Finnish DMP evaluation guidance + General Finnish DMP guidance

Published 30.4.2021

By Tuuli working group





This evaluation guide is created by the Tuuli working group:

Sirpa Aalto (Univeristy of Oulu), Minna Ahokas (CSC), Jari Friman (Tampere university library), Siiri Fuchs (CSC), Tuija Korhonen (University of Helsinki), Mari Elisa Kuusniemi (Tuuli office /University Helsinki), Kati Laakso (Aalto university), Mietta Lennes (FIN-CLARIN & The Language bank of Finland / University of Helsinki), Soile Manninen (Tuuli office /University Helsinki), Mikko Ojanen (University Helsinki), Jukka Rantasaari, *chair* (Turku university library), Mika E. Virtanen (University of Oulu), Qingbo Xu (Hanken School of Economics)

The guide is loosely based on Data Management Plan Evaluation Rubric by Science Europe published in Practical Guide to the International Alignment of Research Data Management - Extended Edition (2021), pages 31 – 49.

Localisation work is done based on: General Finnish DMP guidance (Version 2020). Zenodo. http://doi.org/10.5281/zenodo.3630309

General tips for evaluators

This guide gives some general tips for evaluators. It can be used when evaluating DMP by students, peer reviewing or when evaluation is conducted by a data steward. The working group hopes you develop the guidance further in order to meet your specific needs and policies.

Ideally data management plan will be read and evaluated together with the research plan. In the DMP context, 'data' is understood as a broad term. Data covers all the information and material research results are based on (like codes, software, notes, etc).

Please cite (example):

Tuuli working group. (2021, April 30) Finnish DMP evaluation guidance. Zenodo. http://doi.org/10.5281/zenodo.4729832

O. Project details

	Evaluation guidance 2021
	Excellent Project details coversheet is included in the plan containing information of the Project title Funder (if applicable) Grant number (if applicable) Funding instrument (if applicable) (eg. Lippulaiva) Principal investigator Organization Other contributors ORCIDs of all contributors Date of updates Abstract of the project
	Satisfactory Describes the Title, Principal investigator, Date of latest update and Abstract of the project.
	Poor No background information of the project is given

1. General description of the 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? Additionally, give a rough estimate of the size of the data produced/collected.

General Finnish DMP guidance 2020

Briefly describe what types of data you are collecting or producing. In addition, explain what kinds of already existing data you will (re)use. List, for example, the types of texts, images, photographs, measurements, statistics, physical samples or codes.

Categorise your data in a table or with a clear list, for example:

- A. data collected for this project,
- B. data produced as an outcome of the process,
- C. previously collected existing data which is being reused in this project,
- D. managerial documents and project deliverables, and so on.

The categorisation follows the license policy of your data sets. For example, briefly describe the license according to which you are entitled to (re)use the data. The categorisation can form a general structure for the rest of the DMP.

List the file formats for each data set. In some cases, the file formats used during the research project may differ from those used in archiving the data after the project. List both. The file format is a primary factor in the accessibility and reusability of your data in the future. In the DMP, what is important is to describe the required disk space, not how many informants participated in the project. A rough estimation of the size of the data is sufficient, for example, less than 100 GB, approx. 1 TB or several petabytes.

Tips for best practices

- Use a table or bullet points for a concise way to present data types, file formats, the software used and the size of the data.
- Examples of file formats are .csv, .txt, .docx, .xslx and .tif.
- Make sure to describe any special or uncommon software necessary to view or use the data, especially if the software is coded in your project.
- You can also estimate the increase in data production or collection during the project for a specific time period: "The project is producing/collecting approximately 100 GB of data per week."
- AVOID OVERLAPS WITH THE RESEARCH PLAN! Data analysis and methodological issues related to data and materials should be described in your research plan.

Evaluation guidance 2021

Excellent

- Clearly lists data sets/types in categories. The list is in line with research plan and the rest of the DMP.
- Categorisation acknowledges different data handling needs (for example sensitive/confidential data).
- List of data types includes:
 - A) existing data which is being reused and description where/from whom the data comes from. (Licences or other agreements needed are described in section 2.2.)
 - B) new data produced/collected,
 - C) data produced (derivates) as an outcome of data analysis of this project.
- Data types and formats are clearly outlined in a table or a list with bullet points.
- Explains why certain formats have been chosen and indicates if they are in open and standard format. If a proprietary format is used, it explains why.
- Explains if any special or uncommon software are needed to view or use the data.
- Clearly describes how data volume or its accumulation has been calculated.

Satisfactory

- Describes or lists what data types will be used or generated and their associated data formats.
- List of datatypes is in line with the research plan and the rest of the DMP (no data set is missing compared to research methods or other data sources).
- Provides information about the estimated data volume or how the data volume accumulates.

- Provides no or little details on what data types will be generated and does not provide a valid reason for this omission-
- Only lists/describes the kinds of data without specifying the formats.
- Only lists formats, without specifying the kinds of data.
- Does not provide an estimate of data volume.

1.2 How will the consistency and quality of data be controlled?

General Finnish DMP guidance 2020

Explain how the data collection, analysis and processing methods used may affect the quality of the data and how you will minimise the risks related to data accuracy.

Data quality control ensures that no data is accidentally changed and that the accuracy of the 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.

Tips for best practices

- Transcriptions of audio or video interviews should be checked by someone other than the transcriber.
- Analog material should be digitised in the highest resolution possible for accuracy.
- In all conversions, maintaining the original information content should be ensured.
- Software-producing checksums should be used.
- Organise training sessions and set guidelines to ensure that everyone in your research group can implement quality control and anticipate the risks related to the quality of the data.

AVOID OVERLAPS WITH THE RESEARCH PLAN! Issues related to data analysis, methods and tools should be described in your research plan, that is, do not include, for example, instrument calibration descriptions here.

Evaluation guidance 2021

Excellent

- Clearly recognises possible error sources during the data lifecycle, to ensure the quality of data.
- Describes appropriate practices (data capture, validation/monitoring, versioning, logs, etc) chosen for each tasks of the data handling to ensure high quality of data.

Satisfactory

• Describes the approach taken to ensure data quality control during the data lifecycle

Poor

• Provides no information or only a vague mention on how data quality is controlled and documented during the lifetime of the project.

2. Ethical and legal compliance

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

General Finnish DMP guidance 2020

All types of research data involve questions of rights and legal and ethical issues. Demonstrate that you are aware of the relevant legislation related to your data processing. If you are handling personal or sensitive information, describe how you will ensure privacy protection and data anonymisation or pseudonymisation.

Tips for best practices

- Check your institutional ethical guidelines, data privacy guidelines and data security policy, and prepare to follow the instructions that are given in these guidelines.
- 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 you will remove personal or sensitive information from your data before sharing data to ensure privacy protection).
- Will you process personal data? If you intend to do so, please detail what type of personal data you will collect.
- All data related to an identified or identifiable person is personal data. Information such as names, telephone numbers, location data and information on the congenital diseases of the individual's grandparents is personal data.
- Office of the Data Protection Ombudsman (https://tietosuoja.fi/en/processing-of-personal-data)
- AVOID OVERLAPS WITH THE RESEARCH PLAN! Details of the ethical issues, the ethical committee statements and the use of laboratory animals should be described in the research plan.

Evaluation guidance 2021

Excellent

- Clearly indicates whether personal data, sensitive or confidential data, copyrighted data or other legally restricted data will be processed (collected, used and/or shared) as part of the project.
- Identifies the legal requirements and ethical practices that apply to processing the data.
- If applicable, explains how compliance with applicable legislation will be ensured during the project, for example by some of the following means:
 - · identifying the Data Controller
 - identifying the legal basis for processing personal data according to GDPR
 - considering data minimization, anonymization or pseudonymization, when appropriate considering the need for data encryption or other appropriate technical safeguards.
- If applicable, provides details of ethical issues that may affect data storage, transfer, use, sharing and/or preservation, and demonstrates that adequate measures are in place to manage ethical requirements.
- If applicable, mentions whether ethical review is being pursued. In case an ethical statement has already been obtained, refers to the relevant committee and documents.

Satisfactory

- Clearly indicates whether personal data, sensitive or confidential data, copyrighted data or other legally restricted data will be processed (collected, used and/or shared) as part of the project.
- Identifies the legal requirements and ethical practices that apply to processing the data.

- Clearly fails to notice that personal data, sensitive or confidential data, copyrighted data or other legally restricted data will be processed (collected, used and/or shared) and that the relevant legislation and ethical practices should be taken into account in the project.
- Does not discuss legal and ethical issues or potential risks to the data subjects and does not provide a sufficient explanation why they are not relevant for the research project.
- Personal data will be processed but there is no clear purpose for doing so.
- Does not discuss any safeguards for protecting personal data and does not provide an explanation for not using safeguards, if applicable.
- Describes safety measures that are clearly disproportionate with respect to the relevant legislation and ethical practices and the types of data in question.

2.2 How will you manage the rights of the data you use, produce and share?

General Finnish DMP guidance 2020

Describe how you will agree upon the rights of use related to your research data – including the collected, produced and (re)used data of your project. Here, you can employ your categorisation in the first question. Each of these categories involves different rights and licenses. Describe the transfer of rights procedures relevant to your project. Describe confidentiality issues if applicable in your project.

Tips for best practices

- Check your organisational data policy for ownership, the right of use and the right to distribute.
- Have you gained consent for data preservation and sharing?
- Agreements on ownership and rights of use should be made as early as possible in the project life cycle.

Consider the funder's policy.

It is recommended to make all of the research data, code and software created within a research project available for reuse, e.g., under a Creative Commons (https://creativecommons.org/choose/),

GNU (https://www.gnu.org/licenses/gpl-3.0.en.html) or MIT license

(<u>https://opensource.org/licenses/MIT</u>), or under another relevant license.

Evaluation guidance 2021

Excellent

Data rights and agreements

- Identifies the owner or rights holder of the data and the rationale concerning data ownership.

 Identifies the licenses or other terms and conditions that will apply to data re-use or describes how this issue will be settled.
- If applicable, explains how intellectual property rights will be managed and what agreements are required, e.g., for transfer of rights.
- In the case of multi-partner projects and multiple data owners, explains how their roles and responsibilities regarding the data are addressed in the consortium agreement.

Data protection

- If none of the restrictions mentioned below are applicable, this is stated clearly.
- Explains how the lawful processing of personal data will be ensured both during the project and regarding potential re-use of the data (e.g., compatibility with the legal purpose of processing the data).
- Explains how data confidentiality and non-disclosure will be ensured, to an appropriate extent.
- Mentions how the project complies with the funder's data sharing policy and if it does not, explains why this is not possible.

Satisfactory

- Identifies the owner or rights holder of the data (or mentions how this issue will be addressed at an early stage).
- Describes, at least on a general level, how the legal and ethical issues, stated in 2.1, will be taken into account when processing and sharing the data; or provides an explanation of why no special measures are required.
- On a general level, describes the agreements concerning the rights to use the data, but possibly without specifying them according to different types of data (as listed in section 1.1).
- On a general level, describes the terms and conditions that will apply to the processing and re-use of the data.

- Does not identify the owner or rights holders of the data.
- Does not discuss the effects that legal issues can have on the processing and sharing of the data (or a subset of it), and does not provide a good explanation for not doing so.
- In case of a multi-partner project, does not address the legal and ethical roles and responsibilities of the project participants and does not provide a good explanation for not doing so.

3. Documentation and metadata

3.1 How will you document your data in order to make the data 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?

General Finnish DMP guidance 2020

Data documentation enables data sets and files to be discovered, used and properly cited by other users (human or computer). Documentation includes essential information regarding the data, for example, where, when, why and how the data were collected, processed and interpreted. Without the proper documentation, your data is useless. Describe the tool, such as Qvain, that you will use to describe your data sets. Do not mention metadata standards if you do not use them. You can anticipate the open accessibility of your data and its description already here. However, a detailed description of which part of your data can be set openly available will be included in Section 5 below.

AVOID OVERLAPS WITH THE RESEARCH PLAN! The data-level documentation (https://www.ukdataservice.ac.uk/manage-data/document/data-level.aspx) and details about experiments, analytical methods and the research context belong to the research plan. In the DMP you should concentrate on the study-level documentation (https://www.ukdataservice.ac.uk/manage-data/document/study-level.aspx).

Tips for best practices

- Describe all the types of documentation (README files, metadata, etc.) you will provide to help secondary users to find, understand and reuse your data.
- Following the FAIR (https://www.force11.org/group/fairgroup/fairgrinciples) principles will help you ensure the Findability, Accessibility, Interoperability and Re-usability of your data.
- Know the minimum requirements for data documentation; see, for example, Qvain Light (https://www.fairdata.fi/en/qvain/qvain-light-user-guide/).
- Use research instruments, which create standardised metadata formats automatically.
- Identify the types of information that should be captured to enable other researchers to discover, access, interpret, use and cite your data.

Evaluation guidance 2021

Excellent

- Clearly outlines the documentation needed to verification and enable data re-use.
- Lists the metadata standards used for each data type.
- Describes how the documentation protocol is agreed (and documented), if no standard is available for a data type.
- Refers to documentation requirements of a data repositories/archives planned to use.
- Outlines who is/are responsible for the documentation during the data lifecycle (collection, analysis, storing, publishing, etc.)

Satisfactory

- Clearly outlines the documentation needed to to verification enable data re-use.
- Indicates how the data will be organised during the project (for example naming conventions, version control strategy and folder structures).
- Mentions common data documentation elements like, a 'readme' text file, file headers, code books, lab notebooks.

- Provides little or no details on the metadata that will accompany the data.
- Provides no information, or only a very vague mention of documentation, without providing any detail or explanation.

4. Storage and backup during the research project

option.

4.1 Where will your data be stored, and how will the data be backed up?

General Finnish DMP guidance 2020	Evaluation guidance 2021
Describe where you will store and back up your data during your research project. Explain the methods for preserving and sharing your data after your research project has ended in more detail 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.	 Excellent Clearly describes: why a certain storage solution and backup strategy has been chosen during the active research phase. Explains if, and why it is not the (preferred) home institution's storage. home institution's, CSC's or other service provider's suitable data storing places and backup policies with version management system for each data type in terms of data security and privacy, performance, capacity, usability and sustainability. storing physical data
Show that you are aware of the storing solutions provided by your organisation. Do not merely refer to IT services. In the end, you are responsible for your data, not the IT department or the organisation.	Satisfactory Describes: • The location where the data will be stored during the research activities, recognizing the possible limitations caused by types of data.
 Tips for best practices The use of a safe and secure storage provided and maintained by your organisation's IT support is preferable. Do NOT USE external hard drives as the main storing 	 If home institution's or CSC's storing services are not used: How often and how backups will be performed. Even a link to a service description can be sufficient.
	 Poor Provides no information or very vague reference to how data will be stored and backed up during the project. Does not show adequate understanding on the systems suitable for storing different types of data.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

General Finnish DMP guidance 2020	Evaluation guidance 2021
It is essential to consider data security issues, especially if your	Excellent
data include sensitive data, personal data, politically sensitive	Clearly explains:
information or trade secrets. Describe who has access to your	Who will be responsible for controlling access to the data platform(s).
data, what they are authorised to do with the data, or how you	How the access will be controlled.
will ensure the safe transfer of data to your collaborators.	what kind of appropriate safeguards, e.g. log file system is used, if any.
	Clearly describes the access control of physical locations, if data is stored in researcher's own premises.
Tips for best practices	
 Access controls should always be in line with the level of 	Satisfactory
confidentiality involved.	Describes:
	Who will be responsible for controlling access to the data platform(s).
	How the access will be controlled.
	Poor
	 Provides little or no details on who will control the access to the data platform(s) during the research, and how.
	 Provides little or no details about data protection and risk management, or the explanation is too vague, (especially when sensitive data are involved).

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?

General Finnish DMP guidance 2020

Describe whether you will make openly available or publish all your data or only parts of the data. If your data or parts of the data cannot be opened, explain why.

• In the case of sensitive data, which cannot be opened, describe the opening of its metadata. Describe the secured preservation procedure of sensitive data in Section 5.2.

The openness of research data promotes its reuse.

Tips for best practices

- You can publish a description (i.e., the metadata) of your data without making the data itself openly available, which enables you to restrict access to the data.
- Publish your data in a data repository or a data journal.
- Check re3data.org (https://www.re3data.org/) to find a repository for your data.
- Remember to check the funder, disciplinary or national recommendations for data repositories.
- It is recommended to make all of the research data, code and software created within a research project available for reuse, for example, under a Creative Commons
 (https://creativecommons.org/choose/), GNU
 (https://www.gnu.org/licenses/gpl-3.0.en.html) or MIT license (https://opensource.org/licenses/MIT), or under another relevant license.
- Consider using repositories or publishers, which provide persistent identifiers (PID) to enable access to the data via a persistent link (e.g. DOI, URN).
- AVOID OVERLAPS WITH THE PUBLICATION PLAN! The research article publication does not equal data publication. The data journal is a publication forum specialised in publishing research data.

Evaluation guidance 2021

Excellent

- Clearly describes how the data and/or metadata, or software will be made discoverable and shared.
- Specifies when data will be shared and under which license.
- Includes the name of the repository, data catalogue, or registry where data will or could be shared.
- Data will be shared in repositories, catalogues or registries which provide PIDs.
- Explains why a chosen data repository or archive is an ideal solution for opening the type of data.
- Clearly explains, if applicable, why data sharing is limited or not possible, and who can access the
 data under which conditions (for example, only members of certain communities or via a sharing
 agreement).
- Explains, where possible, what actions will be taken to overcome or to minimise data sharing restrictions.
- Clearly indicates which specific tools or software (for example specific scripts, codes, or algorithms developed during the project, version of the software) potential users may need to access, interpret, and (re-)use the data.
- Provides information, if relevant, on any protocol to access the data (for example if authentication is needed or if there is a data access request procedure).
- Anticipates how the data can be re-used in other contexts.
- Ensures that the data, metadata and software are shared under licenses which provide maximum reusability and openness.

Satisfactory

Describes

- which data and/or metadata or software will be shared
- the preliminary schedule and place for sharing.
- which data and/or metadata or software will not be shared, with the reason for not sharing.

- Provides little or no details on how and when data and/or metadata or software will be shared, or the explanation is not adequate or technically viable
- Misunderstands the difference between open access publishing of research articles and sharing of the data.

5.2 Where will data with long-term value be archived, and for how long?

General Finnish DMP guidance 2020

Briefly describe what part of your data you will preserve and for how long. Categorise your data sets according to the anticipated preservation period:

- A. Data to be destroyed upon the ending of the project
- B. Data to be archived for a verification period, which varies across disciplines, e.g., 5–15 years
- C. Data to be archived for potential re-use, e.g., for 25 years
- D. Data with long-term value to be archived by a curated facility for future generations for tens or hundreds of years

Describe which part of the data you will dispose of after the project and how you will destroy the data. Describe the access policy to the archived data. Consider using archives with a curation policy.

Tips for best practices

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

Evaluation guidance 2021

Excellent

- Specifies and categorize datasets that need different length of preservation:
 - A. data to be destroyed after the project. Describes how the data will be disposed after preservation period.
 - B. data to be archived for a verification period, e.g. 5-15 years.
 - c. data to be archived for potential re-use, e.g. for 25 years;
 - D. data to be preserved and curated for tens or hundreds of years,
- Describes how reliable management, preservation and admission to the datasets will be secured when needed because of verification, agreements or other reasons.
- Acknowledges the impact of legal, ownership, agreements, funders', institutions', and publishers' demands on data preservation.
- Provides the name of the archive or trustworthy repository or the way to curate and preserve data that will be used to make data available for re-use.

Satisfactory

- Describes what part of the data will be archived, where and for how long.
- Acknowledges the difference between storing of data vs. long-term preservation (archiving) of data.
- Acknowledges impact of legal, ownership, agreements, funders', publishers' and institutions' demands on data preservation.

- Does not acknowledge the difference between storing of data vs. long-term preservation (archiving) of data.
- Provides no further information or lacks adequate explanation what is planned about preservation.
- Does not take into account legal or other aspects which can make long-term preservation of the data impossible.

6. Data management responsibilities and resources

6.1 Who (for example role, position, and institution) will be responsible for data management (i.e., the data steward)?

General Finnish DMP guidance 2020	Evaluation guidance 2021
Summarise here all the roles and responsibilities described in the previous answers. Tips for best practices Outline the roles and responsibilities for data management/stewardship activities, for example, data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name the responsible individual(s) where possible. For collaborative projects, explain the co-ordination of data management responsibilities across partners. Indicate who is responsible for implementing the DMP and for ensuring that it is reviewed and, if necessary, revised. Consider scheduling regular updates of the DMP. Finally, consider who will be responsible for the data resulting from your project after your project has ended.	 Excellent Clearly outlines all the roles and responsibilities described in the DMP and names the individuals where possible: e.g. data management / stewardship, data capture, metadata production, data quality, storage and backup, data archiving, and data sharing Clearly states who is responsible for the data resulting from the project after the project has ended. Clearly states the procedure for transferring these responsibilities (in case the person is expected to leave the project). Explains how data management responsibilities are co-ordinated in collaborative projects. Indicates who is responsible for implementing the DMP and updating it during the project Satisfactory Describes data management roles and responsibilities and/or mentions that responsibility will be taken for data management without giving details of who / which processes. Does not clearly state who is responsible of the data resulting from the project after it has ended. Provides some information but does not give a clear statement how data management is taken care of in collaborative projects.
	 Poor Does not discuss responsibility for data management/stewardship activities and/ or does not indicate who is responsible for day-to-day implementation and adjustments to the DMP. Provides no description, in case of a collaborative project, on how data management responsibilities will be co-ordinated across partners.

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

General Finnish DMP guidance 2020

Estimate the resources needed (for example, financial and time) to manage, preserve and share the data. Consider the additional computational facilities and resources that need to be accessed, and what the associated costs will amount to.

Tips for best practices

• Remember to specify your data management costs in the budget, according to funder requirements.

Account for the costs of the necessary resources (for example, time) to prepare the data for sharing/preservation (data curation). Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, the costs of preparing data for deposit and repository charges.

Evaluation guidance 2021

Excellent

- Lists the required resources and facilities for data management (e.g. storing environment, computational facilities, hardware, staff time for preparing data for sharing, deposit, and repository charges) and refers to the specified financial costs in the budget, according to funder requirements.
- Provides estimates of time and money needed to prepare the data for sharing, publishing, preservation (data curation).
- Describes investments to expertise, like how lawyer, data steward, transcription service, IT expert's consultancy is purchased, or are these experts hired to the project.
- Outlines what kind of resources is needed on training and education to upscale data management skills needed.

Satisfactory

- Describes what part of the data will be archived, where and for how long.
- Acknowledges the difference between storing of data vs. long-term preservation (archiving) of data.
- Acknowledges impact of legal, ownership, agreements, funders', publishers' and institutions' demands on data preservation.

- Does not acknowledge the difference between storing of data vs. long-term preservation (archiving) of data.
- Provides no further information or lacks adequate explanation what is planned about preservation.
- Does not take into account legal or other aspects which can make long-term preservation of the data impossible.