Checklist and guidance for a Data Management Plan, v1.0

Please cite as: DMPTuuli-project. (2016). Checklist and guidance for a Data Management Plan, v1.0. Available online: https://wiki.helsinki.fi/x/DZeaCw

<table>
<thead>
<tr>
<th>Question</th>
<th>Example of Answer</th>
<th>Guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Data Collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1 What kinds of data are collected or reused?</td>
<td>Data will be collected using a postal and electronic survey. Survey questions will cover [main subject areas of the research]. The invitation to the survey will be posted to a representative sample of 2,000 Finnish- and Swedish-speaking people aged 39-79 years living in rural areas in Finland. It is expected to have around a 20-40% response rate. Red Cap research data management software will be used for organizing responses.</td>
<td>Give a brief description of your research data including the data you are collecting and any existing data that you will reuse. Outline how the data will be collected. Outline the types of data e.g. survey, interview, observation, machine or instrument collected, models, physical samples that are expected to be used. <strong>Tips for best practices</strong>&lt;br&gt;· You don’t need to explain your methods in a DMP again, you can refer your research plan.&lt;br&gt;· Use standard data collection methods or validated protocols, if these exist for your research field.&lt;br&gt;· By reusing data, produced by you or others, you will avoid duplicating work already done.</td>
</tr>
<tr>
<td>1.2 What file formats will the data be in?</td>
<td>Digital video data files generated will be in MPEG-4 (.mp4) format. MPEG-4 is an International Standards Organization (ISO) specification and this format is readable by most media players.</td>
<td>File format is a primary factor in accessing and reusing your data in the future. List the file formats that data will be in e.g. .csv, .txt, .mp3. <strong>Tips for best practices</strong>&lt;br&gt;· Consider your research field’s and the collaboration partner’s usual methods for presenting metadata.&lt;br&gt;· When listing the data formats you will be using, make sure to include software necessary to view the data.&lt;br&gt;· Use software and formats based on open standards to enable data reuse, interoperability and sharing.</td>
</tr>
</tbody>
</table>

2. Documentation and Quality | | |
| 2.1 How will the data be documented? | The X and Y data collected from this project will be documented using A and B metadata standards. | Data documentation explains terms, variable names, codes or abbreviations used. Consider what information is needed to find, use and |
interpret the data in the future and describe the types of documentation that will accompany the data.

Metadata provides standardized structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection.

**Tips for best practices**

- Consider how the data will be organized during the project mentioning e.g. file naming conventions, version control and folder structure.
- In the beginning of the project you might not know what metadata standards you will be using, but you need to ensure that “all variables will be described and suitable metadata standards will be used, if available”.

| 2.2 How will the consistency and quality of data be controlled and documented? | For quantitative data files, the researchers in the project ensure that missing data codes are defined, that actual data values fall within the range of expected values. Processed data files are reviewed by a supervisory staff member before release. | Data quality control ensures that no data will be lost or accidentally changed during the research process. Quality control of data is an integral part of all research and takes place during data collection, data entry or digitization, and data checking. **Tips for best practices**

- Explain how the data collection methods will affect to the quality of data. Remember that documenting in detail how data are collected provides evidence of data quality.
- Quality control measures can include e.g. using standardized methods and protocols for capturing observations, alongside recording forms with clear instructions, taking multiple measurements, observations or samples and calibration of instruments. |

3. Storage and Backup
| 3.1 How will the data be stored and backed up? | Stored research data is backed up on a daily basis to a separate password protected secure server maintained by X organization. During data analysis, the data will be accessible only by certified members of the project team. The research project will remove any direct identifiers in the data before deposit with [repository]. | Attention should focus on the lifespan and subsequent use of the materials throughout the research process. Researchers should ensure the storing of their data both during and after the research process. Consider who will be responsible for backup and recovery. If there are several researchers involved, create a plan with your collaborators and ensure safe transfer between participants. **Tips for best practices**  
- The use of robust, managed storage provided by your organization’s IT support is preferable.  
- Estimate the volume of the needed storage space.  
- Do you have sufficient storage or will you need to include charges for additional services (if the amount of data is exceptionally big)? Remember to tell that you will specify your data management costs in the budget.  
- Note that using secure network drives is a more professional alternative than using hard drives. |
| 3.2 How will you control access to keep the data secure? | During data analysis, the data will be accessible only by certified members of the project team. The research project will remove any direct identifiers in the data before deposit with [repository]. | During the research project you want to keep your research data safe and secure. You will want to determine who has access to your data and what they are authorized to do with it. Providing unauthorized people with access to the data may be unlawful. Access controls should always be proportionate to the kind of data and level of confidentiality involved. |
| 4. Ethics and Legal Compliance | Information collected can be released without privacy restrictions because it does not constitute private information about identified human subjects; informed consent for full public release of the data will be obtained; the data will be anonymized using an | Researchers obtaining data from humans are requested to maintain high ethical standards and comply with relevant legislation. If you will manage personal or sensitive information, how will you ensure privacy protection and data anonymization? Informed consent is an ethical requirement for most research. It must be considered and implemented throughout the research lifecycle, from planning to publication |
IRB-approved protocol prior to the conduct of analysis. and should include making provisions for sharing and future use of data.

**Tips for best practices**

- If your research is to be reviewed by an ethical committee, outline how you will comply with the protocol (i.e. remove personal or sensitive information from your data to ensure privacy protection before data sharing or use restricted access procedures).

See e.g. [Finnish Advisory Board on Research Integrity](https://www.ffi.fi/index_en.html) for more information about the responsible conduct of research.

<table>
<thead>
<tr>
<th>4.2 How will copyright and Intellectual Property Right (IPR) issues be managed?</th>
<th>The research project will not use any data which is covered by the Copyright, Designs and Patents or any other similar legislation. Every research partner will sign a contract agreeing that data arising from the research project will be made openly available where possible. The intellectual property of the data generated will remain with the researchers, but will also be handed to X organization.</th>
</tr>
</thead>
</table>
|  | State who will own the data and can give permission to reuse the data. If you use third-party data, consider copyright issues and permissions to reuse the data. **Tips for best practices**
- Consider the relevant funder, institutional or departmental policy on copyright or IPR.
- It is recommended that research data created by a research project is available for reuse e.g. under [Creative Commons](https://creativecommons.org) or other relevant-license. The recommended CC-license for open science is [CC-By license](https://creativecommons.org/licenses/by/). |

<table>
<thead>
<tr>
<th>5. Data Sharing and Long-Term Preservation</th>
<th>Data will be available and cited in publication. Researchers will be able to contact the PI for access to data. Data will be maintained in an open XML format to enable open reuse of the data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 How, when, where and to whom the data will be made available?</td>
<td>Describe, whether you will choose all data to be shared or whether you share your data only partially and for how long. Explain if your data or part of it cannot be shared and give reasons for that. Reasons might be confidentiality, trade secrets or ownership issues (license, copyright). Sometimes data cannot be shared due to unreasonable effort required for data sharing (e.g. legacy data or big volumes of analogical data). <strong>Tips for best practices</strong></td>
</tr>
</tbody>
</table>
| **5.2 How and where the data with long-term value will be made available?** | Consider both, data sharing during and after research. The openness and sharing of research data promotes its reuse.

- The openness and sharing of research data promotes its reuse.

- When sharing your data it is recommended that it is available for reuse e.g. under Creative Commons or other relevant-license. The recommended CC-license for open science is [CC-By license](https://creativecommons.org/licenses/by/).

- There are different solutions for long-term preservation and sharing.

- There are different solutions for long-term preservation and sharing.

- Use persistent identifiers (PID) to enable access to the data via a persistent link (e.g. DOI, URN).

| **5.3 Have you estimated costs in time and effort to prepare the data for archiving?** | **Tips for best practices**

- Give a brief description what data to preserve and how long and what data to dispose after the project.

- Remember to check funder, disciplinary or national recommendations for data repositories, data archives or data banks.

- Use persistent identifiers (PID) to enable access to the data via a persistent link (e.g. DOI, URN, Handle).

| **The research data from this project will be deposited with [repository] to ensure that the research community has long-term access to the data. By depositing data with [repository], our project will ensure that the research data are migrated to new formats, platforms, and storage media as required by good practice. The [repository] will generate DOI's [=persistent identifier] enabling access to the data sets via persistent links.** | **The aim of long-term preservation is to store and keep data usable and comprehensible for tens or even hundreds of years. Data selected for long-term preservation will be submitted to a data repository or data archive. Long-term preservation will ensure your data can be found, understood, accessed and used in the future, even for generations.**

**Tips for best practices**

- Staff time has been allocated in the proposed budget to cover the costs of preparing data and documentation for archiving. The [repository]**
<table>
<thead>
<tr>
<th>Preservation and sharing?</th>
<th>has estimated their additional cost to archive the data at [insert cots]. This fee appears in the budget for this application as well.</th>
</tr>
</thead>
</table>
|                          | • Will you need specialist expertise to manage preserve and share the data?  
• Do you have sufficient storage or will you need to include charges for additional services? Consider if additional computational facilities and resources need to be accessed, and what will be the costs associated with this.  
• Consider if additional computational facilities and resources need to be accessed, and what will be the costs associated with this.  
• How will responsibilities for data management and costs be split across partner sites in collaborative research projects? |