Research Data Management at KU Leuven: infrastructure and services

Leen Van Rentergem
ICTS KU Leuven
Welcome in Leuven
Human Health
Medical Technologies
Bio-Sciences & Environment
Matter, Materials & Energy
Nature Unlimited
Manufacturing & ICT
Arts, Religion & Culture
Economy, Law & Society
Human Behavior

> 10,000 Peer Reviewed Publications
149 Spin-offs
130 m€ License Income

> 1850 faculty
Research Data Management

The Challenge

ACCEPTED
Challenge 1: Why RDM

RDM is relevant for the researcher
RDM helps them...

• collect/create **high quality data** → high quality research
• work **more efficiently**
  • *e.g.* develop **consistent procedures** early on, check that you fulfill legal and ethical obligations, improve impact and quality of grant applications, etc.
• **Protect** their research data (e.g., security measures, data loss,...)
• Following their **values** as a researcher:
  • striving for **scientific integrity**
  • ensuring **scientific reproducibility**
Challenge 1: Why RDM

RDM is relevant for institutes

RDM helps them...

- collect/create **high quality data** → high quality research
- **Protect** their research data (e.g., security measures, data loss, …)
- Track and promote **reuse of data**
- Following their **values** as a institution:
  - striving for **scientific integrity**
  - ensuring **scientific reproducibility**
KU Leuven data policy

KU Leuven considers research data as a valuable research output, an asset to KU Leuven and a critical contribution to the knowledge economy. A high standard of research data management is fundamental to both high quality research and academic integrity.

CORE PRINCIPLES OF RESEARCH DATA MANAGEMENT AT KU LEUVEN

1. Research data must be stored and documented in a safe, secure & sustainable way - in this way data can be retrieved and accessed by KU Leuven when needed
   WHAT DOES THIS IMPLY?

2. Metadata of research data must be registered to ensure findability of the data
   WHAT DOES THIS IMPLY?

3. Deletion of data must be justified and documented; documentation relating to the data should not be deleted as this impedes the audit trail
   WHAT DOES THIS IMPLY?

4. After the end of the research, relevant research data must be retained, for a minimum of 10 years in a safe, secure & sustainable way for purposes of reproducibility, verification, and potential reuse; a best effort is expected to make relevant research data available for reuse
   WHAT DOES THIS IMPLY?

5. All research is carried out in accordance with and taking into account existing contractual agreements, legislation, regulations, guidelines or valorization potential
   WHAT DOES THIS IMPLY?
**Challenge 2:** Research ecosystem is complex, research data even more

<table>
<thead>
<tr>
<th>Starting project</th>
<th>Active project</th>
<th>Published Project</th>
<th>Long-term (from during to years beyond the project)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data management plan description</td>
<td>Data Collection</td>
<td>Data Storage</td>
<td>Metadata description</td>
</tr>
<tr>
<td>(Legal) compliance</td>
<td></td>
<td></td>
<td>Data Processing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Publishing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Metadata Publishing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Archiving</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Preservation</td>
</tr>
</tbody>
</table>
RDM infrastructure and tools at KU Leuven
Data flow between data storage systems

**ACTIVE DATA**
- Data Collection
- Data Storage
- Metadata descript.
- Data process

**PUBLISH DATA**
- Data publishing
- Metadata publishing

**LONG TERM STORAGE**
- Data archiving
- Data Preserv.

Other Storage
- GVS
- OneDrive
- Shared drives
- HPC

Regional/domain specific rep.

RDR

Teneo (preservation)

Cold storage (archiving)
Challenge 3: Move in a Connected world
Challenge 4: “FAIR”ification

1. Retrieve non-FAIR data
2. Analyse the retrieved data
3. Define the semantic model
4. Make data linkable
5. Assign license
6. Define metadata for the dataset
7. Deploy FAIR data resource

Questions across multiple sources
Combine with other FAIR data

Findable Accessible Interoperable Reusable
Help and encourage researchers to apply good data management practices.

- Give context to data
- Automate data workflows
- Search data
- Share data
- Secure data

Active data → Published data → Regional/domain specific rep. → Long term storage → Cold storage (archiving) → Teneo (preservation) → iRODS

KU LEUVEN
# Active data platform (iRODS)

## Central Storage
- High Available
- Secure

## Metadata
- Rich metadata templates
- Management interface:
  - Upload/create/modify/share
  - Enforce standardization
  - Combine and apply

## Multiple clients/interfaces
- GUI, CLI, WebDav, NFSRods, Python, Go, etc
- Linux, MacOS and Windows

## Workflow automation
- System wide/project rules:
  - e.g: netCDF metadata extraction)
- Delegation of the implementation/deployment of rules
Challenge 5: The Cost
Sponsoring

DEPARTEMENT ECONOMIE WETSCHAP & INNOVATIE
Vlaamse overheid

KU LEUVEN

Tiering

 Tier 1  Tier 2  Tier 3  Tier 4

Sponsored Critical  Hot  Warm  Cold

Use the right solution

This Storage Guide helps you choose central IT solutions for data storage and management. Based on the needs you indicate in the left column, the possible solutions are shown. Click on a solution to see more details. From the details you can click even further to the relevant service sheet in the ICTS Service Catalogue.

For more information please contact the ICTS Service Desk or the EDM Support Desk.

Clear all filters
What is the classification of your data with regard to confidentiality?
- Yes
- Confidential
- Strictly confidential

Do you want to share data with colleagues?
- No
- Yes, but only within the university
- Yes, with persons outside the university

What type of data do you want to store?
- Research data
- Other

The ability to add metadata is important to me?
- Yes

How much storage space do you think you will need?
- Less than 1TB
- Between 1TB and 5 TB
- More than 5TB

To what extent are your data reproducible?
- Easy
- Difficult
- Not

iRODS
Solution for storing active research data, with corresponding management tools for (meta)data.

Large Volume Storage
Solution for storing large amounts of research data in a cost-efficient manner.

Shared network drive
A shared file system available to all staff via PCs and laptops managed by ICTS KU Leuven or your local KU Leuven IT department.

SharePoint on premise-site
On premise Microsoft solution for storing, managing and sharing documents and files, hosted in the ICTS data center.

SharePoint online-site
Microsoft cloud solution to store, manage and share documents and files.

Teams-site
Microsoft cloud solution for communication and collaboration.

Personal Network Drive
A personal file system available to most staff via PCs and laptops managed by ICTS KU Leuven or your local KU Leuven IT department.

OneDrive for Business cloud storage
Personal cloud storage via KU Leuven Microsoft 365.

Enterprise Box cloud storage

Local storage on managed PC/laptop
Challenge 6: The Expertise

**Expertise**
/eks-per-tydz/
def: Special skill or knowledge in a particular subject. eg. He has expertise in his field of molecular science.
RDM Network @ KU Leuven

RDM Policy initiatives

- **International** (e.g. LERU)
- **National** (e.g. FOSB)

- **KU Leuven context:**
  - Open Science Task Force
  - RDM steering committee
  - Underlying WGs:
    - Policy
    - Infrastructure
    - Advice&Training

RDM Competence Centre

RDM Support Staff

Central & Embedded

- Departement
- Faculty
- Libraries
- Local services (e.g., IT)
- Central Services
- RDM Network KU Leuven Libraries

RDM Support Desk

RESEARCHERS

Strategic/Policy

Operation / Support
Challenge 7: Researcher-first
Data Steward

"I ensure data usage complies with our policies."
The greater the obstacle, the more glory in overcoming it.

~ Moliere