UNIVERSITY OF OSLO

ConcePTION:

An outsider perspective

Luigi Maglanoc

Data Management, IT-Department

How DAPs should «work» with local IT-infrastructure

Advice regarding R-scripts

How DAPs should «work» with local IT-infrastructure

UiO and TSD

- Several VMs per project
  - Split users across VMs/projects, to avoid memory overload
- Adequate specs on VMs (RAM, Cores)
- Linux vs Windows
- Colossus – HPC
- https://www.uio.no/english/services/it/research/sensitive-data/help/hpc/

UNIVERSITY
OF OSLO

Advice to groups writing R-scripts

UNIVERSITY
OF OSLO

Git and Github

- Repository size
    - One large repo
    - Many smaller repos

- Branching
    - Use for developing new code – more transparent and better control
    - Pull requests will enable possibility for «code review»

- (Github actions)

Coding/computing best practices

- Minimum level
  - Create own functions

- Memory management in R
  - Packages
  - Vectorization

- Test driven development (TDD) / checks
  - Test if input/output are the way they are supposed to be
  - Type checking

- Test scripts on «real» data
  - Realistic size
  - With several DAPs (alternative: 1-to-1 meeting)
  - And/or realistic simulated CDM data

- Ideally - more than one person for a given repo
  - Maintenance over time
  - Who to contact when errors?

- Ideally - clear directory structure across repos for related work

- Colloborate / communicate with IT-deparmtent or someone with computer science background
  - Technical consderations regarding computing performance, memory, databases, cloud servers, etc.
  - Consideration of DAP's IT-infrastructure

UNIVERSITY
OF OSLO

Provided instructions for DAPs

- What is necessary to check for relevant steps in a set of scripts
    - Input (e.g. what is required, how to fill out csv files)
    - output (are all outputs made? Contains what you expect?)

- Consider splitting up scripts/steps (related to previous points)

- Clear instruction on what can be, and should be rerun

- Can be easier to assess if something went wrong, rather than discovering much later in process

UNIVERSITY
OF OSLO

Misc

- Clear communication between teams and team members
    - Agreement on what should be done, and how, at each stage
    - Also, across groups

- Project management

- «System» to notify DAPs when and what scripts to use
    - Some existing, but only «enforced» at start

UNIVERSITY
OF OSLO