UH DMP guidance 2018

This is an updated version of the guidance for writing data management plans at University of Helsinki. The structure and the content of the guidance follows the General Finnish DMP guidance. The guidance text is modified and tailored by the University of Helsinki Datasupport.

Please cite as: University of Helsinki Datasupport. (2018, August 9). UH DMP Guidance 2018. Zenodo. https://doi.org/10.5281/zenodo. 1342371

Introduction	
Motivation	Why should you manage your research data and write a data management plan (DMP)? Your DMP should describe how you manage data *during the whole research life cycle* – as well as cover what happens after the active phase of the project. The DMP is a *living document* which should be updated as the research project develops. Research data management and its planning (DMP) is an *integral part of good research practices*. By writing a DMP before your project starts, you will help minimise unexpected problems. The invaluable *advantages of data management planning* include the following: · Reducing the risk of losing data · Saving time and money · Meeting funder and other policy requirements (such as GDPR) · Maintaining/ensuring data integrity · Ensuring that your measurements are reproducible · Helping you overcome *complex ownership and user rights issues* in advance A clearly outlined DMP will also help you *support open access* in order to promote new discoveries and productive future collaborations. In the DMP *data is understood as a broad term* including: · Data collected by various methods (such as surveys, interviews, measurements, imaging techniques etc.) · Data produced during the research (such as analysis results) · Research sources (such as various archive material) · Source code and software Too often data cannot be shared or reused in research, teaching and learning at a later time because of poor data management planning at the beginning of a project. Good luck with your DMP!

1. General description of data	
1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in?	Consider your DMP as a part of your research plan. To avoid redundancy, refer to your research plan in your DMP and vice versa. Data analysis and methodology should be described in your research plan, not here. Briefly describe what types of data you are collecting or producing. Also explain what kinds of existing data you will use, for example, the types of texts, images, photographs, measurements, statistics, physical samples or codes. If you collect and handle sensitive data, please read the guidance HERE. Your answer to this question forms a general structure for the rest of the plan. Categorise your data in the following way and use the same categorisation in all the phases of your plan. 1. Data collected for this project 2. Data produced as an outcome of the process 3. Previously collected existing data reused in this project List the file formats used both during the research project and in archiving the data. Favour software and formats based on *open standards* to enable data reuse, *interoperability* and sharing. Tips for best practices · Use a table or bullet points for a concise way of presenting o data types o file formats (for example, csv, .txt, .docx, .xslx, .tiff) o the software used (especially if the software is coded in your project) o other information Clearly distinguish the data which is produced in this project from the data that has been produced earlier.
1.2 How will the consistency and quality of data be controlled?	*Data quality control* ensures that no data is accidentally changed and that the accuracy of data is maintained over its entire life cycle. Quality problems can emerge due to the technical handling, converting or transferring of data, or during its contextual processing and analysis. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, the peer review of data or representation with controlled vocabularies. Tips for best practices · Transcriptions of audio or video interviews should be checked by someone other than the transcriber. · Digitisation of analog or physical material should be done with sufficient accuracy.

	· In all conversions, maintaining the original information content should be ensured. Software that produces checksums should be used.
2. Ethical and Legal Compliance	
2.1 What ethical issues are related to your data management, for example, in handling sensitive data, protecting the identity of participants, or gaining consent for data sharing?	Describe how you will maintain high ethical standards and comply with relevant legislation when managing your research data. Ethical issues must be considered throughout the whole research data life cycle. For example, following the guidelines regarding informing research participants is considered an ethical requirement for most research. Moreover, if you are handling *personal* or *sensitive information*, describe how you will ensure *privacy protection* and data *anonymisation* or *pseudonymisation*. Tips for best practices · If you collect and handle sensitive data, please read the guidance HERE. · Check whether an ethical review is required of your research project. · If your research is to be reviewed by an ethical committee, outline in your DMP how you will comply with the protocol (e.g., how to remove personal or sensitive information from your data before sharing it to ensure privacy protection). Links to general guides · Data protection in Flamma (in Finnish) · Informing Research Participants (Data management guidelines, FSD) · Research ethics in Flamma · Legal aspects & research integrity (UH guide on research data management) · See the Finnish Advisory Board on Research Integrity for more information about the responsible conduct of research. · See the European Code of Conduct for Research Integrity.
2.2 How will data ownership, copyright and Intellectual Property Right (IPR) issues be managed? Are there any copyrights, licenses or other restrictions which prevent you from using or sharing the data?	Describe who will own the data and how the ownership issues have been agreed upon. Describe who can issue permissions to (re)use it. The University of Helsinki does not own the data researchers have collected, unless this has been clearly agreed upon. Tips for best practices · Agreements on the ownership of the data have to be done inside the research group and between the research partners. · If the agreements have not been made, a complex situation with overlapping ownerships is created. The unclear situation can hinder the use of the data and the publication of results. · When you make agreements on the ownership of data, you prevent possible conflicts about the usage rights to the data.

	· "Principal investigators are responsible for concluding contracts on the ownership and user rights of research data at as early a stage as possible..." [University of Helsinki research data policy] · It is recommended to make all research data, code and software created within a research project available for reuse, e.g., under Creative Commons, GNU, MIT or another relevant license. · Consider the relevant funder, institutional or departmental policy on copyrights or *IPR*. It is often required (especially in EU funded research) that the ownership of data is transferred to the university. Links to general guides · Research Ethics, Ethical Guidelines, Legislation etc. (In Flamma) · Legal aspects & research integrity (UH guide on research data management
3. Documentation & metadata	
3.1 How will you document your data in order to make it findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?	The purpose, origin, time, location, creator, *access conditions* and terms of use of a data collection are explained in your documentation (that is metadata). In addition, describe for example, your file-naming conventions, *version control* and folder structure in your metadata. Also, explain, the terms, variables, codes, abbreviations, units of measurements in the documentation. Tips for best practices · Describe the types of documentation that will accompany the data, for example README files code books, data dictionaries, questionnaires · Use research instruments which create standardised metadata formats automatically. Then your data can be moved from one manufacturer tool to another. · Repositories often require the use of a specific *metadata standard*. Check whether a discipline/community- or repository-based metadata schema or standard (i.e., preferred sets of metadata elements) exists that can be adopted. Links to general guides · Documentation & metadata (UH guide on research data management) · Data Description and Metadata (FSD: Data Management Guidelines) · Disciplinary metadata standards (DCC)
4. Storage and backup during the research project	
4.1 Where will your data be stored, and how will it be backed up?	Describe here, where you will store and backup your data during your research project. Opening, publishing and archiving your data after your research project has ended is explained in Section 5. 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 · Use storage services provided and maintained by University of Helsinki IT services. · Does your project have sufficient storage space. If not, please contact Helpdesk tel. + 358 (0)2 941 55555 or helpdesk@helsinki.fi · Don't use cloud services to store *sensitive data*. · Use OneDrive for Business instead of commercial cloud services (such as Google Drive) that are not supported by IT services. · Remember to state your intention to specify your data management costs in the budget. Links to general guides · Using & Storing (UH guide on research data management) · Research data services at UH · Data protection (section 8) in Flamma (in Finnish)
4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?	Create a brief *data security plan* where you describe your processes of how you access and handle the data safely. Tips for best practices · Use personal or shared network drive to control who can access and use your data. Personal and shared network drives are also backed up. · *Access controls* should always be in line with the level of confidentiality involved. · If you collect and handle sensitive data, please read the guidance HERE. Links to general guides · Data protection (section 3) in Flamma (in Finnish) · UH Helpdesk Tel. + 358 (0)2 941 55555 or helpdesk@helsinki.fi
5. Opening, publishing and archiving the data after the research project	
5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?	If your data or parts of it cannot be opened, explain why. The openness of research data promotes its reuse. Too often data can't be shared because of poor data management planning at the beginning of the research project. Tips for best practices · "As a rule, research data produced under the auspices of the University of Helsinki and related to published research results are open and available for shared use. The discoverability and citability of research data must be ensured." [University of Helsinki research data policy] · Publish your data in a data repository. If you can not open your data, you can publish a description (i.e., the *metadata*) of your data without making the data itself openly available o Check funder, disciplinary or national recommendations for data repositories. o Check re3data.org to find a *repository* for your data.

	○ Early selection of a specific data-sharing repository helps you prevent unpleasant surprises at the end of your research when you deposit your data. ○ Use repositories or publishers which provide *persistent identifiers* (e.g. DOI, URN) to enable access to the data via a persistent link . · It is recommended to make all research data, code and software created within a research project available for reuse under Creative Commons, GNU, MIT or another relevant license. ○ A CC0 waiver is recommended for opening research data and metadata. Links to general guides · Sharing & reuse (UH guide on research data management) · Citing data (UH guide on research data management) · Which License Should I Choose? (UH open access guide) · Publishing data (Open science)
5.2 Where will data with long-term value be archived, and for how long?	Briefly describe what data to archive and for how long – as well as what data to dispose of after the project. Describe the access policy for the archived data. Tips for best practices · Check funder, disciplinary or national recommendations for data archives. Links to general guides · Research data services at the UH · Five steps to decide what data to keep (DCC, UK) · Data disposal (FSD: Data Management Guidelines)
5.3 Estimate the time and effort required for preparing the data in order to publish or to archive it.	Publishing data can require a lot of preparation, for instance selecting, organising and transferring the data. It also takes time and effort to produce metadata, describe the data and anonymise personal data. Determine whether you need help from an expert to manage, preserve and share your data. The better you plan your data management already in the beginning of your research project, the less work you need to do when you open and share your data. Tips for best practices · Specify your data archiving, opening and publishing costs in the budget.