UNIVERSITEIT VAN AMSTERDAM

Data Management Plan

Guidance

Introduction

A data management plan (DMP) serves two key purposes:

1. It provides a framework to help you get the most out of your research data, making it easier to communicate the way you work with collaborators and prompting you to consider aspects of data management which may be new to you; and
2. it demonstrates to funders and commercial partners that you are taking responsibility for data they have funded or supplied.

Data management planning should not be overly time-consuming and the level of detail should be proportionate to the size and nature of the project. For a 6-month pilot study, a few bullet points will suffice, whereas a 10-year project costing several million euros will require more depth.

Saving yourself time

Many elements of a DMP will be the same or similar for every project you undertake – we encourage you to keep note of these and reuse them to save yourself time and reduce the probability of error. You can gather these recurring elements into a standing personal-, group- or department-level DMP.

If you adhere to a department-level protocol regarding data ownership, storage, security and/or documentation, in the relevant sections of the DMP you can refer to this protocol instead of answering each questions in full.

Adapting to the situation

Circumstances change, or simply become clearer, during a research project. If part of a data management plan is no longer appropriate, it can and should be changed. Some of the benefits of a

data management plan are only realised at the end of a project, and these will be lost if the plan is out of date.

Involving the right people

If a data management plan identifies actions or resources that you need help with, you should contact the relevant people as early as possible. This may include colleagues within your department for advice on issues particular to your discipline or to pool resources, and RDM Support for general advice on data management and for information on how data can be catalogued, archived and preserved in the long term.

More information on all aspects of research data management can be found at http://rdm.uva.nl/ (Dutch) or http://rdm.uva.nl/en (English).

Completing the plan

Administrative details

This section records some basic details to ensure that the DMP is associated with the correct project and Principle Investigator.

Data creation/collection

This section collects information about the type of data you will be using, along with where and how it will be obtained. This information will be useful later when you come to archive and possibly publish your data.

Heading	Guidance	Why is this required?
1. What existing sources of data will be used?	If you will be reanalysing existing data (your own or belonging to someone else) using new techniques, or combining multiple sources of data, you should describe this here. If the data you need for your research does not yet exist (i.e. you will be producing new data), describe the gap. If you are reusing existing data, you should also consider what processing will be required to bring it into a useable form. If specific expertise is required for this, you should allow staff time and possibly training in the budget. If you will not be reusing existing sources of data, state that here.	Funding bodies can be reluctant to fund the creation of data which they consider has already been collected. You should demonstrate that you have considered other sources of data.

Heading	Guidance	Why is this required?
2. What are the characteristics of the data?	Give a brief overview of the type of data you will be using and how much (try to give a numerical estimate, in MB/GB/TB), along with what file types will be used and what software will be required. If possible, suggest how this data will grow during the course of the project.	Information about types of data and software will enable the University to plan for the long-term preservation of your data. Information about volume will enable ICT Services to model storage demand and continue to provide a high-quality storage service.
3. Who owns the copyright and intellectual property involved?	Briefly state who owns or will own the data. For single-institution projects ownership will usually lie with the University. In a multi-partner project, you should outline which partner owns what intellectual property and what rights other partners have to use it. This should have been set out in the collaboration agreement. If you are using secondary data, give an idea of the licensing restrictions that apply.	This will be useful to refer to later in your project, if questions arise about what can be done with a particular piece of data and by whom. This is particularly important when archiving or publishing.
4. How will the quality of the data be guaranteed?	Data quality is a measure of how accurate and reliable it is, and whether it is valid for the purpose intended. Outline how the quality of data will be assured, from the original collection, through digitisation/transcription, to checking, validation and cleanup. Consider calibration and both manual (e.g. peer review, double entry) and automatic (e.g. validation rules) checks.	Documenting quality control procedures and sharing them within the project team will help ensure that everyone understands what's expected of them. This will also help when offering data to an archive for long-term preservation, as it gives a simple way of showing compliance with their quality criteria.

Data management, documentation and curation

This section considers how data will be organised and described during the life of the project. It should be written as a day-to-day reference for members of the project team, and will help to ensure that data is still available and comprehensible when you come to write up results for publication. To support this, your data will need to be held safely and securely, in an organised way that the whole project team understands.

Heading	Guidance	Why is this required?
5. What will be the primary storage medium and location?	Describe where you will store your main copy of the data. If secondary copies are stored elsewhere, describe how you will keep these in sync. If you store your data on the University network drive, backup and security will be taken care of for you. See also the section on Data security.	Clarity on which copy is the master and confidence that it is up to date will ensure that no-one wastes time analysing old or incomplete data. It will also support the transparency and integrity of your research.
6. How will files and folders be named?	Describe how the project's files will be organised. You may already have some rules that you follow; in that case just document them here.	Having a clear, simple policy will help you and others find the right files later on, when some time has passed since saving the original files and you come to analyse, archive or share the data.
7. How will the data be described and documented?	Describe how the context required to interpret the data will be recorded. This could be as simple as recording this context in a "readme" document placed in the same location as the data file(s) it describes. Some software, including Microsoft Office, allow such information to be recorded as "properties" in the file itself. If there are established practices for this in your research area, such as codebooks or lab notebooks, you can briefly refer to these.	Being able to accurately interpret data after the passage of time is just as important as being able to find it, and to aid in this some additional description must be held with the data. Considering, early on, how this will be done will save time later.
8. How will file versioning be managed?	If only the latest version of each file needs to be kept, then they can simply be overwritten. A more powerful technique is to include a version number, date and/or author's initials in the filename when saving. This can be combined with the "Track changes" feature, available in many software packages, when collaborating. For some purposes, especially tracking changes made by multiple authors or recording the development of software code, dedicated version control software can be useful.	It is very easy to lose important information by accidentally saving over an existing file. Considering how new versions of a file will be handled will help to prevent this, as well as providing a valuable record of how the work and the thinking around it developed.

Heading	Guidance	Why is this required?
9. What metadata standard and metadata format will be used?	There are a number of general and specialised standards for describing data at various levels. The simplest and most widely applicable is Dublin Core, which describes 15 elements (such as Title, Creator and Subject) that are the bare minimum required to describe any dataset, document or other object. There are also a number of machine-readable forms in which standard metadata can be recorded.	Using well-documented standards means that important information about your data can be understood by the professional curators who will be responsible for its long-term preservation after archiving. Understanding at least the Dublin Core elements and ensuring you record the relevant ones for your data will make archiving and publishing your data easier.
10. How will non-digital data be catalogued, described and stored?	Describe how non-digital data (such as written notes) will be incorporated into the plan so that they, too, can be kept safe and shared where appropriate. This could be as simple as scanning them at regular intervals, though in some cases physical storage options such as a fireproof safe may be necessary. These materials should also be catalogued and, ideally, cross-referenced with the rest of the project's data.	Taking simple actions, such as digitising handwritten notes, enables these non-digital objects to be shared and backed up in the same way as digital data. They will need to be kept in order to respond to queries arising through peer-review and after publication.

Data security

This section considers how to keep your data safe from accidents and malicious attacks. Data or information security comprises two key aspects. First, data must remain intact and available to those who need to use and access it. Second, unauthorised access to the data must be prevented. These two aspects are intertwined, since some backup techniques can lead to unintended release of private information, and unauthorised access to data can be used to destroy or corrupt it rather than to steal.

Heading	Guidance	Why is this required?
11. What is the nature of any security requirements?	If any particular security requirements apply, list them here along with the obligations they require. Such requirements may arise, for example, from legislation, from contractual obligations such as non-disclosure agreements, from a desire to protect an	This section should flag up whether there is a need for additional security precautions beyond what would be regarded as the University's standard level.

Heading	Guidance	Why is this required?
	innovation prior to patenting or because the research is potentially controversial.	
12. What are the main risks to data security?	List the potential risks which may arise from unauthorised access to your data. These risks will vary according to the nature of your research, and may already have been identified as part of ethical clearance or contract negotiations. Consider who might wish to access or damage your research data, for what reason and what harm might result. You may find it useful to also consider the likelihood and impact of each risk.	An understanding of the risks to data security, along with the relative danger that they pose and the likelihood of their being realised, is key to deciding what steps must be taken to protect that data. Some security techniques, such as encryption, bring their own risks, and cost both time and money, so should not be applied without careful consideration.
13. What measures will be taken to comply with these requirements and mitigate the risks?	Consider the probability and potential impact of the risks listed above. This may include both physical security measures (storing materials in a safe/locked filing cabinet, swipe-card access to buildings/labs) and technological measures (restricting file access, strong password policy, encryption, avoiding cloud storage such as Dropbox).	This will serve as a reference to all involved in the project, including support staff at the University.
14. To whom will access be granted / restricted?	List the people or roles, both within and outside the University, that will be allowed access to the data. If appropriate, mention what scope their access will be restricted to and what actions they should be permitted to take (e.g. view, edit, delete). Also specify which person or people will have the authority to grant additional access should the situation change in the future.	This will be useful when setting up storage services, and if security is particularly strict, will provide a reference for members of the team as to who is allowed access to what. It can also be used to show potential participants how widely their private contributions will be seen. It will also indicate whether there is a need for data to be shared outside the university during the life of the project – if so, this will need to be handled in some way.

Heading	Guidance	Why is this required?
15. How will resilience be guaranteed?	Describe how you will ensure your data remains available to you and uncorrupted during the project. It should be backed up regularly and both working copies and backups should be periodically tested to ensure they are free of errors and that the backups can still be restored. Care must also be taken to ensure that backups are not subject to the same or greater risk as working copies; for example, external hard drives are often stolen with the laptops they sit next to, and USB storage devices are usually not password protected. The simplest way to ensure resilience is to use the University's shared network storage system. If it is not possible to use the network storage, you should describe in full your resilience policy.	This will help you assess whether your data would still be safe and available to you in the event of an accident or disaster, and whether there might be simpler or more effective ways to guarantee this.

Data archival and preservation

There are many good reasons for archiving data after the end of a project. You may wish to continue using it for future projects, or make it available so that others can perform novel analyses (see also the following section on publication). You will probably need to refer to it when responding to queries about your publications, and this can often lead to new collaborations if handled well. There may also be legislative reasons why it must be kept for a minimum period.

Heading	Guidance	Why is this required?
16. What data must be retained, for how long and in what format?	Describe the elements of the data that will need to be retained after the end of the project. If the exact data that you will collect is not finalised, describe how you will select data for archival. This could be based, for example, on scientific or historic value, potential for re-use or cost of collection. At an absolute minimum, you should archive all data underlying published work. Record the length of any minimum retention period, which will depend on the policy of your funder and on	It is usually not useful to keep absolutely everything from a project. Setting out in advance what you do and do not need to keep, and for how long, will make these decisions much easier at the end of the project. Your chosen archive/data centre will also require this information so they can manage your data appropriately. This question will also flag up what conversions will need to

Heading	Guidance	Why is this required?
	legislative requirements. If the data must be disposed of after a specified period, note this here too. For each element of data, list an appropriate format for archiving it. This would usually be the format you would use when sharing data with colleagues in your discipline (even if you don't intend to share this particular data). Suitable formats will usually be open standards, and not tied to a particular piece of software. If possible the format should be readable using a simple text editor.	take place (if any) before your data can be archived. Proprietary file formats tied to specific software are usually preferable during a project (because they take advantage of the unique features of that software), but archived data must be protected against the risk of obsolete software.
17. How will long-term preservation be assured?	Specify where your data will be archived after the end of the project. In many disciplines there are well established national and international data centres, archives and databases. You may be able to find a suitable place to archive at the Registry of Research Data Repostories or DataCite. If possible, you should use one of these locations as they have funding and expertise to look after specialised data.	All publicly-funded research is now required to archive data for a minimum period after the end of the project (check with your funder for details), and publishers increasingly require archival of data associated with published articles. Your answer to this question will demonstrate how you intend to achieve this.
18. What metadata / documentation must be deposited with the data?	Describe here any documentation and other information, such as codebooks and details of provenance, which will be necessary for your data to remain useful into the future. This may be as simple as cross-referencing Data management, documentation and curation section of the plan, but check with your chosen archive for specific requirements.	Even if you don't share your data, archiving the information needed to interpret it will help future members of your research team to get up to speed quickly, and aid your collaborators. Most archives will only accept data that is suitable for reuse, which requires contextual information. Listing the requirements here will help you ensure you have all the documentation ready at the end of the project, saving you time.
19. Who will have long-term responsibility	In many cases, responsibility should pass to your chosen archive after the end of the project — contact the archive to	Recording this responsibility in advance will reduce the possibility of confusion later, and

Heading	Guidance	Why is this required?
for the data after the project ends?	check whether this is the case. If you choose to take long-term responsibility yourself, you will need to satisfy reviewers that you have adequate processes and infrastructure to do so.	will also help the University ensure that all data generated here is properly looked-after.

Data publication and access

This section will help you understand whether or not it is appropriate to share your data after the project, and in what form. Published data usually has its own persistent identifier, such as a DOI (Digital Object Identifier), and can be cited as part of the scholarly record. Data that has almost no on-going value when kept private can have enormous value when published, forming the basis of a meta-analysis or providing opportunities to validate novel analytical techniques.

Heading	Guidance	Why is this required?
20. Do you have an obligation or intention to publish / share your data?	Most major research funders have policies on what data must be shared at the end of a project, and many publishers require it as a condition of publication. Many researchers choose to share their data even if they are not obliged to. This question requires a simple yes/no answer, along with a reference to the source of the obligation if relevant.	This will serve as a reminder of any conditions attached to your funding.
21. What secondary use is intended or foreseeable?	Briefly describe any potential future uses of your data that you can foresee. This should include further use by you and your team as well as reuse by the wider academic community.	This will help you to understand what data could usefully be shared, who might be able to use it and how they might access it. Don't forget, though, that the most creative and valuable applications may be ones that require knowledge from a domain outside your own or that aren't yet possible.
22. What difficulties are expected in sharing?	Describe any difficulties which impose limitations on whether, how and when you can share your data. You should also explain what measures will be taken to overcome these difficulties. For example, sensitive data may be more widely shared if anonymised or summarised, and it may	This question is particularly important if you are unable to share any data, or if you intend to impose an embargo or restrictive license conditions. You must justify these restrictions by showing that

Heading	Guidance	Why is this required?
	only be possible to share a subset or processed version of a very large dataset.	there is no other reasonable way to overcome the difficulties.
23. What data will/will not be published as outputs from the project?	Describe which elements of your data will be of value to other researchers. This will be a subset of the data that you consider worth archiving. If no data is to be shared, you should justify clearly why not. It is important to demonstrate that you have considered ways of overcoming any potential difficulties (see below).	Setting out clearly which data you expect to share will mean that you (and your team) can manage it more carefully throughout the project without wasting effort on data which will never be shared. Remember that other researchers often wish simply to understand your publications better (which would lead to more citations and potentially to collaboration), and not always to use the data for their own research.
24. How will potential users discover the data / metadata?	Describe the combination of methods you will use to ensure that data you share will be found by others. Possible ways of doing this include: deposit in an established specialist archive or database in your discipline; submitting to a journal as supplementary information with a paper, or to a dedicated data journal; deposit in a self-archiving system such as or figshare; dissemination via a personal or project website; personal networking at conferences.	There are many benefits to sharing your data, but it takes some effort, and those benefits will not be realised if no-one can find that data. In addition, easily-findable data creates extra opportunities for potential collaborators, students, employees and employers to learn about you. Planning how it can be discovered will help ensure you get the most value from your data (and your publications).
25. How will continued access be guaranteed?	Describe how those who wish to access your data will be able to do so. Many of the options for discoverability mentioned above, as well as many archiving options (see section on Data archival and preservation) also provide for access. If data is made available only by request to you (as the corresponding author on a paper, for example), you should describe how you will ensure you continue to be	This will flag up any preservation actions you need to take near to or after the end of the project.

Heading	Guidance	Why is this required?
	able to receive and respond to such requests. Bear in mind that, should you leave the University of Amsterdam, your University email account will be discontinued. Web links should be checked periodically to ensure they still work as intended.	
26. When will the data be published?	State and justify any embargo on the release of data. It is accepted in most research areas that even if you do plan to share your data you can expect a reasonable period of exclusive use, to give you time to publish or to file patents for example. Most research funders impose limits on this and you should have a strong justification for any embargo.	This will be required when submitting to an archive, to ensure they do not make your data public prematurely. It is also important to discuss this when sharing data privately, to ensure those requesting your data do not release it before the embargo has expired.
27. What license conditions and other restrictions will apply?	Make sure copyright is applicable – data are not always or automatically subject to copyright, while a database could be protected by database right.[1] If copy- or database right is applicable, specify what terms the data will be released under. You can either indicate the license which applies to your data when it is published, or negotiate an *ad hoc* agreement in response to particular requests. You can also use a dual licensing strategy, permitting some rights automatically to all users and agreeing additional rights for specific users (such as collaborators or commercial companies) on request.	Potential users of your data cannot own it; they can only use it under license, according to the terms in that license. You will need to know what license you wish to use when you publish your data, and should be prepared to discuss this when responding to private requests. Setting this out here will save time and confusion later, especially if more than one of the original researchers might want to pass on the data.

Roles, responsibilities and resourcing

This section will summarise who is responsible for carrying out the plan and how any costs will be met.

[1] See for more information the SURF report *The legal status of raw data: a guide for research practice* (2009) or contact RDM Support.

Heading	Guidance	Why is this required?
28. Which named roles or individuals have specific data management responsibilities?	Under the University's RDM Policy the PI has ultimate responsibility for research data management. Set out how this responsibility will be delegated, to research team members, technicians and professional services within the university. If delegating outside your immediate research team, be sure to discuss this with the people and services involved.	Explicitly listing roles and responsibilities will reduce the chance of misunderstandings. It will also flag up the need for involvement by those outside your immediate team.
29. What relevant policies at the University of Amsterdam and partner institutions apply?	This may include: • UvA Research Data Management policy (to be published) • VSNU Nederlandse Gedragscode Wetenschapsbeoefening (2004/20014, pdf) / Netherlands code of conduct for scientific practice (2004/2014) • VSNU Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek (2005, Dutch only) • UvA IT Security policy (Dutch only, UvAnetID login required) • Any standing department, group or personal data management policy or plan which has been published • The research data policy of your funder • Any collaboration or non-disclosure agreement relating to the project • Any similar policies applying to your partners	This will provide a quick reference to check on any data-related obligations as the project proceeds.
30. How will adherence to this plan be checked and / or demonstrated?	You may wish to briefly discuss your data management plan in interim or final reports, or to cover it periodically in project team meetings. This should not be onerous, and should fit with the way you and your project team work.	Some funders explicitly require data management to be mentioned in final reports. Even if not explicitly required, it will reflect well on you when making future applications.
31. When will implementation of this plan be reviewed and by whom?	Describe how often this plan will be reviewed and who will be responsible for doing so.	Your plan will provide no value at all if it is simply written and then filed away. It will need to be checked periodically so that you can be sure that it is still being followed and is still fit for purpose.

Heading	Guidance	Why is this required?
32. What resources will be required to implement this plan?	List whatever extra staff time, training, funding for services and other costs you have identified that are necessary to deliver the plan. If you have identified costs for archiving the data or preparing it for archive, you should allow time to complete these activities so that the cost falls within the project lifespan.	This section will flag up RDM-related costs which will need to be included in the project budget. Some resources will be available from the University, but those that are not will need to be funded. Most funders regard data management costs as allowable as long as they are justified and incurred within the life of the project.