**CONSTRUCTING A DATA MANAGEMENT PLAN**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

 If your research is empirical, you are likely to collect data. Data might include survey responses, biological samples, mineral samples, video recordings, interviews, photographs, simulations, the notebooks you develop, and so forth. Before you collect data, you should develop a research data management plan with your supervisors. This plan stipulates which individuals or organizations own the data and how the data will be organized, stored, shared, and discarded.

 Typically, this plan is written in a document, such as Microsoft Word file. The plan is stored alongside other important documents, such as ethics applications, funding agreements, and collaboration agreements. The researchers, including the candidates, will tend to revisit and refine this document every few months. For other resources, consider

* free software that can help you construct data management plans, Google “Research Data Management Organiser”
* the links from this [CDU webpage](https://libguides.cdu.edu.au/c.php?g=168003&p=1103349) especially the course called Research Data MANTRA

|  |
| --- |
| **Why should I create a research data management plan?** |

 All researchers are expected to construct a data management plan. These plans generate several benefits. Specifically, these plans

* diminish the likelihood of disputes with colleagues, collaborators, or institutions around ownership, access, and similar matters
* can increase the likelihood that researchers attract funds, because reviewers often assess the data management plan: Reviewers strive to ascertain how the researchers will disseminate their data and finding as widely as possible to amplify the impact of their work
* improve budget forecasts: If researchers do not develop a data management plan, they might overlook some of their research costs
* decreases the likelihood that researchers will misplace or misconstrue the data

|  |
| --- |
| **Main contents of this data management plan** |

 In practice, candidates should develop this data management plan in consultation with their supervisors. Nevertheless, the following table could help candidates construct this plan. The first column outlines the key matters this data management plan should address. The second column imparts some insights on how to address these matters

|  |  |
| --- | --- |
| Key matters or topics | Relevant insights |
| Which existing data—data that were collected before the project began—need to be stored and managed for this project | * For example, a candidate might want to use data or information they had acquired before their enrolment
* Researchers should advise the university of these data as soon as possible; otherwise, the university might be granted ownership of these data.
 |
| Which data will be collected or created during the project?  | * Data comprise both the original measures or observations as well as modified, processed, or analyzed measures or observations.
* For example, suppose participants were asked to specify the nation in which they live. Data includes both the responses of participants to this question as well as information that was derived from this question—such as the continent in which participants live
 |
| Which hardware and software will be needed to collect and process the data? |  |
| Who owns the data? | * To clarify who owns the data, first consult the Intellectual Property Policy of CDU
* In general, all data that is collected by staff members of CDU, as part of their employment, are owned by CDU.
* In contrast, the data collected by candidates of CDU are usually owned by the candidate.
* Nevertheless, other agreements might override these regulations. Furthermore, in particular circumstances, the Pro Vice Chancellor of Research can override these regulations.
 |
| In what format will the data be stored—such as which file formats? | * Consider formats that are more likely to be accessible and engaging to relevant audiences. For example, data might include graphical representations of the data
* Data should be de-identified whenever possible
 |
| How will you protect the data—to preclude illicit access | For example* files or storage devices may be password protected
* the data may be encrypted somehow
* the data may be anonyomous; or pseudoynms may be used
 |
| Which meta-data will be retained  | * That is, construct a document that delineates the data—such as the amount of data, the location, date, and time of data collection, the fields or categories of data, and contact information
* This document should also enumerate each data file—and specify how each file can be accessed.
 |
| Who can access the data? | * Often, only the researchers who need to analyze or monitor the data should be able to access these data before this project is completed.
* This information must be specified in the plain language statement that participants receive
 |
| How will this access be restricted to the relevant individuals? | * For example, consider how the necessary passwords are restricted to the relevant people.
 |
| How will the data, such as data files and folders, be organized |  |
| How will data be synchronized, coordinated, shared, and managed across machines and researchers? How will the versions of these data be monitored and maintained correctly? | To manage the data, you might * use a spreadsheet
* use a paper or electronic log book
* specialised data management software
 |
| Where will the data be stored during the project and in the future?  | * Whenever possible, to conserve space, data should be digitized. For example, some pictures could be scanned and then destroyed.
* These data should be stored on the available university facilities, such as Onedrive or eSpace
* To access eSpace, proceed to http://espace.cdu.edu.au
* If data is stored at another organization, assess whether or not this storage complies with university policies.

You might utilise a range of devices to store data, such as* hard disk drives—on a university or personal computer
* external hard drives, memory USB sticks, CD, or DVD
* network storage managed by the university
* university or personal email accounts
* cloud services, such as Dropbox, Google Drive, or Cloudstor
* CDU OneDrive

In the future, after you complete the study, you might store the data in * the CDU data repository
* general data repositories, such as figshare

If possible, store data in the CDU data repository. You can then be certain the data will be accessible to you indefinitively. Furthermore, the dataset is more likely to comply with the FAIR data principle and be* findable—with appropriate search terms
* accessible to the public
* interoperable—in a language that any system can recognise, and
* reusable—so that other researchers can perhaps reanalyse the data if needed

 To store data in the CDU data repository, you could select “RIS” from the portal.cdu.edu.au. Then, after you press the downwards arrow next to “Personal”* locate a button called “New” under “Datasets”
* once you press this button, answer the relevant questions to store your data
 |
| Who will manage the storage of data? | * Typicallty, candidates will manage the storage of data, unless these individuals are working on a funded project
* If the individuals who manage the storage of data leave the university, they might be able to transfer the data to another location—unless this transfer violates other rights or contracts
 |
| How will the data be backed-up? How often will the data be backed-up? | * If possible, utilize a system that backs-up data automatically
* The back-up storage should be housed separately from the main storage of data
* Portable devices such as external hard drives or USB sticks should not be used to store data over extended periods, because these devices can be corrupted or misplaced.
 |
| Which data will be shared? To whom will these data be shared?  | * If your data are widely shared with other researchers, papers that report these data are more likely to be cited widely, according to some studies
* Data may not be shared only with other researchers but also with policy makers, not-for-profit agencies, commercial organizations, and the general public.
* Some journals prefer or oblige authors to share their data. Explore the preferences and policies of the journals to which you plan to submit.
 |
| How will these data be shared? | * This information is discussed in another document on this CDU webpage
 |
| When and how will data be discarded? | * In general, data should be retained at least five years after the research ends and five years after the research is published

Some exceptions should be acknowledged. For example* If the research is a clinical trial on humans—and thus explores the efficacy of some medical treatment—data must be retained at least 15 years after the research is published
* If the research entails modern clinical therapies, such as gene therapy, the data may need to be retained permanently rather than deleted
* If the work affords significant community or heritage value, the data should also be retained permanently
* If the results of some research are disputed or challenged, the data must be retained until the matter is resolved.
* If the results could be high in public interest, commercial interest, or controversy, the data could also be retained for a longer period.
* Funding bodies, commercial sponsors, and specific institutions might also inform these decisions on when data should be discarded.

The procedure that is used to destroy data must not permit recovery of these data later.  |
| Who is responsible for monitoring whether these plans are fulfilled? | * On regular occasions—perhaps several times a year after data collection commences— the principle supervisor should evaluate whether the candidate has complied with these data management plans.
 |
| What will be the cost of these activities to organize, store, share, and discard the data cost?  | * Include the costs of hardware to store data
* Include the costs of research assistants to manage data
 |
| Do you need to consider anything else about data management?  |  |

|  |
| --- |
| **Principles to consider** |

**Which data should I retain?**

Often, you will not be able to retain all the primary data. For instance, you might not retain biological materials, written surveys, or actual recordings. Instead, you might choose to retain the records or information that were derived from these primary data. Nevertheless, you should be more inclined to retain data if

• the data set or setting is especially unique and cannot be replicated in the future

• the data set is especially reliable, valid, and thus usable

• some collection, society, or research initiative would be especially interested in these data

• the data is especially valuable to a community or culture

• the data could generate economic benefits

**Management of confidential data**

 Some data needs to be confidential. For example, data that should be confidential includes

• data that is commercial in confidence

• data in which confidentiality was promised to participants

• data about secret or sacred cultural practices

• data that could be misused, such as the locations of vulnerable species

• data that is regarded as sensitive, according to privacy acts, such as health information

• data or information that is subject to classification regimes and other controls, such as national security information, police records, or materials subject to export controls

In these circumstances, you need to devote special care to preserve the confidentiality of these data. You could even conduct a risk assessment procedure to prevent breaches.