DMP themes: Guiding questions & links

Data collection

Will you be reusing existing datasets and if so, from where?

Will quality assurance processes be adopted?

How often will you collect new data? Will the frequency change over time?

Links:

https://staff.ki.se/finding-and-collecting-data

Data description

How will data be created/collected?

What type of data will be created? For example:

- Observational data: captured in real time, i.e. sensor readings
- Experimental data: from lab equipment, i.e. chromatographs, blots, images, NGS data
- Simulation data: generated by computational models
- Derived/compiled data: generated by data mining/database compilation, etc
- Datasets: data from registries, gene sequence databanks, etc

Is the data created reproducible (e.g. experiment) or irreproducible (e.g. observations)?

Data format

What file formats will be produced? For example:

- Text: e.g. txt, HTML, PDF
- Numeric: e.g. xlsx, csv
- Audiovisual: e.g. jpeg, png, tiff, mp3, mp4, avi
- Simulated: state the model, model type, computer code, output format
- Discipline-specific: e.g. CIF in chemistry
- Instrument-specific: state the equipment and format

Do these formats conform to an open standard or are they proprietary?

Links:

https://staff.ki.se/file-organization-naming-versioning-and-formats

Data volume

What is the expected volume (in GB or TB) of data created within the project?

Metadata & documentation

How will the versioning be handled?

Will established standards/terminologies/ontologies be used to allow inter-disciplinary interoperability?

In case it is unavoidable that you use uncommon or project-specific ontologies or vocabularies will you provide explanations?

Do you use and provide information about variable names, codes abbreviations?

How will metadata (information with purpose to describe, explain and localize the data) be documented? Metadata may include for example information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data and software used to collect and/or process the data.

Will you be using any metadata standards? Which?

Are special tools or software needed for the data to be read and interpreted?

How will you name your files and folders in a systematic and consistent way?

How will you structure your folders?

Will any persistent identifiers (for example DOI: Digital Object Identifiers) be used?

Where will the research process (aim, materials, methods, analyses, results) be documented (will KI ELN be used?)

Links:

https://staff.ki.se/electronic-research-documentation-ki-eln

https://staff.ki.se/what-to-document

https://staff.ki.se/metadata-and-standards

https://staff.ki.se/file-organization-naming-versioning-and-formats

https://staff.ki.se/the-fair-principles

Storage & security

Where will the data be stored? For example, on a central KI IT server, on an institutional server, on KI ELN?

Will all data be stored? For how long?

How is backup organized (frequency and responsibility)?

Can data be recovered after an incident?

How will you make sure that unauthorized access is avoided? Describe how permissions and restrictions will be enforced.

Is the data encrypted?

Are firewalls used?

Links:

https://staff.ki.se/store-and-share-data

https://staff.ki.se/snic

https://staff.ki.se/information-security

Ethics & privacy

Does your data contain confidential or sensitive information? In that case, do you have written consent from the respondents?

Who will have access to confidential/sensitive information during the project? How will you make sure that unauthorized access is avoided? Describe how permissions and restrictions will be enforced.

Are there any ethical issues that can have an impact on data sharing?

Links:

https://staff.ki.se/human-ethics-application

https://staff.ki.se/personal-data-in-research

Intellectual Property Rights

Are there any copyright and/or intellectual property right issues to consider? In that case name persons/entities that hold these rights and note any constraints.

Will data sharing be postponed or restricted due to patenting?

Will data be licensed to regulate its reuse by others?

Preservation

How will long-term preservation of data be assured?

Who is responsible for selecting which data will be archived? Which data should be kept forever and which data can be discarded after 10years?

In which repository or data archive will the data be stored in the long run?

Have you chosen formats suited for long-term sustainability and accessibility that:

- are widely adopted by users
- are independent of specific software, developers, or vendors
- have open technical specifications (i.e. are not owned by a single person or organisation)

When will the data be archived?

Will you use the central KI archive or is there an archive at your institution?

Links:

https://staff.ki.se/archiving-research-data

Data sharing

Do you intend to make all your data available for sharing or will you select certain data to share only (if so on which grounds)?

Are there any ethical/legal issues that can have an impact on data sharing?

When will the data be shared?

If some datasets cannot be shared openly and are kept closed:

- Explain why the data cannot be shared openly, clearly separating ethical/legal issues from voluntary restrictions
- Will an anonymized version of the data be shared?
- Will metadata be shared?

- Will it be possible to share the data upon request?
- Will a data sharing agreement be required?

How will the data be shared? Will a repository be used? Will the group create and maintain a dedicated website?

Have you taken preliminary contact with a repository?

What tools will be needed/software will be needed to work with the data that you share?

Are there any embargo periods for the material?

Will the data be licensed to permit the widest re-use possible? How (e.g. creative commons)?

Links:

https://staff.ki.se/share-data-and-collaborate

https://staff.ki.se/the-fair-principles

https://staff.ki.se/publish-share

https://staff.ki.se/data-repositories

https://staff.ki.se/finding-and-collecting-data

An international list of data repositories is available via http://re3data.org

Roles & responsibilities

Have you ensured that adequate resources for handling the data are available?

Who is responsible for which part of data management?

For collaborative projects you should explain the coordination of data management responsibilities across partners.