

Finnish DMP evaluation guidance

Published 30.4.2021

By Tuuli working group



This evaluation guide is created by the Tuuli working group:

Sirpa Aalto (University 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, May 24) Finnish DMP evaluation guidance. Zenodo.
<http://doi.org/10.5281/zenodo.4762326>

0. Project details

	Evaluation guide
	Excellent Project details coversheet is included in the plan containing information of the <ul style="list-style-type: none">• 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 <ul style="list-style-type: none">• 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

Question	Evaluation guide
<p>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.</p>	<p>Excellent</p> <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> 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.
	<p>Satisfactory</p> <ul style="list-style-type: none"> 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.
	<p>Poor</p> <ul style="list-style-type: none"> 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.

Question	Evaluation guide
1.2 How will the consistency and quality of data be controlled?	Excellent <ul style="list-style-type: none">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 <ul style="list-style-type: none">Describes the approach taken to ensure data quality control during the data lifecycle .
	Poor <ul style="list-style-type: none">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

Question	Evaluation guide
<p>2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)</p>	<p>Excellent</p> <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 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.
	<p>Satisfactory</p> <ul style="list-style-type: none"> • 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.
	<p>Poor</p> <ul style="list-style-type: none"> • 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.

Question	Evaluation guide
<p>2.2 How will you manage the rights of the data you use, produce and share?</p>	<p>Excellent</p> <p><i>Data rights and agreements</i></p> <ul style="list-style-type: none"> 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. <p><i>Data protection</i></p> <ul style="list-style-type: none"> 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.
	<p>Satisfactory</p> <ul style="list-style-type: none"> 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.
	<p>Poor</p> <ul style="list-style-type: none"> 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

Question	Evaluation guide
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?	Excellent <ul style="list-style-type: none">• 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 <ul style="list-style-type: none">• 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.
	Poor <ul style="list-style-type: none">• 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

Question	Evaluation guide
4.1 Where will your data be stored, and how will the data be backed up?	<p>Excellent</p> <p>Clearly describes:</p> <ul style="list-style-type: none">• 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
	<p>Satisfactory</p> <p>Describes:</p> <ul style="list-style-type: none">• The location where the data will be stored during the research activities, recognizing the possible limitations caused by types of data.• 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.
	<p>Poor</p> <ul style="list-style-type: none">• 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.

Question	Evaluation guide
4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?	<p>Excellent</p> <p>Clearly explains:</p> <ul style="list-style-type: none"> • Who will be responsible for controlling access to the data platform(s). • How the access will be controlled. • 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.
	<p>Satisfactory</p> <p>Describes:</p> <ul style="list-style-type: none"> • Who will be responsible for controlling access to the data platform(s). • How the access will be controlled.
	<p>Poor</p> <ul style="list-style-type: none"> • 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

Question	Evaluation guide
<p>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?</p>	<p>Excellent</p> <ul style="list-style-type: none"> • 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.
	<p>Satisfactory</p> <p>Describes</p> <ul style="list-style-type: none"> • 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.
	<p>Poor</p> <ul style="list-style-type: none"> • 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.

Question	Evaluation guide
5.2 Where will data with long-term value be archived, and for how long?	<p>Excellent</p> <ul style="list-style-type: none"> • Specifies and categorize datasets that need different length of preservation: <ul style="list-style-type: none"> 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.
	<p>Satisfactory</p> <ul style="list-style-type: none"> • 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.
	<p>Poor</p> <ul style="list-style-type: none"> • 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

Question	Evaluation guide
<p>6.1 Who (for example role, position, and institution) will be responsible for data management (i.e., the data steward)?</p>	<p>Excellent</p> <ul style="list-style-type: none"> • 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
	<p>Satisfactory</p> <ul style="list-style-type: none"> • 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.
	<p>Poor</p> <ul style="list-style-type: none"> • 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.

Question	Evaluation guide
<p>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)?</p>	<p>Excellent</p> <ul style="list-style-type: none"> • 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.
	<p>Satisfactory</p> <ul style="list-style-type: none"> • Refers to the project budget and mentions some of the required resources (e.g. storing environment, computational facilities, hardware), but does not list the resources or specify the expected financial costs. • Mentions but does not give an estimate of the time and money needed to prepare the data for sharing and preservation (data curation).
	<p>Poor</p> <ul style="list-style-type: none"> • Provides no answer or is vague about the resources required for data management (e.g. resources are not listed and time for data management is not allocated). • Does not refer to the project budget regarding data management resources and costs.