Managing research data – an introduction
SIM course, LTH

MARIA JOHNSSON, UNIVERSITY LIBRARY, 2 MARCH 2018
Presentación del marco de la presentación

1. Introducción a la gestión de datos de investigación (RDM)
2. Planes de gestión de datos
3. Recolección de datos
4. Análisis de datos
5. Compartir datos
6. Archivado y preservación de datos
7. RDM en la Universidad de Lund
8. Ayuda adicional y soporte
Definition of research data management (RDM)

“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.”

1. Introduction to RDM

What is research data?

- Research data are collected, observed or created, for the purposes of analysis to produce and validate original research results
- Both analogue and digital materials are 'data'
- Lab notebooks and software may be classed as 'data'
- Digital data can be:
  - created in a digital form ('born digital')
  - converted to a digital form (digitised)
  - Observational (not possible to recreate)
  - Experimental (should be possible to recreate)
1. Introduction to RDM

Data life cycle

- 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
1. Introduction to RDM

Why share research data and make it open?

• 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
1. Introduction to RDM

Society and funders increasingly expect that data is open

“Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.”

RCUK Common Principles on Data Policy, http://www.rcuk.ac.uk/research/datapolicy/


- Other research funders will require data management plans and increasingly also “open data” in the close future.

- Many scientific journals recommend/demand that the data your paper is based upon is made freely available, i.e. Scientific Data, Nature.com, http://www.nature.com/sdata/
FAIR-principles

• **FAIR** stands for:
  – **Findable**
  – **Accessible**
  – **Interoperable**
  – **Re-usable**

Principernu were crafted by FORCE11, more info: [https://www.force11.org/group/fairgroup/fairprinciples](https://www.force11.org/group/fairgroup/fairprinciples)

Are now widespread, i.e. in the Horizon 2020 program where they are used.
2. Data management plans

What is a data management plan (DMP)?

• Describes the data the researcher will collect or produce during the whole research project

• Describes how the researcher will analyse, archive and share data during and after the research project

• Requirement from many research funders internationally, i.e. Horizon 2020

Tools for creating a DMP

• DMP Online, DMP Tool

• Templates for DMPs

A data management plan is also a help for the continued research work
2. Data management plans

Task 2 in SIM course assessment

1. Select a subject

2. Create a data management plan for selected subject with help of the DMP template

3. Send in your data management plan to Emma-Lisa Hansson, deadline 3 April 2018

Further help

• See the DMP template, links at the end

• Contact Maria Johnsson or Emma-Lisa Hansson
3. Data collection

Data collected, some examples:
Data from: measurements
observations
surveys
registers

Data quality
• Complete
• Structured
• Understandable
3. Data collection

Data security

• Technical systems
  – Hard-drives
  – Internet connection
  – Backup

• Physical systems
  – Locked rooms
  – Safes

Administrative systems

✓ Who has access, key?
✓ Who has password?
✓ Who can read and/or write?

Storage

✓ Collection tools
✓ Cloud based
✓ Backup copies
3. 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
4. Data analysis

- Different phases of analysis, from raw data to cleaned, processed data
- Working files
- Documentation
- Versioning of data
- Data security and backup
- Different tools and techniques
4. Data analysis

Working files

- Several versions of data files
- Variables are added, removed, merged
- Researchers may use different sub-sets of data within a research group
- Recommended to create one master version of data
4. Data analysis

Documentation

• Why, when, what, how?
• Which level?
• What is sufficient for re-use?
• Which metadata?

• Good for your own
• Good for external revision/control

• Variable list
• Code books
4. Data analysis

Versioning of data

• Different versions of data should be labelled clearly, with date and version number

• Major or minor version changes
4. Data analysis

Data security and backup

• Regular backup, i.e. daily

• Storage on different places
4. Data analysis

Different tools and techniques

• Software
• Methods
• Quantitative data analysis
• Qualitative data analysis

• Which tools or methods do you use?
Open data – the culture of sharing data

• What’s in it for me?
  – Produce results quicker
  – Combine data sets to conduct new analyses
  – Facilitate new (cross-disciplinary) collaborations
  – Career strategy: data sets increasingly cited
5. Data sharing

Where to find and share data – some examples

Repositories:

- re3data.org
- Dataverse
- Dryad
- Zenodo
- FigShare
- Mendeley
5. Data sharing

Where to find and share data – some examples

Research infrastructures:
- Swedish National Data Services (SND)
- Environment Climate Data Sweden (ECDS)
- Bioinformatics Infrastructure for Life Sciences (BILS)
- Global Biodiversity information facility (GBIF)
- Worldwide LHC Computing Grid
5. Data sharing

Where to find and share data – some examples

Data journals:

- F1000 Research
- Biodiversity Data Journal
- Scientific Data
5. Data sharing

Limitations in data sharing:

- Data that contain personal data
  - Informed consent, ethical approval
- Intellectual, i.e. embargo period
- Commercial, i.e. patent
5. Data sharing

Different levels of access to data:

• Open access to everybody
• Restricted access to a certain group
• Access upon request to the data producer
6. Data archiving and preservation

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.
6. 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.
6. Data archiving and preservation

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?
7. RDM at Lund University

- Research infrastructures, e.g. ESS, ICOS, Max IV, HumLab
- E-science and e-infrastructure, e.g. eSSENCE, LUNARC
- The Research Board’s working group on research data > draft of a research data policy
- Lund University has joined a new consortia for research data headed by SND (Swedish National Data Service)
- Many seminars and other collaborations:
  - **Compile**, [http://www.compile.lu.se/](http://www.compile.lu.se/)
8. Further help and support

**Lund University Libraries**, common webpage with information, see:

https://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
## List of links

<table>
<thead>
<tr>
<th>Link</th>
<th>Website</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>FOSTER</td>
<td><a href="https://www.fosteropenscience.eu/">https://www.fosteropenscience.eu/</a></td>
</tr>
<tr>
<td>Open Aire</td>
<td><a href="https://www.openaire.eu/">https://www.openaire.eu/</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>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>
<tr>
<td>DCC, Horizon 2020, data management, FOSTER, Open Aire, ANDS, DMP Online, Dryad, GBIF</td>
<td></td>
</tr>
</tbody>
</table>
Thanks!

Maria Johnsson,
Dept of Scholarly Communication, University Library
maria.johnsson@ub.lu.se