Introduction to Research Data Management for Social Scientists

Astrid Recker & Sebastian Netscher
CESSDA Training at the Data Archive for the Social Sciences
GESIS - Leibniz Institute for the Social Sciences
@CESSDA_Data
Introduction
CESSDA, What’s That?

- Consortium of European Social Science Data Archives
  - a pan-European Research Infrastructure
    - providing large scale, integrated and sustainable data services
    - supporting research and co-operation in areas expected to be of great importance
What is CESSDA Training?

• Provides training and consulting in Research Data Management and Digital Preservation
• Promotes strategies and procedures to ensure data quality and long-term availability
• Get in touch:
  – visit our webpage:  http://cessda.net/CESSDA-Training
  – follow us on Twitter: @CESSDA_Data
  – follow our blog:  http://www.cessdatraining.wordpress.com
  – join our workshops
We Are

Astrid Recker
(data sharing and digital preservation)

Sebastian Netscher
(research data management)
And...

... who are you ???

... what's your research about ???

... why are you here ???

???
Outline of the Workshop

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>09.15-10.30</td>
<td>Research Data Management, Data Discovery and Collection (Astrid &amp; Sebastian)</td>
</tr>
<tr>
<td>10.30-10.45</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>10.45-12.30</td>
<td>Data Handling: Documentation, Organization and Storage (Sebastian)</td>
</tr>
<tr>
<td>12.30-13.30</td>
<td>Lunch break</td>
</tr>
<tr>
<td>13.30-15.15</td>
<td>Research Ethics and Legal Compliance (Sebastian)</td>
</tr>
<tr>
<td>15.15-15.30</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>15.30-16.30</td>
<td>Data Sharing and Long-term Preservation (Astrid)</td>
</tr>
<tr>
<td>16.30-17.00</td>
<td>Wrap-Up (Astrid &amp; Sebastian)</td>
</tr>
</tbody>
</table>
Research Data Management
What is Research Data Management (RDM)?

• RDM is like “health care” for your data
  – keeps them safe from harm
  – makes them usable and discoverable
• All strategies, processes and measures to maintain
  – data quality
  – interpretability of research results
  – (re-)usability of research data
• Specifically addresses legal and ethical issues (e.g. informed consent, ownership, licensing)
RDM and the Data Lifecycle

• RDM guides the research process along the data life cycle
  • Entails strategies to
    – process and validate
    – store and protect
    – document and describe
    – preserve and share data.
Why Do We Need RDM?

• It is an integral part of the research process
  – increases the quality of research
  – supports planning and guides research
  – creates transparency and replicability of findings
  – boosts reputation

• It may be required by others
RDM may be required by...

- Your institute, e.g. by your working contract or project agreements
- Funding agencies, e.g. to ensure reusability
- Journals, claiming your data before publishing your article
- Your supervisor…

⇒ check for such conditions
Without RDM …

… your data remain an undiscovered mystery
- neither understandable for you in the future
- nor for anyone else who tries to reuse them

… your data is lost for yourself as well as for the research community, in disadvantage of research
Data Management Planning

Research
Archiving & registering
Data analysis
Data collection
Study planning
The Data Management Plan (DMP)

- A DMP is a systematic documentation of RDM
- It describes your strategies to
  - process and validate
  - store and protect
  - preserve and share your data throughout the data lifecycle
How to Write a DMP?

• A DMP uniquely relates to your research project
  – start right at the beginning of your research project
  – document what you did and why
  – frequently up-date and adapt it

• A DMP is not just a plan, it is the implementation of a (research) plan
Where to Get a DMP?

- Rely on the data life cycle
- Various online templates and tools, e.g.
- Use the template provided in this workshop
A First Look at the DMP

- Have a look at the DMP template in your folder
- Familiarize yourself with its sections and subsections
- Consider …
  ... parts you would be able to fill in off the top of your head
  ... parts you would have to look for further information
  ... obstacles or problems you encountered in the past

_individual work - メディカル time: about 15 minutes
The Structure of the DMP Template

• Consists of seven parts
  – Cover page and general information on the project
  – Six sections (with subsections):
    1. data collection \(\Rightarrow\) session 1
    2. documentation and metadata \(\Rightarrow\) session 2
    3. storage, organization and security \(\Rightarrow\) session 2
    4. ethics, legal compliance and Intellectual Property Rights \(\Rightarrow\) session 3
    5. preservation and sharing \(\Rightarrow\) session 4
    6. responsibilities and resources
Discovering and Collecting Data
Research Idea and Research Data

• Research idea
  ⇒ strategy to test this idea
  ⇒ definition of data needed
    – qualitative and/or quantitative data
    – individual and/or macro data
    etc.

• Decision on
  – whether to reuse existing data
  – or to collect new data
Discovering Research Data

• Reuse of existing data
  – a lot of research data available
  – saves time and effort

• Check for reuse of existing data
  – networks, papers and conferences
  – repositories,
    ⇒ e.g. [www.re3data.org](http://www.re3data.org)
  – CESSDA Data Catalogue
    ⇒ [http://cessda.net/CESSDA-Services/Resources/Data-Catalogue](http://cessda.net/CESSDA-Services/Resources/Data-Catalogue)
Collecting Research Data

- No existing data available for reuse
  ⇒ need to collect data
- Data collection process
  – use of standards and methodologies
  – quality assurance
  – documentation
Further Readings