NITC DATA MANAGEMENT PLAN – GUIDELINES AND CONTENT

 *Adopted from* <https://nsf.gov/eng/general/ENG_DMP_Policy.pdf>

# Overview

Under federal guidelines, Principal Investigators (PIs) are required to submit a Data Management Plan (DMP) **maximum of two pages** for all NITC research proposals. This plan must address the management of all digital data that result fully or in part from activities supported by NITC. Below are important points to consider when developing your DMP. The subsequent section will provide detailed guidelines for each DMP component. A DMP template is included at the end of the document.

## What data are included?

Research data are formally defined as “the recorded factual material commonly accepted in the scientific community as necessary to validate research findings” by the U.S. Office of Management and Budget (1999). The DMP should focus on digital data, because these data will be archived and made available to the public as required under DOT’s Public Access Plan.

The basic level of digital data to be archived includes (1) analyzed data and (2) the metadata that define how these data were generated. In other words, these are data that are or that should be published in theses, dissertations, refereed journal articles, supplemental data attachments for manuscripts, books and book chapters, and other print or electronic publication formats.

• Analyzed data are (but are not restricted to) digital information that would be published, including digital images and tables of the numbers *used* for making published graphs or summary tables.

• Necessary metadata are (but are not restricted to) descriptions or suitable citations of experiments, apparatuses, computational codes, and computer calculation input conditions.

## What data are not included at the basic level?

Not included in the basic level are preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues (Office of Management and Budget statement, 1999). Raw data fall into this category as “preliminary analyses.”

## What about issues restricting data sharing?

The DMP should include any data-management issues as well as the conditions that might prevent or delay the sharing of data.

Some proposals may use restricted data. For example, projects may purchase data from data vendors who legally prohibit PIs from sharing data on an open platform or involve agencies or research centers that place restrictions on data sharing. If this is the case, PIs will need to provide details about the data source and data sharing restrictions in the DMP.

PIs are required to submit data and metadata with the project’s final report. However, PIs may request a proprietary period for pending or planned publications, and NITC’s research administrator will withhold the project’s data product from public release from the time the project’s final report is published on NITC’s website until the end of the approved proprietary period.

Projects collecting sensitive information will be required to comply with university policies governing the collection and handling of sensitive information (e.g., personal identifiers), consistent with Institutional Review Board (IRB) procedures and in accordance with relevant federal and state privacy laws and regulations. If sensitive information is collect, investigators will be responsible for de-identifying data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset prior to submitting data for archiving to NITC.

Any additional exceptions to the basic data-management policy should be discussed with the NITC Research Administrator before submission of such proposals.

# Contents of the Data Management Plan

The DMP should clearly outline the rights and obligations of all parties as to their roles and responsibilities in the management and retention of research data. *The plan must also consider changes to roles and responsibilities that will occur should a PI or Co-PI leave the institution.* The activity of maintaining data and summarizing the data product should be included in the proposal. Associated costs should also be included in the project’s budget (see data submission requirements below to estimate time needed to prepare data for submission). Portland State University (PSU) will provide long-term storage for submitted data and metadata. Alternative data archives for large datasets will need to be approved by NITC and are handled on a case-by-case basis. Specific components that need to be included in the DMP are listed below.

1. Title of the data collection project, PI, Co-PI(s) and Institution(s).
2. Describe the types of data, samples, and other materials to be produced in the course of the project.
3. Standards to be used for data and metadata format and content.
4. Methods and policies for providing access and enabling sharing.
5. Policies for re-use, re-distribution, and the production of derivatives.
6. Methods for archiving and preserving access to data and materials.

Descriptions along with sample text for the specific sections of the DMP can be found in the template at the end of this document.

# Submitting Data Products

Please submit the data product for your project with the finalized version of your project’s final report. NITC will not be able to publish the final report and closeout the project until the data product and metadata are received. Metadata must provide adequate detail to ensure that the data product, all of its components, and its context are fully described.

In addition, a word document that includes the following information must be included:

* Title (e.g., Data from: “Name of Final Report”)
* Abstract summarizing the study, methodology, & findings
* Name of PI and Co-PI(s) and Institutions
* Description of dataset (e.g., These data support a final report published on NITC’s website “Title of Project” (2016))
* Instructions for how to use the enclosed tools or dataset
* A list of files that are included with data product that specifies the data format and provides a one sentence description of the data or information contained within each file
* Copyright and proprietary restrictions (if appropriate)
* Disclaimers (if appropriate)
* Recommended Citation (e.g., Dill, Jennifer and Hagedorn, Hau, “Data from: Title of project” (2017)

# Convention for Naming Files and Citations

To streamline the data archiving process across projects, we ask you to adhere to the following naming convention:

* Files included with each data product will need to specify the project identification number (e.g., NITC762), the content of the file (e.g., Data), and a number designation based on the total number of files submitted with each data product (e.g., NITC762Data1.doc, NITC762Data2.xlsx).
* Identify closely related files with the same number and an additional letter (e.g., NITC762Data2A.xlsx, NITC762Data2B.txt).
* In cases where data are downloaded from publically available regional, state, or national data repositories (e.g., US Census Bureau, EPA, US National Climate Data, Office of the Treasurer-Tax Collector), investigators are responsible for citing the source(s) of the data appropriately. For example, the following acknowledgement is suggested: Data were downloaded from the Bike-Ped Archive maintained by PSU and can be accessed through the PSU website (http://bp.its.pdx.edu/).

# References and Additional Resources

* Office of Management and Budget, Circular A-110, September 30, 1999. White House Website, OMB Home. <http://www.whitehouse.gov/omb/circulars/a110/a110.html>
* PSU Library: Manage Your Research Data: Documentation & Metadata (includes Dublin Core Metadata example). <http://guides.library.pdx.edu/data/metadata>
* DataONE: metadata resources. <https://www.dataone.org/best-practices/metadata>
* Cornell University: Guide to writing “readme” style metadata. <https://data.research.cornell.edu/content/readme>

NITC Data Management Plan Template

Please submit your DMP along with your proposal on-line at: <http://ppms.trec.pdx.edu>. For further information or questions, please contact Brendan Williams, brendan.williams@pdx.edu.

## Title of the data collection project, PI, Co-PI(s) and Institution(s).

## Describe the types of data, samples, and other materials to be produced in the course of the project.

This section should provide a general overview of the nature of the data or other materials produced under the proposed project. These could include data characteristics such as observational, experimental, reference, derived, simulated, and/or other. The data types referenced could include data generated by computer, data collected from sensors or instruments, images, audio files, video files, reports, surveys, and/or other. If the project will be collecting data of a sensitive nature, this should be noted here and will be reflected in subsequent sections of the plan. Include the responsible parties involved.

Example language for this section:

This project will produce *\_quantitative and/or qualitative\_* type of data in the following formats: *\_.csv, ascii, and/or other*. Textual and non-numeric data may include *PDFs, Word document (docx) files, image files, and/or others*. The data are characterized as *observational, experimental, reference, derived, simulated, and/or other*. The data types referenced include data generated by *computer, data collected from sensors or instruments, images, audio files, video files, reports, surveys, patient records, and/or other*. The data *are/are not* of a sensitive nature. The *PI and/or Co-PIs* will be responsible for the collection, management and sharing of all data generated by the project.

## Standards to be used for data and metadata format and content.

This section should identify the format in which the data will be stored as well as the types of metadata that will be included. These standards will often be determined by the accepted best practices of your discipline and should reference any relevant existing university standards. Your DMP should address how you will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future. If you are unable to use platform-independent and non-proprietary formats, you should specify the standards and formats that will be used and the rationale for using those standards and formats.

Example language for this section:

Raw data will be stored in the following formats: *FileMaker Pro v.7 files, Microsoft Access files, VSSIM, PTV Viswalk, and/or other*. Final data will be exported from *FileMaker Pro v.7 files and Microsoft Access* and will be archived as a *\_csv file, PDF file, and/or other*. Data generated with *VSSIM and PTV Viswalk* will be archived in the platform-specific file formats to retain their functionality.

The types of metadata that will be included are *Dublin Core metadata, ISO metadata standard, and/or other*, which conform to the standards in this field.

## Methods and policies for providing access and enabling sharing.

Projects collecting sensitive information will be required to comply with university policies governing the collection and handling of sensitive information (e.g., personal identifiers), consistent with Institutional Review Board (IRB) procedures and in accordance with relevant federal and state privacy laws and regulations. If sensitive information is collected, investigators will be responsible for de-identifying data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset prior to submitting data for archiving to NITC. Your DMP should address these issues and outline the efforts you will take to provide informed consent statements to participants, the steps you will take the protect privacy and confidentiality prior to archiving your data, and any additional concerns (e.g., embargo periods for your data). If necessary, describe any division of responsibilities for stewarding and protecting the data among PIs or other project staff.

Example language for this section:

Ethical and privacy issues including HIPAA and IRB will be addressed by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Sensitive data will be protected by \_\_\_*describe how you protect sensitive data by for example, using encryption, password protection, limited access to datasets and/or other*. Data will be de-identified by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. There will/will not be any restrictions placed on the data.

Restrictions will be placed on the data via \_\_\_\_\_\_\_\_\_\_ and access will be granted by\_\_\_\_\_\_\_\_\_\_.

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

Similar to the preceding section, here you should identify any potential limitations on the data requestor’s ability to re-use or re-distribute the data or materials. This could be because of the sort of sensitivity identified above or because the source of some of the data collected for the project has placed limitations on the re-distribution of those data and materials. Be sure to review any applicable agreements with such parties to determine the extent of re-use and re-distribution you are allowed to guarantee in your DMP.

In some rare cases, PIs will want to license the re-use or distribution of the data. If this is the case, the following applies. DOT uses the standards laid out in OMB Memorandum M-13-13, which indicates a strong preference for the use of Creative Commons licenses. PIs will have the option of setting conditions on the re-use of their materials by affixing a permission and copyright license statement, including a Creative Commons License to their work*. If you will be enforcing terms of use or a requirement for data citation through a license, indicate as much in your DMP. Describe any other legal requirements that might need to be addressed.*

Example language for this section:

Data will be available on or after the following date\_\_\_\_\_\_\_\_\_\_\_\_. Foreseeable uses/users of the data are\_\_\_\_\_\_\_\_\_. Findings from the data will be published *in final report, peer-reviewed article, technical report and/or other.*

Some data will be obtained from \_data vendor\_ and cannot be shared directly due to sharing restrictions, but detailed data parameters will be disclosed with the final data product to allow the purchase of the same dataset by interested parties.

Data re-use and re-distribution *will/will not be* limited to \_\_\_\_\_\_\_\_ and licensed by\_\_\_\_\_\_\_\_\_\_. (Note: This will not apply to most projects.)

## Methods for archiving and preserving access to data and materials.

This section identifies the means by which the data and materials will be stored to enable future access and sharing. NITC will archive project data and other related files in a PSU’s data archive. Projects generating large datasets have the option of using an alternative data archive.

If PIs elect an alternative data archive to PDX Scholar for data storage, they need to outline in their DMP how their proposed data archive(s) meet(s) DOT criteria (see <https://ntl.bts.gov/publicaccess/evaluatingrepositories.html>). Specific points to addressed include a documented plan for long-term preservation of the archive’s holdings and if adequate funding, staffing, and a system of governance are in place to support its mission of digital archiving. The PI also needs to discuss how the archive will enable the users to discover and use the data, and refer to them in a persistent way through proper citation (e.g., does it provide a persistent URL?). Another point to consider is if the data archive possesses a technical infrastructure that explicitly supports the tasks and functions described in internationally accepted archival standards like Open Archival Information System (OAIS).

Example language for this section:

Data will be permanently archived using PDX Scholar (<http://pdxscholar.library.pdx.edu>), a document and data archive that is curated, administered, and maintained by Portland State University Library. Backups are run off site and will be moved to Amazon S3 Cloud backup in the future. In addition to a DOI, each data product will also be assigned a persistent web address. Access to the data through PDX Scholar will be retained in perpetuity.

The project’s final report will be made available on NITC’s website and PDX scholar. Other publications will be made available in accordance with relevant copyright agreements with publishers.