Data Management Plan
for projects at the University of Vienna

Recommended Repository: Phaidra

Version: 2.0 (Including Feedback from the first pilot phase from May/June 2015)
Edited by: Susanne Blumesberger, Paolo Budroni, Ralf Pausz, Barbara Sánchez Solís
Editing of online version and English translation: José Luis Preza Díaz
Adapted for the University of Vienna
Date: 06.10.2015

e-infrastructures
austria

CONTENTS

1. Administrative Data

The purpose of Administrative Data section is to provide basic information on the research project allowing identifying the project, people responsible for it and means of contacting them. This section is NOT repeating any information about the project itself, e.g. project description that can be found in other documents like project proposal, description of work, etc.

Title of the project:

Author of this document (please include Name, Tel.-number, and E-Mail address):

Contact Person for the Data Management Plan: (please include Name, Tel.-number, and E-Mail address):

Version of the Document and Date:

2. Data Collection

The purpose of Data Collection section is to identify datasets that are used and created during the project. This description is not limited to the data that needs to be later archived and preserved – this is specified later in the Selection and Preservation section. By identifying data used in the course of the project the researchers will better estimate they requirements for software and hardware infrastructure needed to run the project.

a) What type and amount of data will you generate?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- In what formats do you produce your data?
  - *Examples: Text documents (i.e. DOC, ODF, PDF, TXT etc.), Structured Text (i.e. HTML, JSON, TEX, XML etc.), Tables (i.e. CSV, ODS, XLS, SAS, Stata, SPSS etc.), Databases (i.e. MS Access, MySql, Oracle etc.), Images (i.e. JPEG, SVG, PNG, GIF, TIFF etc.), Audio (i.e. MP3, WAV, AIFF, OGG etc.), Video/Film (i.e. MPEG, AVI, WMV, MP4 etc.), Source code (i.e. CSS, JavaScript, Java etc.), Configuration data (i.e. INI, CONF etc.), Software applications*

- What is your approximate (current or expected) amount of data (in order of gigabytes, megabytes, terabytes or petabytes)?

• How big are the individual largest files?

b) How will the data be collected or created?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.
• How is your data collected or produced (description of processes)? • Do you use special software? • Do you use special hardware? • Reuse of the data: Is the choice of technology, the formats and the metadata (descriptive, contextual, provenance, technical or other metadata) suitable to ensure subsequent use?

3. Documentation

The purpose of Documentation section is to describe practices that will be taken during the research process that facilitate correct interpretation and provenance collection of data created during the research, including the processes taken place, the contextual frame and their contextual interpretation, the data structure, the relationships to other entities, and which changes and procedures are necessary to prepare and analyze the data. Good documentation should address the questions Why, Who, What, Where, When and How.

a) What documentation will accompany the data?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.
• In which form do you document your data (e.g. laboratory notebooks, field notes, audio files, and so on)? • Do you document your data in a specific format?

4. Metadata

The more extensive your data are described, when you deposit them in a long time archiving system, the easier will be to find them and reuse them. Standardized Vocabularies and Classifications (like ÖFOS, Eurovoc, ACM or Getty) will help to make the data more visible and reusable..

a) What metadata will accompany the data?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- How is your metadata structured?
- Do you have the necessary information to archive data? (Example Phaidra: Title, Description, Person and Role, License)
- Do you use metadata standards? If so, which?
- Do you use metadata made by other Researchers? Do you have the rights to use it?
- Who is responsible for the metadata?

5. Ethics and Legal Compliance

The purpose of Ethics and Legal Compliance section is to identify any issues affecting the way the sensible data can be processed, stored and published.

a) How will you manage any ethical issues?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- Is ethically questionable material included?
- Do you include personal data?

- If so, is anonymization provided? How and where will the anonymization be documented?
- Are there any limitations regarding image size or resolution?
- Should access be limited to a particular target group?
- Do you have written permission to publish the data? (i.e. audiovisual materials from the people depicted, or the soundtracks)

b) How will you manage copyright and Intellectual Property Rights (IPR) issues?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- Is the legal situation concerning copyright, exploitation, and individual rights clarified?
- Storing data in a Repository must comply with the Terms and Conditions of the service.
  - *This applies to "Phaidra", the institutional Repository of the University of Vienna (in particular Section 7 – Exploitation Rights): https://phaidraservice.univie.ac.at/en/legal-issues-in-phaidra/terms-of-use/*
  - *Quote: „The user insures that they own the exploitation rights in particular the making available right and the distribution right of the media items uploaded by them, and is entitled to grant the following rights of use to the service provider, or that they are entitled under a statutory provision to upload the media object in question."*

- Has the project responsible the necessary permissions to store the data in a Repository?
- May your digital objects be displayed on the internet? May the metadata be displayed on the internet?
- Are there any embargo periods?
- May the project managers store the data in a repository?
- Terms of Licenses: What licenses are provided (e.g. Creative Commons license, General Public license, and so on)?
  - *More about CC-Licenses: http://creativecommons.org/about*

6. Storage and Backup

The purpose of Storage and Backup section is to describe how the data will be secured during the course of the project. It focuses on actions ensuring that no data is lost and that only authorized users have access to it. Please note: section "Selection and Preservation" describes how the data is secured

in the long-term, especially after the end of the project. This section focuses on how data is managed during the project.

a) How will the data be stored and backed up during the research?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- Do you have sufficient storage at your disposal?
- Will you need to include charges for additional services?
- (technical advice, implementation of a website for the project, implementation of a CMS ,etc.)
- How will the data be backed up?
- Who will be responsible for backup and recovery?
- How will the data be recovered in the event of an incident? Do you have emergency plans?

b) How will you manage access and security?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- How sensitive is your data?
- Is there a risk that the data could be illegally accessed and manipulated?
- How will you control access to keep the data secure?
- How will you ensure that collaborators can access your data securely?
- If creating or collecting data in the field, how will you ensure its safe transfer into your main secured systems?
- Who will have access to the data?
- Is password protection provided?
- Who is responsible for the security and access to the data? If possible, please include the contact information
- Will all the data be archived long term?
- Once the project has ended, what will happen to the data that will not be archived long term?

7. Selection and Preservation

The purpose of Selection and Preservation section is to provide information on data that needs to be secured in a long term. These will be likely a subset of data specified in section Data Collection. Also in this section the researchers should estimate how these actions can be funded and how much it will cost. They should receive estimations from the repositories in which they decide to deposit their data.

a) Which data should be retained, shared and/or preserved?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- Describe what data should be stored long term.
- In what formats are these data available?
- How long should the data be stored in the repository?
- Do you need a persistent identifier? If so, you need a specific one (e.g., DOI, Handle, URN, and so on)?
  - *Definition of Persistent Identifier:*
    Most publications have an ISBN number, which is a worldwide standard for publications. With the ISBN you can be sure that when you order a publication, you will receive the right publication. The Persistent Identifier for electronic publications (published online on the internet) functions similarly. The PI allows different digital objects (i.e. PDFs, Images, Audio, Animations, etc.) to be identified and found.
    The common internet addresses (Uniform Resource Locators-URL) are not adequate as „ISBN for digital objects" because they change overtime. Stable, globally unique identifiers are for digital repositories inevitable. A Permanent Identifier is important for reliable referencing and for safe linking of metadata.

  - *The Repository of the University of Vienna – Phaidra – supports the long term archiving and the provision of a persistent identifier. Please see:*
    https://phaidraservice.univie.ac.at/en/phaidra/policy/

    "Once in Phaidra, always in Phaidra": All objects stored in Phaidra receive a permanent ID (Persistent Identifier - PI). The PI behaves like the signature of a publication in a library. The PI is always available worldwide and can be quoted or shared with others. The link of the PI provides the most important data of the published object, which can be viewed in the browser and downloaded as needed.

- Are there any plans to delete the data after a certain time?
  - *Please note that after storing the data in Phaidra, deletion is not possible. For certain data sets that require to be deleted, please use a temporary archiving system.*

b) What is the long-term preservation plan for the dataset?

put your answer here…

Guidance to the question
Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- In which repository or archive will the data be held?
- Please enter the name of the repository and the operator (including address)
- What costs if any will your selected data repository or archive charge?
- Please indicate the cost per year
- Do costs arise for the preparation of the data for archiving (e.g. legal questions for clarification of the rights or technical solutions for conversion)?
- What are the yearly or total costs of the project? Please take into account the costs and responsibilities after the end of the project
- Do costs arise for the preparation of the data for archiving? (i.e. clarification of legal issues or technical solutions for data conversion)
- Do costs arise for re-use (e.g. visualization) of data?

For members of the University of Vienna:
[Phaidra](#) has been the comprehensive, all-university Digital Asset Management System for permanent archiving functions of the University of Vienna since 2008. It enables instructors, researchers, administrative and organizational units as well as individual users to store their publications for research and teaching, as well as to document and archive them for longer periods. It supports electronic publishing, archiving of images, displaying of collections, creation of eBooks, saving and streaming of videos and much more. All objects contain a permanent digital signature, and the objects can be described in multiple languages.
More on the subject: https://phaidraservice.univie.ac.at/en/phaidra/
Contact information:
  - *Phaidra – [Local phaidra@univie.ac.at](mailto:phaidra@univie.ac.at)*
  - *Technical questions: support.phaidra@univie.ac.at*

8. Data Sharing

The purpose of Data Sharing section is to describe which data, how and in what form will be shared with other stakeholders or systems (i.e. Europeana). The issues described in Ethics and Legal Compliance section have impact on decisions described in this section.

a) How will you share the data?

put your answer here…

Guidance to the question

Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- How should the data be found online?
- Is it necessary to grant restricted/differentiated access rights?
- Do you want or have to publish your data with Open Access?
- Which operating licenses (e.g. Creative Commons License, General Public License, and so on) are planned?
- Must embargo periods be taken into account?
- Should others be able to cite your data?

b) How will the data be used after completion of the project?

put your answer here…

Guidance to the question

Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- How should the data after project completion be re-used (e.g. visualized)?
  - *The re-use of data includes: re-use from particular target groups or machines, the provision of data to other repositories, the linking of data and the visualization of data in different contexts.*

- Is there an agreement between the project partners (e.g. in relation to target group-specific representations)?
- Are you planning follow up projects?

9. Responsibilities and Resources

The purpose of Responsibilities and Resources section is to identify people responsible for implementing this data management plan. Furthermore, it summarizes additional resources required to deliver this plan, e.g. resources needed to ingest the data into a selected repository (personnel, infrastructure, money, time).

a) Who will be responsible for data management?

put your answer here…

Guidance to the question

Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- Who is responsible for implementing the DMP?
- Who is responsible for ensuring the DMP is reviewed and revised?
- Are there any conditions from the funding agencies to be taken into account?

b) What resources will you require to implement your data management plan?

put your answer here…

Guidance to the question

Below you can find a list of auxiliary questions that can help you in answering the question above. You do not have to consider all of these questions. It all depends on the kind of the project you are running.

- Do you need additional resources (software, legal consulting, technical consulting, and so on) to manage your data and to prepare for long-term archiving?
- If yes, please specify your requirements. (if possible, please include estimates of required personnel).