open by default

a data management strategy

Lars Schöbitz

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

Characteristics

Data at Global Health Engineering

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

```
knitr::include_graphics("img/kiteezi-landfill-kampala-uganda.jpg")
```


Photo: Kiteezi Landfill, Kampala, Uganda

Pain Points

Lost and overwhelmed

- terminology
- licensing
- security
- repositories

- terminology (underlying data, raw data, processed data, metadata, etc.)
- licensing (CC-BY, MIT, public domain, code, data)
- security & privacy (phone numbers, GPS locations of houses)
- repositories (documentation, upload, maintenance)

Computational notebooks

- 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

- acts alone (until becoming a manager/professor)
- exposure to tools (100s throughout research career at different institutions)
- own stack of tools (builds their own individual worksflows -> not team oriented)
- the way to do it (doesn't have time to learn what would be most effective for a team)

Data Management Plan

Yes! :)

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