**Data Management Plan Guidance and Templates**

**Table of Contents**

A. What is a Data Management Plan (DMP)?.................................................................................1

B. Getting started with your DMP...................................................................................................1

C. Sections of a DMP.......................................................................................................................2

D. Agency Guidelines and VCU Resources....................................................................................8

E. NSF DMP Template....................................................................................................................8

**A. What is a Data Management Plan (DMP)?**

A Data Management Plan (DMP) is a written document detailing data that will be acquired or produced during a research project; how the data will be managed, analyzed, and stored; and how the data will be handled, protected, shared, and preserved during and after the project has been completed.

**B. Getting started with your DMP.**

The questions below provide a starting point for beginning to think about and formulate a DMP:

* + What is the purpose of the research?
  + What is the data? How and in what format will the data be collected? Is it numerical data, image data, text sequences, or modeling data?
  + How much data will be generated for this research?
  + How long will the data be collected and how often will it change?
  + Are you using data that someone else produced? If so, where is it from?
  + Who is responsible for managing the data? Who will ensure that the data management plan is carried out?
  + What documentation will you be creating in order to make the data understandable by other researchers?
  + Are you using metadata that is standard to your field? How will the metadata be managed and stored?
  + What file formats will be used? Do these formats conform to an open standard and/or are they proprietary?
  + Are you using a file format that is standard to your field? If not, how will you document the alternative you are using?
  + What directory and file naming convention will be used?
  + What are your local storage and backup procedures? Will this data require secure storage?
  + What tools or software are required to read or view the data?
  + Who has the right to manage this data? Is it the responsibility of the PI, student, lab, or funding agency?
  + What data will be shared, when, and how?
  + Does sharing the data raise privacy, ethical, or confidentiality concerns?  Do you have a plan to protect or anonymize data, if needed?
  + Who holds intellectual property rights for the data and other information created by the project? Will any copyrighted or licensed material be used? Do you have permission to use/disseminate this material?
  + Are there any patent- or technology-licensing-related restrictions on data sharing associated with this grant?
  + Will this research be published in a journal that requires the underlying data to accompany articles?
  + Will there be any embargoes on the data?
  + Will you permit re-use, redistribution, or the creation of new tools, services, data sets, or products (derivatives)? Will commercial use be allowed?
  + How will you be archiving the data? Will you be storing it in an archive or repository for long-term access? If not, how will you preserve access to the data?
  + Is a discipline-specific repository available?
  + How will you prepare data for preservation or data sharing? Will the data need to be anonymized or converted to more stable file formats?
  + Are software or tools needed to use the data? Will these be archived?
  + How long should the data be retained? 3-5 years, 10 years, or forever?

**C. Sections of a DMP.**

***1) Roles and Responsibilities***

* List staff/organizational roles and their responsibilities for carrying out the data management plan (DMP); name specific people where possible. Include a description of time allocations, training requirements, and contributions of non-project staff, as appropriate.
* Indicate when and/or how often adherence to your DMP will be checked and/or demonstrated. Include the names of the person(s) responsible for adherence to your DMP.
* Indicate who/which roles will assume responsibility for carrying out the DMP should personnel changes occur or if the PI leaves the institution. Describe the process for transferring responsibility for the data.
* Indicate who will have primary responsibility for how the data will persist over time when the original personnel are no longer associated with the project.

*Text Examples:*

* "The project will assign a qualified data manager certified in disclosure risk management to act as steward for the data while they are being collected, processed, and analyzed."
* "All research data collected as part of this project is owned by the University. The Principal Investigator of this project will take responsibility for the collection, management, and sharing of the research data."
* "Day-to-day quality assessment will be the responsibility of the Lab Director who in turn is overseen by the Project Director."

***2) Costs***

* Check with the funding agency to determine where in the proposal to include costs related to data management.
* Include any costs for data management services.
* Include any anticipated income from licensing data.
* Include any costs for managing data during the course of the project as well as after the project is complete.

*Text Examples:*

* "Staff time has been allocated in the proposed budget to cover the costs of preparing data and documentation for archiving. The [repository] has estimated their additional cost to archive the data at [insert dollar amount]. This fee appears in the budget for this application as well."
* "The cost model is twice the current cost of storage. At $1,850/usable TB, costs are estimated at $3,700/usable TB for the storage hardware for indefinite data retention."

***3) Types of Data***

* Provide a short description of the data that will be generated in the research project (e.g., samples, physical collections, software, curriculum materials, and other materials to be produced during the course of the project). Include an estimate of the amount of data and content of the data (if possible).
* Describe the data types will you be creating or capturing (e.g. experimental measures, observational or qualitative, model simulation, processed etc.).
* Describe how the data will be created or captured.
* Indicate if you will be using existing data and describe the source of that data. Also describe the relationship between the data you are collecting and the existing data that you are integrating into the project.

*Text Examples:*

* "The associated data types will be captured using X survey software and analyzed using X data analytics tool."
* "Over the course of the project, data will be collected and entered into two relational databases."
* "Over the course of the project, data will be generated from sensors and recorded in X format."
* "This project will produce public-use nationally representative survey data for the United States covering Americans' social backgrounds, enduring political predispositions, social and political values, perceptions and evaluations of groups and candidates, opinions on questions of public policy, and participation in political life."
* "This project will generate data designed to study the prevalence and correlates of DSM III-R psychiatric disorders and patterns and correlates of service utilization for these disorders in a nationally representative sample of over 8000 respondents. The sensitive nature of these data will require that the data be released through a restricted use contract."
* "Few datasets exist that focus on this population in the United States and how their attitudes toward assimilation differ from those of others. The primary resource on this population, [give dataset title here], is inadequate because..."
* "Data have been collected on this topic previously (for example: [add example(s)]). The data collected as part of this project reflect the current time period and historical context. It is possible that several of these datasets, including the data collected here, could be combined to better understand how social processes have unfolded over time."
* "For quantitative data files, the [repository] ensures that missing data codes are defined, that actual data values fall within the range of expected values and that the data are free from wild codes. Processed data files are reviewed by a supervisory staff member before release."

***4) Data Formats and Metadata***

* Indicate which file formats you will use for your data, and why you will use those formats.
* Describe any contextual details (metadata) that are necessary to make the data you capture or collect meaningful to you and others.
* Describe how you will create or capture these contextual details.
* Describe the form that the metadata will take (i.e., which metadata standards, if any, will be used). Where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies.
* Explain why you have chosen particular standards and approaches for metadata and contextual documentation (e.g., recourse to staff expertise, Open Source, accepted domain-local standards, widespread usage).
* Describe the procedures that will be used to ensure data quality.

*Text Examples:*

* "Research data will be stored using X file formats. Related files in different formats will be linked by file naming conventions, e.g...”
* "Metadata will be generated to describe the data generated in X format and will be stored alongside the data. X metadata standards will be applied during the creation of the metadata."
* "Data will be stored in a CVS system and checked in and out for purposes of versioning. Variables will use a standardized naming convention consisting of a prefix, root, and suffix system. Separate files will be managed for the two kinds of records produced: one file for respondents and another file for children with merging routines specified."
* "Data will conform to best practices and standards from the X community."
* "Internal calibration (for geophysical data), instrument calibrations, duplicate samples and field blanks (for hydrochemical data) will be recorded and tested against collected/recorded data to ensure their validity. Qualitative descriptions (lithological data) will be validated through comparative descriptions of collected materials."
* "The clinical data collected from this project will be documented using CDISC metadata standards."
* "Digital video data files generated will be processed and submitted to the [repository] in MPEG-4 (.mp4) format."

***5) Access, Sharing, and Privacy***

* Describe which data will be shared and how you will make the data available including any resources needed such as equipment, systems, expertise, etc.
* Indicate when you will make the data available (including any factors such as embargo periods for political, commercial, patent reasons, or complying with publishing policies).
* Describe the process for gaining access to the data. Be aware that some journals require data access upon publication. Indicate if you will use a repository, if the data will be available as supplementary files with the publication, or if the data will be available upon request.
* Indicate if you anticipate that the original data collector, creator, or principal investigator retains the right to use the data before opening it up to wider use.
* Indicate if there are any ethical and privacy issues related to sharing the data. If so, describe how those will be resolved if the data is shared (e.g., removing any personally identifying information in the data, working with institutional ethical committees, resolving potential conflicts by way of formal consent agreements).
* If applicable, describe what you have done to comply with your obligations in your IRB Protocol.
* Describe how the dataset will be licensed if rights exist (e.g., list any restrictions or delays on data sharing needed to protect intellectual property, copyright or patentable data).

*Text Examples:*

* "Data will be posted on a website within three months of the grant closing. Data will be contributed to X public database. Data will be submitted to supplementary materials sections of peer-reviewed journals."
* "Data will be available and cited in publication. Researchers will be able to contact the PI for access to data. Data will be maintained in an open XML format to enable open re-use of the data."
* "The main output from this project is field data. We recognize that these data are the property of X and hence we will be asking their permission to license these data to Y for use in their exploration program."
* "Our project will generate a large volume of data, some of which may not be appropriate for sharing since it involves a small sample that is not representative. The investigators will work with staff of the [repository] to determine what to archive and how long the deposited data should be retained."
* "X and third party copyright will be protected. The PI will be responsible for ensuring that all project members are aware as to the ownership of data and who may access them and under what conditions. Online access to the data will be password protected."
* "There is an agreement regarding the right of the original data collector, creator, or PI for first use of the data. The specified embargo period associated with the data being submitting extends from [date] until [date]. The embargo will be lifted by [date]."
* "This project will generate data linked to administrative records, so the data will be distributed through a restricted data use agreement managed by [repository]. Through this mechanism, users will apply to use these files, create data security plans, and agree to other access controls."
* "The principal investigators on the project and their institutions will hold the intellectual property rights for the research data they generate but will grant redistribution rights to [repository] for purposes of data sharing."
* "Our research group has been trained in human subjects protection and only trained project staff operating under the IRB approval for the project will have access to the confidential individually identifiable data, and all data will be aggregated or anonymized for publication."
* "The following language will be used in the informed consent: The information in this study will only be used in ways that will not reveal who you are. You will not be identified in any publication from this study or in any data files shared with other researchers. Your participation in this study is confidential. Federal or state laws may require us to show information to university or government officials [or sponsors], who are responsible for monitoring the safety of this study."
* "For this project, the principal investigators will request expedited IRB review compliant with procedures established by the [University] campus IRB. Research activities envisioned present no more than minimal risk to human subjects."
* "During data analysis, the data will be accessible only by certified members of the project team. The research project will remove any direct identifiers in the data before deposit with [repository]."

***6) Policies and Provisions for Re-Use***

* Indicate if any permission restrictions need to be placed on the data.
* Describe which communities/groups are likely to be interested in the data.
* Describe the intended or foreseeable uses and users of the data.
* Indicate if there any reasons not to share or re-use data (e.g., ethical, non-disclosure, etc.).

*Text Examples:*

* "The data gathered will use a copyrighted instrument for some questions. A reproduction of the instrument will be provided to [repository] as documentation for the data deposited with the intention that the instrument be distributed under "fair use" to permit data sharing, but it may not be re-disseminated by users."
* "The project team will create a dedicated Web site to manage and distribute the data because the audience for the data is small and has a tradition of interacting as a community. The site will be established using a content management system like Drupal or Joomla so that data users can participate in adding site content over time, making the site self-sustaining. The site will be available at an .org location. For preservation, we will supply periodic copies of the data to [repository]. That repository will be the ultimate home for the data".
* "Users of field data should acknowledge and/or offer co-authorship to the investigators who collected the data."
* "The data to be produced will be of interest to demographers studying family formation practices in early adulthood across different racial and ethnic groups."
* "In addition to the research community, we expect these data will be used by practitioners and policymakers."

***7) Data Storage and Preservation***

* Describe the long-term strategy for maintaining, preserving and archiving the data. Describe the procedures that your intended long-term data storage facility has in place for preservation and backup.
* Indicate which [archive, repository, central database, or data center](https://www.lib.ncsu.edu/data-management/repos) you have identified as a place to deposit the data.
* Describe the transformations that will be necessary to prepare data for preservation and/or data sharing (e.g., data cleaning, normalization, or removing personally-identifying information where appropriate).
* Describe the metadata and/or documentation that will be submitted alongside the data or created when the data is deposited in order to make the data discoverable and/or reusable.
* Describe the set of conventions that you will use for naming data files and folders, and how you intend to manage multiple versions of files.
* Indicate any related information that will be deposited (e.g., references, reports, research papers, fonts, the original bid proposal, etc.).
* Indicate how long will/should the data be kept beyond the life of the project. Many grant funders suggest that the minimum data retention period for research data should be 3 years after conclusion of the grant award or 3 years after the data is released to the public (whichever is later).

*Text Examples:*

* "The research data from this project will be deposited with the institutional repository on the grantees' campus."
* "The research data from this project will be deposited with [repository] to ensure that the research community has long-term access to the data."
* "By depositing data with [repository], our project will ensure that the research data are migrated to new formats, platforms, and storage media as required by good practice."
* "In addition to distributing the data from a project Web site, future long-term use of the data will be ensured by placing a copy of the data into [repository], ensuring that best practices in digital preservation will safeguard the files."
* "[Repository] will place a master copy of each digital file (i.e., research data files, documentation, and other related files) in Archival Storage, with several copies stored at designated locations and synchronized with the master through the Storage Resource Broker."
* "The data will be processed and managed in a secure non-networked environment using virtual desktop technology."
* "The data files from this study will be managed, processed, and stored in a secure environment (e.g., lockable computer systems with passwords, firewall system in place, power surge protection, virus/malicious intruder protection) and by controlling access to digital files with encryption and/or password protection. De-identifed files will be deposited with [repository] whose security policy has been written according to best practices."
* "Our research project will generate data from a large national sample. These data will be retained by [repository] as part of their permanent collection."

**D. Agency Guidelines and VCU Resources**

Many funding agencies require a DMP with every proposal and each agency or directorate has individual policies for data management. Visit the links below to access specific guidance or visit [SPARC](http://datasharing.sparcopen.org/compare?ids=). The Scholarly Publishing and Academic Resources Coalition (SPARC) has compiled an excellent resource with information about the data management and data sharing requirements from all the federal funding agencies.

* [Department of Defense](http://www.dtic.mil/dtic/pdf/dod_public_access_plan_feb2015.pdf)
* [Department of Energy](http://www.energy.gov/downloads/doe-public-access-plan)
* [National Institutes of Health](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm)
* [National Science Foundation](http://www.nsf.gov/bfa/dias/policy/dmp.jsp)

VCU provides resources and guidance for faculty on the handling, transmission, and storage of information and data. Visit the [VCU Data Management System](https://dms.vcu.edu/) for assistance. Faculty can also create ready-to-use data management plans for specific funding agencies using the [DMPTool](https://guides.library.vcu.edu/dmptool).

**E. NSF DMP Template**

**Data Management Plan**

**Types of data produced**

[Enter content here, then remove the Guidance prior to submission]

**Guidance:**

Types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.

Provide a description of the data you will collect or re-use, including the file types, dataset size, number of expected files or sets, and content. Data types could include text, spreadsheets, images, 3D models, software, audio files, video files, reports, surveys, patient records, etc. Consider the following:

* What data will be generated in the research?
* What data types will you be creating or capturing?
* How will you capture or create the data?
* If you will be using existing data, state this and include how you will obtain it.
* What is the relationship between the data you are collecting and any existing data?
* How will the data be processed?
* What quality assurance & quality control measures will you employ?

**Data and metadata standards**

[Enter content here, then remove the Guidance prior to submission]

**Guidance:**

Standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

Describe the format of your data and how it will be documented. Think about what details (metadata) someone else would need to be able to use these files. For example, you may need a "readme file" to explain variables, structure of the files, etc. Consider these questions:

* Which file formats will you use for your data, and why?
* What form will the metadata describing/documenting your data take?
* How will you create or capture these details?
* Which metadata standards will you use and why have you chosen them? (e.g. accepted domain-local standards, widespread usage)
* What contextual details (metadata) are needed to make the data you capture or collect meaningful?

**Policies for access and sharing**

[Enter content here, then remove the Guidance prior to submission]

**Guidance:**

Policies for access and sharing; Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

Explain how and when the data will become available. Will data be accessible on a web page, by email request, via open-access repository etc.? If there is an embargo period for sharing the data, make sure you provide details explaining this delay (e.g. publisher, political, commercial, patent reasons). And if the data is of a sensitive nature - human subject concerns, potential patentability, species/ecological endangerment concerns - that public access is inappropriate, address here the means by which granular control and access will be achieved (e.g. formal consent agreements; anonymizing data; restricted access, only available within a secure network). Consider these questions:

* How will you make the data available? (Include resources needed to make the data available: equipment, systems, expertise, etc.)
* When will you make the data available?
* What is the process for gaining access to the data?
* Will access be chargeable?
* How long will the original data collector/creator/principal investigator retain the right to use the data before making them available for wider distribution?
* Are there any embargo periods for political/commercial/patent reasons? If so, give details.
* Are there ethical and privacy issues? If so, how will these be resolved?
* What have you done to comply with your obligations in your IRB Protocol?
* Who will hold the intellectual property rights to the data and how might this affect data access?

**Policies for re-use, redistribution**

[Enter content here, then remove the Guidance prior to submission]

**Guidance:**

Policies and provisions for re-use, re-distribution, and the production of derivatives.

Explain how the policies outlined in the previous question can be applied to the re-use and re-distribution of your data. Identify who will be allowed to use your data, how they will be allowed to use your data and whether or not they will be allowed to disseminate your data. If you are planning on restricting access, use or dissemination of the data, you must explain in this section how you will codify and communicate these restrictions. Consider the following:

* Will any permission restrictions need to be placed on the data?
* Who is likely to be interested in the data?
* What and who are the intended or foreseeable uses the data?

**Plans for archiving & preservation**

[Enter content here, then remove the Guidance prior to submission]

**Guidance:**

Plans for archiving data, samples, and other research products, and for preservation of access to them.

Provide a description of your long-term strategy for archiving and preserving the data you plan to generate/use. Consider the following:

* What is the long-term strategy for maintaining, curating and archiving the data?
* Which archive/repository/database have you identified as a place to deposit data?
* What procedures does your intended long-term data storage facility have in place for preservation and backup?
* How long data will/should be kept beyond the life of the project?
* What data will be preserved for the long-term?
* What transformations will be necessary to prepare data for preservation / data sharing?
* What metadata/ documentation will be submitted alongside the data or created on deposit/ transformation in order to make the data reusable?
* What related information will be deposited?

***This document was prepared using content adapted from:***

* NC State University
  + [Elements of a Data Management Plan](https://www.lib.ncsu.edu/data-management/how_to_dmp)
  + [Data Management Plan Examples](https://www.lib.ncsu.edu/data-management/dmp_examples)
* [University of Virginia](https://data.library.virginia.edu/data-management-plan-templates/)
* [MIT](https://libraries.mit.edu/data-management/plan/write/)