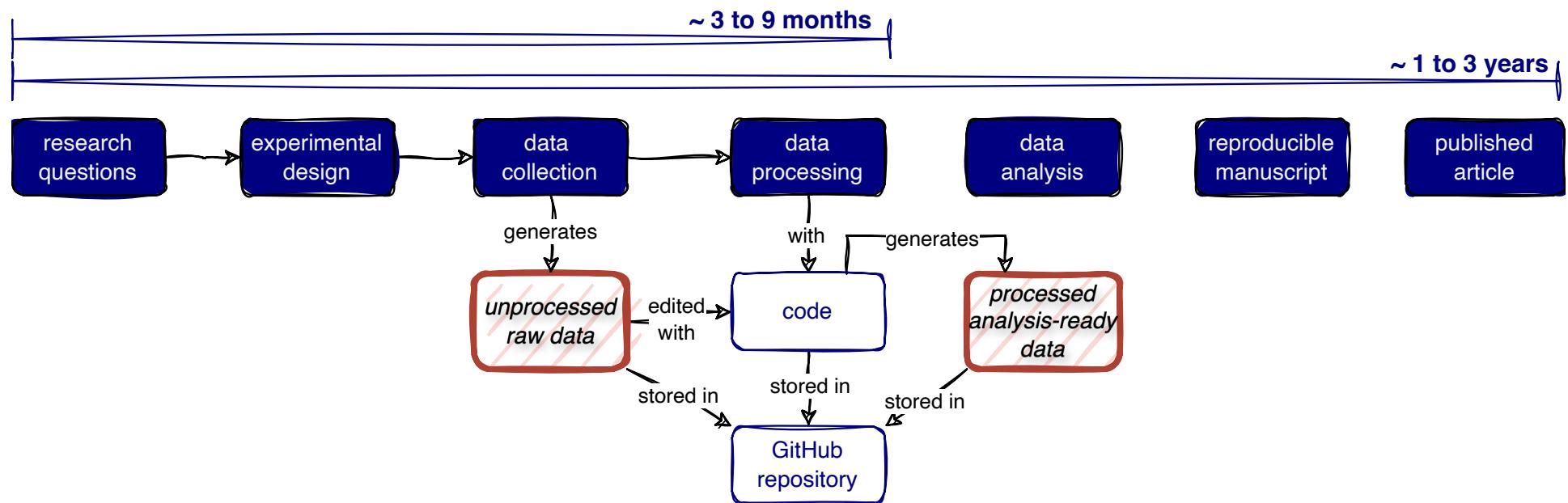
reproducible  
manuscript

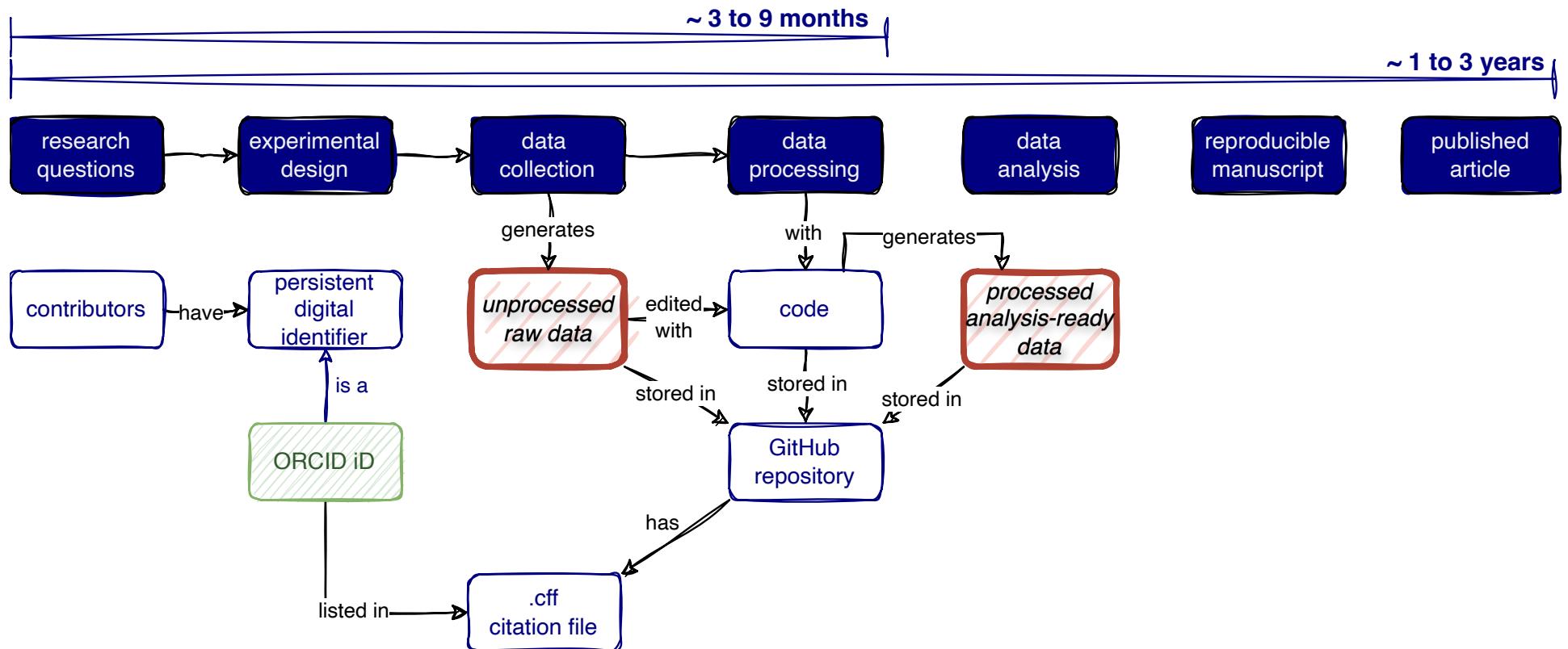
published  
article

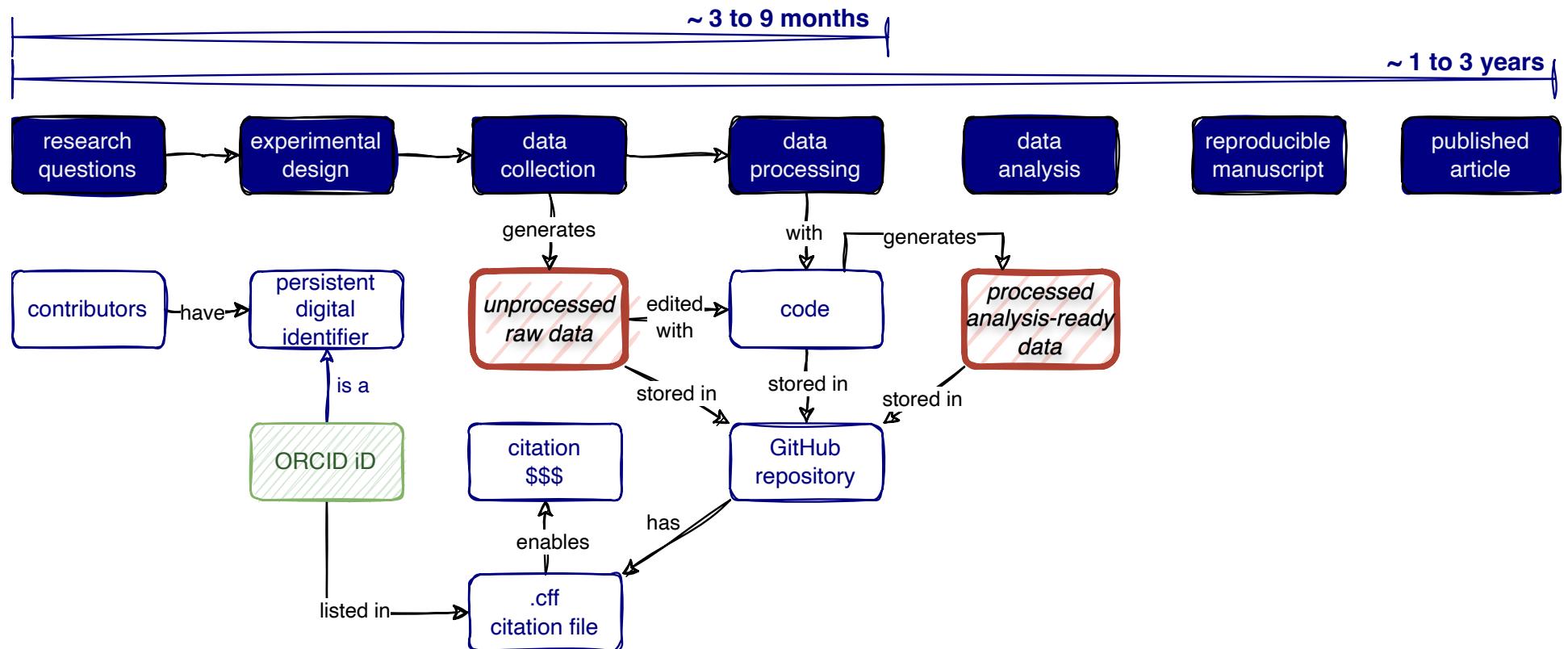


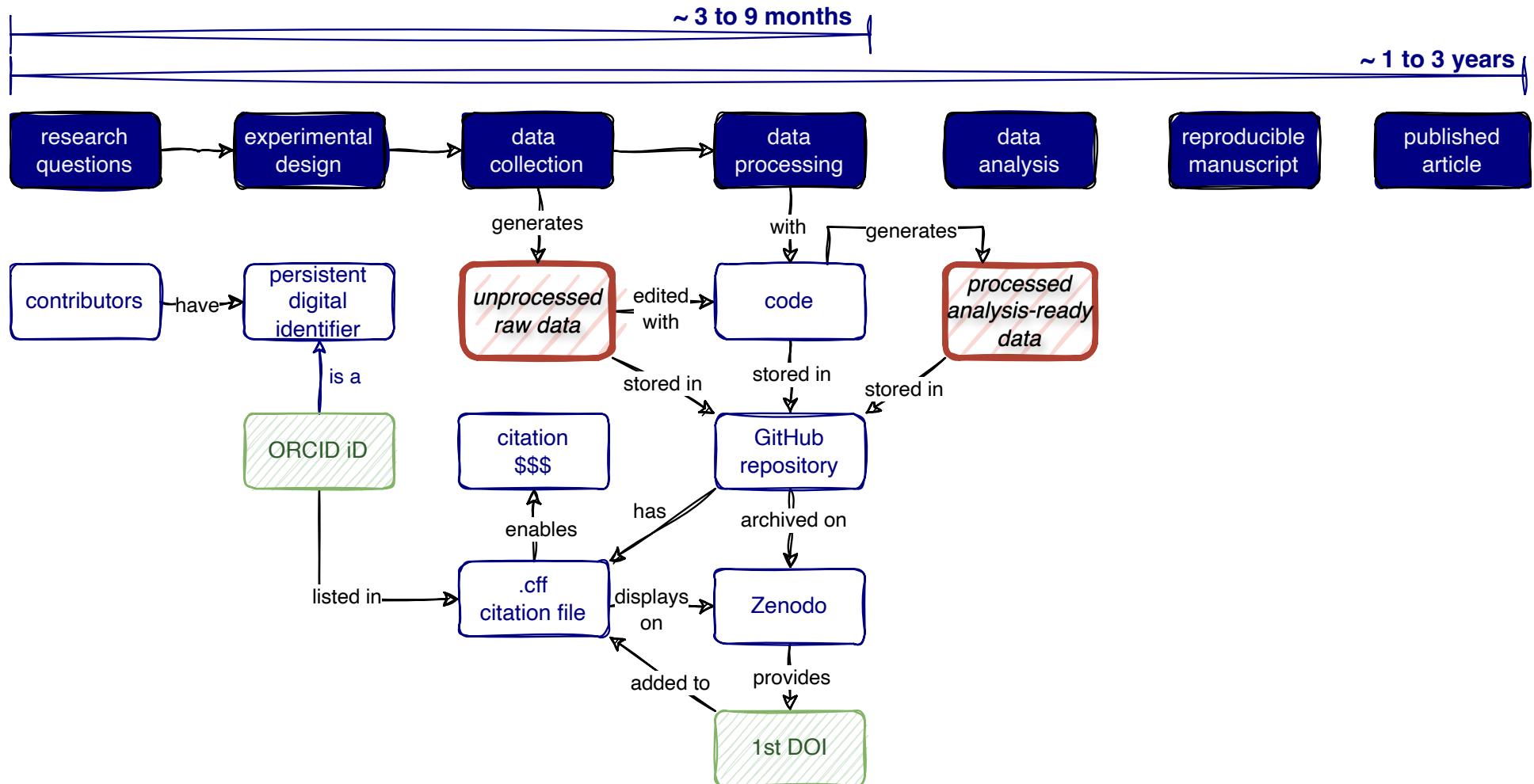


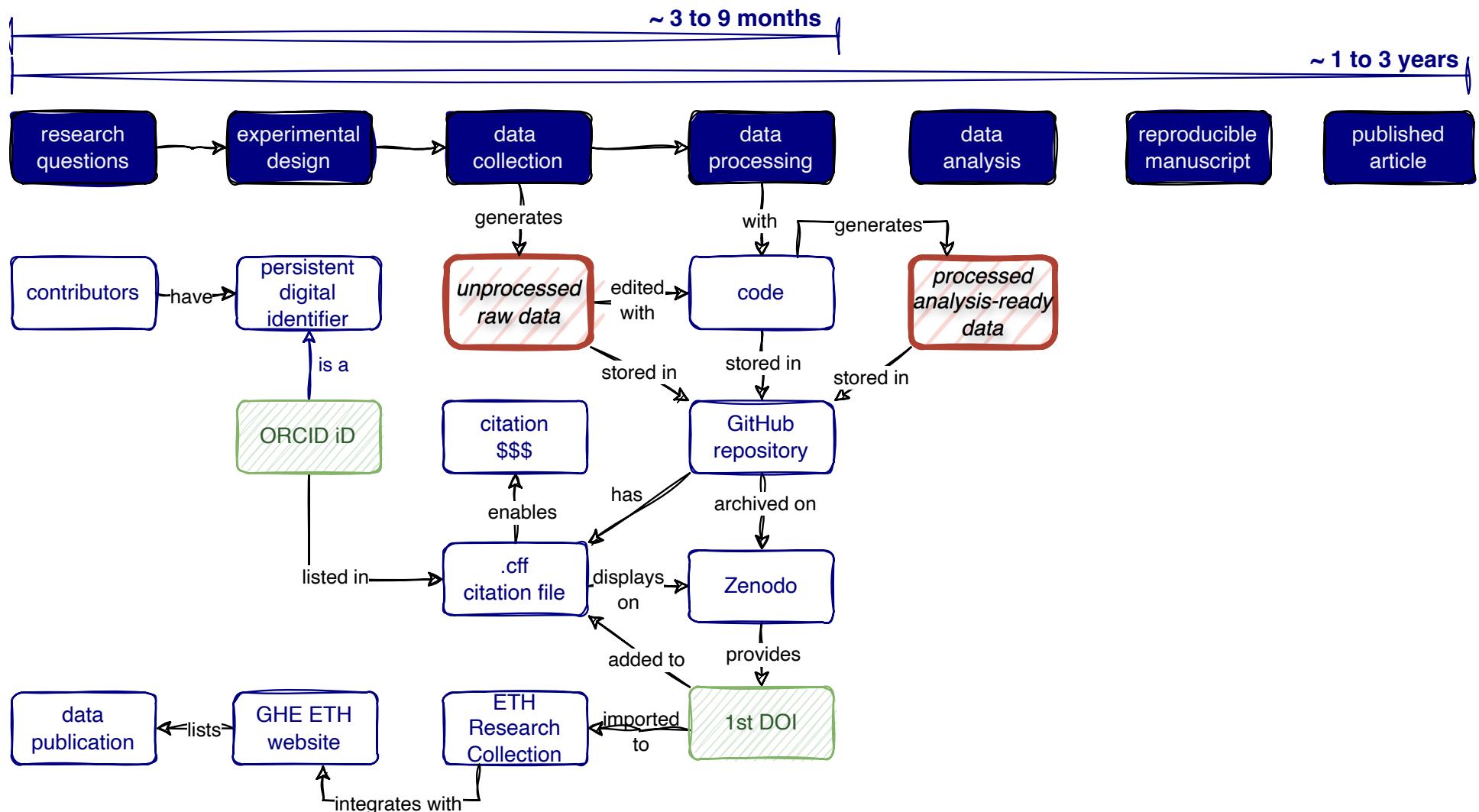


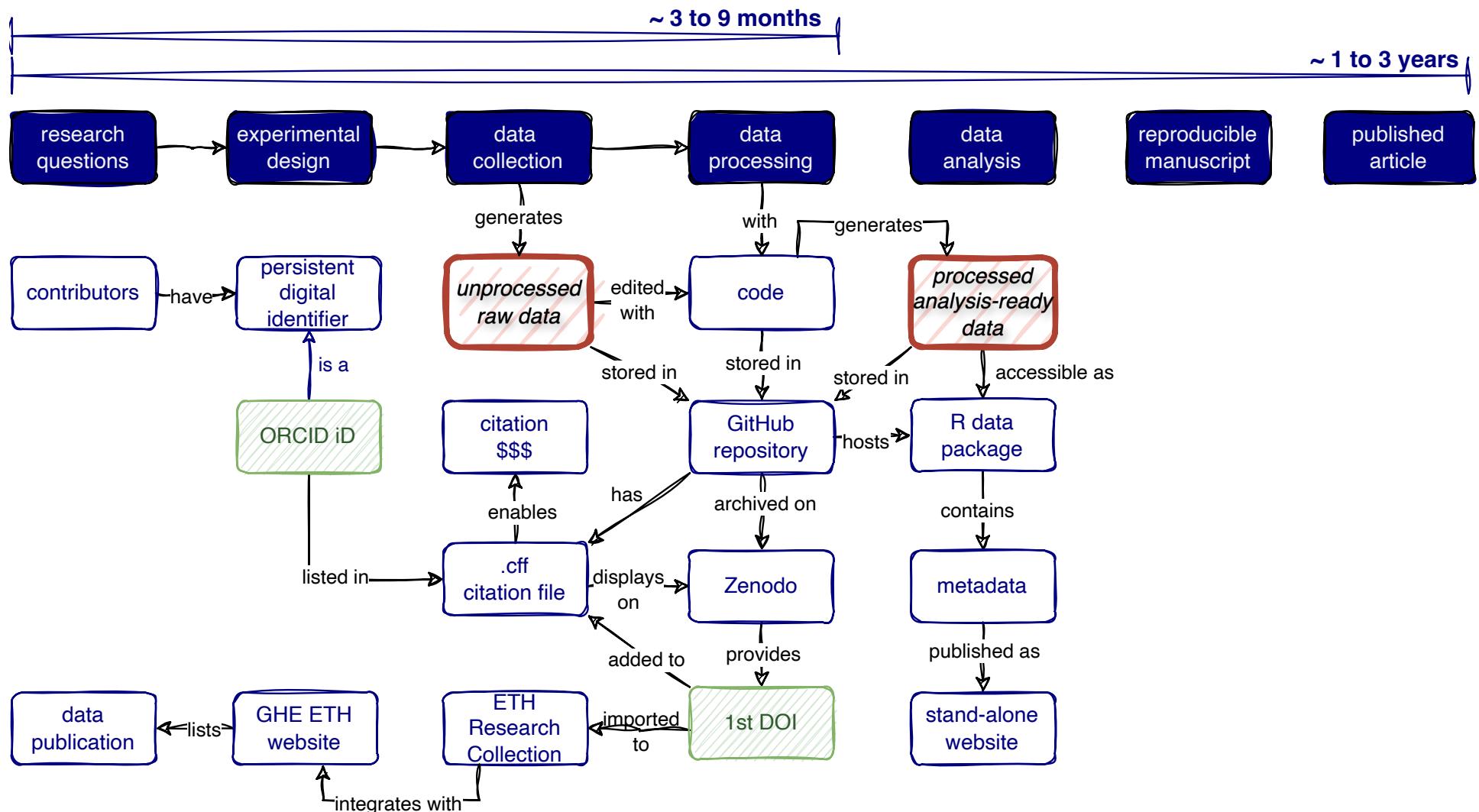


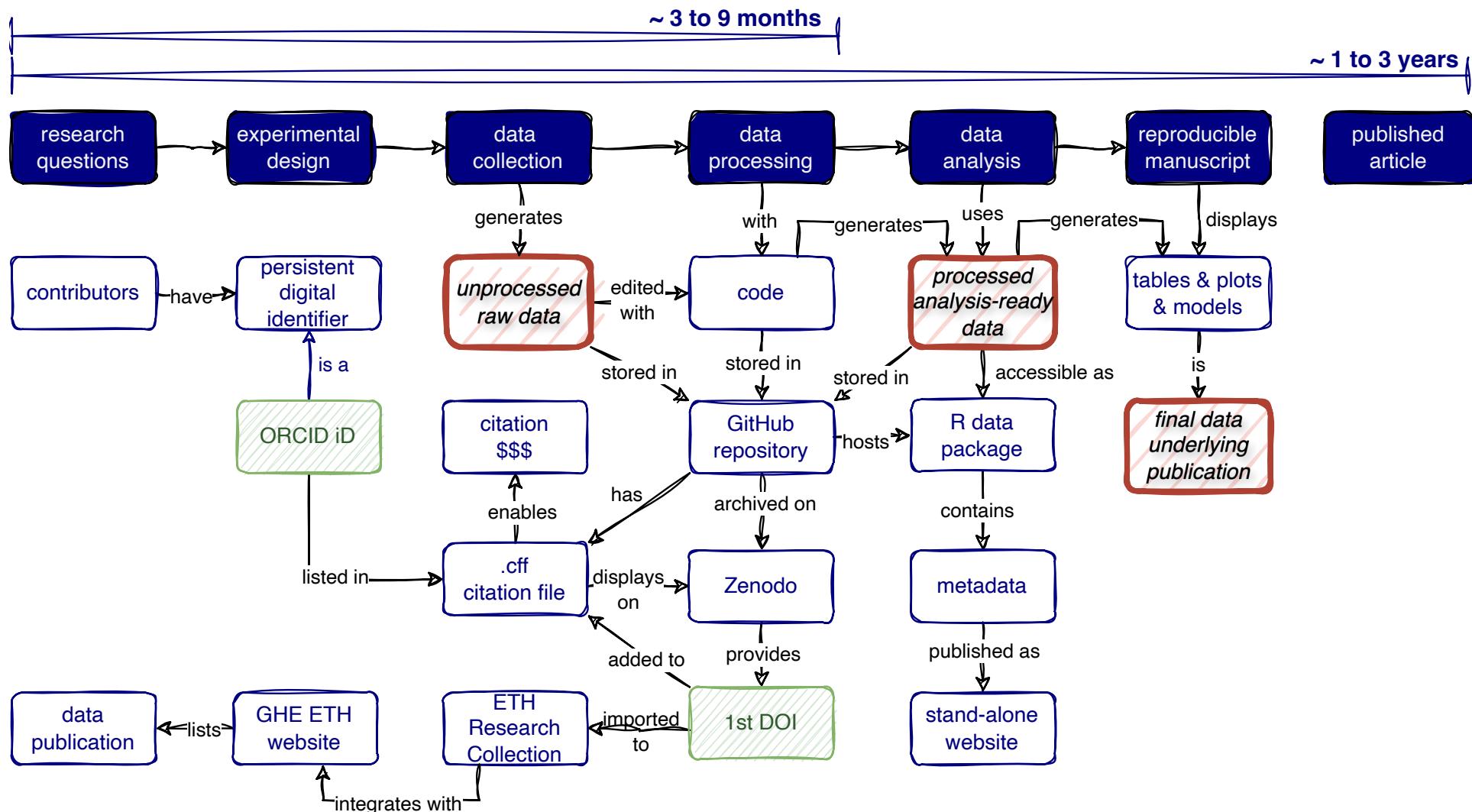


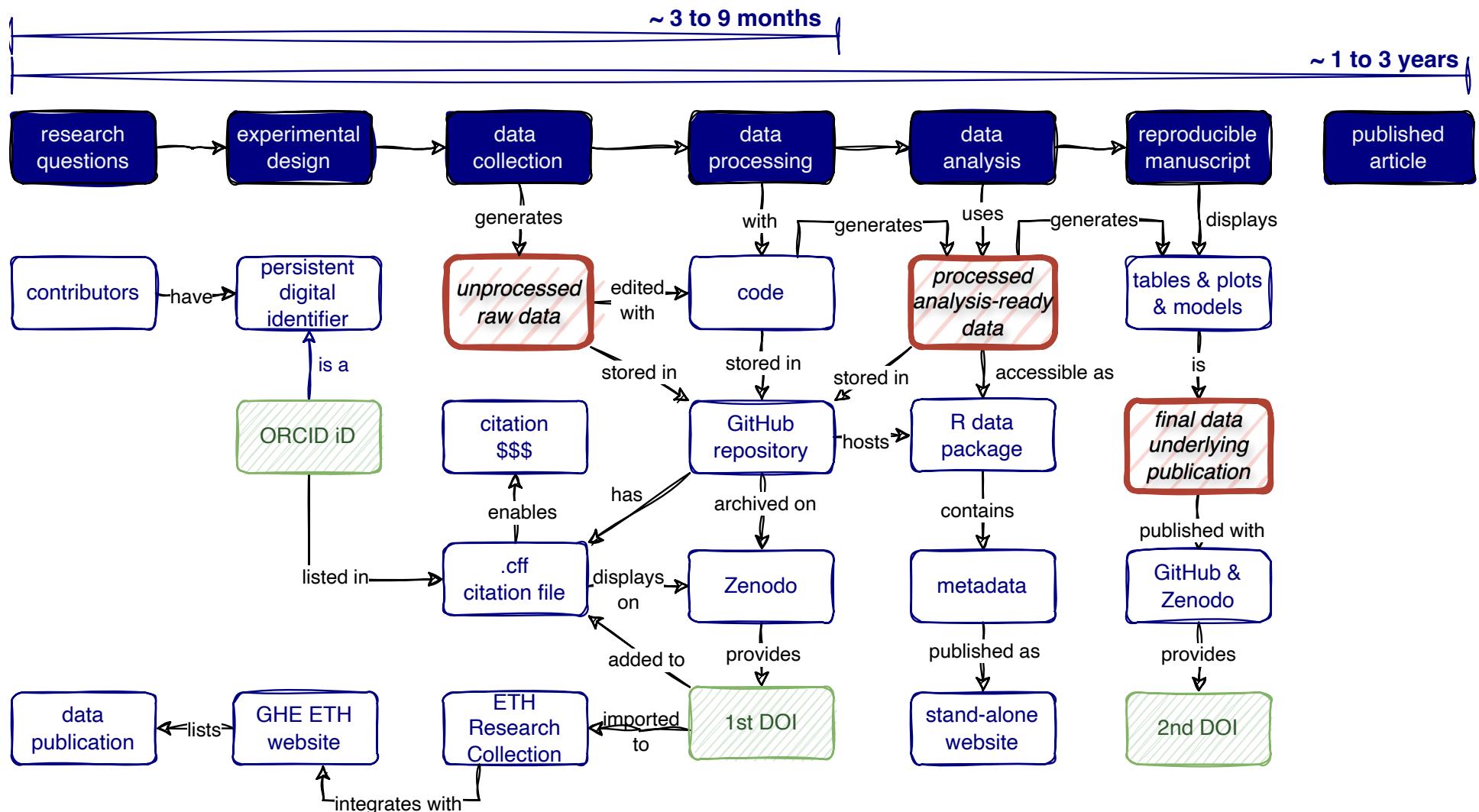


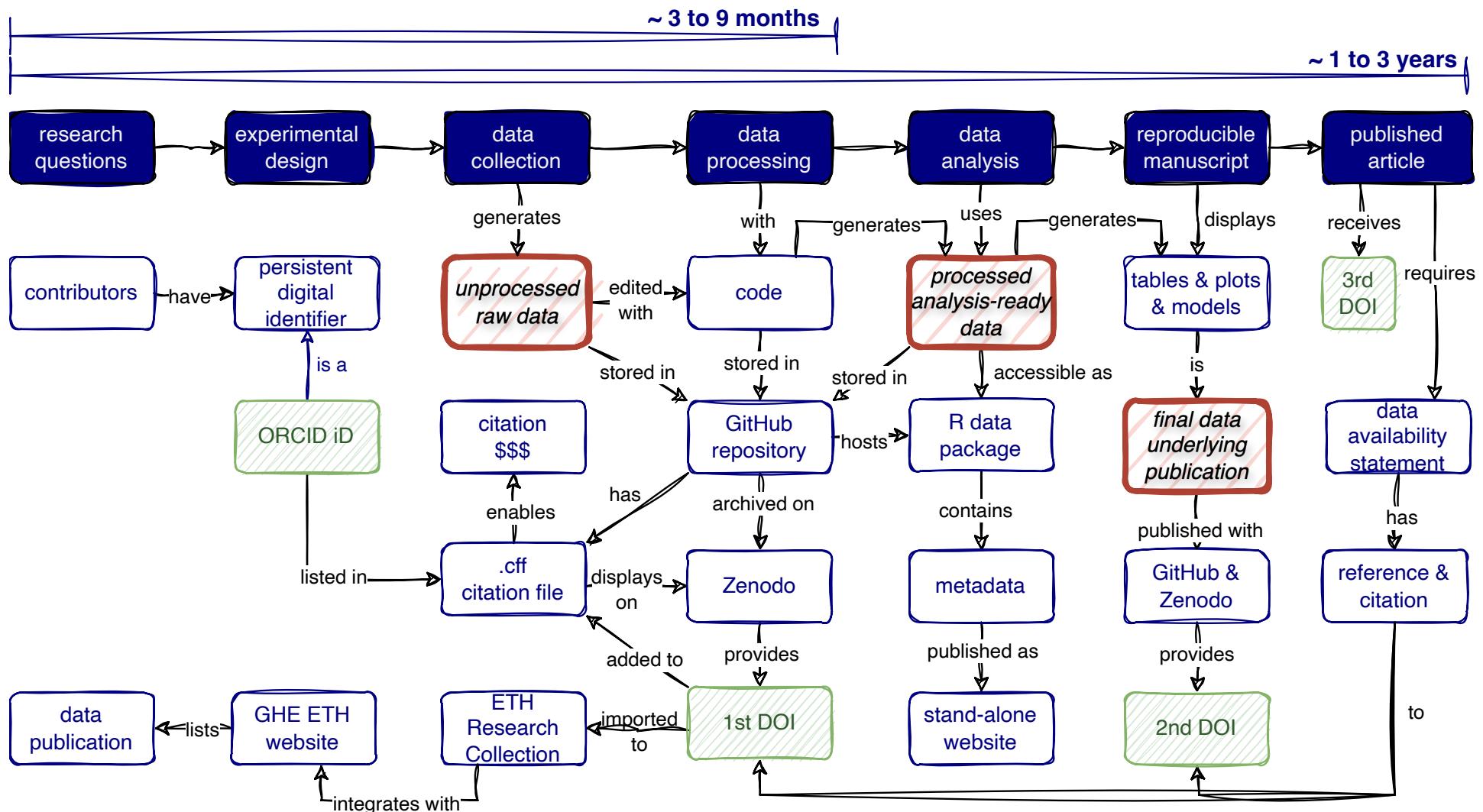


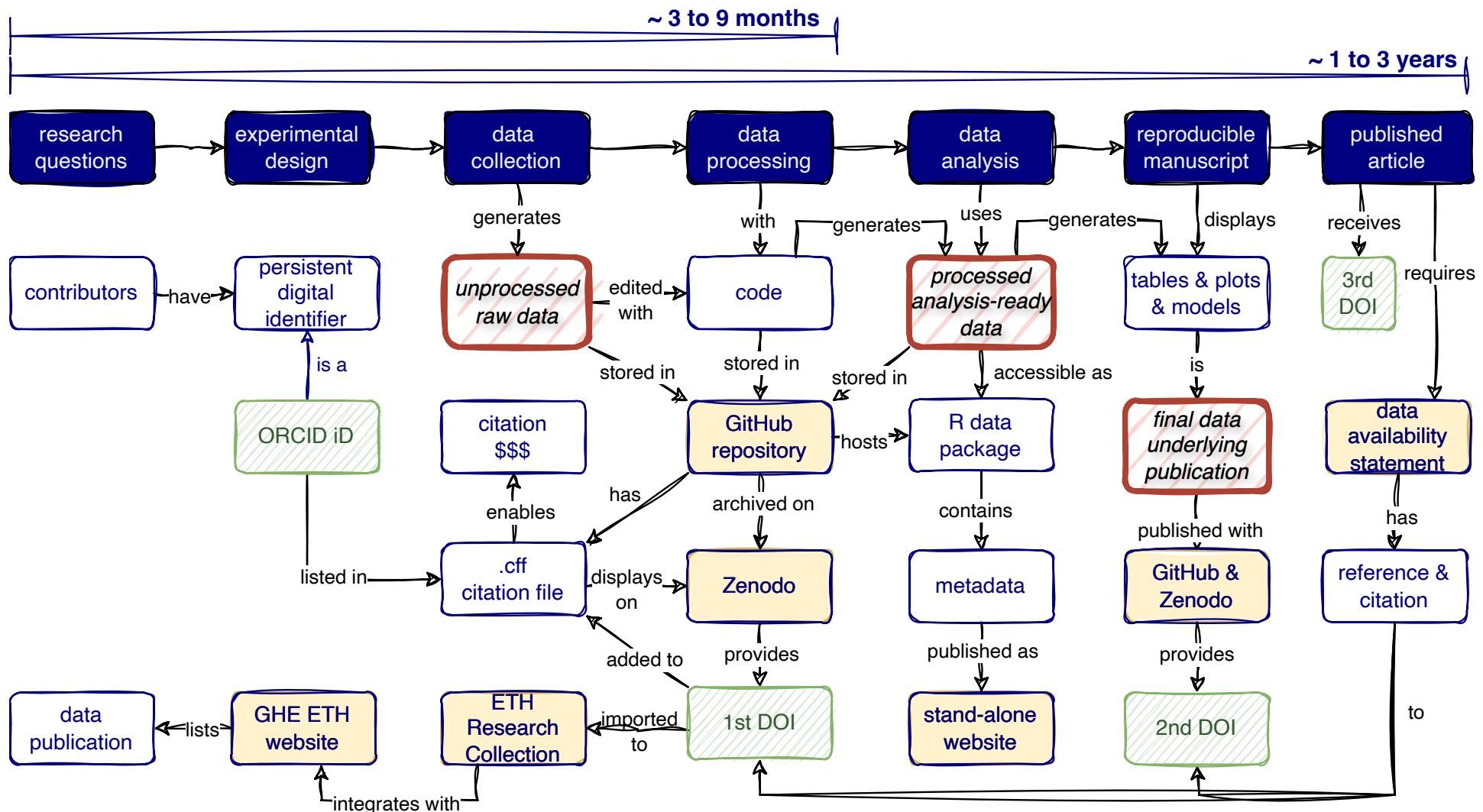












# Data steward for WASH R&D Center

- A fully funded 2-year position, hopefully extended to 5-years with 3rd party funding
- Job announced soon (2024-10-25: due date for submission of application)
- Start date 15th January 2024 (to be discussed)
- Going through a 12-month programme together with data steward at NGO BASEflow in Malawi

# Data steward activities (WP1)

- Activity 1.3: Identify how ethical approval for data collection differs for types of organizations (university, NGO) and types of data (quantitative, qualitative).
- Activity 1.4: Identify current data management practices and develop a draft data management strategy for organization.
- Activity 1.5: Publish at least 10 datasets of two different types that are available to the organization, following openwashdata data publishing workflow.

# Hands-on workshop (end Oct / beginning Nov)

At the end of the workshop, participants will be able to:

1. Describe how data published using the `washr` package follows FAIR principles compared to data shared in an appendix of a PDF or DOCX document.
2. Follow step by step instruction to create an R data package using the `washr` package.
3. Understand the difference between human-readable and machine-readable documentation.

# News

# Support us: 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>.