Outline

• Short introduction to research data management (RDM) and open science

• RDM in the EU, Sweden and Lund

• RDM in practice:
  – data management plans (DMPs)
  – describing the data with high quality metadata
  – high performance computer processing
  – choosing a trusted repository for data publishing

• Discussion (though please ask questions at any time)
Research data are collected, observed or created, for the purposes of analysis to produce and validate original research results

• Analogue as well as digital resources

• Can include other resources needed to understand the data, e.g. lab notebook, coding scheme, software

“Research data management is an explicit process covering the creation and stewardship of research materials to enable their use for as long as they retain value.”

Introduction to RDM

Why open science?

• to avoid unnecessary duplication of collection effort
• to validate results earlier if required
• so research is visible and has impact
• so data can be reused by other researchers
• to get credit when others cite the research
• society and funders increasingly expect that data is open
Introduction to RDM

What is involved in RDM?

• Data Management Planning
• Creating data
• Documenting data
• Accessing / using data
• Storage and backup
• Sharing data
• Preserving data

Source: http://www.ed.ac.uk/information-services/research-support/research-data-service
# RDM in the EU and Sweden

## EU
- Open science cloud
- EUDAT RDM services
- OpenAire
- Open access publishing and data sharing required in Horizon 2020
- DMPs are required from January 2017

## Sweden
- No national guidelines, policies, resources
- Wide range of actors, few with any mandate
  - VR, SUHF, URFI, SNIC, Sunet, SCC… …
- Attempt to coordinate the universities’ efforts: Swedish National Data Service
RDM at Lund University

• The Research Board’s working group for research data – RDM policy, coordinate support functions, intelligence work, expert advice

• Research infrastructures, e.g. ESS, ICOS Carbon Portal, Max IV, HumLab

• E-science and e-infrastructure, e.g. eSSENCE, LUNARC

• Seminars series, e.g. Digital Tools in the Humanities, REACH
RDM in practice...
Data management plans

- Describe the data the researcher will collect or produce during the whole research project
- Describe how the researcher will analyse, archive and share data during and after the research project
- Requirement from many research funders internationally, e.g. Horizon 2020

Tools for creating a DMP
- DMP Online, DMP Tool
- Templates for DMPs

A data management plan provides guidance throughout the research process
Data management plans

• Topics from the H2020 template
• **Summary**: purpose of data collection, objectives of project, types and formats of data, re-use of data … …
• Making data FAIR
  – **Findable**: metadata provision, standards, vocabularies; use of PIDs (Persistent Identifiers), naming conventions..
  – (Openly) **Accessible**: which data will be openly accessible, how, when, tools needed, metadata storage…

Tip: Check out the FAIR principles – they’re prevalent in the open science discourse
Data management plans

Topics from the H2020 template – Making data FAIR cont’d

– **Interoperable**: data and metadata vocabularies, standards, methodologies chosen to facilitate interoperability; mapping between project’s ontology and other common ontologies

– **Re-usable**: licenses for widest reuse possible; when and for how long data will be reusable, motivate restrictions; data quality assurance…

Tip: To create your own DMP, create an account in DMP Online
Data management plans

Topics from the H2020 template cont’d

- **Allocation of resources**
  - Costs for making data FAIR, how costs are covered; responsibilities for RDM in the project, costs and value of long-term preservation

- **Data security**: recovery, storage, transfer of sensitive data

- **Ethical aspects**: ethics review, ethics deliverables…

- **Other**: use of other RDM procedures external to H2020

Tip: Contact the personal data controller, Johanna Alhem, for advice on managing data that contains personal data
Describing data with high quality metadata

Metadata: data about other data – enables identification, location and retrieval of data

- **Administrative**: IP rights, versions, modifications
- **Structural**: how data are organised
- **Semantic/Topical**: the topic/aboutness of the data, e.g. subject, categories, topics, keywords
- **Provenance**: creation, attribution, version history, stewardship
- **System**: creation/deposit time, owner, storage location, data retention period
- **Use**: user access, user tracking
Describing data with high quality metadata

DataCite metadata scheme

<table>
<thead>
<tr>
<th>Mandatory</th>
<th>Recommended</th>
<th>Optional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifier</td>
<td>Subject</td>
<td>Language</td>
</tr>
<tr>
<td>Creator</td>
<td>Contributor</td>
<td>Alternate ID</td>
</tr>
<tr>
<td>Title</td>
<td>Date</td>
<td>Size</td>
</tr>
<tr>
<td>Publisher</td>
<td>Resource Type</td>
<td>Format</td>
</tr>
<tr>
<td>Publication year</td>
<td>Related identifier</td>
<td>Version</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>Rights</td>
</tr>
<tr>
<td></td>
<td>GeoLocation</td>
<td></td>
</tr>
</tbody>
</table>

Tip: Provide as much metadata as possible to promote findability and reusability

Image source: Project THOR

@mtholder motivating git: You mostly collaborate with yourself, and me-from-two-months-ago never responds to email.
@swcarpentry
Describing data with high quality metadata

• Important aspects
• Consider the interoperability of your data and other data, repositories and tools
• Metadata standard suitable for the data and discipline http://rd-alliance.github.io/metadata-directory/standards/
• Ontologies/controlled vocabularies suitable for describing the topical metadata

Tip: Ask a librarian for advice – we’re metadata experts
High performance computing

- Cloud computing – virtual machines
- High-throughput - parallel computing
- Storage during active projects
- Archiving
High performance computing

SNIC – Swedish National Infrastructure for Computing

• Large-scale computation and data storage (for active projects)

• Available for all scientific disciplines from all Swedish universities

• Open application procedures to support the best Swedish research
  – Small, medium, large allocations for computing and storage (Swestore)

EGI – European Grid Infrastructure
High performance computing

- **LUNARC** – LU’s HPC centre
- SUPR service for resource allocation applications [https://supr.snic.se/](https://supr.snic.se/)
  - Rounds are announced on website
- Current examples from Lund
  - Total energy calculations of As-Sn compounds, Masoomeh Ghasemi
  - Physicochemical properties of intrinsically disordered proteins, Marie Skepö
  - Carbon cycle data assimilation on Bethy high resolution, Mousong Wu

### Rounds

<table>
<thead>
<tr>
<th>Round</th>
<th>Deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNAC Medium, 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNAC Small C3SE, 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNAC Small HPC2N, 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNAC Small Lunarc, 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNAC Small NSC, 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNAC Small UPPMAX, 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNAC Medium Swestore 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNAC Small Swestore 2017</td>
<td>—</td>
</tr>
<tr>
<td>DCS 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNIC SENS Medium 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNIC SENS Small 2017</td>
<td>—</td>
</tr>
<tr>
<td>UPPNEX 2017</td>
<td>—</td>
</tr>
<tr>
<td>C3SE Local, 2017</td>
<td>—</td>
</tr>
<tr>
<td>LIU Local, 2017</td>
<td>—</td>
</tr>
<tr>
<td>L-SENS, 2017</td>
<td>—</td>
</tr>
<tr>
<td>LU Local, 2017</td>
<td>—</td>
</tr>
<tr>
<td>Mosler 2017</td>
<td>—</td>
</tr>
<tr>
<td>PDC Industry 2017</td>
<td>—</td>
</tr>
<tr>
<td>PDC Local 2017</td>
<td>—</td>
</tr>
<tr>
<td>Scilifelab Storage 2017</td>
<td>—</td>
</tr>
<tr>
<td>UPPMAX courses 2017</td>
<td>—</td>
</tr>
<tr>
<td>UPPMAX Storage 2017</td>
<td>—</td>
</tr>
<tr>
<td>SNIC Science Cloud 2017</td>
<td>—</td>
</tr>
</tbody>
</table>
Sharing data

• Different levels of access to data
• Open access to anybody
• Restricted access
• Access upon request to the data producer/steward

Tip: Consider any limitations to sharing the data, e.g. ethical, intellectual property, sensitive data before sharing.
Where to share and find data

Registry of ~3000 data repositories (e.g. domain specific):
re3data.org

General repositories
- Dataverse
- Dryad
- Zenodo
- FigShare
- Mendeley

Data journals
- F1000 Research
- Biodiversity Data Journal
- Scientific Data

Tip: Choose a repository that provides DOIs or other PIDs (persistent identifiers) to facilitate data citation
Aspects to consider when selecting a repository

• Citations: DOIs, PIDs
• Data security
• Metadata fields
• File formats
• Volume and size limitations
• Licences
• Archiving
Further advice and support

• **e@LU project** provides help and advice, see: http://meta.blogg.lu.se/about-the-elu-project-on-metadata-ontologies-and-provenance-for-research-data/

• Some **faculty libraries** provide help advice, see: http://www.lub.lu.se/en/services-and-activities/research-data

• **University Library** provides help and advice, see: http://www.ub.lu.se/en/publish/research-data

• For other parts of **LU’s support organisation**, see: http://www.lub.lu.se/en/services-and-activities/research-data/contacts-and-research-data-initiatives
### List of links

<table>
<thead>
<tr>
<th>Link</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCC</td>
<td><a href="http://www.dcc.ac.uk/">http://www.dcc.ac.uk/</a></td>
</tr>
<tr>
<td>FAIR principles</td>
<td><a href="https://www.force11.org/fairprinciples">https://www.force11.org/fairprinciples</a></td>
</tr>
<tr>
<td>DMP Online</td>
<td><a href="https://dmponline.dcc.ac.uk/">https://dmponline.dcc.ac.uk/</a></td>
</tr>
<tr>
<td>Open Aire</td>
<td><a href="https://www.openaire.eu/">https://www.openaire.eu/</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Link</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dryad</td>
<td><a href="https://datadryad.org/">https://datadryad.org/</a></td>
</tr>
<tr>
<td>Zenodo</td>
<td><a href="https://zenodo.org/">https://zenodo.org/</a></td>
</tr>
<tr>
<td>FigShare</td>
<td><a href="https://figshare.com/">https://figshare.com/</a></td>
</tr>
<tr>
<td>Mendeley</td>
<td><a href="https://www.mendeley.com/">https://www.mendeley.com/</a></td>
</tr>
<tr>
<td>SND</td>
<td><a href="https://snd.gu.se/">https://snd.gu.se/</a></td>
</tr>
<tr>
<td>ECDS</td>
<td><a href="https://ecds.se/">https://ecds.se/</a></td>
</tr>
<tr>
<td>BILS</td>
<td><a href="https://bils.se/">https://bils.se/</a></td>
</tr>
</tbody>
</table>

### FAIR principles
- [https://www.force11.org/fairprinciples](https://www.force11.org/fairprinciples)

### DMP Online
- [https://dmponline.dcc.ac.uk/](https://dmponline.dcc.ac.uk/)

### Open Aire
- [https://www.openaire.eu/](https://www.openaire.eu/)

### Horizon 2020, data management
Thanks for your attention!

maria.johnsson@ub.lu.se
monica.lassi@ub.lu.se
Extra material…
Data collection

File structure and organization

• A specified folder structure helps to find files
  – Not too complex
  – Short name of folders

• File naming is important to identify files
  – Create a system for file naming
  – Document system in DMP
  – Avoid too long and complicated file names

ProjectX_170101_rawdata_version_1.xls  ProjX_2017_raw_5647466566665433.xls
GOOD ARCHIVING – THE KEY TO FINDING INFORMATION

The archives at Lund University consist of compiled documents of the public authority, analogue and digital, regardless of age or where the documents are stored.

What does archiving entail and why must we archive?

• To archive means to systematically care for, preserve and organise the University’s public documents so that they are searchable and accessible to us internally as well as to the public.

The Swedish Archives Act specifies three main reasons for archiving within the public sector:

1. The right to access public documents,
2. the information required for administration of justice and other administration, and
3. the information required by research.
Data archiving and preservation

Swedish Archives Regulation, RA-FS 1999:1

• As a public authority Lund University must follow this regulation.
• All research records belong to the university.
• All research records should be properly archived and managed.
• The head of department is responsible for the records management – but in practice it lies on the researcher to describe and prepare research records for archiving.
Important to note further:

- Do your research data include personal data?

- File formats for archiving of research data, should be common and non-proprietary file formats

- Retention of research data. What is the case for your research data?