Education and Training for Research Infrastructures

Laurence Horton, Alexia Katsanidou
GESIS – Leibniz Institute for the Social Sciences
Education and Training for Research Infrastructures

• Introducing WP7
• Who we are
• What we are working on
• How we fit into DASISH
• Our milestones
• Introducing our first training module and workshop
DASISH WP7

• Education and training
  – To establish joint domain for training and education
  – Inspire new research approaches using infrastructures
  – Discuss the role of infrastructures in methodologies
WP7 Structure

Education and training

Training modules
• University of Göttingen*
• Austrian Academy of Sciences
• King’s College London
• Max Planck Institute for Social Law and Social Policy
• Norwegian Social Science Data Services
• University of Bergen

Workshop program
• GESIS – Leibniz Institute for the Social Sciences*
• Austrian Academy of Sciences
• King’s College London
• Max Planck Institute for Social Law and Social Policy
• Norwegian Social Science Data Services
• University of Göttingen
Topics

• Access policies and licensing for archives and repositories.
• Persistent Identifiers
• Access and authentication
• Metadata
WP7 Milestones

- M12: First module
- M15: First workshop

Developing and releasing training modules

Presenting workshops at significant SSH conferences
## Access Policies and Licensing for Archives and Repositories training module

<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
<th>Contributor</th>
<th>Reviewer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Who is your audience?</td>
<td>UGOE</td>
<td>GESIS</td>
<td>OEAW</td>
</tr>
<tr>
<td>2. The European legal environment on IPR</td>
<td>UIB</td>
<td>GESIS</td>
<td>UGOE</td>
</tr>
<tr>
<td>3. Comparing existing license schemes</td>
<td>UIB</td>
<td>KCL</td>
<td>UGOE</td>
</tr>
<tr>
<td>4. Consent and the archive/repository</td>
<td>UGOE</td>
<td>OEAW</td>
<td>UIB</td>
</tr>
<tr>
<td>5. Secure data licenses</td>
<td>NSD</td>
<td>KCL</td>
<td>GESIS</td>
</tr>
<tr>
<td>6. Embargo policies</td>
<td>OEAW</td>
<td>GESIS</td>
<td>NSD</td>
</tr>
<tr>
<td>7. Subsequent data reuse</td>
<td>UGOE</td>
<td>NSD</td>
<td>MPiSOC</td>
</tr>
</tbody>
</table>

http://training.dasish.eu/training/1/index.html
Access Policies and Licensing for Archives and Repositories training module

With information technology being used by many researchers nowadays, the storage and accessibility of research data has become a much debated topic across all scientific disciplines. The online training module attempts to create awareness amongst both researchers as data producers and users as well as repository operators for the needs of dedicated policies and license definitions and regulating the access to and the re-use of research data, while putting a focus on issues from social sciences and humanities.

The module is organized in 7 chapters, which can be worked through chronologically or separately. Here is a brief description of the title, contents and focus of each of the chapters.

Chapter 1: Sharing and licensing research data
This chapter provides an introduction into the topic and describes how basic questions are addressed, along with some examples.

Chapter 2: Overview of European IP
Here the concept of intellectual property rights in general and copyright in particular, as well as national differences are presented.

Chapter 3: Comparing existing license schemes
In this chapter some example license schemes from social sciences and humanities are described, explained and compared.

Chapter 4: Data submission and agreements
The effect of data sharing issues on the processes of creating, submitting and preserving data are the focus of this chapter.

Chapter 5: Secure data services
This chapter covers the implications of sensitive data on data sharing, and the technical and logical measures that are in use in social sciences and humanities to address these problems.

Chapter 6: Embargo policies
Here embargo on research data as a separate aspect of restricting access are addressed, and the current state and future use briefly described.

Chapter 7: Facilitating data re-use
In this final chapter, some aspects of making data accessible and re-useable are covered, addressing the possibilities of archives/repositories to support researchers (both data producers and re-users) in these matters.

Terminology
Since diverging terminology is often an obstacle on the way towards correctly understanding the intention and meaning of text, here we attempt to clarify some terms which are used throughout the chapters, and how we would like them to be understood.

Archive / Repository / Data center / Preservation service etc.
We use these terms more or less synonymously, as the context terms depend not only on the scientific discipline and research context, but also on the overall services provided by the respective institution. What we basically mean when using these terms are all facilities or institutions that archive or store research data and provide access to it in some form.

Researcher:
Unless explicitly stated in the text, the term includes all groups producing, depositing, accessing and re-using research data in a scientific context.

Data producers/depositors:
The term subsumes all people involved in creating data and submitting them to an archive/repository.

Research community:
The term is used as a reference to sub-communities of the worldwide scientific community, each of them concerned with a specific research area. In the context of DASIS and of this table, the term in most cases refers to the communities in the Social Sciences and humanities.

What the training module looks like
Why license research data?

So if the decision for sharing research data is confirmed by so many beneficiaries, why don’t we upload all our research data to a commercial platform like Google, YouTube, Dropbox, etc? As you might know from your own experience, this is actually being done by many researchers to some extent, in order to facilitate easy exchange of data and research results amongst co-workers. But can these providers, as easy as their usage appears and is designed to be, also serve as a basis for achieving the other benefits listed above?

Consider a researcher who stores all their data and intermediate analyses in a cloud storage provider. When the researcher (or even more important: the funder of his research) thinks about how their data is being accessed and used, they basically have three options:

1. restricting access completely, granting access only on a more or less personal basis and upon request. This involves trust that the commercial provider will reliably prevent unauthorized parties from accessing data.
2. delegating the control of access and re-use to policies offered by the cloud storage provider, e.g. through implementation of a registration procedure, accepting terms of use etc.
3. allowing access to practically everybody.

However, each of these options has drawbacks, of which only some are mentioned below:

1. major administrative overhead for the researcher, without gaining reliable control on who has access to it
2. the provider might make decisions not regulated by its Terms of Use, change them or needs to close down
3. others might use his data, not reflecting possible subsequent error corrections, using them for originally unintended purposes or accidently misinterpreting them

Summing up, researchers who want to be able to both continue their scientific work and make it transparent by sharing their data, while still retaining enough reliable control on who uses their data and how, require institutions or services that provide the reliability of controlled and monitored access. Furthermore, such institutions should facilitate the choice of applicable re-use and access policies by offering understandable policies and license templates that suit the needs of the researchers, or that can easily be modified to comply with their requirements.

Table 1.1: Concerns about sharing data and possible solutions:

<table>
<thead>
<tr>
<th>Concern</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inappropriately used due to misunderstanding of research purpose or parameters</td>
<td>• Provide rich metadata at least on Abstract, Purpose, Use Conditions</td>
</tr>
<tr>
<td>Security and confidentiality of sensitive data</td>
<td>• Only make the metadata accessible</td>
</tr>
<tr>
<td>Lack of acknowledgement</td>
<td>• Specify a required data citation</td>
</tr>
<tr>
<td>Loss of data insight and advantages in comparison for funding</td>
<td>• Create second, public version with generalized metadata</td>
</tr>
</tbody>
</table>

Source: [dataset]

These proposed solutions are effective as suggestions or requirements for policies in data submission (resubmit), access and re-use. Regarding access and re-use, there are several prevalent licensing options and types. These are covered and compared in chapter 3 of this tutorial.
Access Policies and Licensing for Archives and Repositories training module

Quiz: Access Policies and Licensing

1. For each of the following groups, mark which benefit they can expect from research data being shared: a) the public
   □ people can improve their lifestyle more easily
   □ people can contact each other more easily

2. For each of the following groups, mark which benefit they can expect from research data being shared: b) the research funders
   □ can frame researchers more easily for data manipulation
   □ can direct their investments on a more substantial basis

3. For each of the following groups, mark which benefit they can expect from research data being shared: c) the research community
   □ can build and extend the work of others better and avoid reinventing the wheel
   □ can become more focused by establishing enclaves at research centers

4. For each of the following groups, mark which benefit they can expect from research data being shared: d) the data producer
   □ can gain more recognition and expect more money from research funders
   □ can generate more publications with less effort

5. There are many concerns regarding sharing research data from the viewpoint of researchers. Of the following, check one concern which can not be remedied by using a trusted repository.
   □ inappropriate use due to misunderstanding of research purpose or parameters
   □ extensive extra work due to detailed specification of metadata and access conditions
   □ compromising the security and confidentiality of sensitive data
   □ lack of acknowledgement / credit for data creation
   □ loss of data insight and advantages in competition for funding

6. When is the best time for a researcher to think about legal and ethical issues with regard to your research data?
   □ shortly before transferring the data into a repository/archive
   □ during the analysis of the data
   □ in the planning stage of his/her research, before the data is created
   □ a researcher does not need to take these issues into account
   □ none of the above

• Example of interactive tutorial
First training workshop

• Access Policies and Licensing for Archives and Repositories @ IASSIST 2013.
  – Laurence Horton (CESSDA)
  – Anje Müller Gjesdal and Gunn Inger (CLARIN)
  – Matteo Romanello (DARIAH)
  – Trond Kvarme (ESS)
  – Eric Balster (SHARE)
  – Claudia Engelhardt and Timo Gndat (Training module)
Access Policies and Licensing for Archives and Repositories workshop
Access Policies and Licensing for Archives and Repositories workshop

Access Policies and Licensing workshop attendees by infrastructure type