open by default

a data management strategy

Lars Schöbitz

lschoebitz@ethz.ch

Global Health Engineering - ETH Zurich
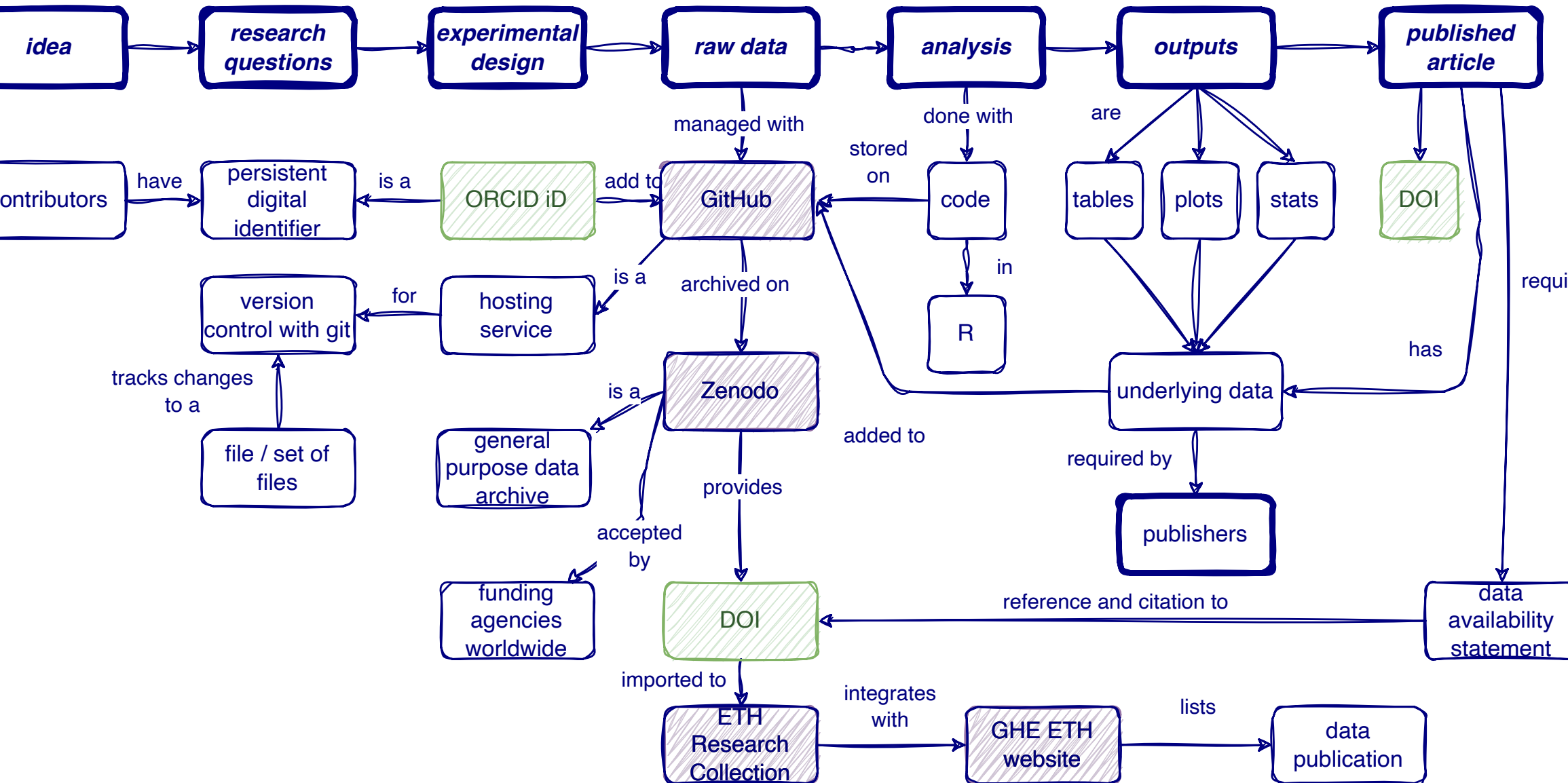
June 12, 2023

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** is a **ORCID iD** add to **GitHub**

raw data managed with **GitHub**

analysis done with **code**

outputs are **tables**, **plots**, **stats**

published article → **DOI**

GitHub is a **hosting service** for **version control with git**

tracks changes to a **file / set of files**

code in **R**

GitHub stored on / code

GitHub archived on **Zenodo**

Zenodo is a **general purpose data archive**

Zenodo accepted by **funding agencies worldwide**

Zenodo provides **DOI**

outputs → **underlying data**

underlying data required by **publishers**

published article requires / has **underlying data**

underlying data added to GitHub

DOI reference and citation to **data availability statement**

DOI imported to **ETH Research Collection** integrates with **GHE ETH website** lists **data publication**

Characteristics

Data at Global Health Engineering

- small (few MBs)

- tabular

- non-sensitive

- topics

  - waste management

  - sanitation

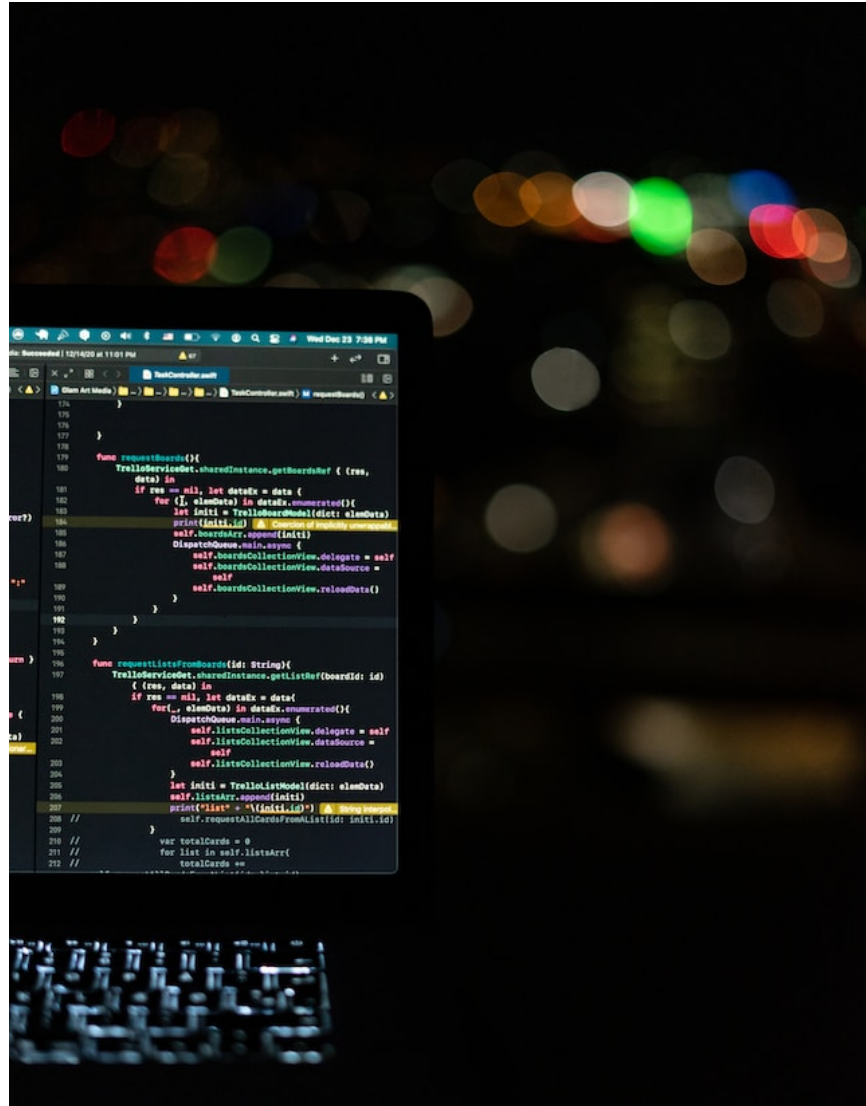 - air quality

  - 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.