DSN Kick-off

Connecting people and topics on data stewardship at ETH Zurich

Lars Schöbitz

lschoebitz@ethz.ch

Global Health Engineering - ETH Zurich

Hello! 👋

Lars Schöbitz

- Environmental Engineer
- Open Science Specialist at Global Health Engineering
- RStudio certified instructor for Data Science with R
- Twitter: @larnsce

My role - Open Science Specialist

- research data management
- reproducible workflows
- mindset for Open Science
- research communication
- teaching data science tools
- proposal writing

Data at Global Health Engineering

- small (few MBs)
- tabular
- non-sensitive
- topics
  - waste management
  - sanitation
  - air quality
  - etc.

Problem statement

Appendix A. Supplementary data

The following is the supplementary data to this article:

Download : Download Word document (152KB)

Multimedia component 1.

Problem statement

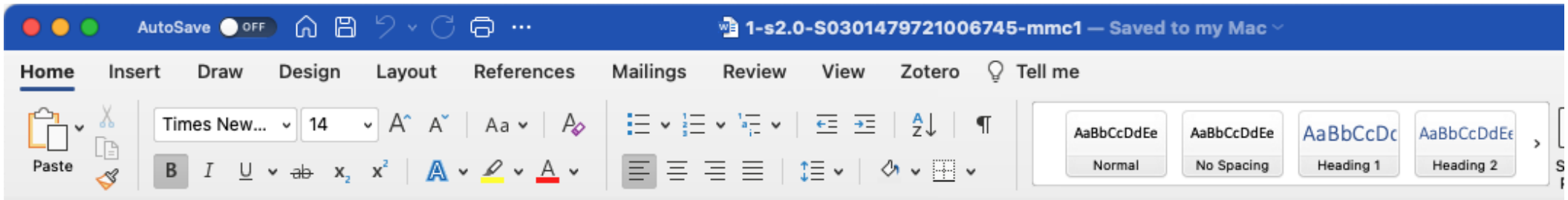
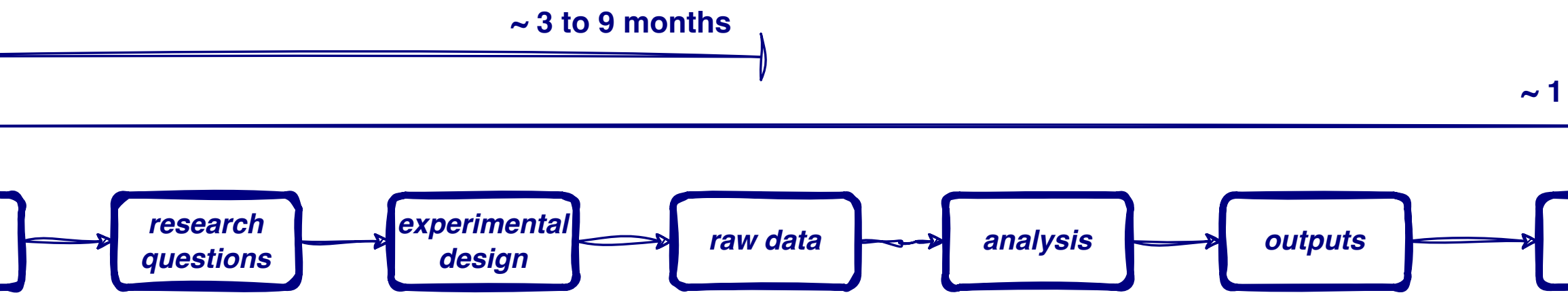


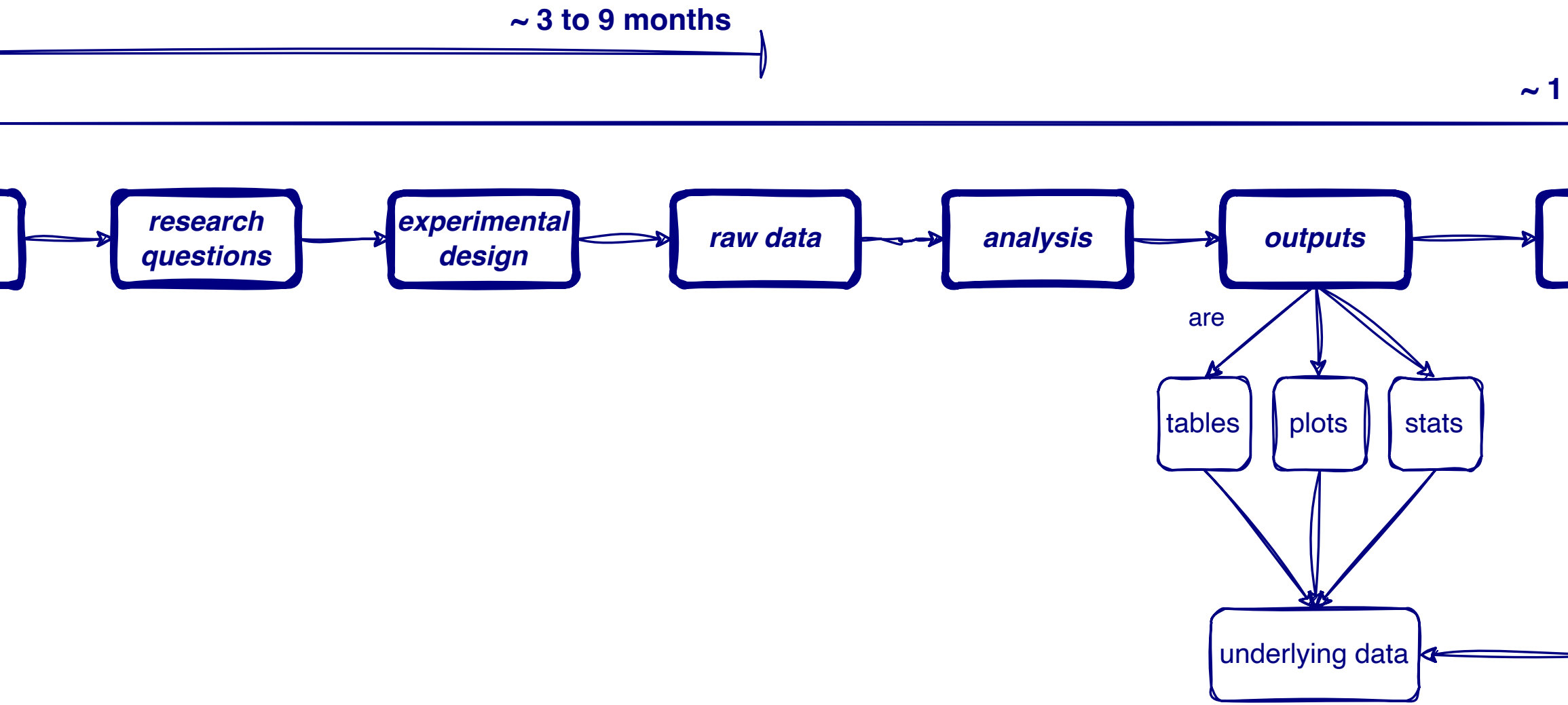
Table 1. The number of onsite sanitation facilities alphabetized by country per Service Type. The Service Type Density is calculated by dividing the population using country population. Thus, the Service Type Density covers users of sewers, onsite facilities, and open defecators. The number of facilities required for Open Defecat facilities will not be shared between households. Urban and rural proportions of the onsite facilities are provided. "ND" indicates no data available from JMP.

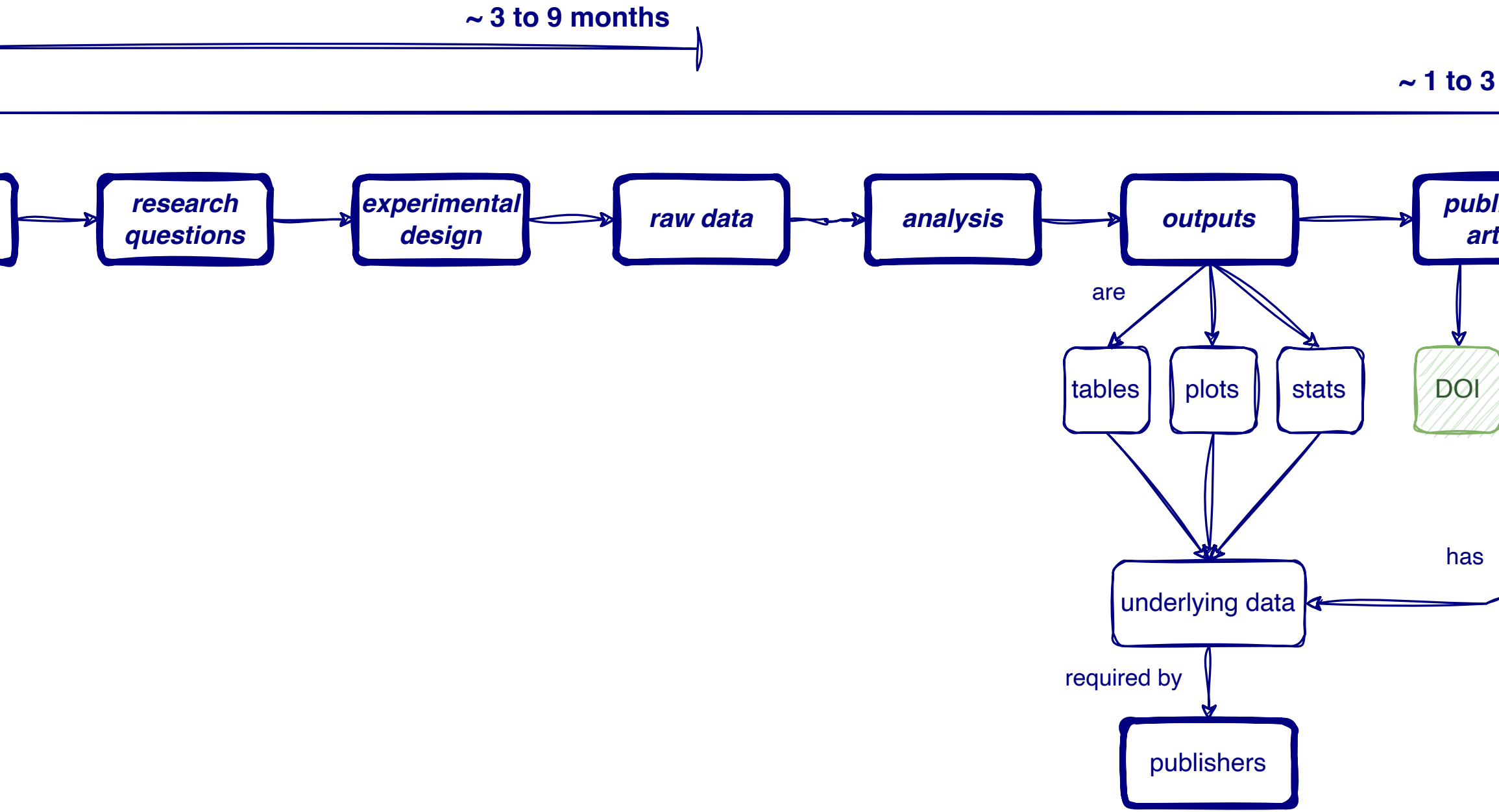
	Mechanized				Non-Mechanized				Unemptiable				
Country	**No. of Facilities**	**Service Type Density**	**Urban**	**Rural**	**No. of Facilities**	**Service Type Density**	**Urban**	**Rural**	**No. of Facilities**	**Service Type Density**	**Urban**	**Rural**	**No. of Facilities Required**
Afghanistan	367,345	10%	66%	34%	1,599,720	41%	31%	69%	1,351,493	34%	12%	88%	563,238
Algeria	482,751	6%	27%	73%	465,887	6%	18%	82%	255,876	3%	63%	37%	64,417
Angola	1,054,451	21%	97%	3%	1,817,690	36%	75%	25%	516,612	10%	44%	56%	1,232,071
Anguilla	2,875	72%	100%	0%	1,056	26%	100%	0%	15	0%	100%	0%	22

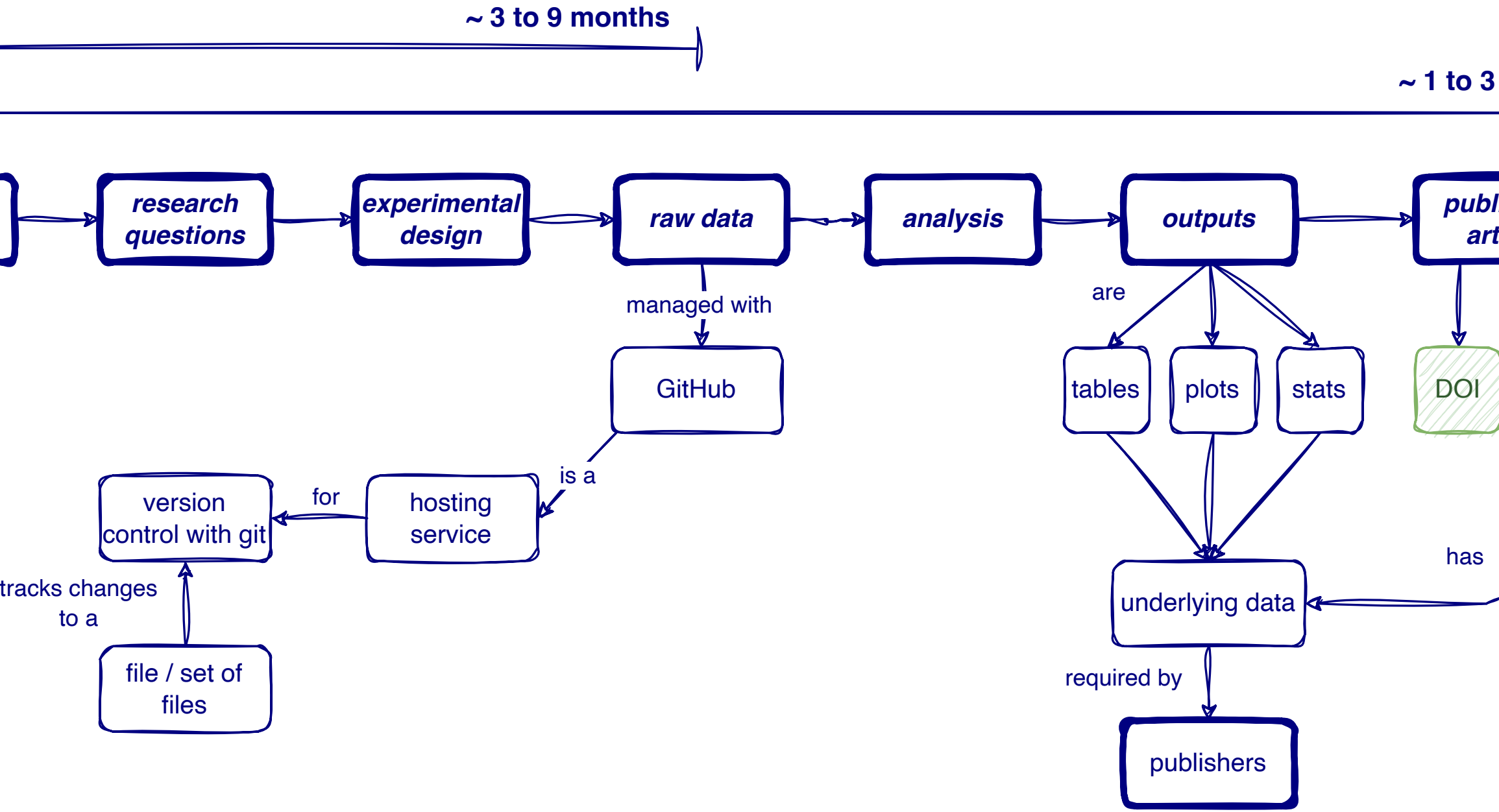
Aspired RDM workflow at GHE

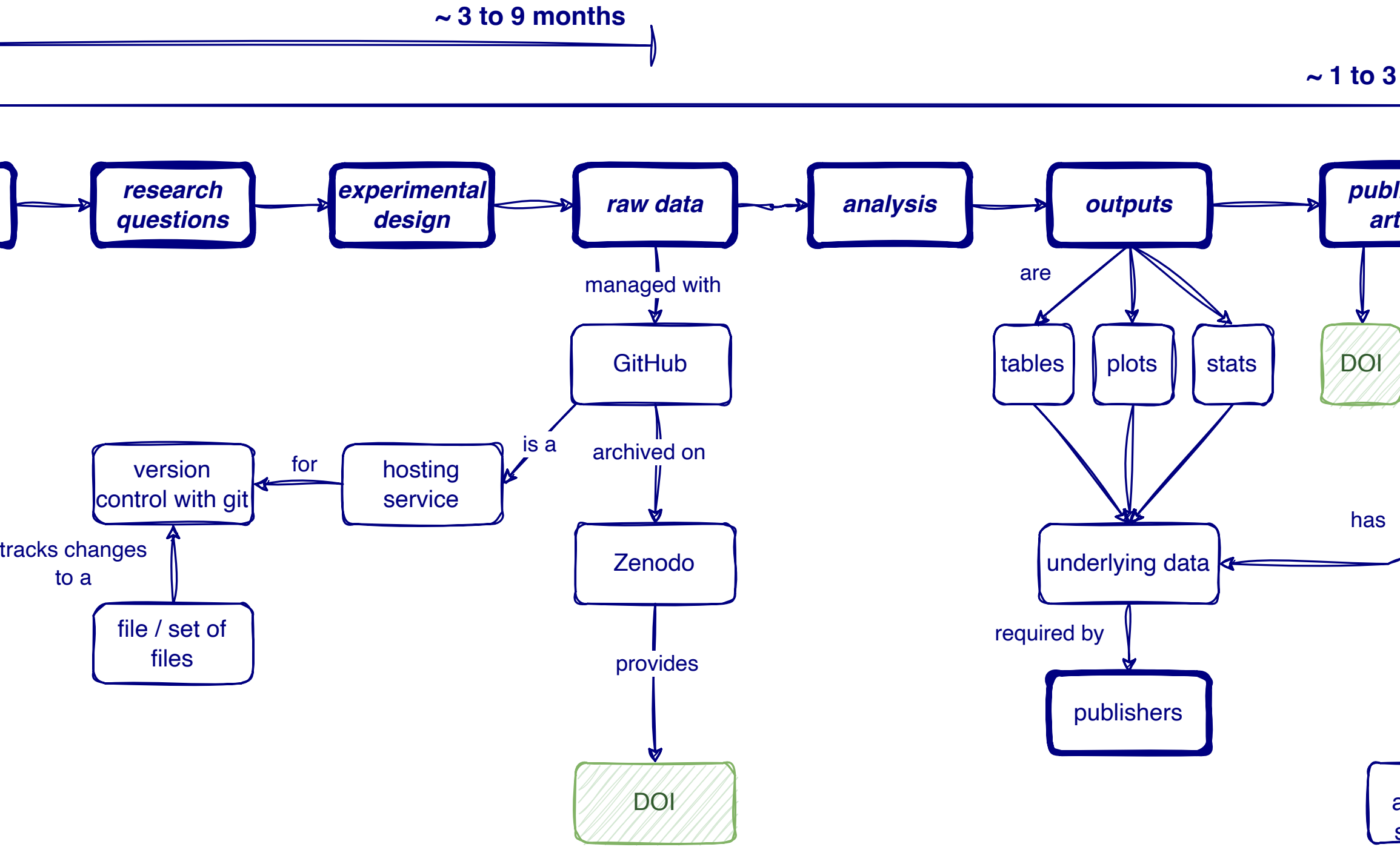
research questions
experimental design
raw data
analysis
outputs

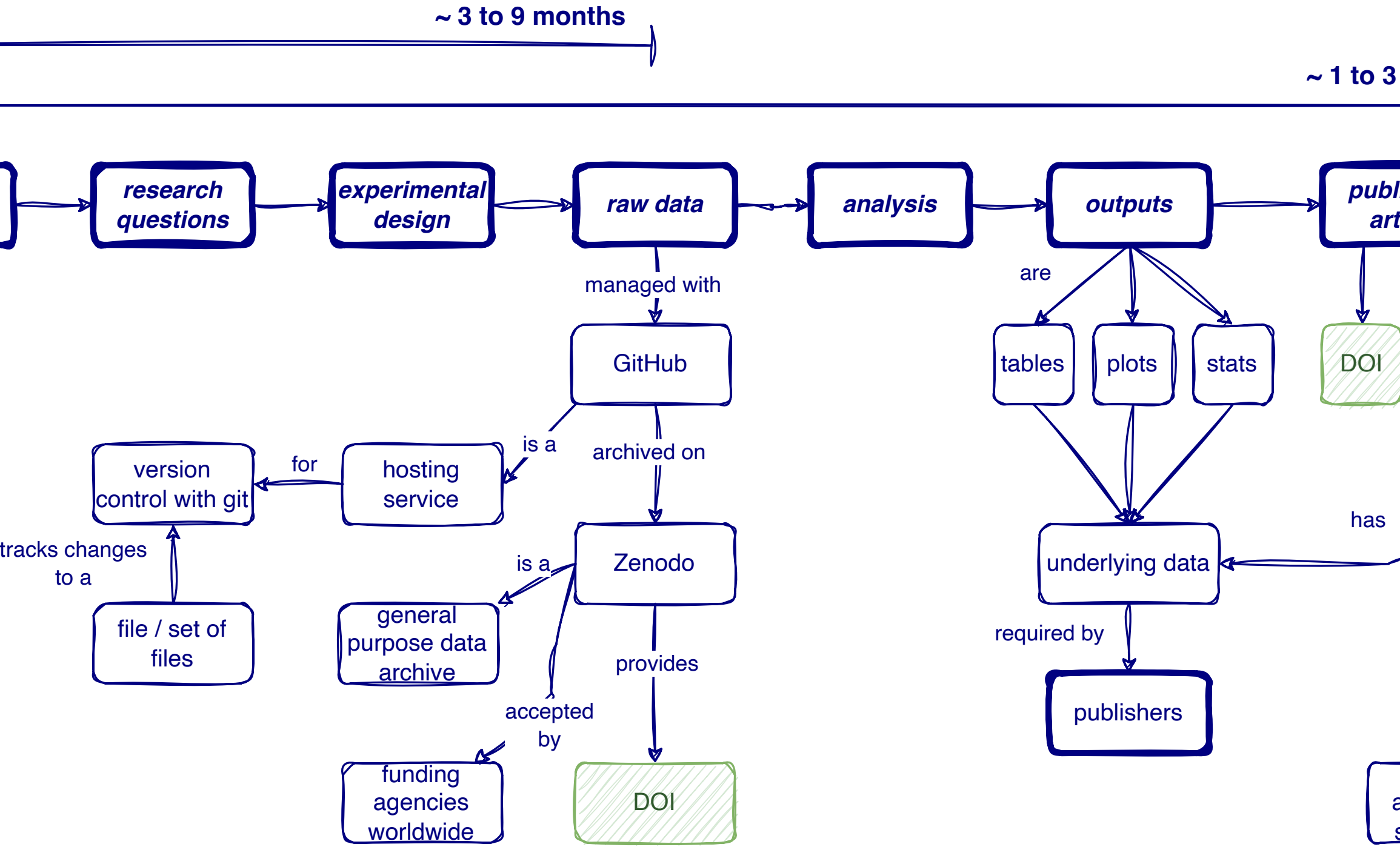

~ 3 to 9 months
~ 1
research questions
experimental design
raw data
analysis
outputs


~ 3 to 9 months
~ 1
research questions
experimental design
raw data
analysis
outputs
are
tables
plots
stats
underlying data


~ 3 to 9 months
~ 1 to 3
research questions
experimental design
raw data
analysis
outputs
publ
art
are
tables
plots
stats
DOI
has
underlying data
required by
publishers

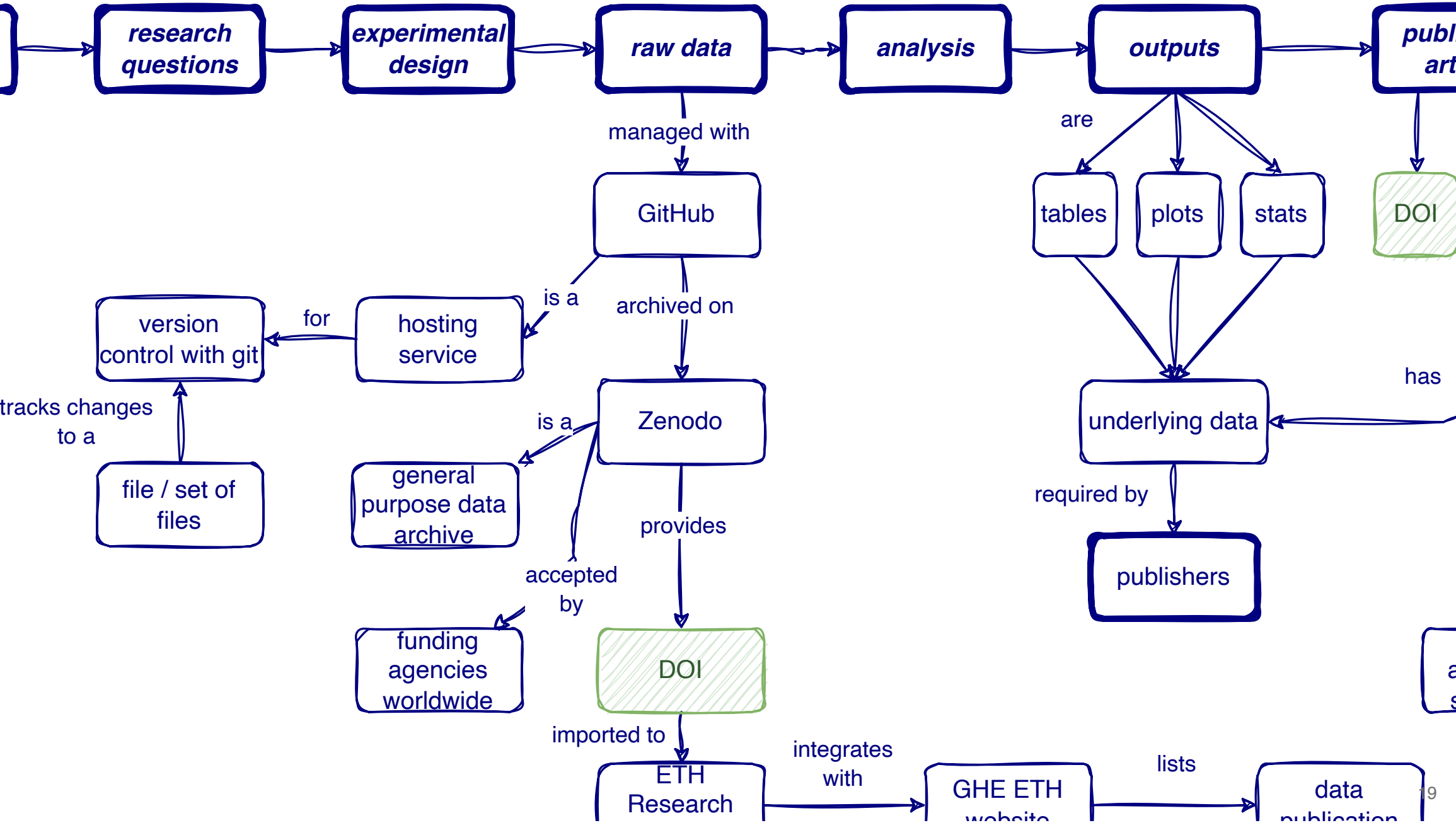

~ 3 to 9 months
~ 1 to 3
research questions
experimental design
raw data
analysis
outputs
publ art
managed with
GitHub
are
tables
plots
stats
DOI
is a
for
hosting service
version control with git
tracks changes to a
file / set of files
has
underlying data
required by
publishers


~ 3 to 9 months
~ 1 to 3
research questions
experimental design
raw data
analysis
outputs
publ
art
managed with
are
GitHub
tables
plots
stats
DOI
is a
archived on
for
version control with git
hosting service
has
tracks changes to a
Zenodo
underlying data
file / set of files
required by
provides
publishers
DOI


~ 3 to 9 months
~ 1 to 3
research questions
experimental design
raw data
analysis
outputs
publ art
managed with
are
GitHub
tables
plots
stats
DOI
is a
archived on
for
version control with git
hosting service
tracks changes to a
file / set of files
has
underlying data
is a
Zenodo
general purpose data archive
provides
accepted by
required by
publishers
funding agencies worldwide
DOI

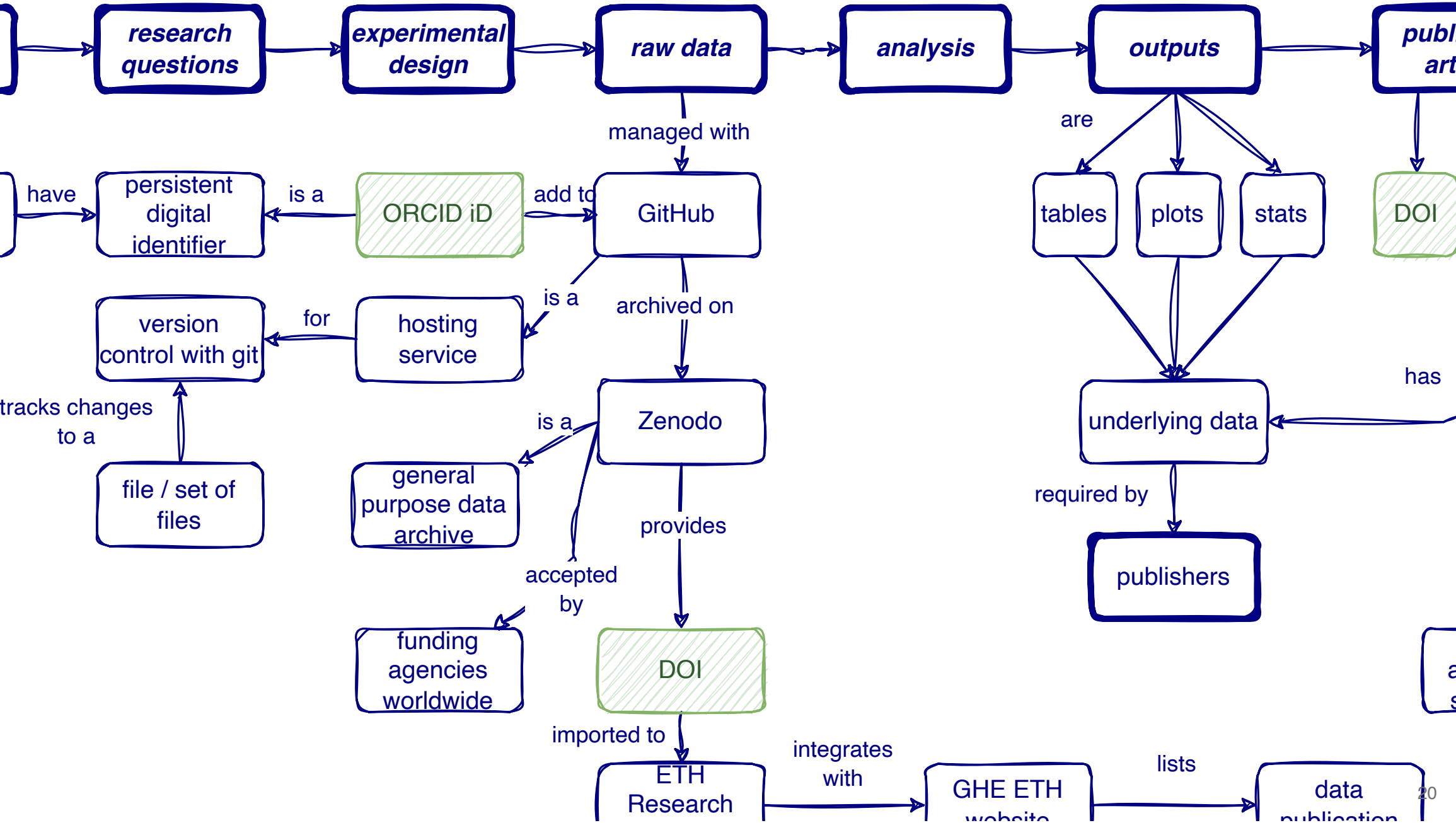
~ 3 to 9 months

~ 1 to 3



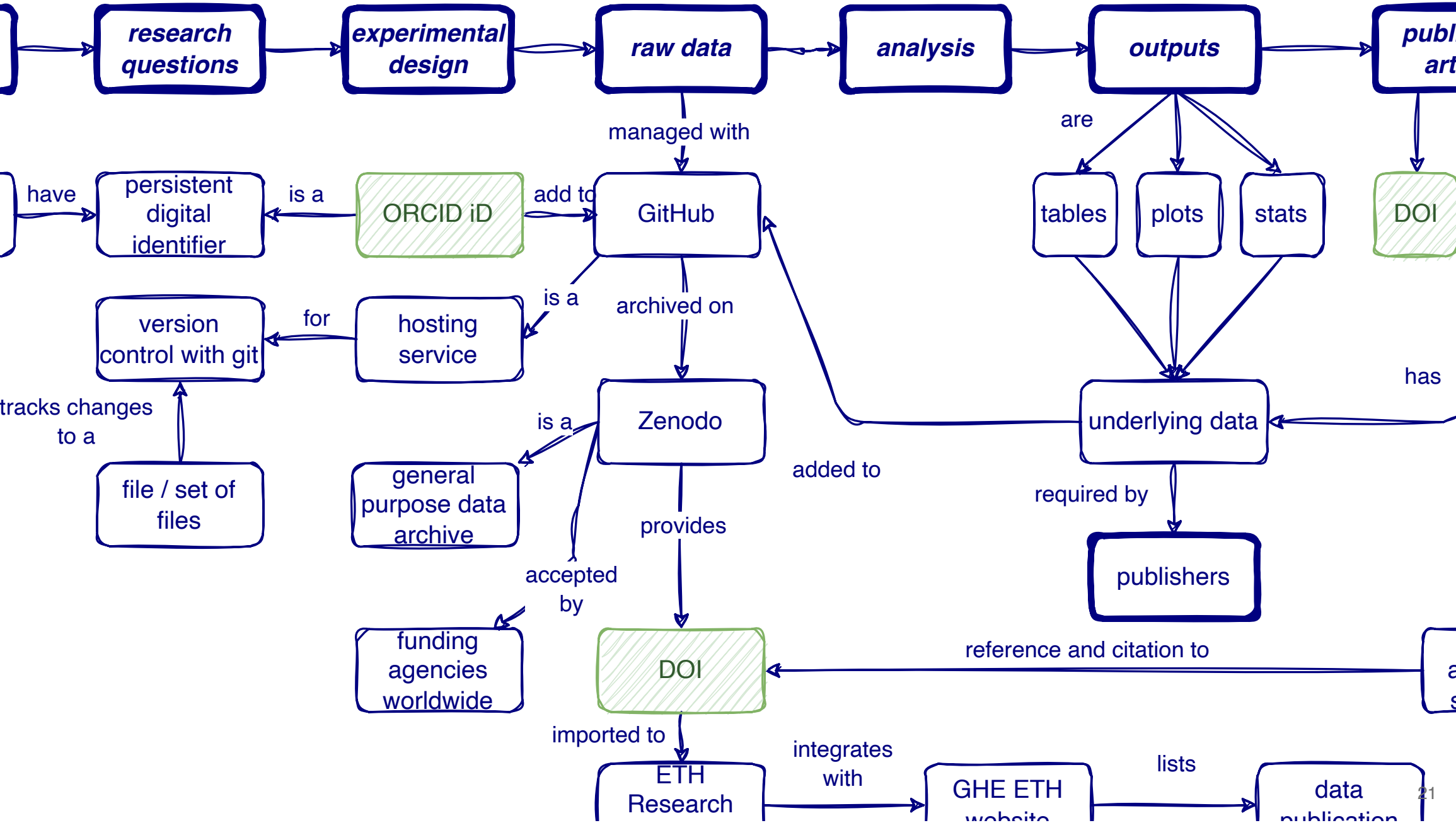
~ 3 to 9 months

~ 1 to 3



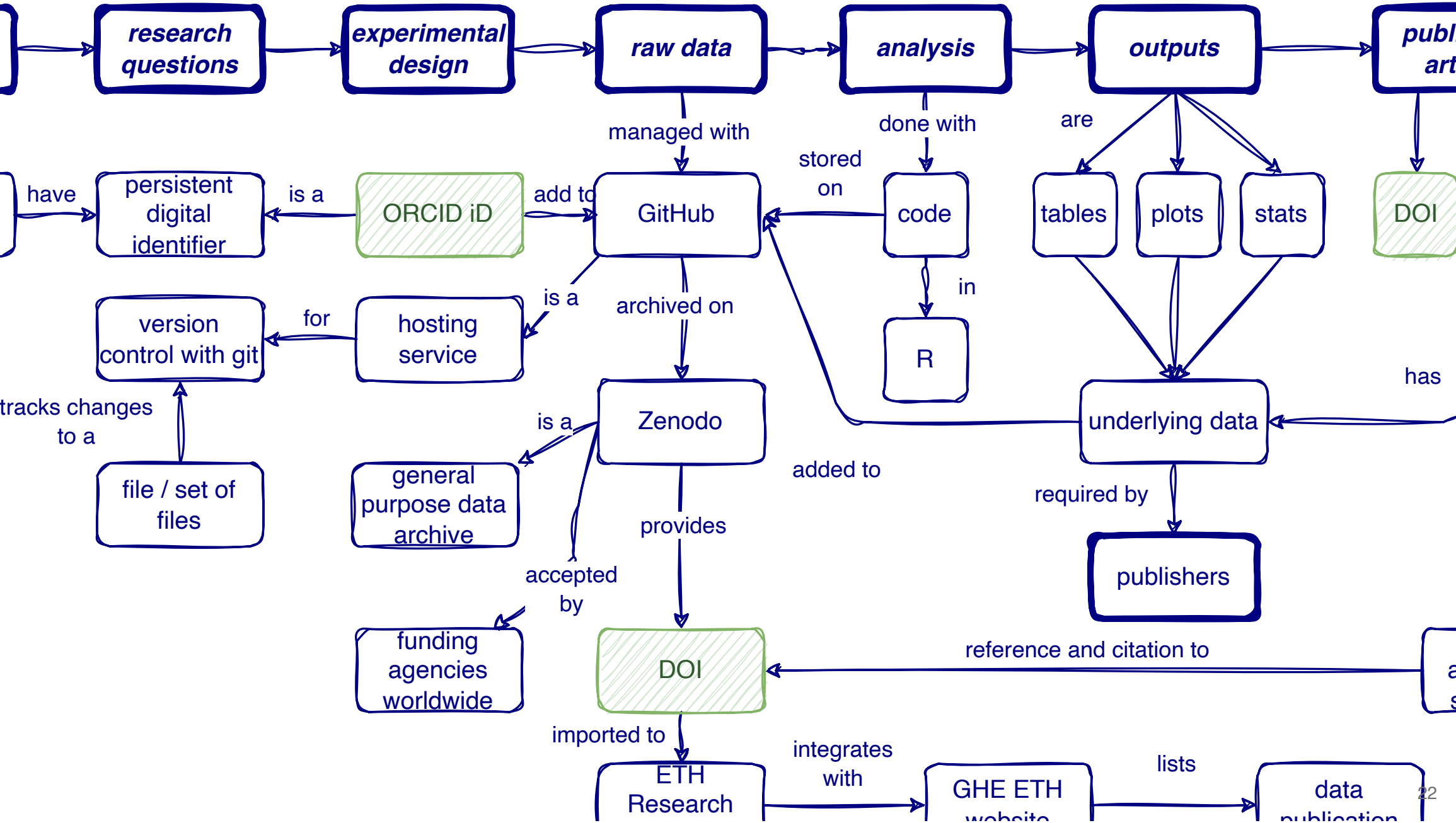
~ 3 to 9 months

~ 1 to 3



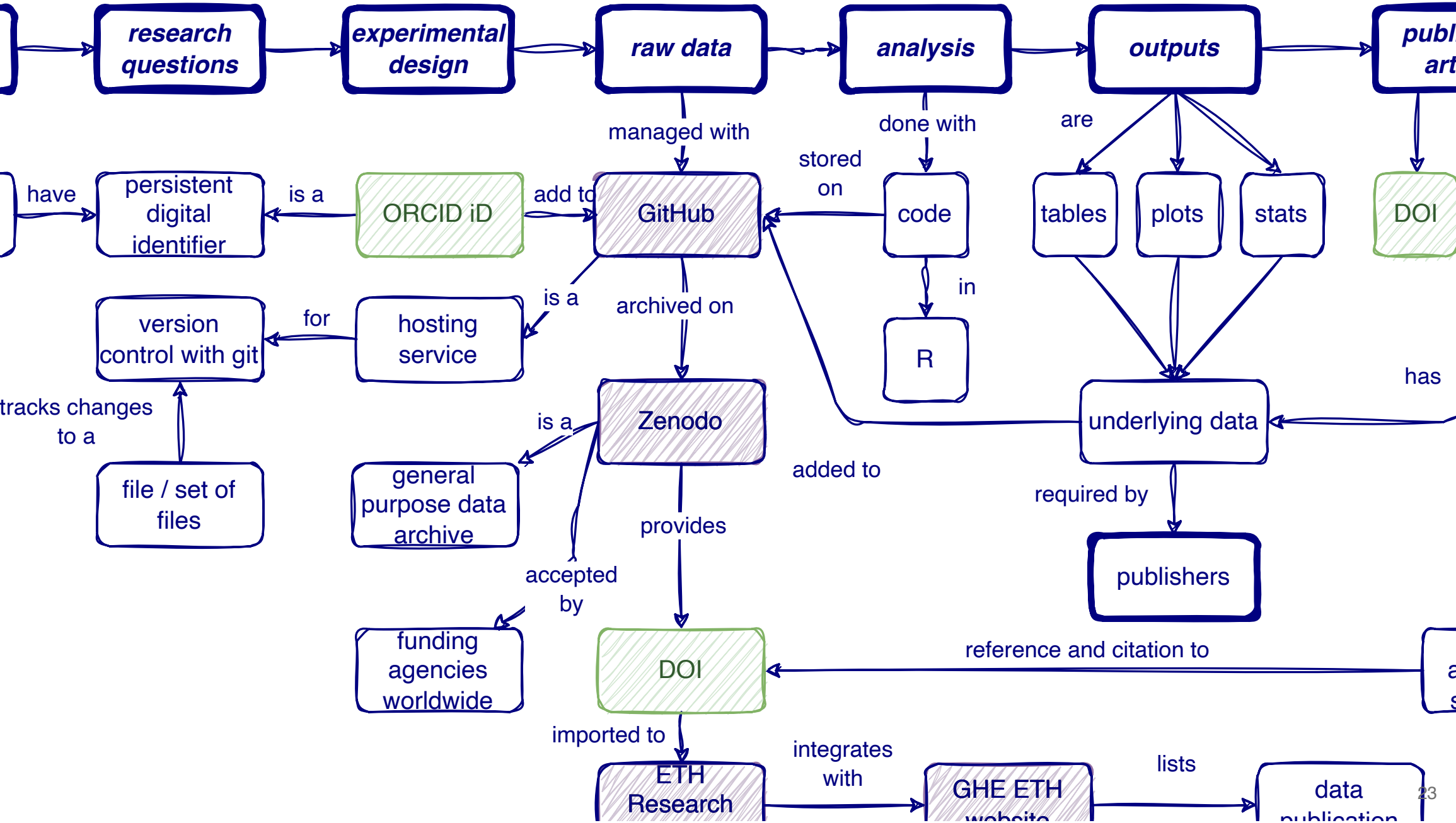
~ 3 to 9 months

~ 1 to 3

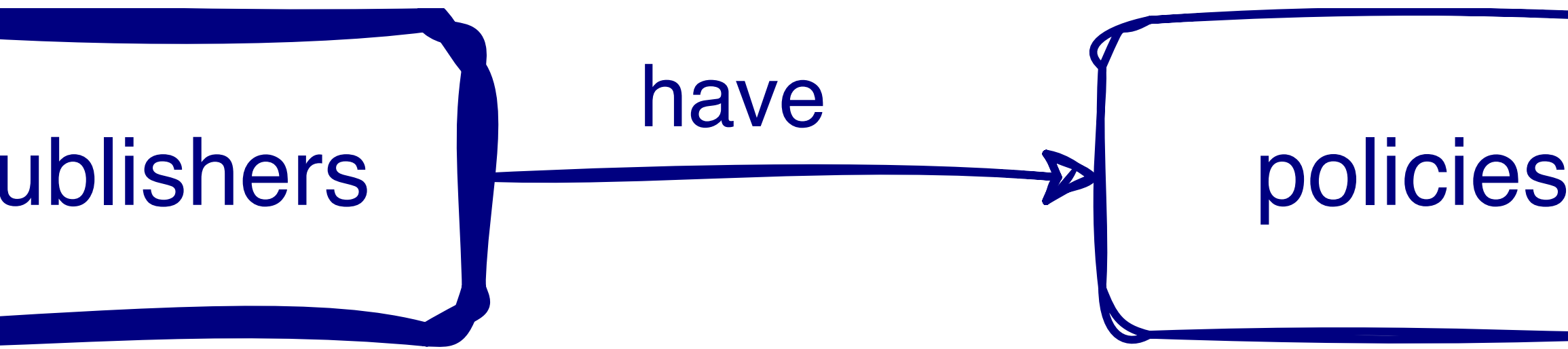


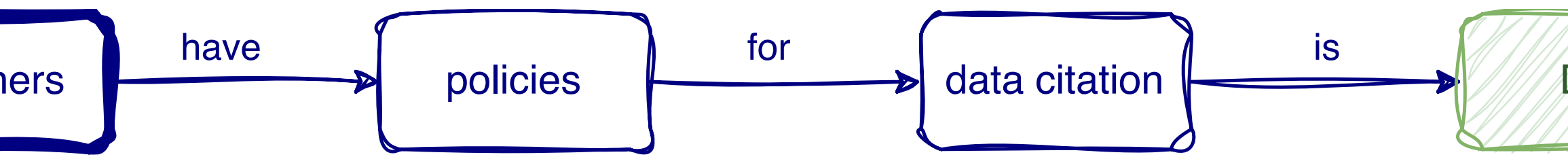
~ 3 to 9 months

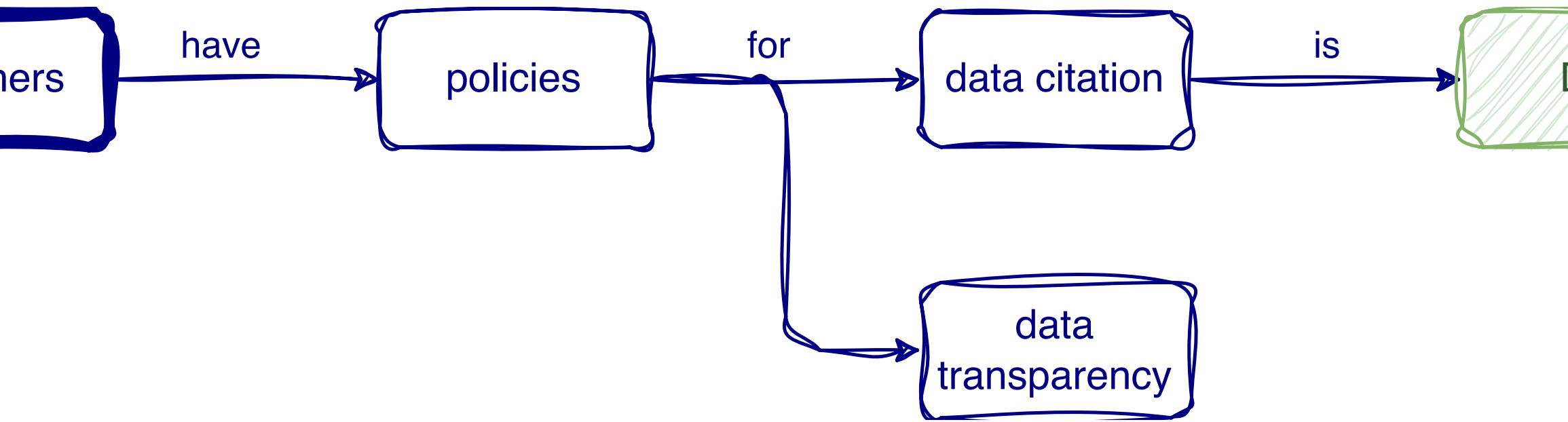
~ 1 to 3

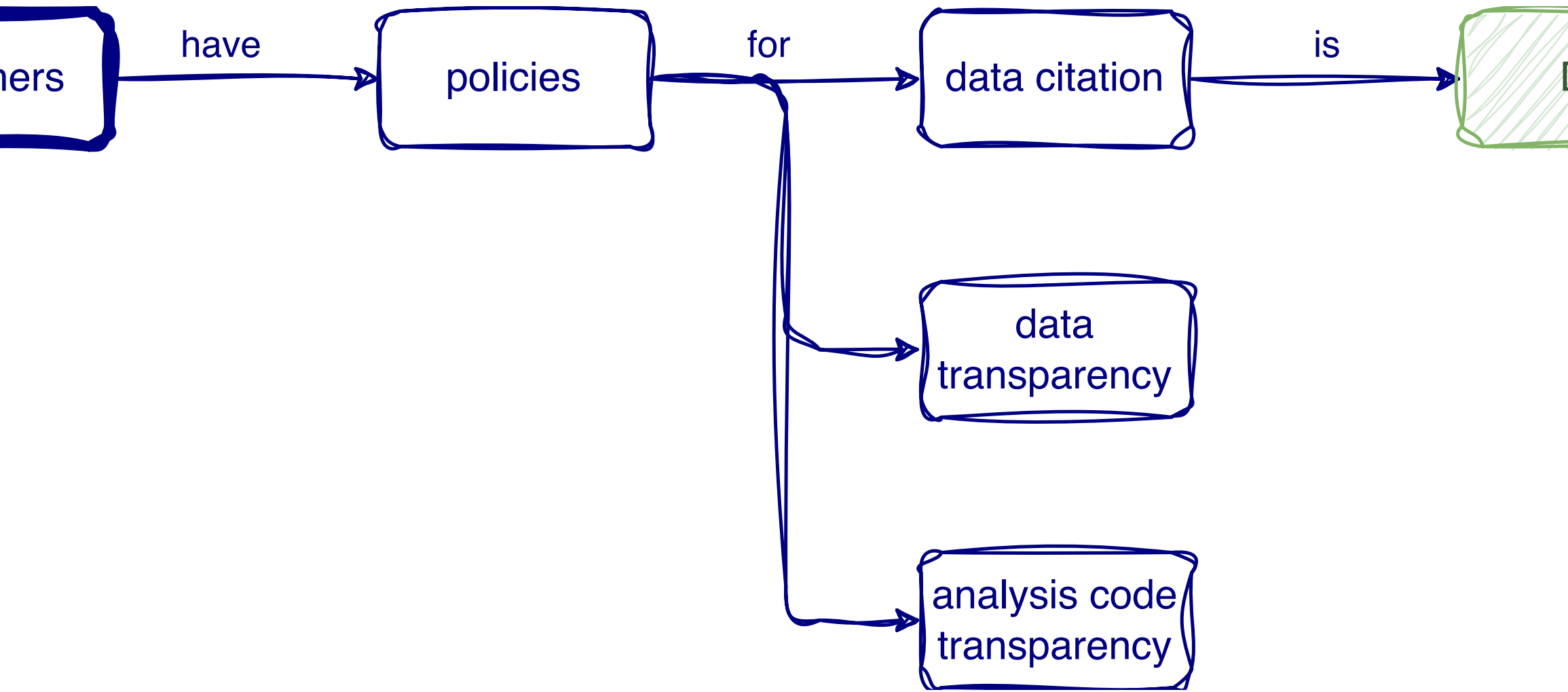


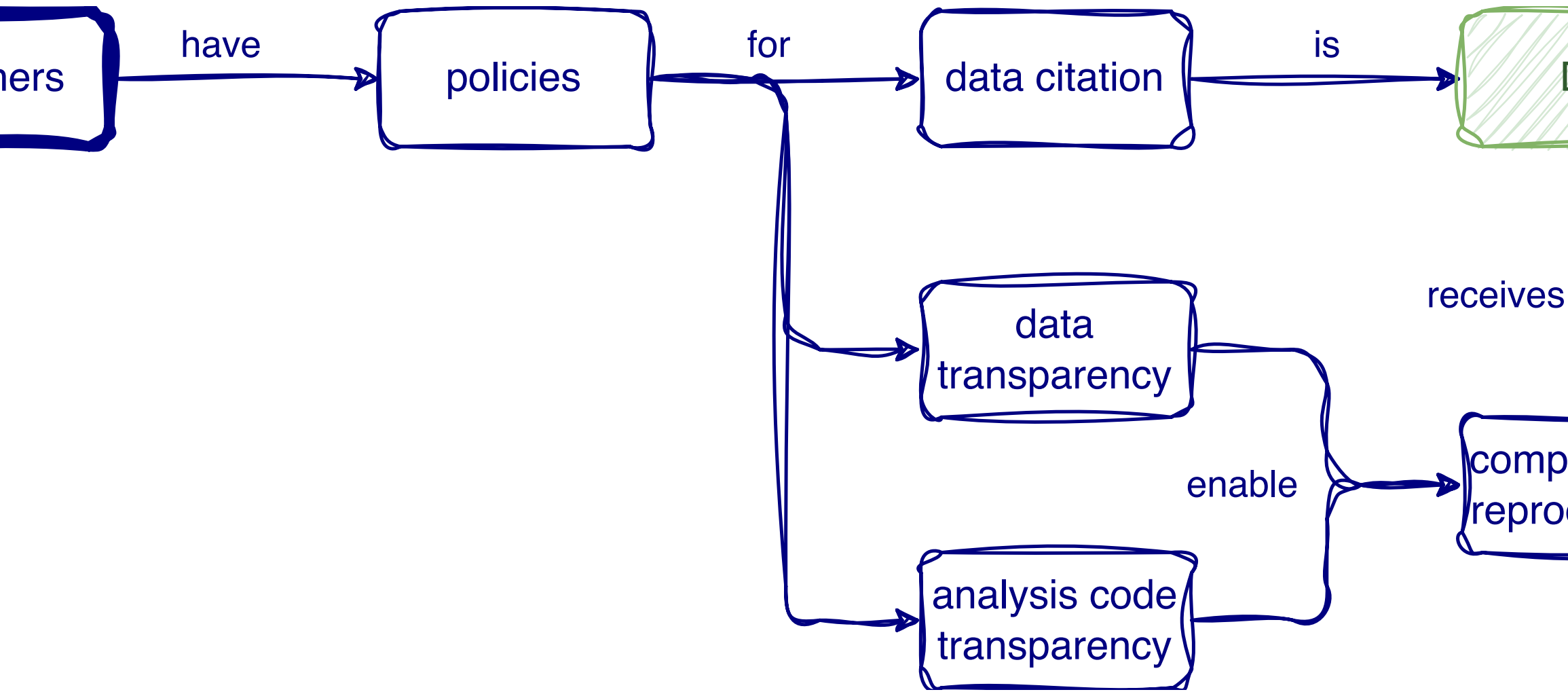
Publisher policies

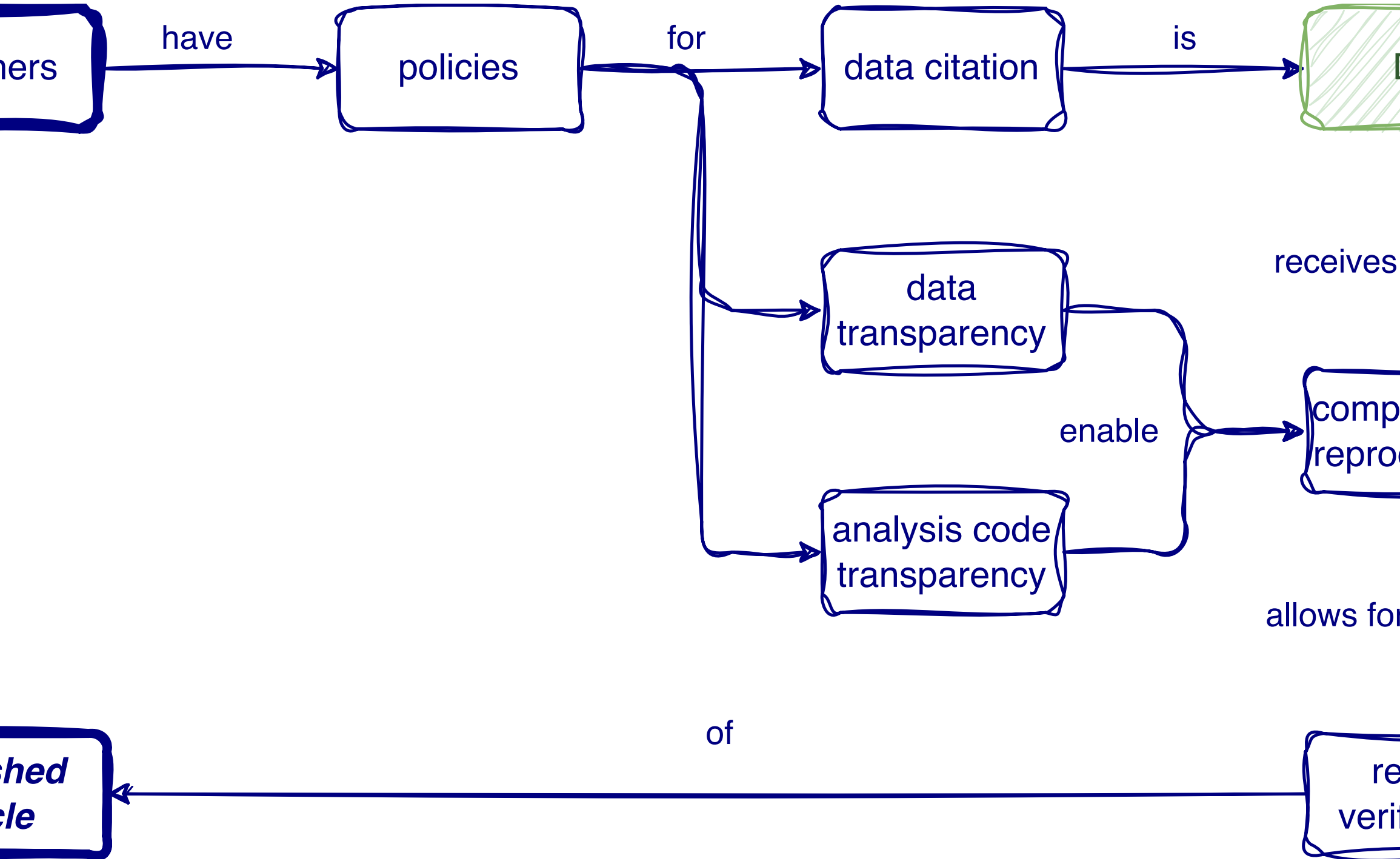

ublishers
have
policies

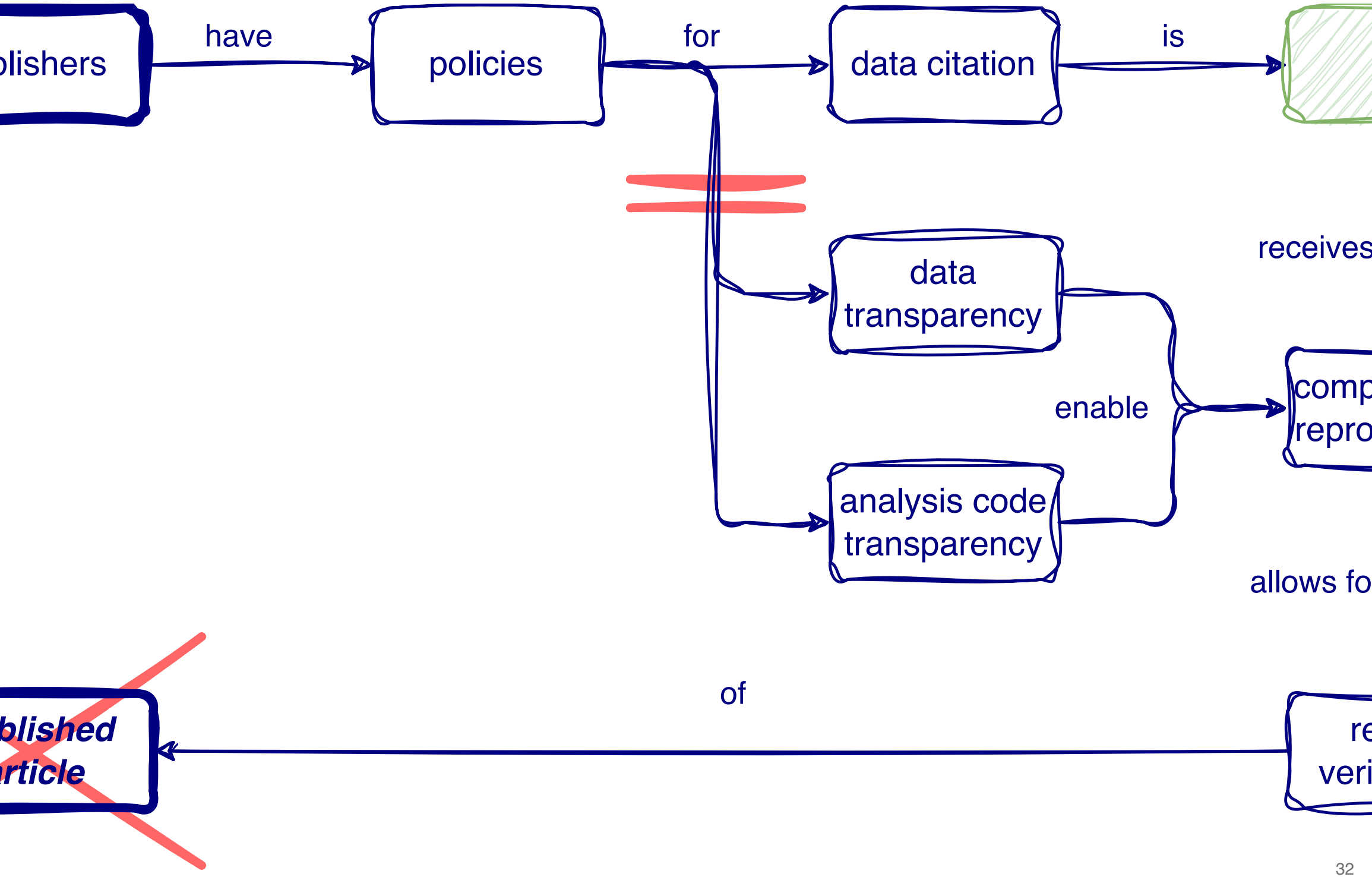

have
policies
for
data citation
is


have
policies
for
data citation
is
data
transparency


have
policies
for
data citation
is
data
transparency
analysis code
transparency


have
policies
for
data citation
is
data
transparency
receives
enable
analysis code
transparency


have
policies
for
data citation
is
receives
data
transparency
enable
analysis code
transparency
allows for
of

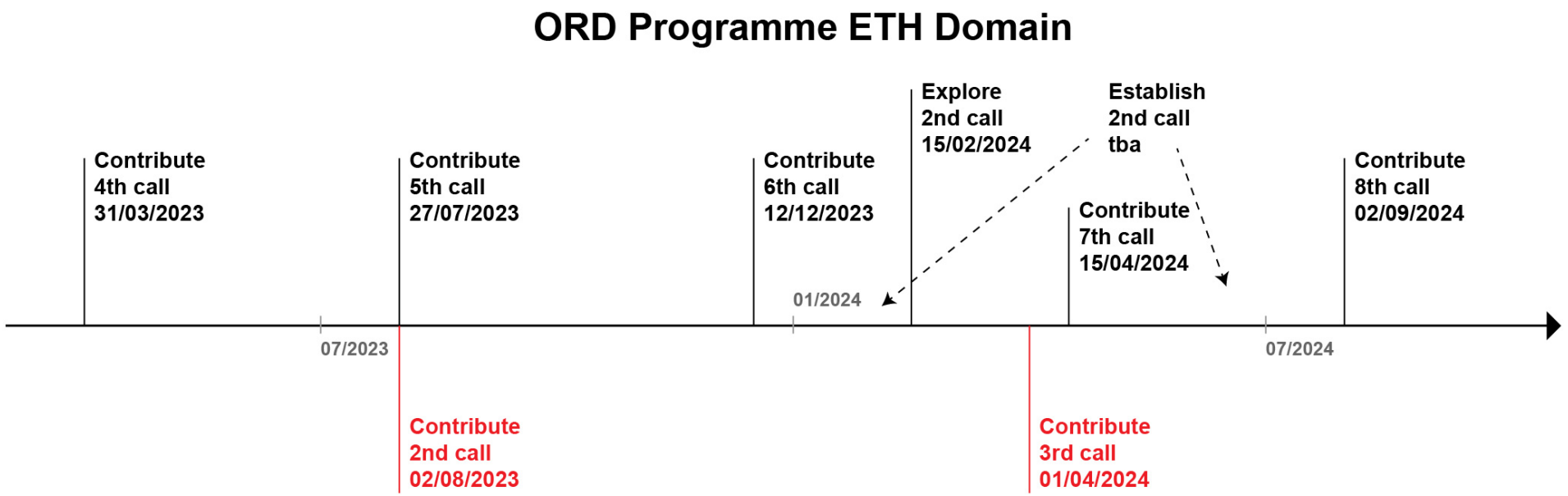

olishers
have
policies
for
data citation
is
receives
data
transparency
enable
comp
repro
analysis code
transparency
allows fo
of
blished
rticle
re
veri

Ongoing activities

Open Research Data (ORD) funding programmes

- Explore project (150k): https://openwashdata.org/pages/proposal/
- Contribute project (30k): https://github.com/openwashdata/proposal-eth-ord-scheme-contribute/blob/main/proposal-eth-ord-scheme-contribute.md

Upcoming funding opportunities



Data stewardship activities

- ETH4D data management support
- ETH4D skills course (HS 2023) - 0.5 ECTS: Open Research Data - Tools and a workflow for the modern researcher

Thanks!

Thanks

- Slides created via revealjs and Quarto: https://quarto.org/docs/presentations/revealjs/
- Access: Source code of slides on GitHub
- Download: Slides as PDF

This material is licensed under Creative Commons Attribution Share Alike 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.