W2: Teaching an introductory workshop in digital preservation

40th IASSIST Conference
Toronto, June 3-6, 2014

Instructors:
Laurence Horton (The London School of Economics and Political Science)
Dr. Alexia Katsanidou (GESIS – Leibniz Institute for the Social Sciences)
Dr. Astrid Recker (GESIS – Leibniz Institute for the Social Sciences)
Teaching an introductory workshop in digital preservation

Workshop at the 40th IASSIST Conference
Toronto, June 3-6, 2014

Alexia Katsanidou | Astrid Recker
Archive and Data Management Training Center
GESIS - Leibniz Institute for the Social Sciences
@archivetraining, archive.training@gesis.org

Laurence Horton
The London School of Economics and Political Science
@laurencedata, L.Horton@lse.ac.uk

What we do

• We offer workshops and consultations for an international audience
• Research Data Management for social scientists
  – topics: data management plans, licensing, anonymization, ethics & consent, etc.
• Digital Preservation
  – topics: introduction to digital preservation, policies, licensing, trusted repositories, audit & certification

About us

• CESSDA Archive and Data Management Training Center at GESIS - Leibniz Institute for the Social Sciences
• established in 2011 as part of the GESIS Data Archive for the Social Sciences
• http://www.gesis.org/en/admtc
Round of introductions

• We would like to hear about you and your background!
• What is your motivation to participate in this course?

About the course

• Original setting
  – "First steps towards digital preservation"
  – 2 days
  – 8-10 participants with no prerequisite knowledge of digital preservation

Course content & form

• focus on the organizational dimension of digital preservation: "policies, procedures, practices, people" (Digital preservation management tutorial)
Course content & form

- 6 modules consisting of presentations and exercises

Course materials

- Presentations in .pptx format
  - License: cc-by
  - Note that some of the images use a cc-by-sa license
- Slide-by-slide notes for each module, including
  - Exercises
  - References

Accessing the course materials

- Materials are available for download from GESIS’s online learning platform ILIAS
- 2 step procedure:
  - register for ILIAS
  - request to join our course to access the materials
ILIAS registration process

- Go to [http://ilias.gesis.org/](http://ilias.gesis.org/)
- To change the language, use the drop-down menu "Choose your language" at the bottom of the screen.
- If you are using ILIAS for the first time, select "New Account Registration".
- Complete the form.
- Submit by clicking the "Register" button. Please note that you have to wait for our administration to verify your account before you can access the materials as this is not an automatized process (so if you register during the weekend, access will not be possible before Monday).
- Go to "Repository" (top menu) and select folder "Workshops"
- Request to join the course "An introductory workshop in digital preservation"

1. Log into your ILIAS account

2. Select “Repository”
3. Select “Workshops”

4. Request to join “An introductory workshop in digital preservation”

Time to play!

- Four sample exercises
- Take 15-20 minutes to work on the exercises & discuss them
  - Difficulties?
  - What are learning outcomes in your opinion?
Original workshop sample program

Day 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:30-13:00</td>
<td>Registration</td>
</tr>
<tr>
<td>13:00-13:30</td>
<td>Session 1: Welcome and introduction</td>
</tr>
<tr>
<td>13:30-15:30</td>
<td>Session 2: What is digital preservation and why do we need it?</td>
</tr>
<tr>
<td>15:30-16:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>16:00-18:00</td>
<td>Session 3: Introduction to the OAIS Reference Model</td>
</tr>
<tr>
<td>18:15</td>
<td>Reception</td>
</tr>
</tbody>
</table>

Day 2

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00-10:30</td>
<td>Session 4: Preserving information for a designated community</td>
</tr>
<tr>
<td>10:30-11:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:00-13:00</td>
<td>Session 5: Important Policy Documents</td>
</tr>
<tr>
<td>13:00-14:30</td>
<td>Lunch</td>
</tr>
<tr>
<td>14:30-16:00</td>
<td>Session 7: Licensing for preservation and re-use</td>
</tr>
<tr>
<td>16:00-16:30</td>
<td>Coffee break</td>
</tr>
<tr>
<td>16:30-17:30</td>
<td>Short presentations:</td>
</tr>
<tr>
<td></td>
<td>• The GESIS Data Archive acquisition policy (Oliver Watteler, GESIS Data Archive)</td>
</tr>
<tr>
<td></td>
<td>• The GESIS Preservation policy (Natascha Schumann, GESIS Data Archive)</td>
</tr>
</tbody>
</table>

Day 3

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00-11:00</td>
<td>Session 8: Trusted digital repositories (Astrid Recker)</td>
</tr>
<tr>
<td></td>
<td>The session will conclude with a presentation by Natascha Schumann on the Data Seal of Approval.</td>
</tr>
<tr>
<td>11:00-11:30</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:30-13:00</td>
<td>Session 9: Wrapping up</td>
</tr>
<tr>
<td></td>
<td>(incl. Q &amp; A with Reiner Mauer, GESIS Data Archive)</td>
</tr>
</tbody>
</table>
Module descriptions and learning objectives
Module 1: What is digital preservation and why do we need it?

Module content

The purpose of this module is to introduce participants to the topic of digital preservation and to highlight some of the challenges and pitfalls of digital preservation. It specifically covers the following aspects:

- BBC Domesday Book Project case study
- Definitions: digital preservation, digital objects levels, how long is long-term?
- the different dimensions of digital preservation (the “three-legged stool”)
- challenges in each of these dimensions

The module is designed to serve as an introduction to the entire workshop. Many of the issues addressed will be dealt with in more detail in subsequent modules.

Learning objectives

After this session, participants will have a first general idea of what digital preservation is and what its goals are. They will be aware of the three dimensions of digital preservation (technology, organization, resources) as well as of important challenges to the preservation of digital objects in these dimensions.

Module 2: A short introduction to the OAIS (Open Archival Information System) Reference Model

Module content

This module will introduce participants to the OAIS Reference Model and its terminology and core concepts. A particular focus is put on the functional and the information model as well as the concept of Preservation Description Information.

Learning objectives

After this session, participants will be familiar with

- the general content and structure of the model,
- the abstract nature of any reference model and implications this has for implementation,
- central elements of the OAIS reference model,
- the OAIS functional entities,
- OAIS information packages
- the concept of Preservation Description Information
Module 3: Preserving information for a Designated Community

Module content

As digital preservation is always preservation for a particular user group (which may be very narrow and homogeneous or very broad and heterogeneous), this session will further explore the OAIS concept of the Designated Community. The latter is defined as “[a]n identified group of potential Consumers who should be able to understand a particular set of information” (CCSDS, 2012, p. 1-11). This module looks at the definition of the concept “Designated Community” and explores processes of monitoring it and communicating with it. Emphasizing that digital preservation does not exist in a vacuum but in a supply-demand relationship, the module concludes by addressing the question of how to meet the future demands of the Designated Community.

Learning objectives

After completing this module, participants will know and understand the OAIS concepts of the Designated Community and its Knowledge Base as well as the need for monitoring them. They will be familiar with different forms and techniques of monitoring the Designated Community to detect changes in its needs and Knowledge Base.

Module 4: Policy documents to support digital preservation

Module content

This module gives an overview of important policy documents to support the preservation of digital resources in archives and repositories. Thus, the focus of the module is once more on the organizational dimension of digital preservation (see Module 1). After a short introduction to different kinds of policy documents (including preservation policies), the acquisition or collection policy and its application are discussed in more detail.

Learning objectives

After completing this module participants

- understand the dual role of policies as supportive framework for digital preservation and an important tool for external stakeholder communication
- are familiar with different types of policy documents: their scope, function/purpose, and relevance to digital preservation
- understand that policies are not isolated documents but relate to each other
- are aware of important aspects to consider when designing an acquisition policy
- are familiar with selected criteria supporting the appraisal of resources.
Module 5: Licensing for preservation and reuse

Module content

This module introduces the concept of Intellectual Property Rights, copyright, and licensing. It elaborates on how they relate to data, data archiving, and reuse. It discusses how licenses are used to address issues of privacy and data protection, the problems of open data and public domain when it comes to social science data. The module continues to outline the main components of reuse licenses through a comparison of CESSDA data archive licenses. The module touches on the problems of enforcing licenses before moving to look at template options such as Creative Commons and Open Data Commons. The module concludes by introducing the problems raised by template licenses inviting the question as to whether we can produce a clear, concise standard that allows for variation in Intellectual Property Law without unnecessarily restricting access.

Learning objectives

The objective of this session is to provide an introduction to the basic concepts and importance of Intellectual Property Rights. Participants should understand:

- How Intellectual Property Rights relate to data acquisition and dissemination.
- Why archives cannot accept data that lacks clear ownership or does not have permission for archiving.
- How personal data effects data reuse licenses.
- Why template licenses may not be appropriate for research data.
- Have an understanding of the basic components of a data reuse license.
- Be aware of community discussions regarding the problems of enforceability, and attribution stacking.

Module 6: Trusted digital repositories

Module content

This module focuses on the concepts of trust and trusted archives or repositories. First, we will discuss a general definition of trust and introduce the concept of external stakeholder trust. We discuss what trustworthiness is and which factors contribute to our perception of someone as being trustworthy. We then discuss how archives and repositories can demonstrate their trustworthiness to their stakeholders. To conclude, selected tools for self-assessment, audit and certification purposes will be introduced.

Learning objectives

After completing this module, participants understand the concept of (interpersonal) trust. They will be familiar with different building block of trustworthiness and will have understood which factors influence trust. They will be aware of measures that archives and repositories can take to actively build trust in the services they offer.
Sample presentation
Preserving information for a Designated Community

Alexia Katsanidou, Astrid Recker
Archive and Data Management Training Center
GESIS - Leibniz Institute for the Social Sciences
@archivetraining

This work is licensed under a Creative Commons Attribution 3.0 Unported License.

Supply and demand

Designated Community

An identified group of potential Consumers who should be able to understand a particular set of information (CCSDS 2012: p. 1-11)
The Designated Community...

- can be composed of multiple user communities
- is defined by the Archive
- and its definition can change over time
- has a knowledge base

Knowledge Base

“A set of information, incorporated by a person or system, that allows that person or system to understand received information.”

(CCSDS 2012, p. 1-12)

Knowledge Base

- Familiarity with file formats, software packages
- Ability to access media
Knowledge Base

• Ability to understand the information object
  – dependent on ability to understand documentation, metadata, context information

Knowledge Base

• Preferred communication channels
  – How do users get in touch with the archive?

Describing the Designated Community

• Who are the users of the preserved digital objects, now and in the future?
• What are their expectations of the archive?
• What is their knowledge base (technical and content-related)?
• How big is the group and how often will they use the archive?
Describing the Designated Community

• How and for which purposes do they expect to use the preserved objects?
  – perceive the object (e.g. looking at a picture, listening to audio)
  – analyse the object or extract information from it (e.g. searching in a database or a text document)
  – re-use and adapt (parts of the) object (e.g. re-using illustrations from a report)
  – running the object (e.g. playing a computer game)

(see nestor 2012, pp. 20-21)

Community watch

• User communities and their knowledge bases can change
• Keep an eye out for:
  • emerging community standards (media, software, file formats),
  • new methods, or
  • changed meaning of concepts that could impact the usability of digital objects

...because the knowledge how to use resources can be lost
Staying in touch

- User surveys, focus groups
- Being part of the community
  - conferences
  - active research
- Teaching and workshops
- Regular communication with data producers and archive users (e.g. help desk)

Staying in touch - GESIS

- Frequent user surveys on service offers and their quality for all departments
- Employees are encouraged to carry out their own research and publish regularly
- Employees are encouraged to teach in university
- Scientific and User Advisory Boards
- GESIS workshops and schools
- Data Archive “help desk” and Acquisition communicate with data producers and consumers

Back to supply and demand
Meeting future demand

Digital preservation entails changes of the digital objects

Object features may be altered or lost completely due to technical issues or cost considerations

• Decision which features to preserve should be based on the expected needs of the designated community (use scenarios)

Meeting future demand

1. Determine important characteristics of the digital objects to be preserved
2. Create scenarios for use
3. Decide which characteristics are essential for a given user community and its needs
References and further reading

Sample notes
Module 3: Preserving Information for a Designated Community

<table>
<thead>
<tr>
<th>Workshop title</th>
<th>First steps towards digital preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module number</td>
<td>3</td>
</tr>
<tr>
<td>Module title</td>
<td>Preserving information for a Designated Community</td>
</tr>
<tr>
<td>Duration</td>
<td>90 minutes in the original workshop.</td>
</tr>
<tr>
<td>Author</td>
<td>Archive and Data Management Training Center</td>
</tr>
<tr>
<td>Publication date</td>
<td>2014-05-09</td>
</tr>
<tr>
<td>Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Contact information</td>
<td>Archive and Data Management Training Center</td>
</tr>
<tr>
<td></td>
<td>GESIS – Leibniz Institute for the Social Sciences</td>
</tr>
<tr>
<td></td>
<td>Unter Sachsenhausen 6-8</td>
</tr>
<tr>
<td></td>
<td>50667 Köln</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
</tr>
<tr>
<td></td>
<td><a href="mailto:archive.training@gesis.org">archive.training@gesis.org</a></td>
</tr>
</tbody>
</table>

### Slide Notes

#### Slide 1: Title slide

Archives and repositories do not operate in a vacuum. That is, digital preservation is not an end in itself, but we preserve digital objects for use (be it use by the public or use by a small, restricted group of users). Therefore, the services (and digital objects) that archives and repositories supply must be orientated towards the demand of the users. If this demand is not met, the collections will not be used, and the efforts made to preserve the digital assets were ultimately futile.

#### Slide 2: Supply and demand

The OAIS reference model defines the Designated Community as "[a]n identified group of potential Consumers who should be able to understand a particular set of information" (CCSDS, 2012, p. 1-11). This is one of the central concepts of the OAIS reference model (see Module 2 of this workshop). The Designated Community and its needs and preferences should be the reference point for all preservation activities, decisions, and services. Thus, the ultimate goal is to maintain the understandability and usability of the preserved resources for the Designated Community. This implies that the archive/repository has to make sure that the metadata (e.g. the representation information, see Module 2 of this workshop) make sense to the Designated Community and contain all information needed by this community to use the resource.

According to the definition, the Designated Community consists not only of actual, current users, but of "potential users" – that is, it may include groups that may have an interest in using the preserved information (i.e. the digital assets) in the future.

Imagine, for example, a company archive which preserves the digital records of a particular company. While the primary motivation for doing so may be to maintain this information because it is valuable for the operations of the company, it is also thinkable that the preserved information is interesting for future historians. These historians are therefore a future potential user group and could be (if the company archive decides to include them) part of the Designated Community of that archive.

#### Slide 3: Designated Community

The OAIS reference model defines the Designated Community as "[a]n identified group of potential Consumers who should be able to understand a particular set of information" (CCSDS, 2012, p. 1-11). This is one of the central concepts of the OAIS reference model (see Module 2 of this workshop). The Designated Community and its needs and preferences should be the reference point for all preservation activities, decisions, and services. Thus, the ultimate goal is to maintain the understandability and usability of the preserved resources for the Designated Community. This implies that the archive/repository has to make sure that the metadata (e.g. the representation information, see Module 2 of this workshop) make sense to the Designated Community and contain all information needed by this community to use the resource.

According to the definition, the Designated Community consists not only of actual, current users, but of "potential users" – that is, it may include groups that may have an interest in using the preserved information (i.e. the digital assets) in the future.

Imagine, for example, a company archive which preserves the digital records of a particular company. While the primary motivation for doing so may be to maintain this information because it is valuable for the operations of the company, it is also thinkable that the preserved information is interesting for future historians. These historians are therefore a future potential user group and could be (if the company archive decides to include them) part of the Designated Community of that archive.

#### Slide 4: The Designated Community...

According to the OAIS reference model, an archive has only one Designated Community – however, it is possible that this consists of different subgroups with differing needs. For example, an archive or repository could have a Designated Community composed of experts (e.g. social science
The Designated Community is defined by the archive/repository. However, this definition can be influenced by regulations or legal requirements. For example, an archive may “inherit” the Designated Community from its parenting organization.

As the definition of the Designated Community is not only a description of the status quo but a projection into the future, it is possible that the definition will change – e.g. when the mission of the archive/repository changes, or when new user communities emerge which were not originally considered by the archive.

An important aspect of the definition of the Designated Community is the description of its Knowledge Base (see following slides).

| Knowledge Base (1) | The OAIS reference model defines “Knowledge Base” as follows: “A set of information, incorporated by a person or system, that allows that person or system to understand received information” (CCSDS, 2012, p. 1-12). The question that the archive/repository has to answer is, what does a user need to know in order to access and understand this resource, and how much of this required knowledge can the archive or repository assume they have. The gap between required knowledge and existing knowledge (i.e. the Knowledge Base) needs to be bridged with the help of additional information and guidelines.

For example, imagine an archive preserving digitized editions of texts written in Middle High German. The current Designated Community for this archive, consisting of historians and researchers in the history of German language and literature, is able to read and understand this language. Accordingly, the archive does not need to provide any additional information to help understanding the resources. However, 25 years later the archive finds that hardly any of the consumers are still capable of reading Middle High German because it is no longer taught in universities. It seems therefore that the Knowledge Base of the Designated Community has changed – if before a knowledge of Middle High German could be taken for granted, this is no longer the case. In this case the archive has to consider providing additional information (representation information in the terminology of OAIS) making it possible for the Designated Community to understand the resources preserved by the archive. For example, the archive could consider making dictionaries and grammars for Middle High German available to enable the researchers to use the resources.

| Knowledge Base (2) | In considering the Knowledge Base of its Designated Community, the archive/repository on the one hand needs to consider whether its users will be capable of accessing the digital resources offered. If we offer text documents for download in .odt format, can we be certain that the Designated Community will know which software to use to open these files, and will the software be readily available to it? If an archive decides to distribute its digital resources on DVDs, they need to make sure that their Designated Community still has the players to run these.

**Discussion point**

- How can an archive/repository resolve the following conflict: the archive has determined that the best possible file format for the preservation of its assets is an open format (such as .odt), but the de facto standard of its Designated Community is a proprietary format (e.g. .doc). What should the archive do?
### Knowledge Base (3)

At the same time, the Designated Community has to be capable of understanding the information preserved by the archive – i.e. the users have to be able to “make sense” of it.

This ability is dependent on the existence of sufficient metadata and context information (as well as the understandability of this metadata). For example, the archive/repository should pay attention that the metadata/documentation for a resource does not use technical terms not (or no longer) understood by the Designated Community.

### Knowledge Base (4)

Finally, in considering the Knowledge Base of its Designated Community, an archive/repository should also think about its users’ preferred communication channels (including their preferred ways of retrieving the preserved resources).

### Describing the Designated Community (1)

A description of the Designated Community should consider the aspects mentioned on this and the next slide.

### Describing the Designated Community (2)

An important aspect in defining the Designated Community is considering how and for which purposes they expect to use the digital resources preserved by an archive or repository. Often, preserving a digital object with its full functionality and original “look and feel” is not possible due to cost restrictions and technological issues (this issue is discussed in more detail below). Knowing which functionality the Designated Community needs and expects can help to decide in which form the digital objects need to be preserved.

For example, the Designated Community of the GESIS Data Archive wants to analyze and re-use the preserved data with the help of common statistical packages. Thus, it would not make any sense to preserve this data only in PDF/A format – even though it may be a viable preservation format – as this would make it impossible to use the data in the way required by the Designated Community.

### Meeting future demand (1)

As just mentioned, in digital preservation what goes in is not necessarily what goes out. For example, format migrations may be necessary which alter the available functionality or the look and feel of a resource.

When it comes to considering such changes to a preserved resource, it is important to be aware of the needs of the Designated Community.

### Meeting future demand (2)

To determine these needs, it can be helpful to create different use scenarios. These can make it easier to decide which route to take in the preservation process (see following slide).

### Meeting future demand (3)

The procedure described on this slide is reminiscent of procedures for the identification of significant properties of digital objects (or object types), for example, as developed by the InSPECT project (see Grace, Knight, et Montague, 2009). However, because the concept of significant properties is rather complex and would by itself be enough to fill an entire module, the decision was made to omit it at this point.

Instead, this slide again emphasizes the importance of matching supply and demand: if the archive (supply side) has to change the digital objects in the preservation process, it must know what the demands of its users are. Otherwise the archive runs the danger of preserving digital information in a form which is useless to its Designated Community.

### Community watch (1)

Because the Designated Community is not static, it is important for archives or repositories to monitor their user communities to detect changes that
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 15 | Community watch (2) | Only if we monitor our Designated Community can we be certain not to miss important developments and changes in a community’s Knowledge Base (e.g. the loss of certain knowledge and capabilities).
|   |   | So can we be sure today, for example, that our Designated Communities still know how to use a Hollerith card or microfiche? Or how to run programs using DOS commands? |
| 16 | Staying in touch | By communicating regularly with its Designated Community, an archive or repository can make sure that important developments or changes in the community and its Knowledge Base do not happen unnoticed. However, it is important to attempt to collect information from these communication processes as systematically as possible – for example, by documenting questions coming in through the help desk or hotline. |
| 17 | Staying in touch – GESIS | These are examples of how GESIS and the GESIS Data Archive interacts with its Designated Community. |

References


Sample exercise 1: “Designated Communities”

Choose a digital object from your work context and describe

- the Designated Community for this object now, in 5, 20 and 100 years.
- the expected uses: How does the Designated Community use the object now and in 5, 20, and 100 years.

Alternatively, complete the exercise for the following digital object: The computer game "SimCity", released for Amiga and other platforms in 1989.

Image: Screenshot of Micropolis, a liberated game based on the port of the original SimCity to Unix by Tomhannen (http://en.wikipedia.org/wiki/File:Micropolis_-_big_city.png) (GPLv3)
Sample exercise 2: “Policy documents to support digital preservation”

We can’t take everything. We have to be selective. This is basis for selection is an acquisitions policy. Think about your organization in answering these questions:

<table>
<thead>
<tr>
<th>Does your organization have a defined mission statement?</th>
<th>😊</th>
<th>😐</th>
<th>🙁</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>The mission statement determines the scope of the collection (usually in very general terms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Are you aware of all legal requirements determining which material has to be preserved or cannot be accepted by your archive or repository? | 😐 | 😐 | 😐 | Funding mandates, data protection laws, Freedom of Information requirements. |

| Has your organization defined its designated community, now and in the future? | 😐 | 😐 | 😐 | It’s important to know who is offering objects, who wants to use them, how, and in what ways? This determines what you accept and reject. |

| Did your organization assess its core areas of expertise? | 😐 | 😐 | 😐 | What can you cope with? If there is no expertise in handling audiovisual formats, the Archive cannot accept these materials |

| Are you familiar with the collection profiles of other archives in your subject field? | 😐 | 😐 | 😐 | If a certain type of content isn’t preserved anywhere else, it might be a reason to consider it for your archive or repository. Alternatively, if certain materials are already covered by other archives, there may be no need for you to archive them |

| Have you defined concrete criteria to guide the act of appraisal? | 😐 | 😐 | 😐 | Do you have a written policy and guidelines, or a checklist to match against submissions? |

| Do these criteria consider | 😐 | 😐 | 😐 | |
| --- | --- | --- | --- |
| • content | • format | • quality | • uniqueness/historical value |
| • relevance | • economic value |

Share with the rest of the group what your organization does well, what needs to be improved, and what areas you haven’t addressed.
Sample exercise 3: Intellectual property issues
This exercise is intended to draw out issues related to intellectual property – which often are the main obstacle to preventing archiving of good quality, relevant data. Two scenarios are presented, both drawn from real cases. The scenarios are intended to stimulate discussion within the group. The possible outcomes can be used as prompts for discussion: Why yes, why not yes etc. so all aspects of the scenario are explored.

Scenario one
A project has collected data from a Thompson-Reuters database of patent registrations. Can the project archive the data in an archive? If not, the project wonders if the archive is able to create a catalogue record with a note indicating the project has collated this data and would be willing to share it with others if they were asked. Is this an acceptable compromise?

Discuss the possible outcomes:

- Yes
- No
- Maybe

Discussion points:

YES - if the copyright holder (TR) grants permission.

NO - an archive could not advertise a project’s data with the knowledge that project could breach copyright by sharing that data. The archive cannot even refer potential users to a third party that may be violating intellectual property rights.

Scenario two
A project has linguistic corpus data based on an extract of text from one of the Harry Potter books. The data consists of the text extract and audio recordings of participants reading the text.

Can you

A) Go ahead and archive the data?
B) Go ahead and archive it only for non-commercial re-use only
C) Ask the depositor to apply for permission from the publisher of Harry Potter

Discussion points

A) No, there are clear intellectual property issues here. Furthermore, the concept of “fair dealing” may not be relevant under your national law.

B) No, commercial or non-commercial usage makes no difference to the intellectual property issue when it comes to archiving and reuse.

C) Yes.
Sample exercise 4: Trusted digital repositories

You work for the repository “Noah's Ark,” which collects data and publications from the discipline of evolutionary biology. Founded recently, the repository has already made good progress in building up a user community. A user survey, designed to measure the trust users place in the repository, has made it apparent, however, that the repository scores low on the following points:

5. Data and publications are stored more safely in Noah's Ark than on my computer.
   - 47% of users answered “I don't know”.

11. Noah's Ark understands my needs and concerns as a researcher in evolutionary biology.
   - 30% of users disagreed.

16. It is easy to understand what Noah’s Ark does to keep my data and publications safe.
   - Only 12% of users agreed.

21. Noah’s Ark was recommended to me by someone whom I value highly because of his/her professional expertise in the field of evolutionary biology.
   - Only 5% of users answered “Yes”.

You are part of a working group developing a strategy of how to address the issues identified. Which concrete measures can the repository take to further increase the trust its stakeholders place in it?

Work on one of the issues in small groups.