Maastricht UMC+
DataHub

First Line Research Data Management for the Life Sciences

A case study.

Paul van Schayck
Maarten Coonen

About DataHub Maastricht

Maastricht UMC+

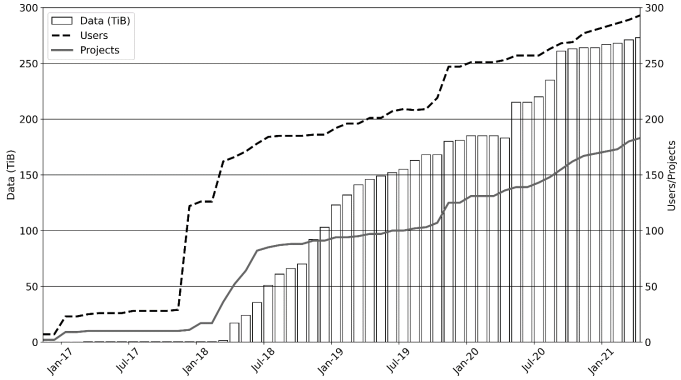
DataHub

- Research Data Management support
- Life sciences faculty and university medical centre
- Started ~5 years ago

DataHub Maastricht Services

Maastricht Data Repository

- Project/dataset level
- Access control
- High volume
- Data publishing (PIDs)
- Data "tiering"
- Links with other domain RDM platforms



iRODS

Other services:

- Pseudonymisation (electronic health records)
- Data catalogue (Ontoforce DISQOVER)
- Clinical research data warehouse (OMOP)
- DMPOnline

About me

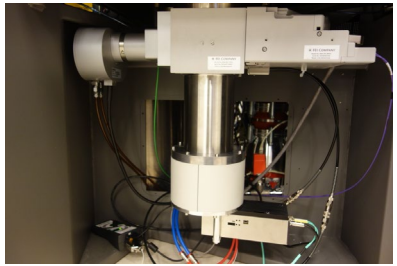
Maastricht UMC+
DataHub

Developer for the Maastricht Data Repository

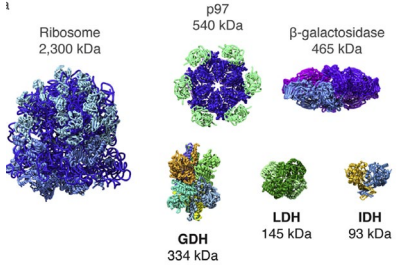
Data Steward

M4I

PhD student in Structural Biology



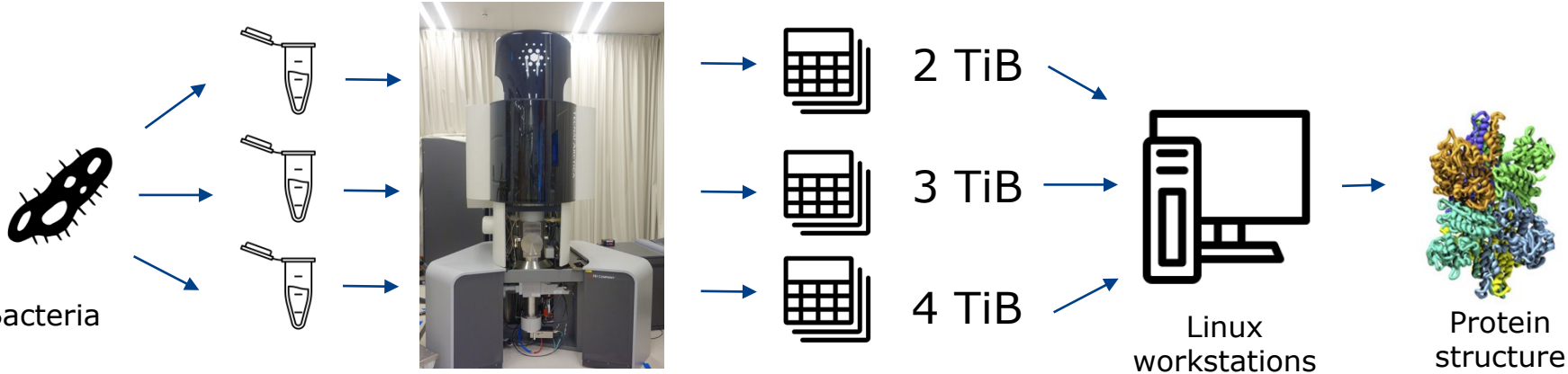
Timepix3 cryo-EM detector



Macromolecular structures

Life sciences: a data intensive example

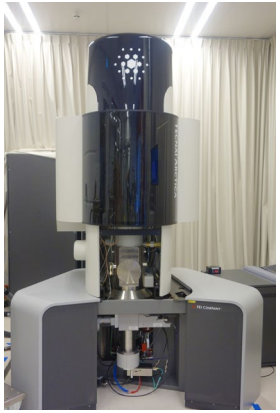
Abril



Life sciences: a data intensive example

Abril

Bacteria



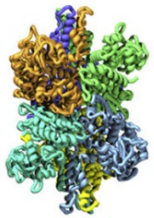
Microscope

2 TiB

3 TiB

4 TiB

Linux workstations



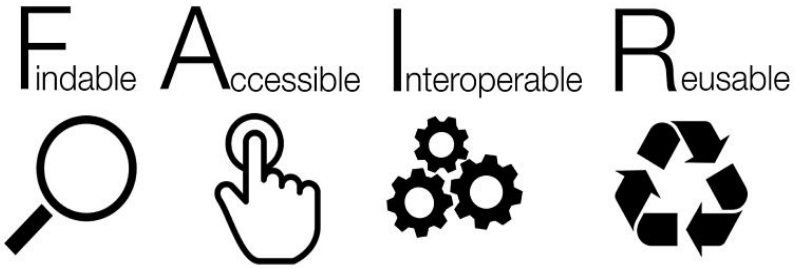
Protein structure

Life sciences: characteristics

- Small groups
- Data intensive
- Multi-modality
- Diverse sub-domains

- Dubbed a “small science”

High standards top-down



DataHub to the Rescue!

Maastricht UMC+

DataHub

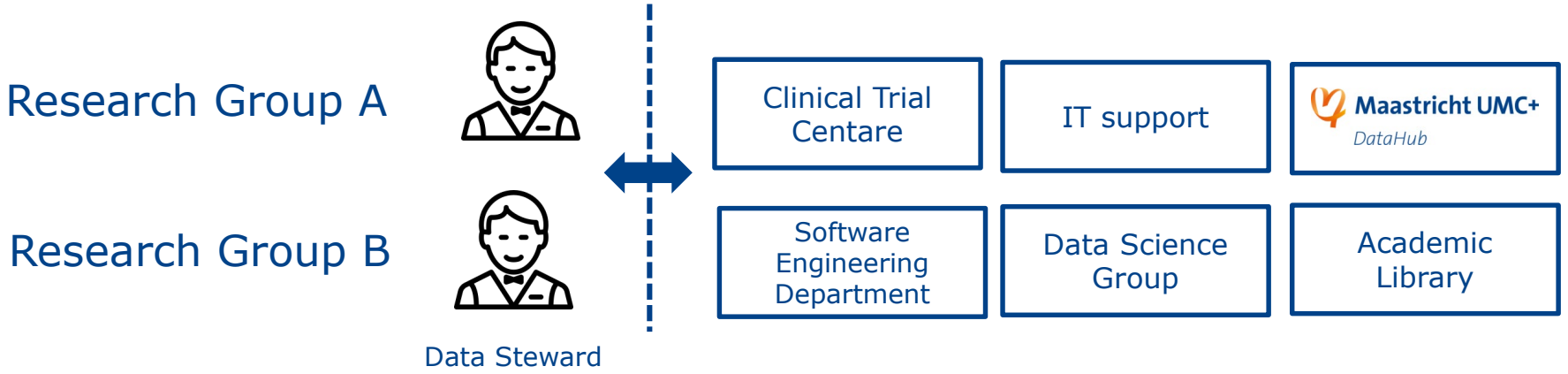
Lessons learnt by DataHub

Reflect on ~5 years DataHub:

1. Organisational
2. Technological

Disciplinary Data Stewards

- Best placed directly at the research group
- Stakeholders in DataHub Scrum process



How to prioritise support?

- Time and manpower are a limited resource
- Context switching between different projects
- How to steer?

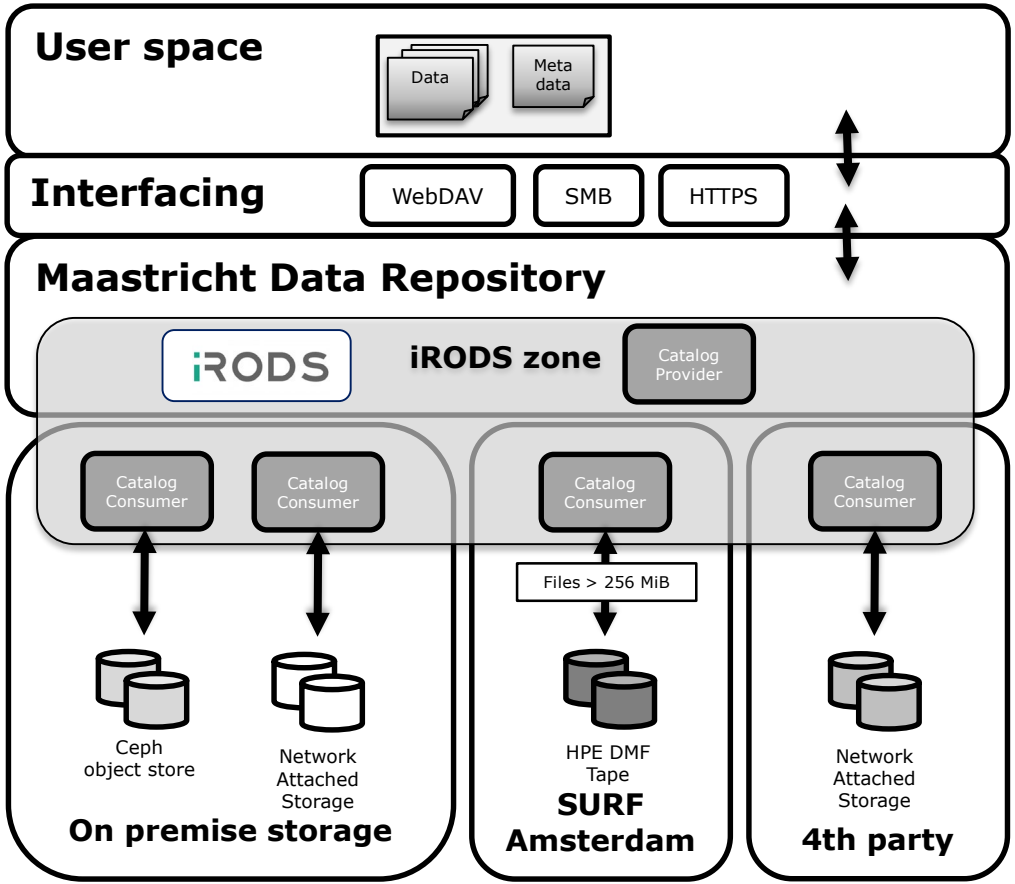
Prioritise efforts on the most reusable data

- Domain Specific Repositories
- Data Stewards
- Development in Scrum process

"If Data Sharing is the Answer, what is the Question?"
Christine L. Borgman

Reducing storage costs through tiered storage

Datasets stored mixed: small files locally and big files remote on tape



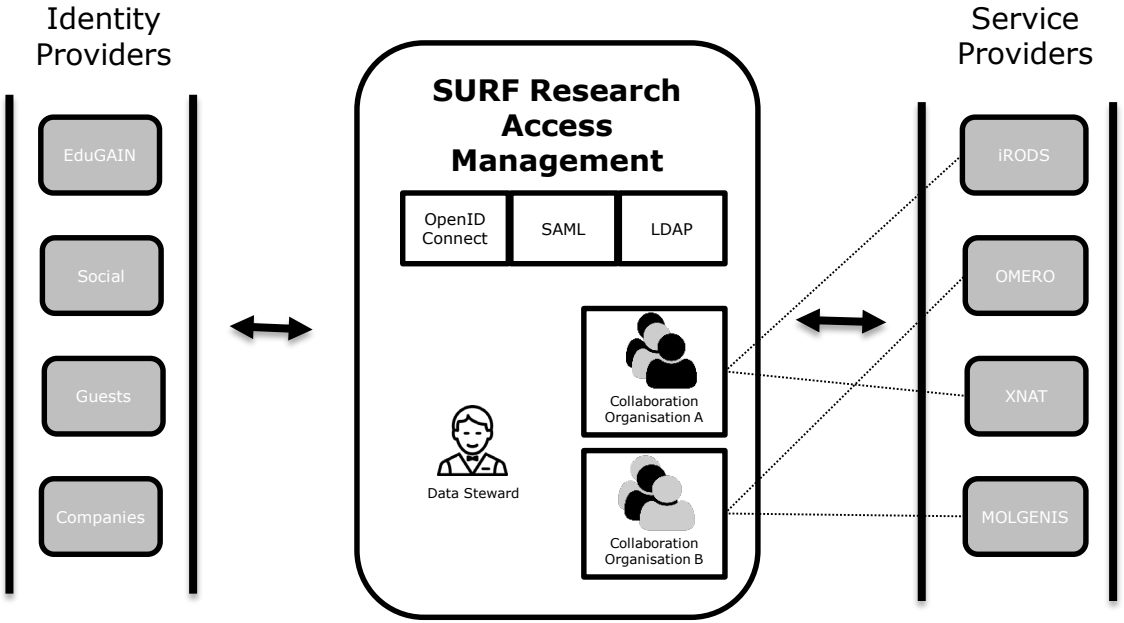
Supporting diversity: Domain RDM tools

Microscopy

Biomedical images

Molecular Genetics

Offering cross-institutional access



elixir

EUROPEAN OPEN SCIENCE CLOUD

eduTEAMS

Outlook

1. Organisational
   - Further embedding of data stewards
     - Mijke Jetten, 14:40 this afternoon
   - Appraisal and reappraisal of data
2. Technological
   - Flexible metadata schema’s

Visit us at **https://datahubmaastricht.nl**

Maastricht UMC+
DataHub

Paul van Schayck
Data Engineer
Data Steward

DataHub Maastricht
Paul-Henri Spaaklaan 1
6229 EN Maastricht
The Netherlands

T +31 6 27 07 16 54
E p.vanschayck@maastrichtuniversity.nl