open by default

a data management strategy

Lars Schöbitz

lschoebitz@ethz.ch

Global Health Engineering - ETH Zurich
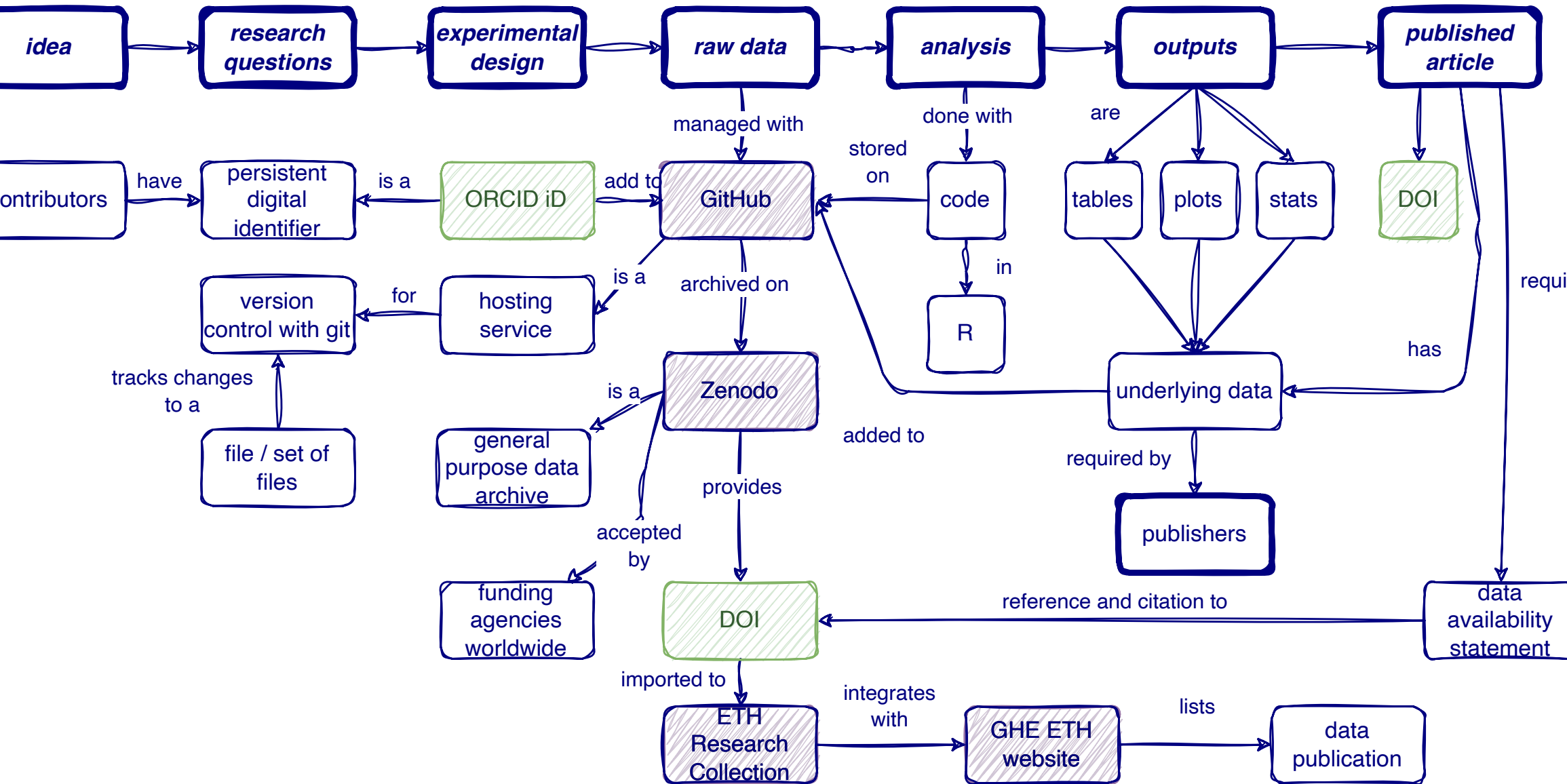
June 12, 2023

Project

My role - Open Science Specialist

- research data management

- reproducible workflows

- mindset for Open Science

- research communication

- teaching data science tools

- proposal writing

Data Management Strategy

~ 3 to 9 months

~ 1 to 3 years

idea → research questions → experimental design → raw data → analysis → outputs → published article

contributors — have → persistent digital identifier

ORCID iD — is a → persistent digital identifier

ORCID iD — add to → GitHub

raw data — managed with → GitHub

analysis — done with → code

code — stored on → GitHub

code — in → R

outputs — are → tables, plots, stats

published article — has → DOI

GitHub — is a → hosting service

hosting service — for → version control with git

version control with git — tracks changes to a → file / set of files

GitHub — archived on → Zenodo

Zenodo — is a → general purpose data archive

Zenodo — accepted by → funding agencies worldwide

Zenodo — provides → DOI

tables, plots, stats — underlying data

underlying data — added to → GitHub

underlying data — required by → publishers

published article — requires → underlying data

underlying data — has → data availability statement

DOI — imported to → ETH Research Collection

ETH Research Collection — integrates with → GHE ETH website

GHE ETH website — lists → data publication

data availability statement — reference and citation to → DOI

https://global-health-engineering.github.io/website/1-slides/

5

Characteristics

Data at Global Health Engineering

- small (few MBs)

- tabular

- non-sensitive

- topics

  - waste management

  - sanitation

  - air quality

  - etc.

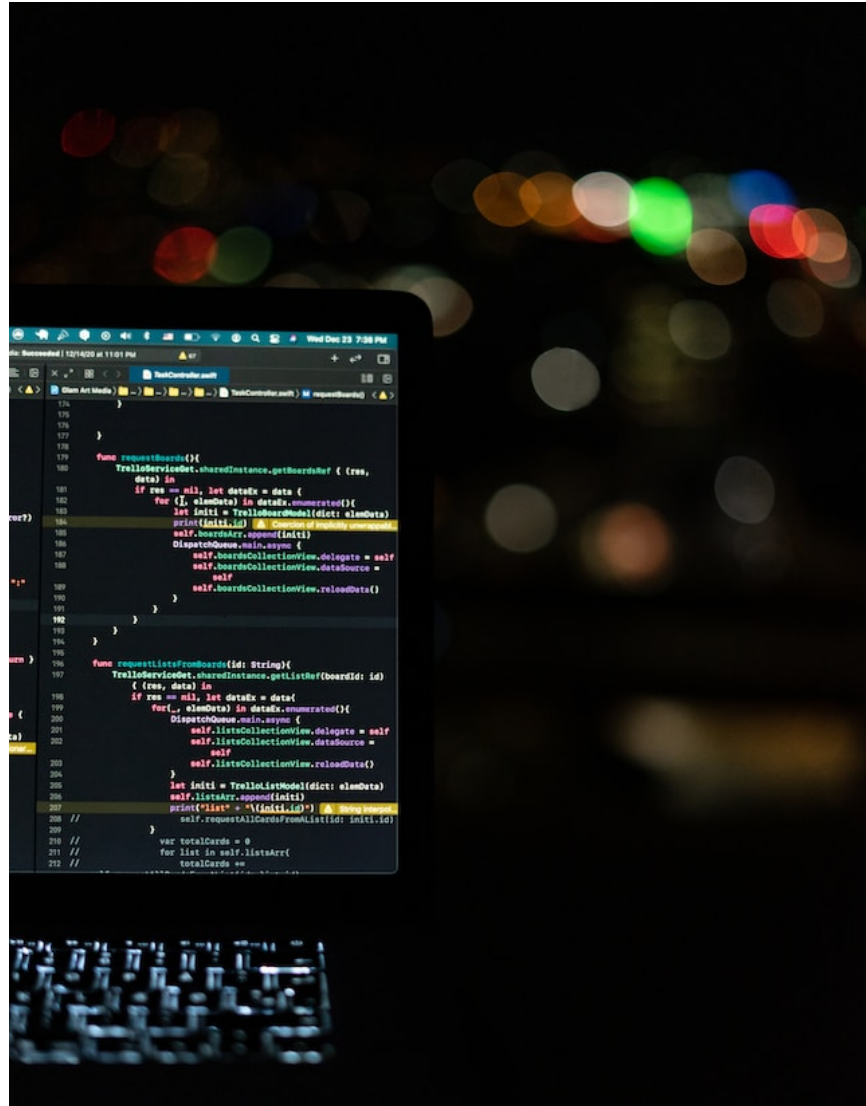# Pain Points

Lost and overwhelmed

- terminology

- licensing

- security

- repositories

Computational tools

- intimidating

- hard

- not everyone's business

Scientist as a lone ranger

- acts alone

- exposure to tools

- own stack of tools

- the way to do it

DMP? 👍

https://github.com/Global-Health-Engineering/dmp-sor4d

Thanks! 🌻

Thanks

Slides created via **revealjs** and **Quarto**:
https://quarto.org/docs/presentations/revealjs/

Source code: GitHub

Access: HTML

Download: PDF

This material is licensed under Creative Commons Attribution Share Alike 4.0 International.