02.09.2022

Draft for CASCB Data Management Plan

Ilja Werner, Lena Dreher

Research Data Managers

Universität
Konstanz

Centre for the Advanced Study
of Collective Behaviour
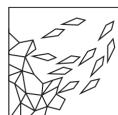
Introductory remarks

Data Management Plans (DMPs) are integral part of research practice. It raises questions any research project will have to face at some point in its life cycle. A DMP helps gain and keep an overview of these questions and requirements for keeping your data safe, usable and findable.

A DMP is a handbook for how you plan to generate, store, back-up, publish, archive and generally work with your data, even after the project ends. Therefore, is a living document that can be and will have to be revised and updated during the research project: not every detail can be known at the beginning of a project. So when details become clearer or plans change or additional questions have to be dealt with, the DMP should be updated with the new information or plans. Only then, it can be a useful guide throughout the project.

This template incorporates the ideas of FAIR research data management. Keeping your data FAIR means to make sure it is always:

- Findable (where do you store it, are naming standards in place, metadata standards, persistent identifiers when publishing, …?)
- Accessible (who needs access to the data during the project, where, how and when will back ups be done, are metadata accessible, …?)
- Interoperable (what hardware/software/other tools do I need, do my publications refer to my data as well, …?)
- Reusable (long-term archiving, what and where, what data format, what licence, …?)

If you have any questions regarding research data management or would like feedback for or assistance with your data management plan, please do not hesitate to contact
Ilja Werner (ilja.werner@uni-konstanz.de) or
Lena Dreher (lena.dreher@uni-konstanz.de).

Project name:
Project number:

Data summary:

(Data files created or used in the project)

Name:	
Origin:	
Owner:	
Format:	
Description:	
Purpose for the project:	
Utility for others:	
Contains personal data:	
Contains sensitive personal data:	
Access restrictions:	
Repository/Archive:	
Embargo period:	
Licence:	

Project Description:

Data description and collection

Which data will be generated?

Types (numeric, text, images, video, …):	
Formats (file formats, proprietary formats should be convertible into open formats):	
(Estimated) Volume (number of files, total file size):	
Does the project re-use existing data? If not, have you considered this option?	

Documentation:

Are you following a metadata standard? If not, which information will be attached to the data to describe them? (Consider what is necessary for others to find, understand and potentially re-use your data, e.g. author, affiliation, project, funder, date, sample, proband, hardware used, software required, …)	
How do you document the definition of variables and details of methodology? And where do you store the documentation?	
Which tools (e.g. software) are required to use the data?	
Are quality controls in place and if so, how do they operate? (What do you do to make sure your data files are never broken?)	
How do you organise the files, what is the naming convention?	
Who needs access to the data during the project (names and affiliations)?	
Where/how can these people access the data?	

Storage and Backup:

Where will data be stored during the project?	
Outline of the lifespan of the data (from data collection, after 3 months, after 6 months, after project end, long term, e.g. on collection data goes to what hardware? After how many days will you transfer it to which other storage for analysis? Where will the data go from there and when, …)	
Where will your data be backed up to? (We recommend 3 different copies stored in at least two separate locations and one version in the cloud)	
What data will be backed up? Have you considered back up of e.g. documentation, settings/calibration data, keys for encryption or de-pseudonymisations?	
How frequently will that data be backed up? (time schedule)	
Who is responsible for doing the back ups?	
How will data be protected from unauthorised access?	

Legal and ethical requirements:

If personal data is collected, what kind? (Please do not forget to fill in the university's VVT form prior to the collection of personal data. The research data managers can assist you with that.)	
If you re-use data: do you hold all the rights required for re-using and sharing? (e.g. intellectual property rights?)	
Are there any legal restrictions to subsequent publication or re-use of your data? (Patents pending, non-disclosure agreements, …)	
Are there any ethical reasons restricting publication or re-use of your data?	
Are there any other legal contracts with projects staff or third parties?	

Data sharing and long-term preservation:

Which datasets will be published and made available?	
Datasets that will be needed to validate results of any text publications of the project:	
Additional datasets that might be useful for other researchers	
Where will these datasets be published? (Which repository or similar?)	
Under which licence will the data be published?	
If there are any datasets that cannot be made freely available, can they be made available with an embargo?	
If there are any datasets that cannot be made freely available and not with embargo, where will the metadata be published?	
For datasets not shared: How is long-term preservation assured and for how long?	

Responsibilities and resources:

Who is responsible for data management in your project?	
Who is responsible for keeping the DMP up to date?	
Who is responsible for keeping track of legal requirements?	
Who is responsible for the data after the end of the project?	
Which resources will you need for your RDM? (e.g. hardware, software, technical expertise, …)	
Which costs will you have for RDM activities (e.g. for hardware, software, archiving services, …) and how will they be covered?	