

Utrecht, Netherlands | June 13 - June 15

### NEUROIMAGING RESEARCH DATA LIFE-CYCLE MANAGEMENT

Hurng-Chun (Hong) Lee, Robert Oostenveld, Erik van den Boogert, Eric Maris





### Outlines

- Lifecycle RDM: objectives and challenges
- The method
  - the RDM protocol
  - Donders Research Data Repository (DRDR) usage of iRODS
- Strength and weakness of DRDR
- Future focuses

### O

### Lifecycle RDM

- RDM spans the entire research lifecycle
- Objectives:
  - long-term data preservation
  - scientific-process documentation
  - data publication





## Challenges

- large institute with heterogeneous scientific-administrative workflows
  - 600 researchers in PI groups  $\rightarrow$  more than 150 projects per year
  - 3 centres with 4 administrative domains
- data complexity:
  - text, audio/video, imaging or signal data, etc.
  - sensitive data
  - size ranges from a few large (>2GB) files to a huge amount of small files (<1MB)</li>
- user expectation



data acquisition conception of research data analysis data publication

5















### Organisational Unit (OU)





## The data repository

- a iRODS-based ICT system implementing the workflow defined by the protocol
- a single data-management system enabling internal collaboration, and external data sharing





### Storage resources





### Collection namespace

- namespace reflects administrative hierarchy
- metadata in KVU triplets
- role-based authorisation with iRODS groups





## Management rules

• a RPC-like interface for collection management



# User provisioning and authentication



- authentication via a national federated IdP
- user is provisioned upon sign-up to the management portal
  - IdP attributes are stored as user KVU-triplets in iRODS
  - setup PAM authentication on OTP (one-time password) for data access



### Event logging

- essential user actions are logged as events
- non-blocking way streaming events to Elastic stack via "filebeat"





### User interfaces

- separating data-access from collection management
- web portal for collection management
- WebDAV (Davrods) for "easy" data transfer
- file stager: a service "intelligently" managing bulk file transfer between a local storage and the repository



# Strength and weakness

It fits to a combined scientific-administrative workflow of a large and heterogeneous institute

It provides sufficient functionality for

- 1. sharing data for publication
- 2. implementing Data Management Plan (DMP)

It has weak integration with data analysis facility

It doesn't implement standard way of organising collection content



### Future focuses

- seamless integration with computing (dataanalysis) facility
- FAIR-ness of published data collections
- adoption beyond neuroimaging



### Summary

- We structured a RDM workflow
  - which covers the entire research lifecycle
  - in which both researcher and administrator take part of responsibility
  - which is specified by protocol; implemented by a iRODSbased digital repository

### https://data.donders.ru.nl