How to write a Data Management Plan

Transcript of video

By Leif Longva, UiT The Arctic University of Norway

ORCID: 0000-0001-6638-8317

Date: December 2022

In this video I will go in detail through the elements you should include in your Data Management Plan.

The DMP will be a useful roadmap to help you through your research project.

And the process of writing the DMP will itself be of great help. It will make you think closely through each step of your project. The DMP will thus be a good management tool for you. And it will be a management tool specifically tailored to your own project.

Here is a list of information usually included in a Data Management Plan. These are typical elements to include, regardless of discipline and type and size of project. Notice that there are separate video sessions on many of these topics, and you are recommended to watch those before you go through how to write a Data Management Plan.

So let us look closer at what to include for each of these elements.

First, you should include general information about the project. This is useful for many reasons. One reason may be that your DMP will be a nice introduction to your project for any peer or colleague or others, who you want to discuss your project with.

You may include the following:

Project name and project number, to avoid ambiguity on which project the DMP is for. The project name and number will also enable precise searches for all kinds of outcomes from your project.

Funders supporting your project. A unique project number is usually given by the funder. Normally a funder will require that funding information is stated, in the DMP as well as in the metadata when you archive your data.

How is your project related to other projects? Is it part of a bigger, overarching project? It may be useful information to place your project within the bigger picture. Those who find your project

interesting, will likely be interested in related projects.
What is the project period – beginning and end?

A short, easy to read description of your project.

List the name of the project leader (this might be you), as well as the name and affiliation of all project partners. People have their networks, and listing the names of your project partners may lead to fruitful new connections for yourself.
Clear documentation of the responsibilities and rights within the project, is important. Be explicit, and make sure it is clearly defined and documented who is responsible for every task related to research data management within the project. And make sure that every project partner and every member involved is aware of and understands their tasks and responsibilities, to eliminate risks of confusion.

Do not give rights to more people than what is needed. This reduces the risk of someone carelessly messing things up.

For example, who is responsible for the Data Management Plan itself? And who is responsible for revising the plan as the project proceeds? It may commonly be the project leader, or someone who is awarded this task.

Who should have access to your data? You may want to let a number of project members have view access to your data, but maybe only you should have the right to manage or edit the data. Do not risk your data getting messed up! Your data may need structuring into files and folders, and being tagged with keywords and descriptions. This should be done consistently, to avoid confusion. And it is therefore perhaps wise to not distribute this task to several persons.

What about the rights to your data? Who owns the data? This is important, and needs to be explicitly stated at an early stage. Are the data yours to own? Or is it your employer who has the ownership? The latter is common. And in any case, your employer may have requirements on what you should do with the data, for instance in terms of archiving them in an open archive.

Do you make use of data from some external entity, either a partner in the project or not? In that case you should enter into a formal agreement with this external entity, and this should clearly state who owns the data, and what you are allowed to do with the data, including terms of archiving and

other distribution.

Your project's resource requirements are of course a crucial part of your planning. At an early stage you need to think through your resource needs, for the activities ahead. If you need to apply for funding of your project, the resource requirements, in terms of funds, is of course vital for your application. And if funding is granted only in part, you need to revise your plan: Are there alternative funding sources you may turn to, to fill the funding gap? Or do you need to revise and reduce or reorganise your planned activities in your project? If so, will the planned outcome of the project still be achievable?
And remember that your resource requirements comprise more than the funding. There may be lots of resources needed, that do not involve funds. This may include vital resources at your home university or institution, such as computer hardware and software, and support staff. Make sure you include your need of such resources in your Data Management Plan, and also include documentation in your DMP as soon as you have verified that they are in fact available at the time when you will need them. The same goes for human resources that you may need, without requiring extra funding –'in-kind' contributions to your project. Be explicit in your DMP and towards the persons in question, so you do not risk any misunderstanding on this.

The act of collecting or generating data is a core activity. The Data Management Plan will help you think through all steps involved here:
What kind of data will be collected or generated? Observations? Simulations in a lab? Interviews? Do you use some data sources, like a text corpus or raw data that already exists?

You should look carefully for existing data on your subject, and examine whether such data may be reused in your project. How to search for data is something you should include in your Data Management Plan. And then, later, you may revise your plan with documentation on how you searched and what you found.

What type of data will be collected? Numbers in a spreadsheet? Sound or movie recordings? Text?

What standards or methods will be used for collecting or generating the data? You need to describe closely the methodology you will use, and you need to plan the activities carefully to ensure that you follow the described methodology. The methodology may be vital for the quality of your data.

When do you plan to do the data collecting or generating?

Do you need some extra computer facilities or expertise to handle the data? We already raised the issue of resource requirements for your project. Planning your data collecting phase should help you to reveal your resource needs on this.
Losing data can be disastrous to your project. So the issue of storage and backup is obviously a crucial issue for you to consider and to plan for.
Be careful to think through, and include in your Data Management Plan, how you intend to store and take backups of your data. The same goes for all notes and information to document your activity, while collecting or generating your data.
The general recommendation for data storage is 'here, near and far', meaning that you should have three copies of your data: For instance the laptop you use daily, plus some other storage facility nearby you (for instance a second computer), and then somewhere remote, which may be in the cloud. In your planning you should investigate how you can comply with recommendations like these. Will you be able, say, to transfer data files to a secure storage facility in the cloud or back at your institution? How safe is your computer, or the cloud storage?
Do not risk running out of storage space. Think through your needs, and plan how you may expand storage space if needed.
Include in your DMP to contact your IT Department or other support services, after you have unravelled your needs, to inquire how they can assist you.

How will your data be documented, and what metadata will you use? A Readme file is a common method for documenting your data. The Readme file is a document where you describe your dataset so that any peer or outsider may understand and interpret your dataset correctly. This includes information on how data was collected or generated, a guide to the data files and what is found where, and any other information to help outsiders understand correctly what information your dataset holds. So if someone wants to scrutinise what you did in your project, and the conclusions you drew from your data analysis, it is important that you avoid any misunderstanding regarding your dataset.

In addition to the Readme file, you will also enter metadata for your dataset, when you, in the end of your project, upload your dataset to an archive. The metadata, along with the Readme file, should add up to comprise good documentation of your dataset.
But the metadata also serves another purpose, namely the findability of your data. Search engines will search through your metadata, if you have archived your data in a searchable repository.
So it is good advice to investigate the metadata schema of the archive you plan to use, when you in

the end reach the stage of archiving your data. Look for information on the archive's policy on how to document your data, and what metadata are required and recommended. And above all, you should think closely through what metadata is needed to give good information on all aspects of your dataset. This way you will know what information to gather, and what metadata you should collect while doing the data collecting or generating activity.

What file formats will you use? You may choose easy to use formats in the collection phase – say an excel spreadsheet. But in the archiving phase you need to make sure to use preferred formats. So, commonly, you may need to convert the data files from the formats you use when collecting or generating the data, to preferred formats suited for archiving. So the file format issue is also something to include in your planning and in your Data Management Plan.

And finally, you should think through and plan the file names and the folder structure, so that it is easy to get an overview of your dataset, and to see what data are found where. These things too are part of the documentation of your dataset.

If you deal with persons in your research, or by other means make use of personal data, you need to be careful to follow ethical rules and regulations concerning such data. You should include in your DMP how you will treat these considerations:
You need to make sure the persons involved are well informed on what data you will collect, and how these data will be used during the project and how they will be handled after the project ends.
You must make sure all persons participating in your research take part voluntarily.
Data should only be used for its original purpose. Some exceptions do apply. You should confer with the GDPR regulations, or relevant authorities, if some other use is considered. And normally, you need to document consent from the persons in question to your planned use of the data.
Data involving personal information need to be anonymised. One way to do this is by aggregating your data sample to avoid persons being identifiable.
You should check at your institution for assistance on how to do these things right, if you do include persons and information on persons in your research. And it is also likely that you need to apply for approval from a national or local authority to carry out such a research project. All this should be made clear in the planning phase of your project, and your effort to clarify this belongs in your data management plan.

Then, finally, is the question of how your data should be archived. Archiving is what you do when you have done all the analysis you intended, and are ready to hand your data over for long term

preservation, and, if possible, for others to find and reuse. For some highly sensitive data there may be a requirement to destroy the data after you finish your data analysis. But normally it is a question of open or restricted archiving. Perhaps your data may be archived openly if properly anonymised or aggregated? While the raw data are archived behind access restrictions? The statement from the European Commission that research data should be “As open as possible, as closed as necessary”, gives a nice lead.
So first you need to consider and clarify whether the nature of your data implies that there must be restrictions on its accessibility. You should confer with the appropriate authority at your institution or nationally, to settle this. And how to go about this is something you should include in your Data Management Plan.

Next is the question of choosing the best archive for your data – restricted or unrestricted. You may look for an archive that is more or less tailored for the subject of your research. Or you may look for a cross disciplinary archive that also covers your subject and kind of data. It may be that your home institution runs such an archive. One thing to look for is whether the archive is certified.
The Core Trust Seal is a well reputed seal that indicates that an archive has good data curating routines and is trustworthy and long term financially sustained.
You should put effort into finding a reliable archive, and make your choice of archive at an early stage, so you know what requirements to keep in mind regarding documentation and metadata. Include the considerations on how to choose an archive in your Data Management Plan. And update your plan as soon as your choice is made.
Perhaps your data need some processing before they can be openly archived? Or maybe you need to convert your files from the formats used in the collecting phase to a preferred format for archiving? These considerations too, may be entered into your DMP.

And finally, how do you plan to license your data for reuse. Remember the statement “As open as possible, as closed as necessary”, which means that the license should be as open as possible. And in any case, explain your choice in your Data Management Plan.
Now, please pause the video, read the exercise, and take a few moments to evaluate your own planning. This is purely for your own benefit, and you are not asked to submit your answers anywhere.