

SST Educational Workshop

Data Stewardship

Lars Schöbitz

lschoebitz@ethz.ch

Global Health Engineering - ETH Zurich

November 17, 2023

Hello!

Lars Schöbitz



- Environmental Engineer
- Open Science Specialist at Global Health Engineering
- **RStudio certified instructor** for Data Science with R

Data Stewardship

The Vision?

An active community that applies FAIR principles (Wilkinson et al. 2016) to data generated in the greater sector of the Swiss Society of Toxicology.

- Findable
- Accessible
- Interoperable
- Reusable

A set of technical principles that can be applied to data.

The Opportunity

Journal Articles

The screenshot shows a web browser window with a purple header bar containing three dots and a red 'E' icon. The main title in the header is "The role of emptying services". The address bar shows the URL: scencedirect.com/science/article/pii/S0301479721006745?via%3Dihub#appsec1. The page itself is from ScienceDirect, featuring the Elsevier logo and the journal "Journal of Environmental Management". The article title is "The role of emptying services in provision of safely managed sanitation: A classification and quantification of the needs of LMICs". It lists authors: Nicola Greene^a, Sarah Hennessy^a, Tate W. Rogers^a, Jocelyn Tsai^b, Francis L. de los Reyes III^b, and provides links to "View PDF" and "Download full issue". On the left, there's a sidebar with links like "Outline", "Highlights", "Abstract", etc. On the right, sections include "Part of special issue", "FSM5 Conference", "View special issue", "Recommended articles", and "View PDF". At the bottom right is a "FEEDBACK" button.



Who has published a scientific article in a journal?

 hands up 

Who has published data as supplemental material?

Journal Articles

The screenshot shows a web browser window with a purple header bar. The title bar says "E The role of emptying services" and the address bar shows the URL "sciencedirect.com/science/article/pii/S0301479721006745?via%3Dihub#appsec1". The main content area displays a journal article. On the left, there's a sidebar with links: "2. Materials and methods", "3. Results and discussion", "4. Conclusions", "Funding acquisition", "Author contributions", "Funding sources", "Declaration of competing inter...", "Appendix A. Supplementary data" (which is underlined and highlighted in red), "References", and "Show full outline ▾". The main content area has a blue button "View PDF" and a link "Download full issue". The text in the main area discusses the influence of certain factors on the work reported in the paper, mentions "Appendix A. Supplementary data", provides a download link for supplementary data, describes a multimedia component, lists references, and includes a citation for Bassan and Strande (2013). At the bottom right, there's a "FEEDBACK" button with a speech bubble icon.

2. Materials and methods
3. Results and discussion
4. Conclusions
Funding acquisition
Author contributions
Funding sources
Declaration of competing inter...
Appendix A. Supplementary data
References
Show full outline ▾

[View PDF](#) Download full issue

appeared to influence the work reported in this paper.

Appendix A. Supplementary data

The following is the supplementary data to this article:

[Download : Download Word document \(152KB\)](#)

Multimedia component 1.

References

Bassan and Strande, 2013 M. Bassan, L. Strande
Characterization of Faecal Sludge in Dry and Rainy Seasons

<https://www.sciencedirect.com/science/article/pii/S0301479721006745?via%3Dihub#appsec1> (2013)

FEEDBACK

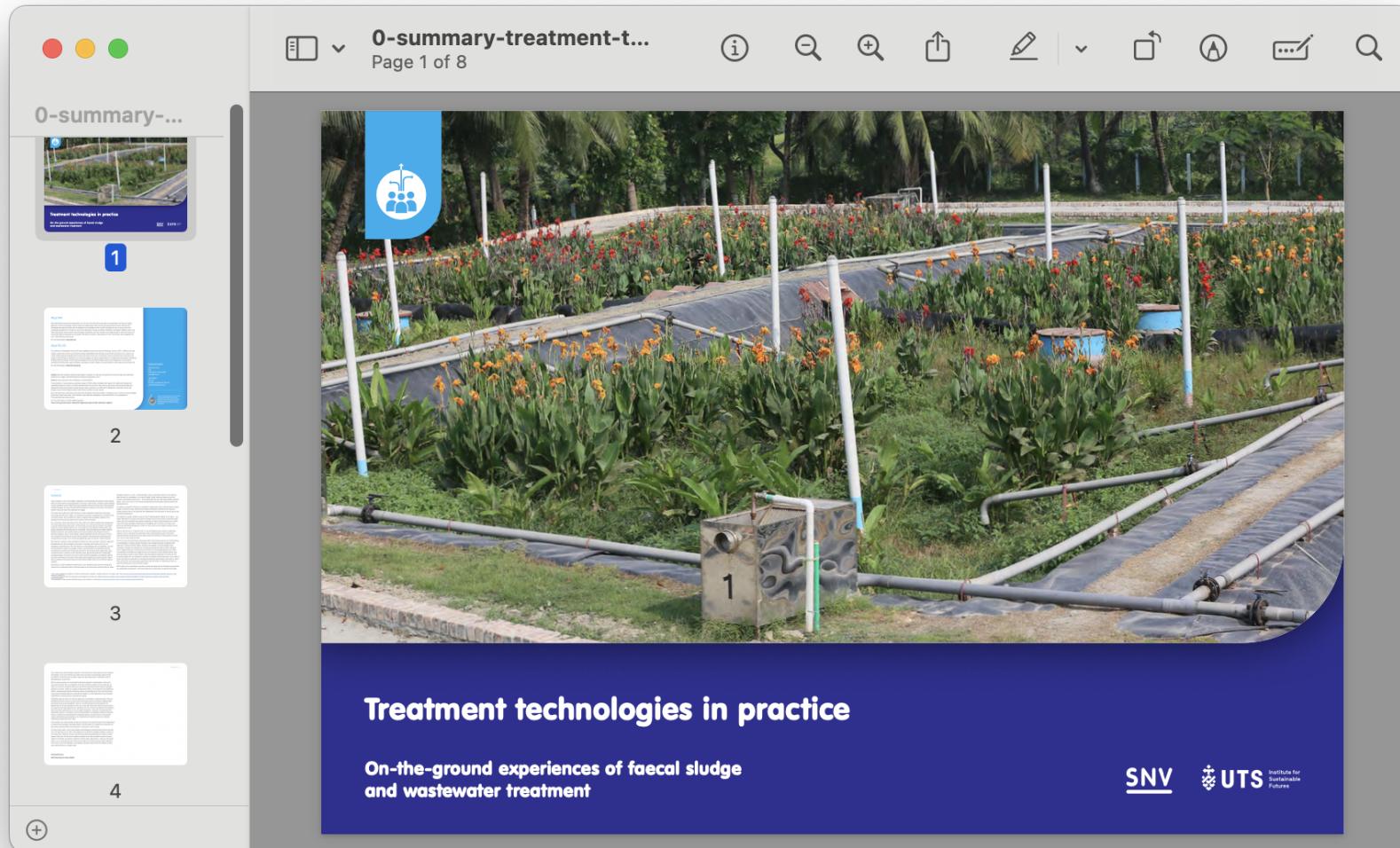
Journal Articles

The screenshot shows a Microsoft Word document titled "1-s2.0-S0301479721006745-mmcl — Saved to my Mac". The ribbon menu is visible at the top, showing Home, Insert, Draw, Design, Layout, References, Mailings, Review, View, Zotero, Tell me, Comments, Editing, Share, and other options. The main content area contains a table titled "Table 1. The number of onsite sanitation facilities alphabetized by country per Service Type. The Service Type Density is calculated by dividing the population using the specified service by the total country population. Thus, the Service Type Density covers users of severs, onsite facilities, and open defecators. The number of facilities required for Open Defecators is estimated assuming the facilities will not be shared between households. Urban and rural proportions of the onsite facilities are provided. "ND" indicates no data available from JMP." The table has four main sections: Mechanized, Non-Mechanized, Unemptiable, and Open Defecation. Each section contains data for No. of Facilities, Service Type Density, and Urban/Rural proportions. The table includes data for 10 countries: Afghanistan, Algeria, Angola, Anguilla, Argentina, Armenia, Azerbaijan, Bangladesh, and Belize. The "Unemptiable" section is underlined in red.

Country	Mechanized				Non-Mechanized				Unemptiable				Open Defecation			
	No. of Facilities	Service Type Density	Urban	Rural	No. of Facilities	Service Type Density	Urban	Rural	No. of Facilities	Service Type Density	Urban	Rural	No. of Facilities Required	Service Type Density	Urban	Rural
Afghanistan	367,345	10%	66%	34%	1,599,720	41%	31%	69%	1,351,493	34%	12%	88%	563,238	13%	0%	100%
Algeria	482,751	6%	27%	73%	465,887	6%	18%	82%	255,876	3%	63%	37%	64,417	1%	18%	82%
Angola	1,054,451	21%	97%	3%	1,817,690	36%	75%	25%	516,612	10%	44%	56%	1,232,071	20%	3%	97%
Anguilla	2,875	72%	100%	0%	1,056	26%	100%	0%	15	0%	100%	0%	22	1%	100%	0%
Argentina	3,266,830	25%	100%	0%	1,178,771	9%	100%	0%	ND	ND	ND	ND	215,594	2%	100%	0%
Armenia	129,305	18%	2%	98%	46,930	6%	2%	98%	42,431	6%	3%	97%	ND	ND	ND	ND
Azerbaijan	710,244	33%	41%	59%	477,267	22%	22%	78%	100,844	5%	0%	100%	3,357	0%	0%	100%
Bangladesh	1,809,175	8%	90%	10%	14,713,956	58%	30%	70%	7,439,355	29%	22%	78%	ND	ND	ND	ND
Belize	53,989	56%	46%	54%	30,453	32%	33%	67%	2,199	2%	10%	90%	804	1%	44%	56%

Page 2 of 8 2547 words English (Ireland) Accessibility: Investigate Focus + 150%

PDF reports



PDF reports

The screenshot shows a PDF document with a table comparing wastewater qualities. The table has three columns: Parameter, Inlet, and Outlet. The Inlet column lists ranges of values, while the Outlet column lists specific target values.

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 ₅	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 ₄	108, 04-568, 72 mg/L	54, 21-150, 50 mg/L
Ammonia, NH ₃ -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

PDF reports

1-mechanised-26conventional-f...

Page 5 of 8 — Edited

i Q Q ↗ P A ⌂

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 ₅	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 ₄	108, 04-568, 72 mg/L	54, 21-150, 50 mg/L
Ammonia, NH ₃ -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



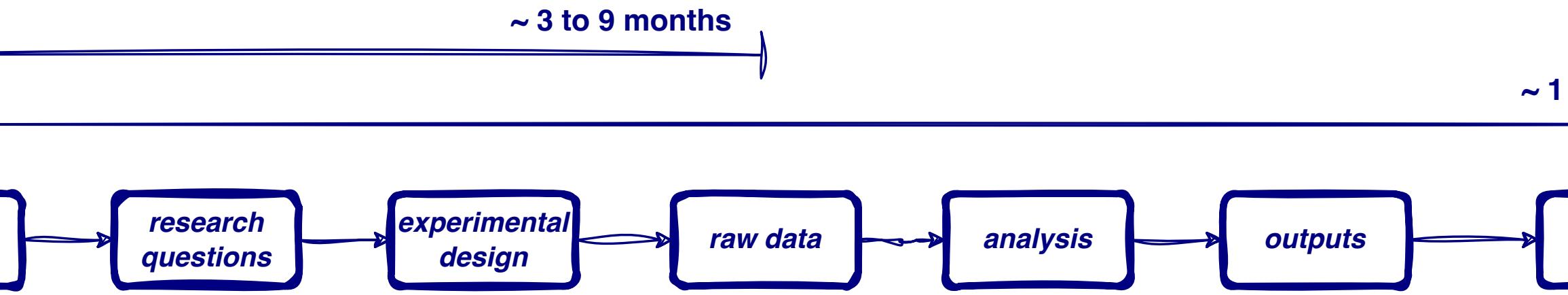
Who has published data, separate from an article/report, and assigned a unique DOI?

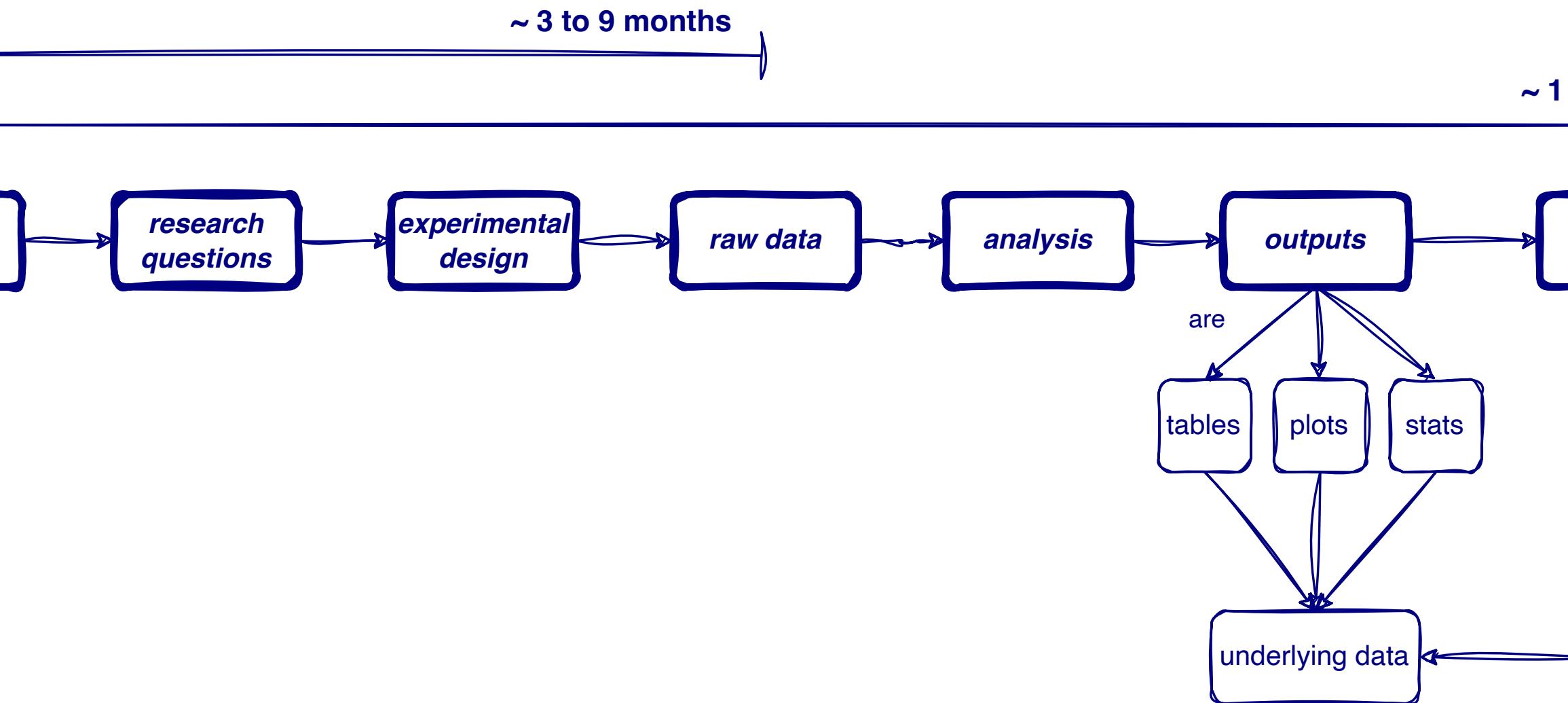


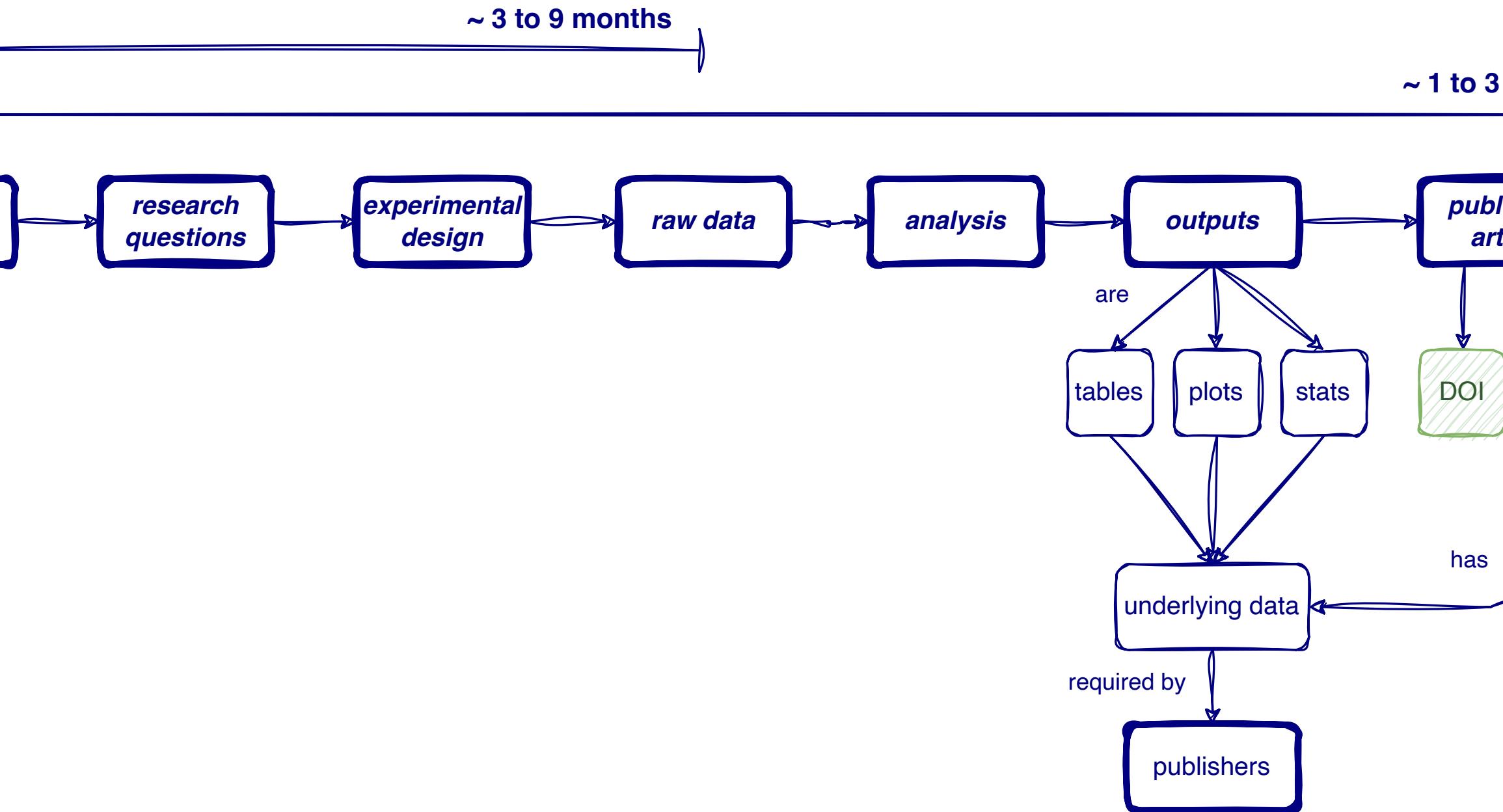
Who has an ORCID iD
(account on: <https://orcid.org/>)

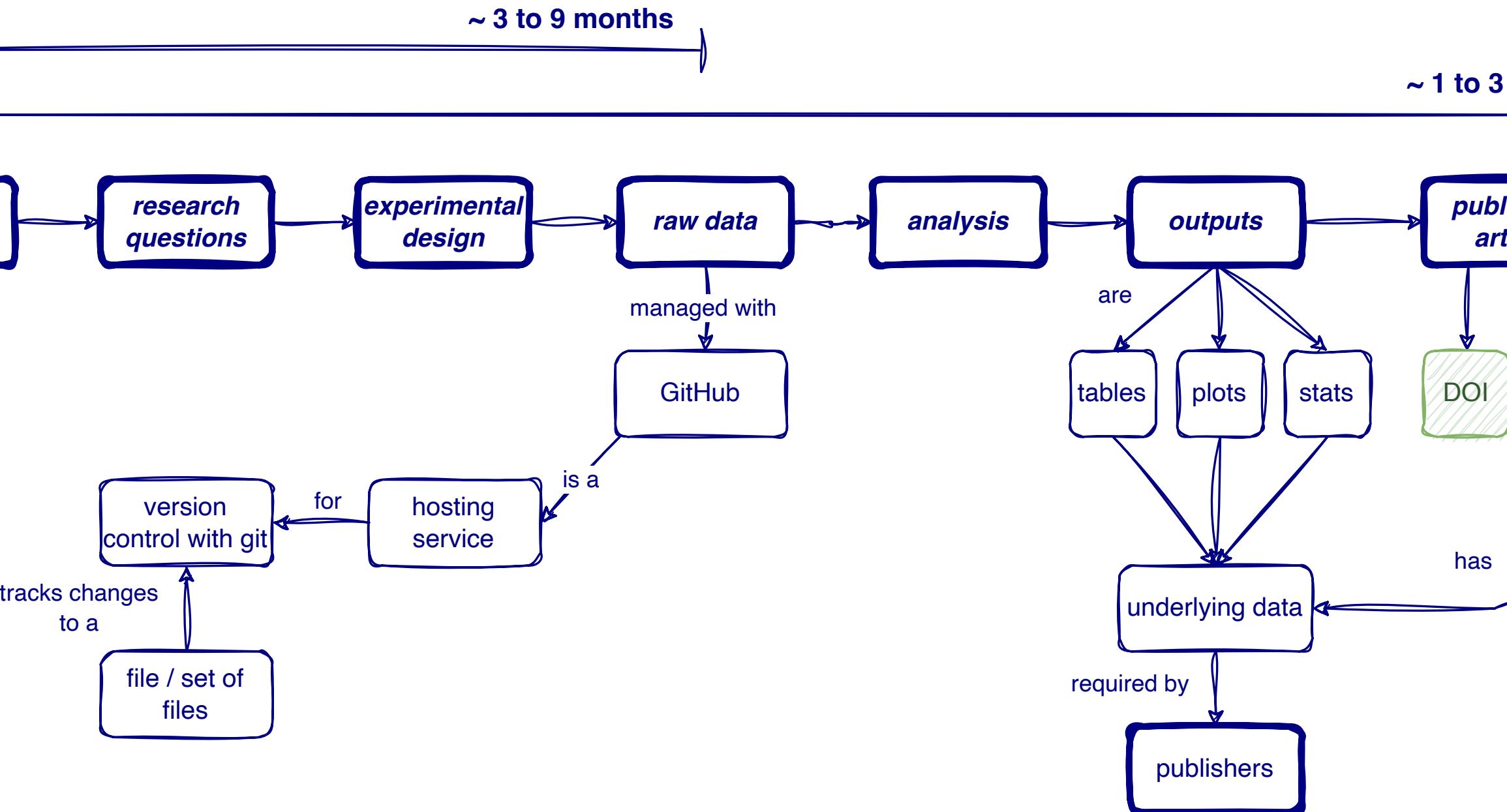
Our RDM workflow at GHE

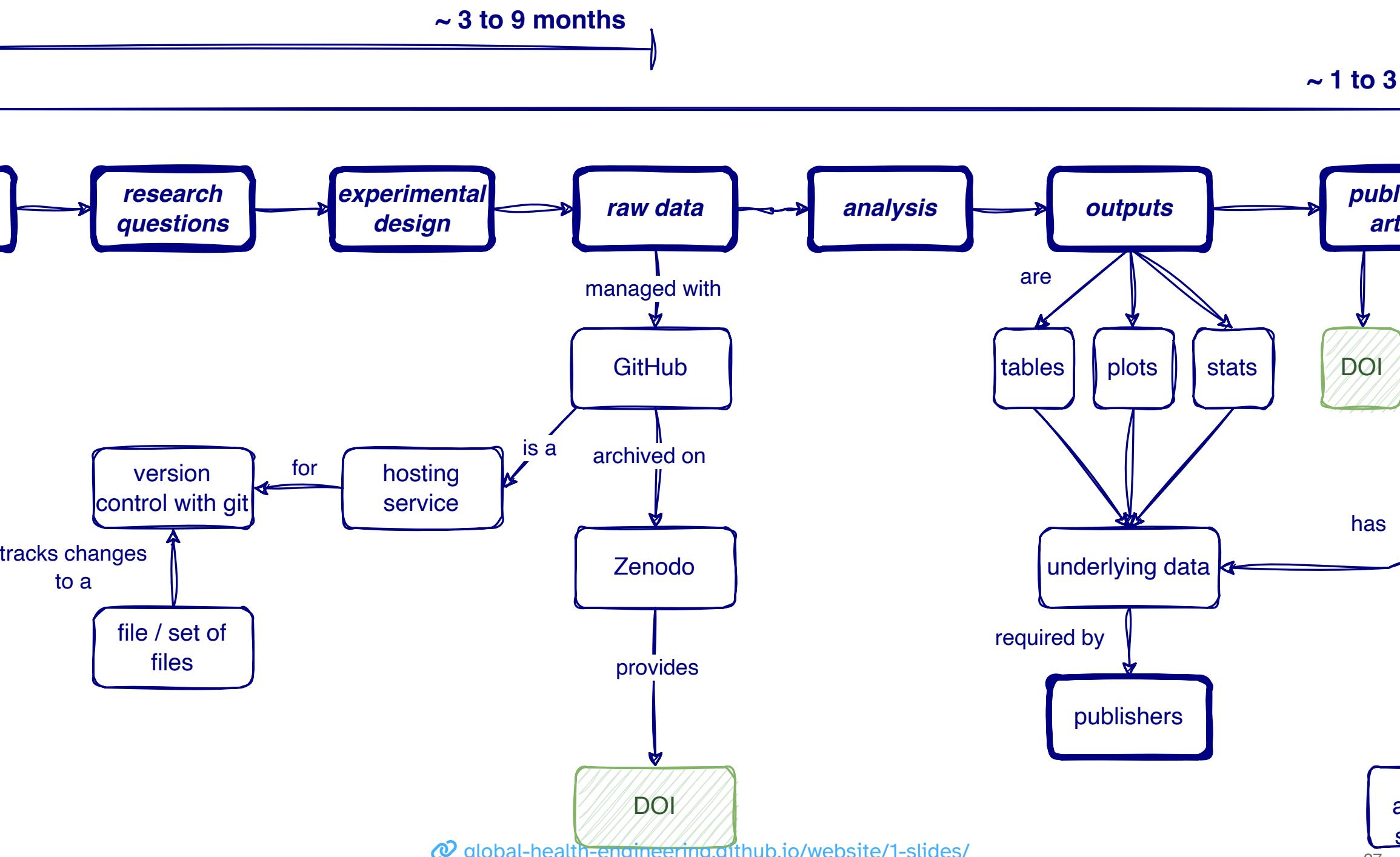






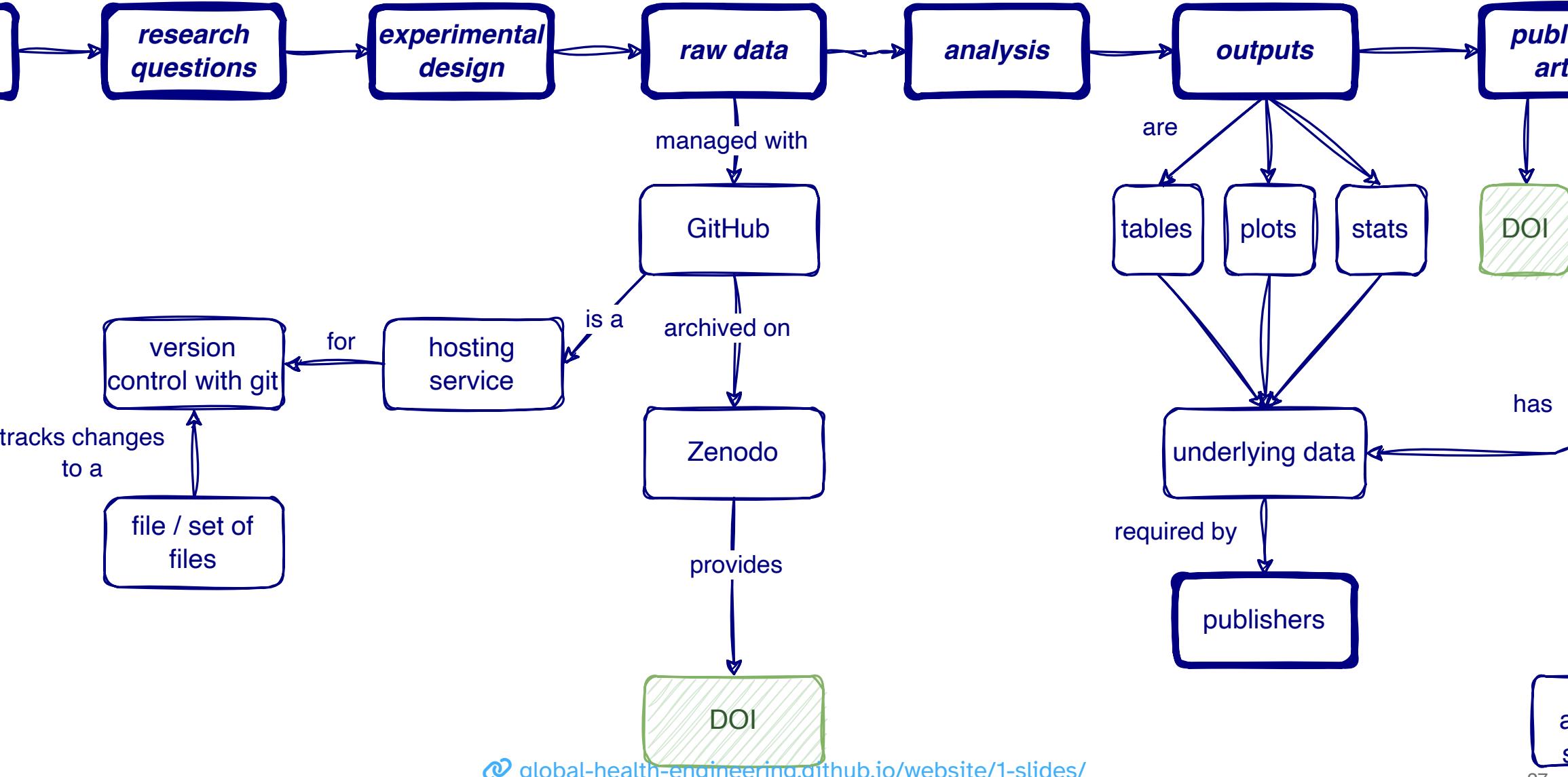






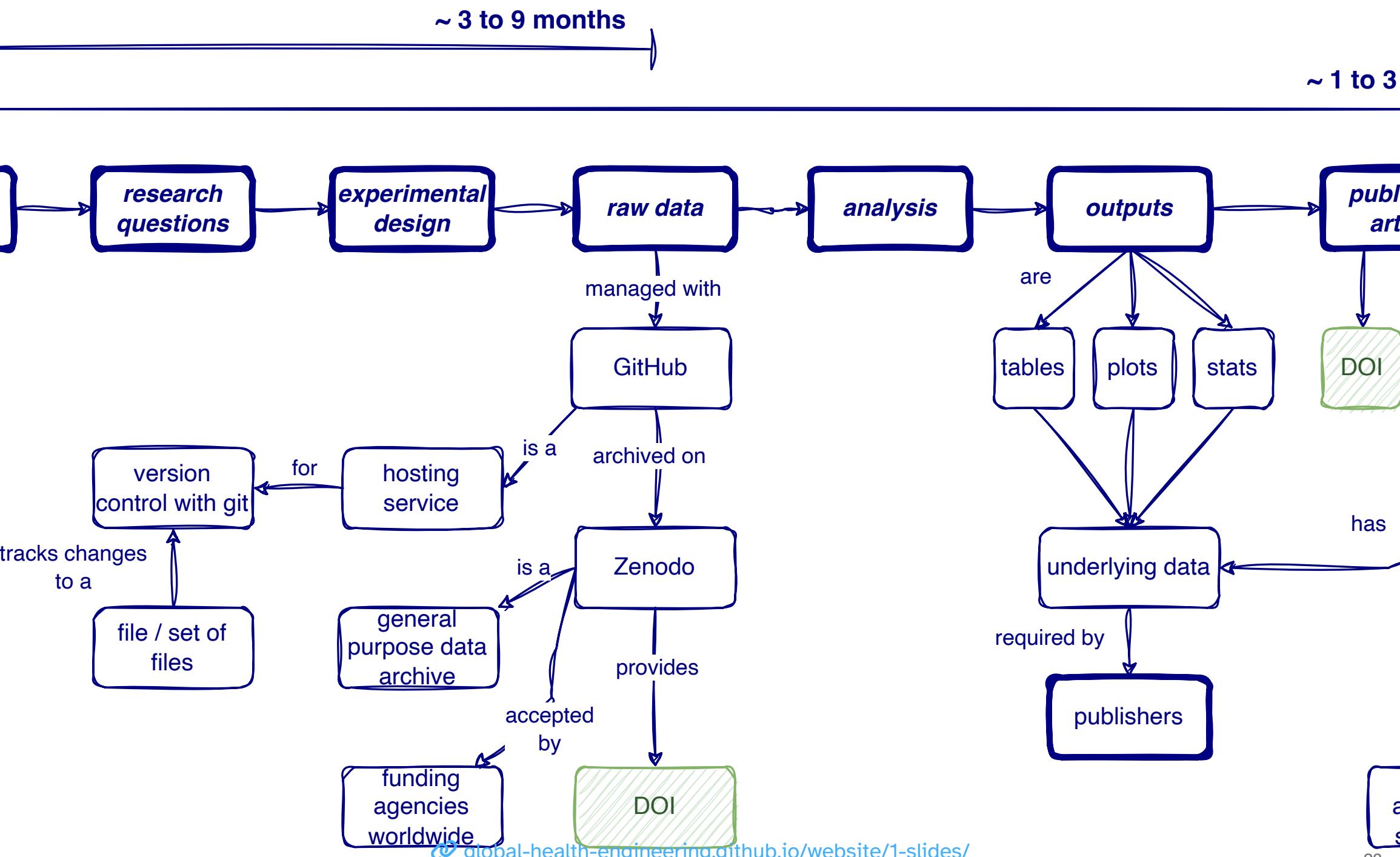
~ 3 to 9 months

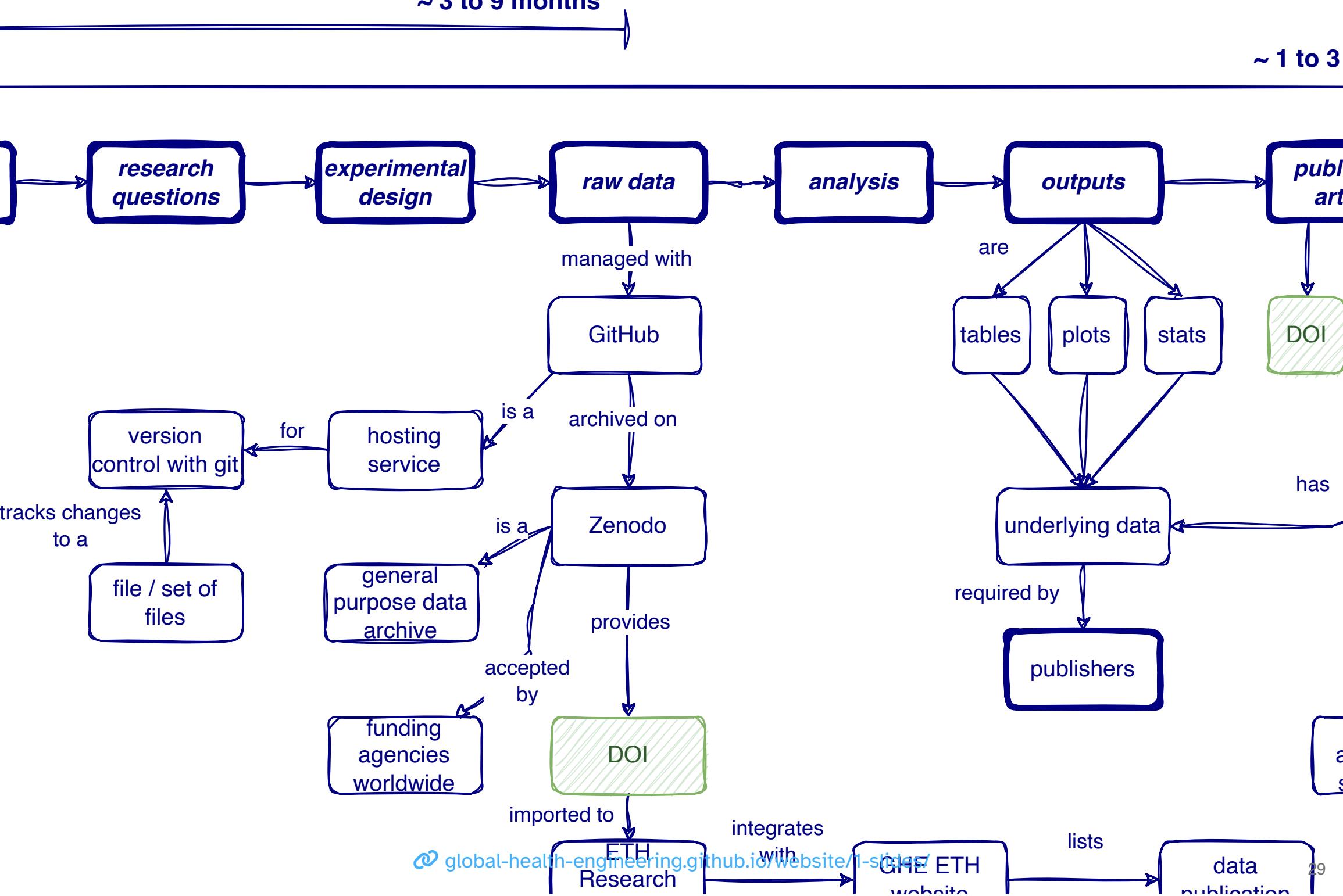
~ 1 to 3

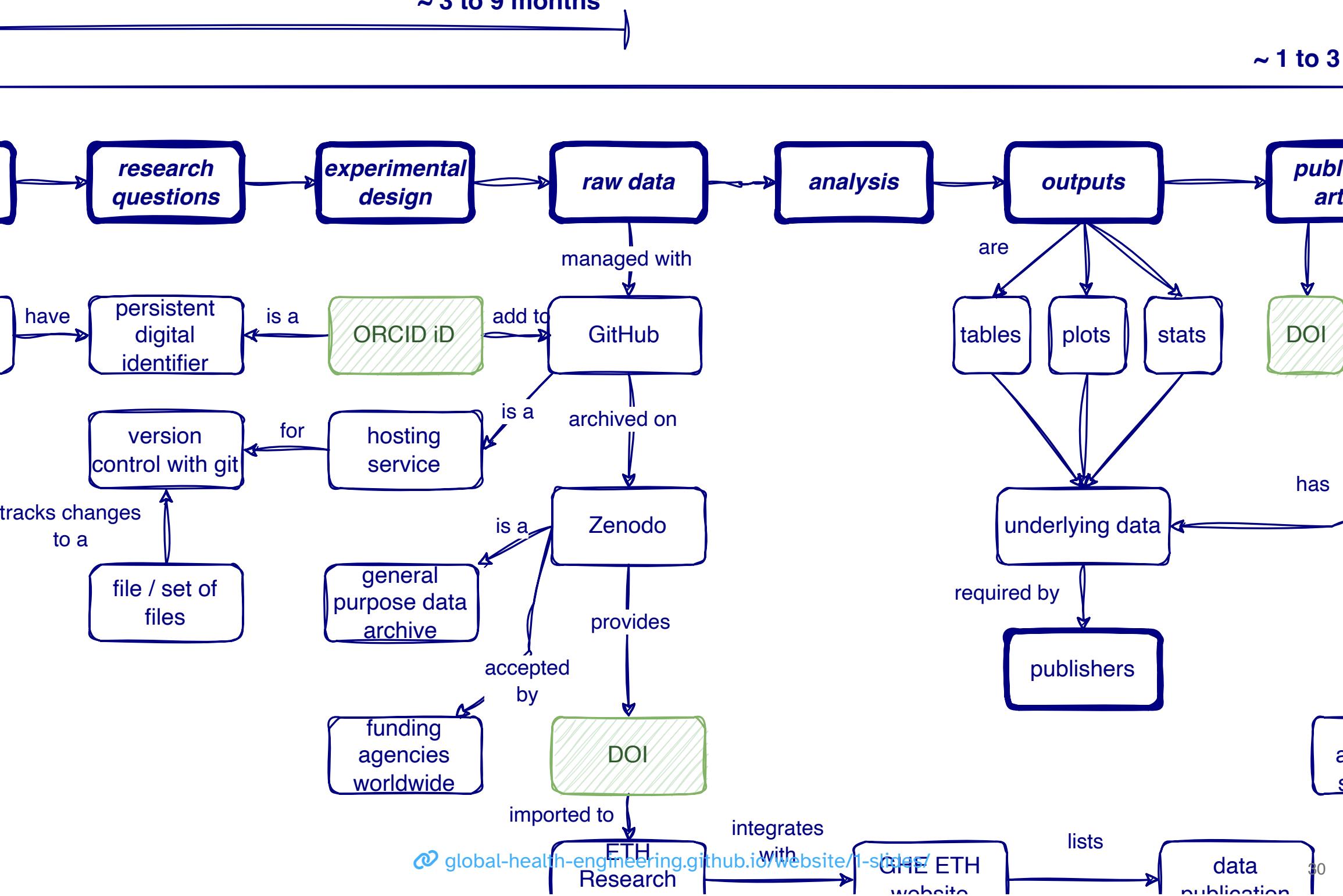


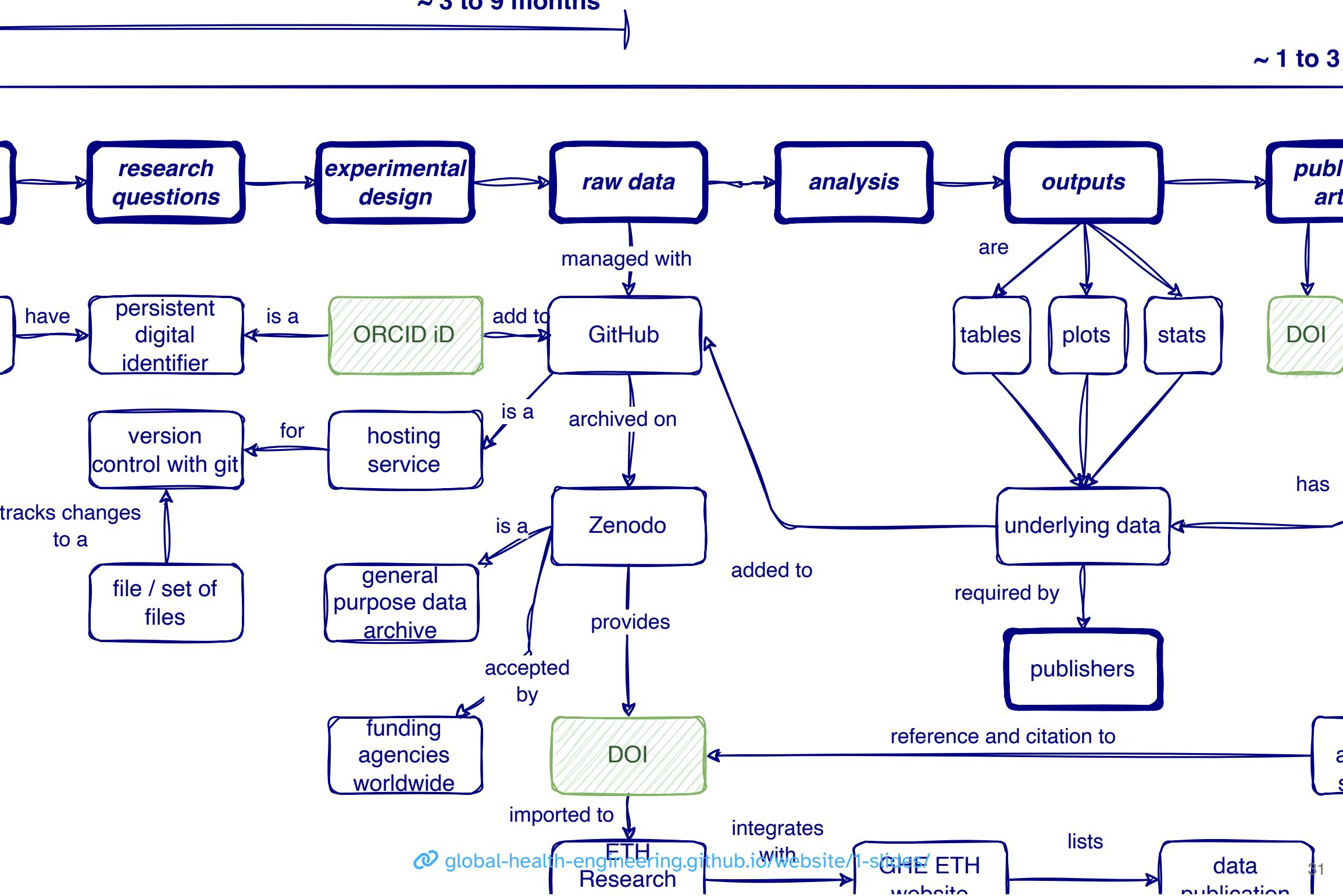
~ 3 to 9 months

~ 1 to 3



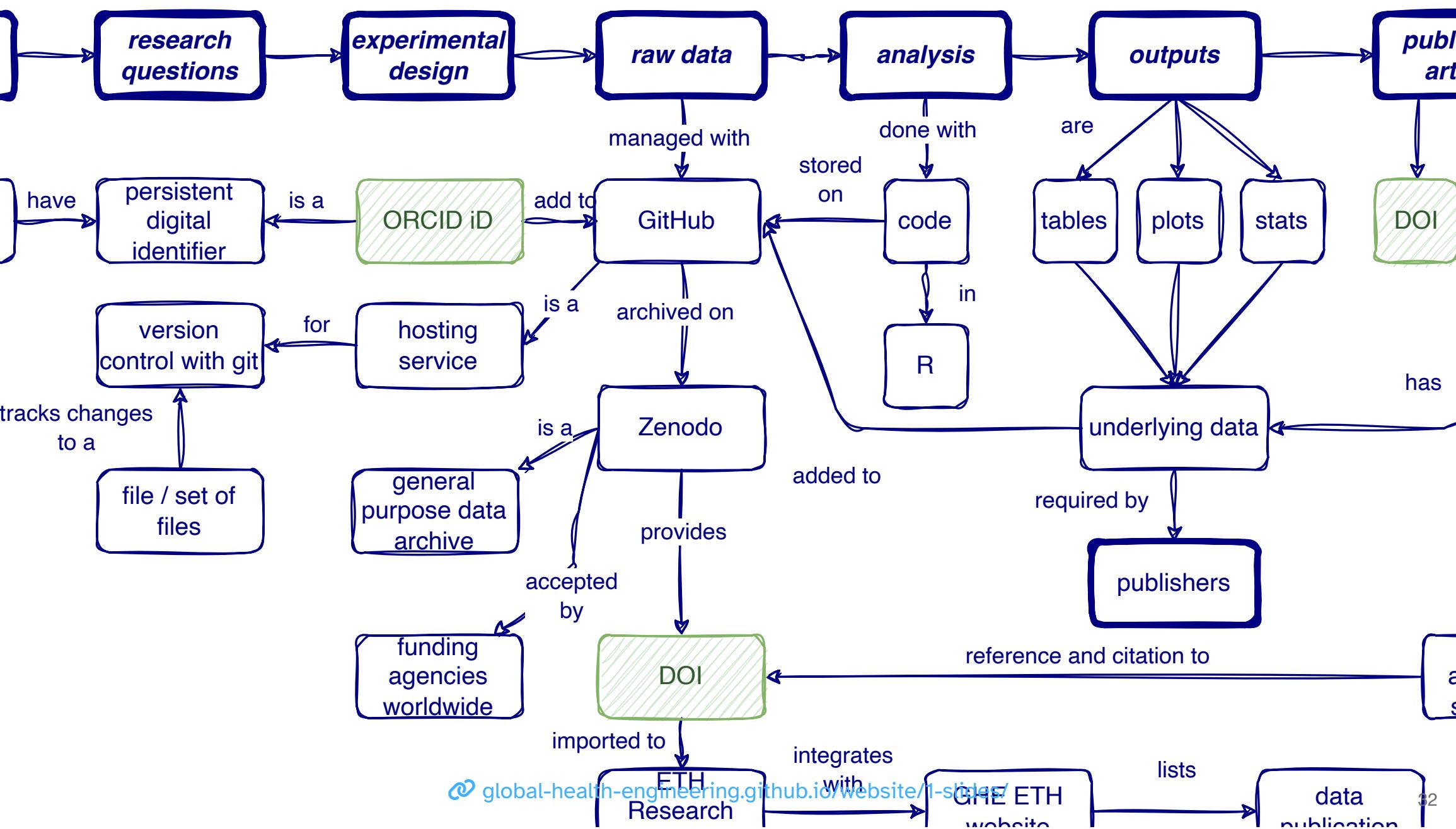


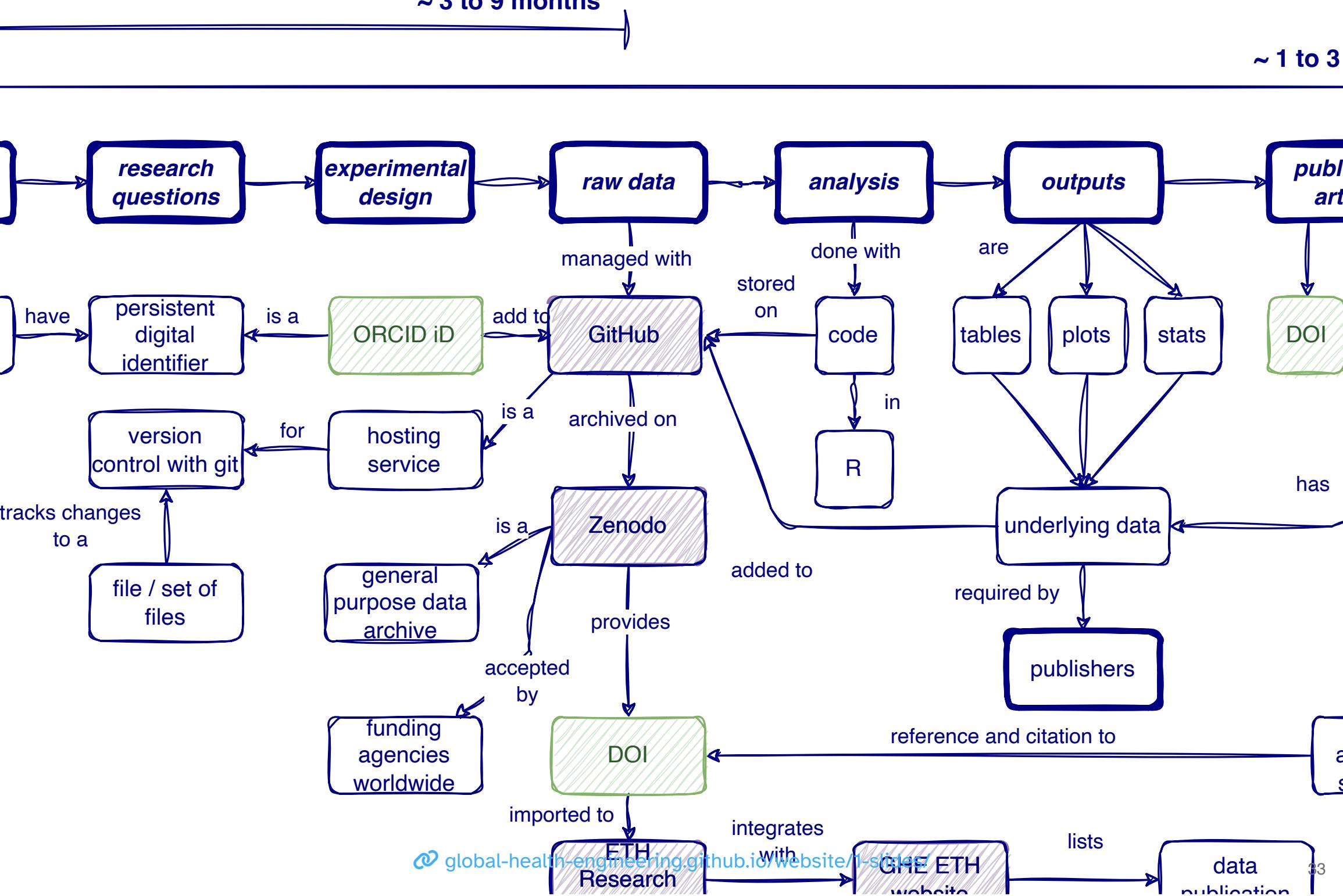




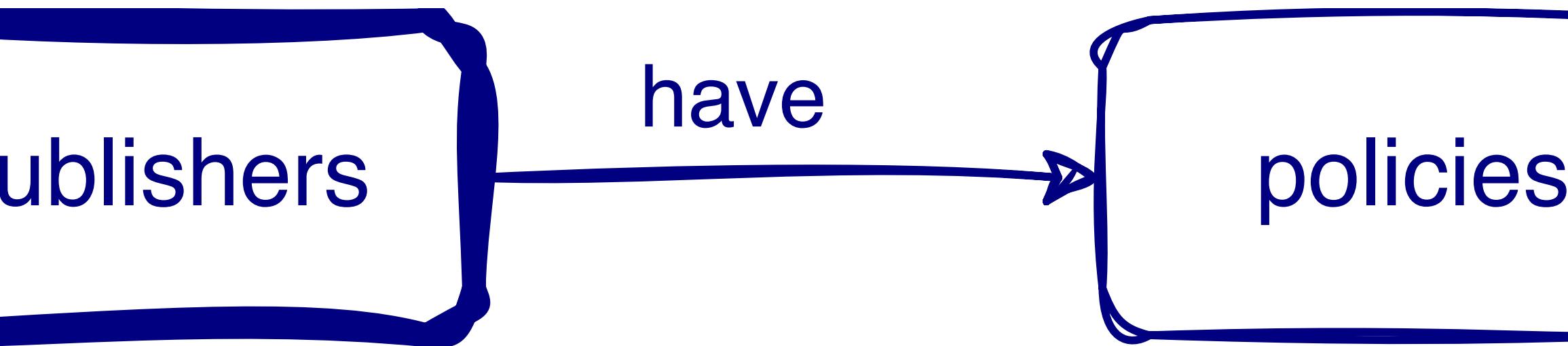
~3 to 9 months

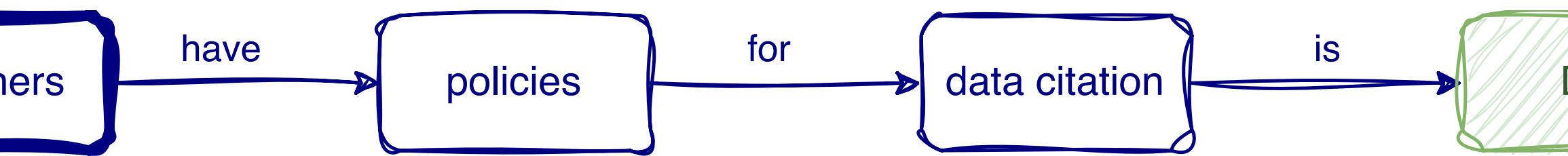
\sim 1 to 3

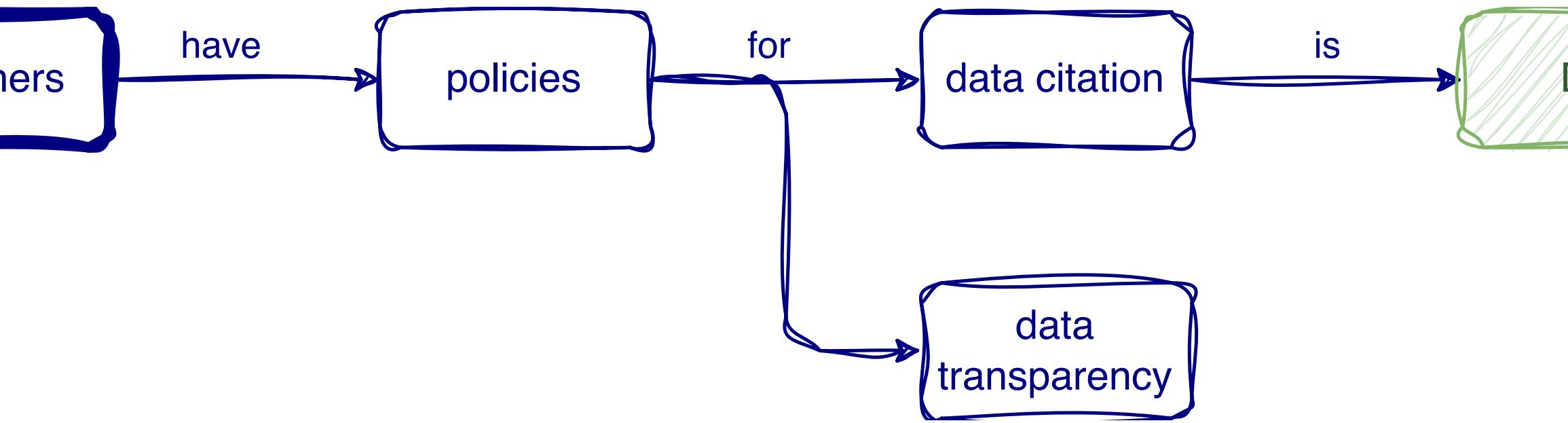


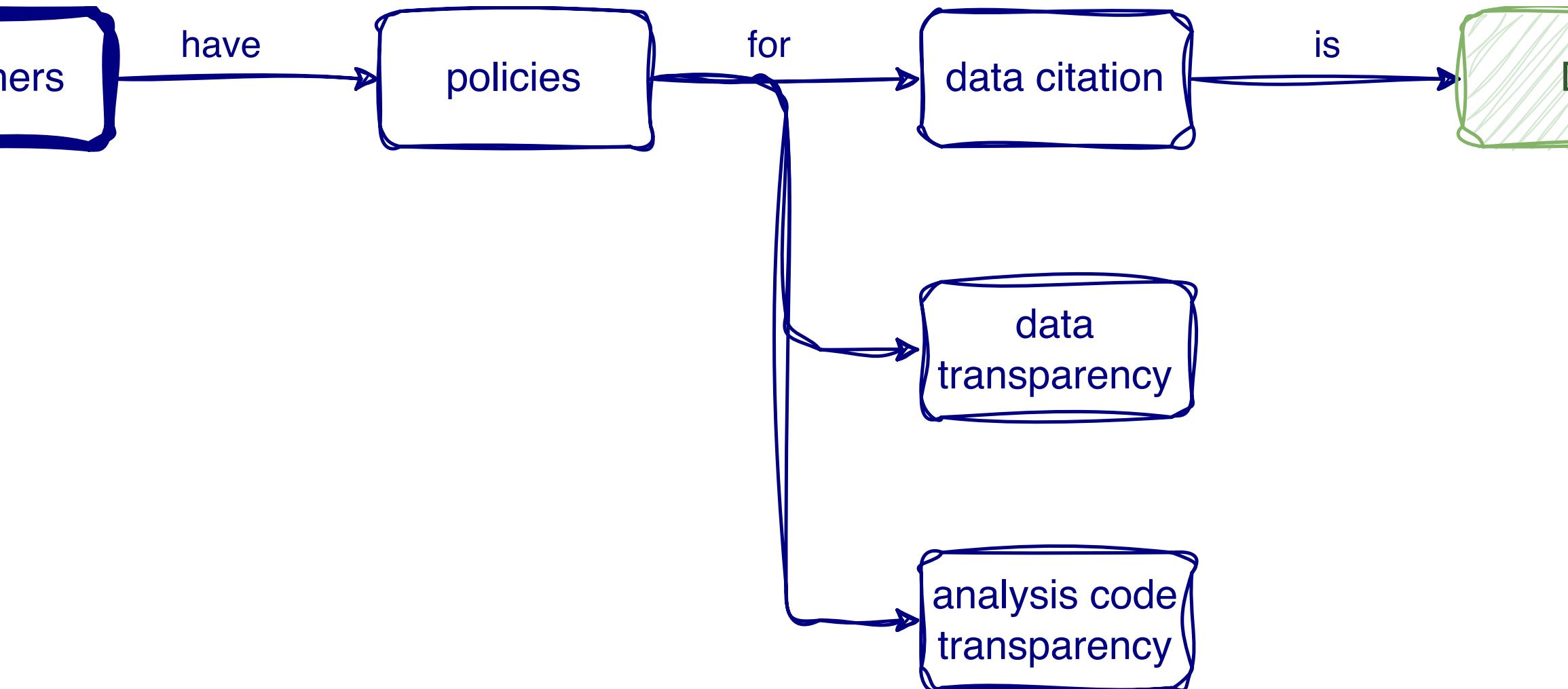


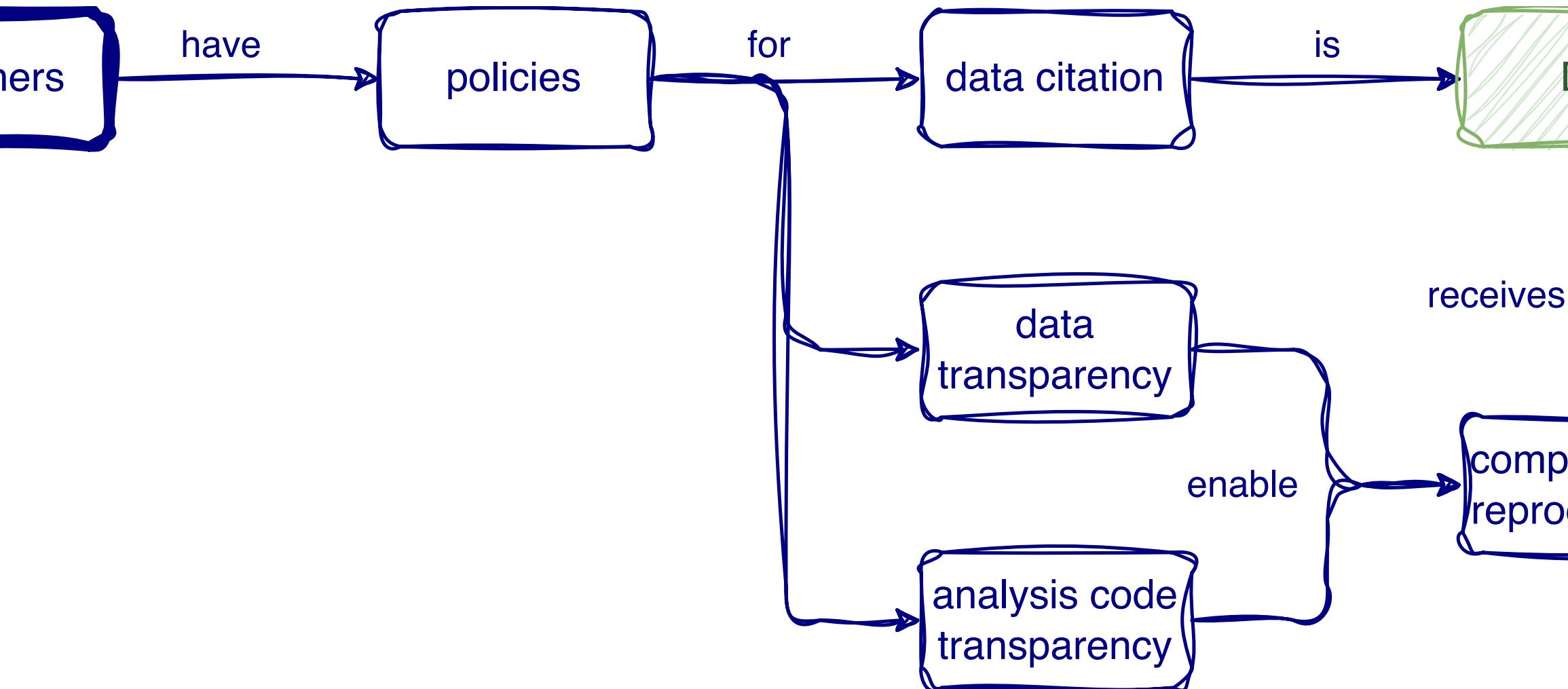
Publisher policies

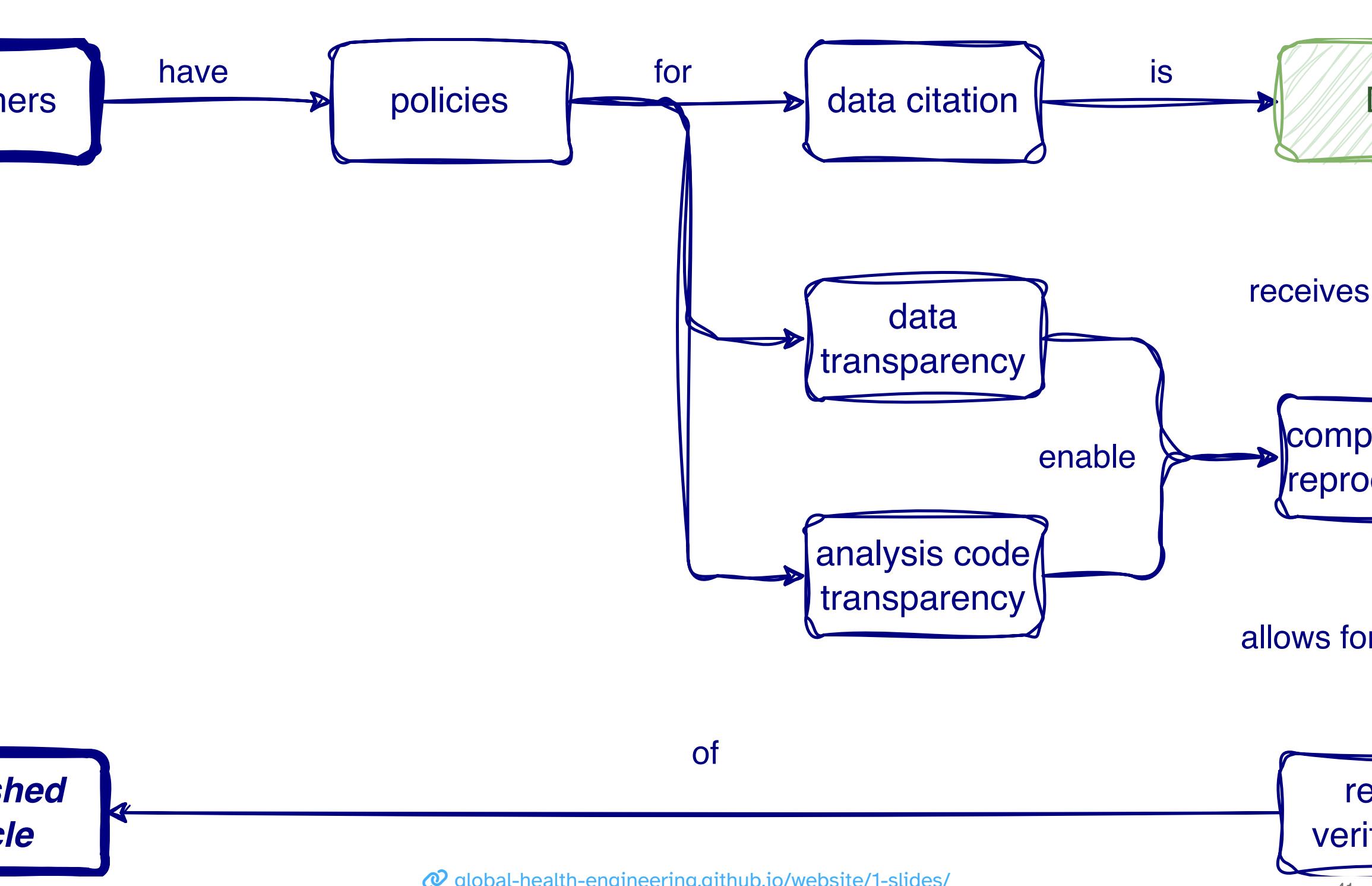


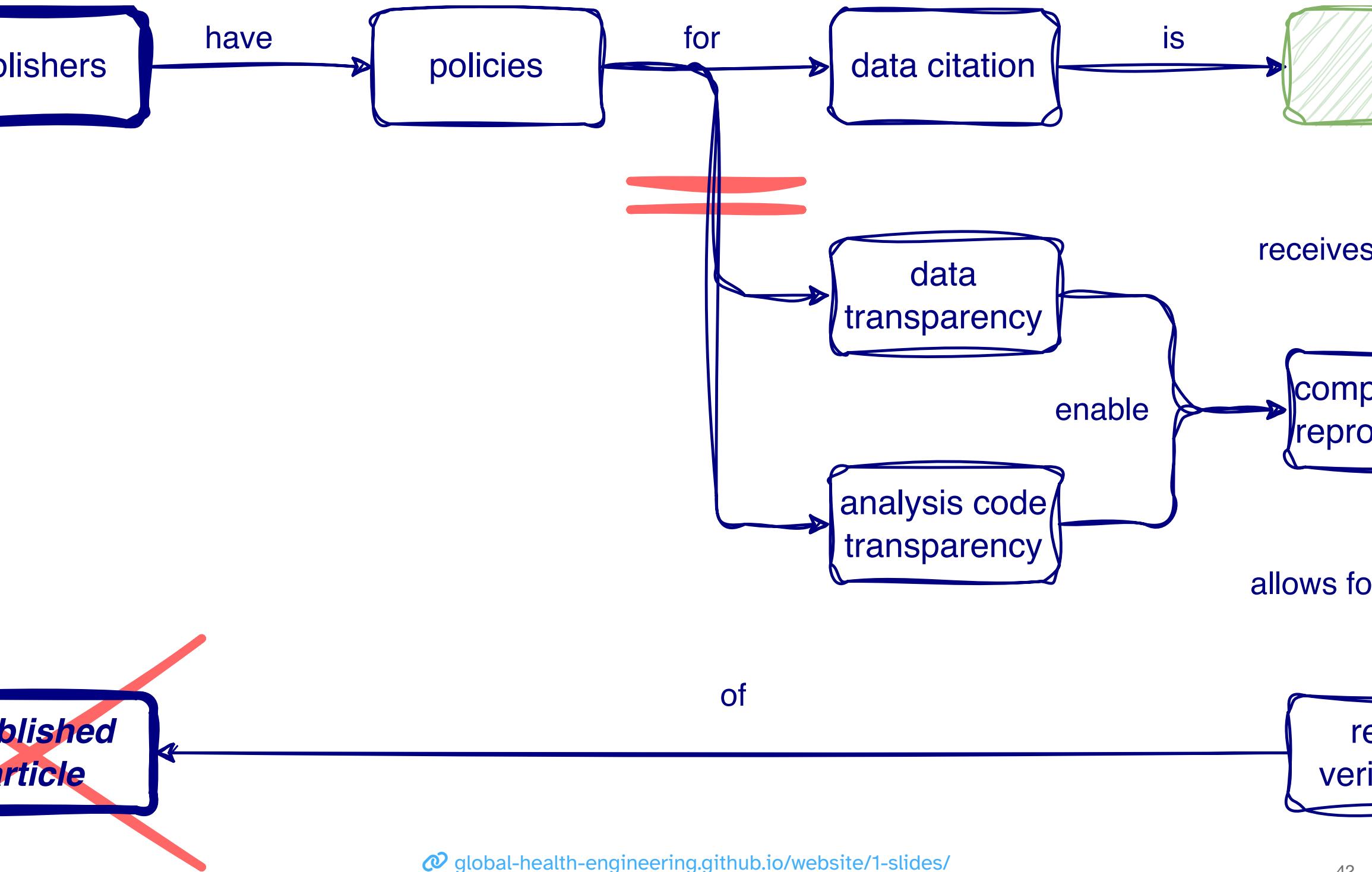












The Challenge

(Technical) Overhead

- Data management workflows
- Tools for scientific computing (R, Python)
- Version control system (Git)
- Project management tools (GitHub, GitLab)
- Communication (Slack, Element)
- Authoring (RMarkdown, Jupyter, Quarto)
- Publishing (Zenodo, Figshare, OSF)

The Start

Funding - ETH ORD Program

- Open Research Data Program of the ETH Board:
<https://ethrat.ch/en/eth-domain/open-research-data/>
- ORD 1st Explore (150k), our funded proposal:
<https://openwashdata.org/pages/gallery/proposal/>
- ORD 2nd Contribute (30k), our funded proposal:
<https://github.com/openwashdata-dev/proposal-eth-ord-scheme-contribute/blob/main/proposal-eth-ord-scheme-contribute.md>
- ORD 2nd Explore (150k) guidelines, due 29th February 2024:
https://ethrat.ch/wp-content/uploads/2023/10/Application-Guidelines_TrackAExplore_2nd_call_FINAL.pdf
- ORD 6th Contribute (30k) guidelines, due 12th December 2023: https://ethrat.ch/wp-content/uploads/2023/10/Application-guidelines_TrackC-Contribute-projects-6th-call-FINAL.pdf

Funding - ETH ORD Program

We draw your attention to the fact that the Contribute calls are suitable for young researchers (PhD and postdoc) who would like to add value to their data within their projects.

Open Science Working Group Email (2023-10-05)

Learning - Open Science Tools

ost-hs23.github.io/website/syllabus.html

Course Overview

Open Science Tools - authoring and publishing workflows for collaborative scientific writing

Details

Sign-up

- [851-0623-00L - Course Catalogue](#)

Dates

- 17th October 2023, 08:30 am to 12:30 pm
- 24th October 2023, 08:30 am to 12:30 pm

Instructor

- [Lars Schöbitz - Global Health Engineering - ETH Zurich](#)

Thanks

Slides created via revealjs and Quarto:

<https://quarto.org/docs/presentations/revealjs/>

Access slides as [PDF on GitHub](#)

This material is licensed under [Creative Commons Attribution 4.0 International.](#)

References

- Greene, Nicola, Sarah Hennessy, Tate W. Rogers, Jocelyn Tsai, and Francis L. de los Reyes III. 2021. “The Role of Emptying Services in Provision of Safely Managed Sanitation: A Classification and Quantification of the Needs of LMICs.” *Journal of Environmental Management* 290 (July): 112612. <https://doi.org/10.1016/j.jenvman.2021.112612>.
- 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>.

