GEN-8001: Take control of your PhD journey

Research data management
Part 1

Helene N. Andreassen, PhD
Lars Figenschou, PhD
Philipp Conzett

University Library, October 11, 2018
PART 1 - GENERAL BACKGROUND AND RATIONALE (13.15 – 14.30)
• Presentation
• Course objectives and layout (Lars)
• General background and rationale for research data management and sharing
  • Reproducibility and transparency (Lars)
  • FAIR data (Philipp)
  • Concerns about sharing data (Philipp) ACTIVITY
• Benefits of Data Sharing (Philipp)
• Expectations and requirements from publishers, funders and the institution (Helene)
• Research data management support at UiT (Helene) ACTIVITY

PART 2 – BEST PRACTICE FOR RESEARCH DATA MANAGEMENT (14:45 – 16:00)
• Qualitative and quantitative data
Course objectives, goals and layout

Objectives

...a glimpse into how research data should be managed

Data management plan at the outset of your project

Show you how to structure, document, and preserve data

...and how you can archive and share your own data – and use others
Course objectives, goals and layout

Goals

To help you crack some codes within “The Lifecycle of RDM” - i.e., how to structure your data

Be more prepared to fulfill present and future requests from research funding agencies and your home institution

 Understand (better) the background and rationale of data sharing

Make you share and re-use data
Course objectives, goals and layout

• **Layout**
  • VERY dynamic...(large variation in skills...!!)
  • Variation in your needs, so let us know!
  • Disruptions are necessary (welcome...)
  • Presentation
  • Tasks and discussions
  • Working with your own (or others) data
  • Tea/coffee/fruit/chat break

• Take home message: **WE ARE DEPENDENT ON ACTIVE PARTICIPATION FROM YOU**
General background and rationale for research data management and sharing
Reproducibility

- DATA
- SOURCE
- STANDARDS
- LICENCES & POLICIES
- PUBLICATIONS
- PEER REVIEW

We need a new ecosystem matching the digital world

Then we can harvest all the fruits i.e., higher degree of reproducibility - within all disciplines
IS THERE A REPRODUCIBILITY CRISIS?

- 7% Don’t know
- 52% Yes, a significant crisis
- 38% Yes, a slight crisis
- 3% No, there is no crisis

1,576 researchers surveyed
FAIR data

- Persistent identifier (DOI)
- Good metadata
- Indexed

- Well-defined & open protocols
- Adequate authentication

- Documentation
- Clear user license

- Open metadata & file formats
- Standard metadata
- Consistent vocabulary

Findable

Accessible

Re-usable

Inter-operable
Concerns about sharing data

Why do researchers have hesitations about sharing their research data?
Typical key worries about data sharing commonly voiced by researcher:
1. “I want to publish my work before anyone else sees my data.”
2. “I have not got the time or money to prepare data for sharing.”
3. “My data are not of interest or use to anyone else.”
4. “Other researchers would not understand my data at all - or may use them for the wrong purpose.”
5. “I am doing highly sensitive research. I cannot possibly make my data available for others to see.”
6. “I have made promises to destroy my data once the project finishes.”

☞ Activity (20 min.): Discuss in groups these six worries, and come up with some counter arguments in favour of data sharing. Add them to the distributed sheet.

(From Corti et al. (2014), pp. 10-12)
<table>
<thead>
<tr>
<th>Reasons not to share data</th>
<th>Counter arguments in favour of sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I want to publish my work before anyone else sees my data.”</td>
<td></td>
</tr>
<tr>
<td>“I have not got the time or money to prepare data for sharing.”</td>
<td></td>
</tr>
<tr>
<td>“My data are not of Interest or use to anyone else.”</td>
<td></td>
</tr>
<tr>
<td>“Other researchers would not understand my data at all - or may use them for the wrong purpose.”</td>
<td></td>
</tr>
<tr>
<td>“I am doing highly sensitive research. I cannot possibly make my data available for others to see.”</td>
<td></td>
</tr>
<tr>
<td>“I have made promises to destroy my data once the project finishes.”</td>
<td></td>
</tr>
</tbody>
</table>
Key worries about data sharing

Reason not to share data:
“I want to publish my work before anyone else sees my data.”

Counter argument in favour of sharing:
Data sharing will not stand in the way of you first using your data for your publications. Most research funders allow you some period of sole use, but also want timely sharing. Also remember that you have already been working with your data for some time so you undoubtedly know the data better than anyone coming to use them afresh.

(From Corti et al. (2014), pp. 10-12)
Key worries about data sharing

Reason not to share data:
“I have not got the time or money to prepare data for sharing.”

Counter argument in favour of sharing:
It is important to plan data management early in the research data lifecycle. Data management ideally becomes an integral part of your research practice, reduces time and financial costs and greatly enhances the quality of the data for your use too.

(From Corti et al. (2014), pp. 10-12)
Key worries about data sharing

Reason not to share data:
“My data are not of Interest or use to anyone else.”

Counter argument in favour of sharing:
They are! Researchers want to access data from all kinds of studies, methodologies and disciplines. It is very difficult to predict which data may be important for future research. Who would have thought that amateur gardeners' diaries would one day provide essential data for climate change research? Your data may also be essential for teaching purposes. Sharing is not just about archiving your data but about sharing them amongst colleagues.

(From Corti et al. (2014), pp. 10-12)
Key worries about data sharing

Reason not to share data:
“Other researchers would not understand my data at all - or may use them for the wrong purpose.”

Counter argument in favour of sharing:
Producing good documentation and providing contextual information for your research project should enable other researchers to correctly use and understand your data.

(From Corti et al. (2014), pp. 10-12)
Key worries about data sharing

**Reason not to share data:**

“I am doing highly sensitive research. I cannot possibly make my data available for others to see.”

**Counter argument in favour of sharing:**

The first thing is to ask respondents and see if you can get consent for sharing in the first instance. Anonymization procedures can help to protect identifying information. If these first two strategies are not appropriate then consider controlling access to the data or embargoing for a period of time.

(From Corti et al. (2014), pp. 10-12)
Key worries about data sharing

**Reason not to share data:**
“I have made promises to destroy my data once the project finishes.”

**Counter argument in favour of sharing:**
Why were such promises made? Always avoid making unnecessary promises to destroy data. There is usually no legal or ethical need to do so, except in the case of personal data. But that certainly would not apply to research data in general. Also consider where you have received this advice. You may need to negotiate with your ethics committee or institutional review board about this agreement.

(From Corti et al. (2014), pp. 10-12)
Concerns about sharing data - summary

Why do researchers have hesitations about sharing their research data?

Real reason (often):
• Lack of understanding of what data sharing actually entails

Real issue:
• Lack of explicit career rewards for creating and sharing of data
• But: Actions are being taken to improve career rewards for data sharing

(From Corti et al. (2014), pp. 10-12)
Benefits of data sharing

Benefits for researchers:

• increases visibility of scholarly work;
• likely to increase citations rates, for example, open access journal articles are cited more;
• enables new collaborations;
• encourages scientific enquiry and debate;
• promotes innovation and potential new data uses;
• establishes links to next generation of researchers.

(From Corti et al. (2014), pp. 11-12)
Benefits of data sharing

Benefits for research funders:
• promotes primary and secondary use of data;
• makes optimal use of publicly funded research;
• avoids duplication of data collection;
• maximizes return on investment.

(From Corti et al. (2014), pp. 11-12)
Benefits of data sharing

Benefits for the scholarly community:
• maintains professional standards of open inquiry;
• maximizes transparency and accountability;
• promotes innovation through unanticipated and new uses of data;
• enables scrutiny of research findings;
• improves quality through verification, replication and trustworthiness;
• encourages the improvement and validation of research methods;
• provides resources for teaching and learning.

(From Corti et al. (2014), pp. 11-12)
Benefits of data sharing

Benefits for research participants:
• allows maximum use of contributed information;
• minimizes data collection on difficult-to-reach or over-researched populations;
• allows participants’ experiences to be understood as widely as ethically possible.

(From Corti et al. (2014), pp. 11-12)
Benefits of data sharing

Benefits for the public:
• advances science to the benefit of society;
• adopts emerging norms such as open access publishing;
• to be, and appear to be, open and accountable;
• complies with openness laws and regulations.

(From Corti et al. (2014), pp. 11-12)
Skills

“The drive towards increased access to and sharing of research data places data management practices in a prominent position within research practices.” (Corti et al., 2014, pp. 12-13)

➤ Follow good practices at all stages of the research data lifecycle!
Requirements and expectations from publishers

Data and Materials Availability after Publication

After publication, all data and materials necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of a Science Journal. After publication, all reasonable requests for data, code, or materials must be fulfilled. Any restrictions on the availability of data, code, or materials, including fees and restrictions on original data obtained from other sources must be disclosed to the editors as must any Material Transfer Agreements (MTAs) pertaining to data or materials used or produced in this research, that place constraints on providing these data, code, or materials. Patents (whether applications or awards to the authors or home institutions) related to the work should also be declared.

Fossils or other rare specimens must be deposited in a public museum or repository and available for research.

Unreasonable restrictions on data, code, or material availability may preclude publication. Problems in obtaining access to published data are taken seriously by the Science Journals and can be reported at science_data@aaas.org.
Requirements and expectations from funders and national policies

**EU, Horizon 2020**
*As open as possible, as closed as necessary*

H2020 Programme: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

**Research Council of Norway**
*Open as standard*

Open access to research data, Research Council of Norway

**Ministry of Education and Research**

Basic principle 1: Research data must be as open as possible, as closed as necessary

Basic principle 2: Research data should be managed and curated to take full advantage of their potential.

Basic principle 3: Decisions concerning archiving and management of research data must be taken within the research community

National strategy on access to and sharing of research data
Principles and guidelines for research data management at UiT The Arctic University of Norway

• Effective as of September 1, 2017

• Apply to all projects that started 01.09.2017 or later

• Apply to all employees, including PhD students

uit.no/researchdata
Principles and guidelines for research data management at UiT The Arctic University of Norway

Purpose

• Clarify responsibility and provide guidance about how the institution and its employees shall manage, share, and archive research data in line with the institution’s administrative, financial, and ethical guidelines.

• Not supposed to be an administrative burden, but part of good research practice
Principles and guidelines for research data management at UiT The Arctic University of Norway

**Researcher responsibilities**

- Write a Data Management Plan (DMP).

- Ensure that research data are securely stored, and archived in UiT Open Research Data or in other reliable external repositories.

- Equip research data with standardized metadata and licenses for access, reuse, and dissemination.

- Make research data openly available for further use provided that there are no legal, ethical, security or commercial reasons for not doing so.
PhD regulations at UiT

(remember what the UiT policy says, and complete with):

“The student shall set up a Data Management Plan for research data in the project within six month according to the current regulations.”

“The committee may request that the candidate submit his/her data and any additional information that complements or clarifies the thesis.”

From: Regulations concerning the degree of Philosophiae Doctor (PhD) at the University of Tromsø - The Arctic University of Norway (UiT).
Research data management support at UiT

Adapted original source:
The University of California, Santa Cruz, Data Management LibGuide, Research Data Management Lifecycle, diagram, viewed May 2, 2016 at <http://guides.library.ucsc.edu/datamanagement>
Research data management support at UiT: training and guidance

1. RDM – An introduction
   How to structure and document
   How to store
   How to share
   How to search and cite
   Agreements
   Licenses
   Data management plan

2. Data collection containing personal data (in cooperation with NSD)

3. Classroom/Skype & Norwegian/English
Research data management support at UiT: information and contact

UiT Research Data Portal
uit.no/researchdata
uit.no/forskningsdata

Support email
research-data@support.uit.no
UiT Open Research Data

Launched on September 1, 2016.

A data archiving service for archiving, sharing, reusing, and citing open research data.

Available for upload to all employees and students at UiT via Feide login.

Available to all for download and reuse.

Visible to main search engines.

opendata.uit.no
Activity: How can we* help you?

*we = The library, the IT department, the research administration, your supervisor, your head of department, etc etc.

1. Turn to your neighbour(s), and write down things you (think you) would like to get help or advice on regarding research data management. It may small things, big things, concrete things, blurry things.

2. When you’re done, hand in your notes to Helene, who is the U. Library member in the group coordinating the work on establishing support services for research data management at UiT.

Why this activity is important: You get to start thinking about the things you need/want to do with your data, and we get valuable (and necessary) input from you researchers.
After the break, we split in two groups.

Those working on (qualitative) data with personal/sensitive information, follow Helene to C1006.
GEN-8001: Take control of your PhD journey

Research data management
Part 1

UiT Research Data Portal: https://uit.no/researchdata
Email: research-data@support.uit.no
References


UiT The Arctic University of Norway. (2018). Regulations concerning the degree of Philosophiae Doctor (PhD) at the University of Tromsø - The Arctic University of Norway (UiT). Retrieved from https://uit.no/om/enhet/artikkel?p_document_id=200477&p_dimension_id=88199