

Data Management Plans

In their revised *Grant Proposal Guide*, the NSF mandates that **all proposals** submitted after January 11, 2011 include a short – no more than two-page – supplementary document detailing the Data Management Plan (DMP) for the proposed project. **NSF requires Rice to be able to guarantee that data (see definition below) from NSF-funded research at Rice be preserved for at least three years beyond the end of the award.** While it is not feasible to develop a comprehensive framework for a DMP that would apply to all of the disciplines supported by NSF, this document is an attempt to provide some guidance about these plans.

This document is based on readings of the NSF *Grant Proposal Guide* and accompanying NSF documentation, including a FAQ list and further directorate-specific guidance. In addition, this document includes templates that have been prepared by Rice faculty that you can modify as necessary to construct a data management plan for your own particular project. The document concludes with a list of web links to several other useful and relevant sites.

Note that the sample plans below allude to Rice University's Data Management Policy. This policy is currently being drafted, and will be adopted officially early in 2011, at which time it will be hosted on the Office of Sponsored Research website.

Recommended Content for Data Management Plans

Any data management plan should accomplish the following:

- *Be as simple as possible while still doing what is necessary.*
- *Describe the data.* Identify the data that will be measured, recorded, calculated, or modeled, as well as the manner in which these tasks will be performed.
- *Present the context of the data.* Explain the rationale for collecting or producing the data, as well as what insights could be gleaned from the data.
- *Explain the nature of the data.* Note that data can include numerous types of information, including field survey reports, laboratory notebook entries, computer files created by data acquisition software, and images produced via microscopy. When explaining the nature of the data, give an estimate the amount of data, and provide the necessary metadata, such as laboratory notebook entries, software codes, etc., that are required to properly interpret the data.
- *Describe the method for preserving and/or curating the data.* Indicate how data will be backed up, as well as how it will be stored both on- and off-site.
- *Discuss the approach for accessing the data, if relevant.* Be aware that when requested, data should be available for sharing within a reasonable period of time; understand that different communities may define “a reasonable period of time” in different ways.
- *State how long the data will be preserved and/or curated.* NSF **requires** data to be preserved **at minimum** three years beyond the end of the award.
- *Clarify ethical and/or privacy issues associated with the data, if relevant.* Explain how these issues will be addressed. If the data will involve human subjects, refer to <http://www.nsf.gov/bfa/dias/policy/> for NSF's policy on this topic.

- Detail intellectual property concerns associated with the data, if relevant. Explain how these concerns will be addressed.
- Provide a final boilerplate statement about Rice's Data Management Policy. Indicate that any issues not specifically referenced in the proposed project's data management plan will fall under the overarching Rice Data Management Policy.

Note that subsequent NSF proposals that refer to these data should discuss ongoing data management under "Results of Prior NSF Support."

Sample Data Management Plan: Physical Sciences and Engineering

The data obtained during the proposed project will consist of measurement records of electronic and optical properties of the nanodevices, as described in the main body of the proposal. These records will consist chiefly of measurements of current as a function of voltage, conduction as a function of time, and Raman spectra as a function of wavenumber obtained via custom software. These data will be recorded via computerized data acquisition software, with essential metadata present either as header in the relevant electronic files, or included along with the indexed laboratory notebook narrative.

These data will provide an experimental look at the detailed dissipative processes at work in nanoscale junctions, as delineated in the main body of the proposal. As such, they will be of interest to the nanoelectronics community, as well as to the condensed matter physics and physical chemistry communities.

The electronic transport and optical spectra will be computer files generally in the form of tab-delimited numbers with header information. These computer files will be accompanied by dated laboratory notebooks, as well as by numerical data files analyzed using MATLAB™ and commercial plotting software, such as OriginLab™. A copy of the custom Raman spectrometer data acquisition software also will be included, since it is essential for interpreting the raw data files.

The electronic data will be preserved in multiple on-site backups in the form of DVDs and RAID hard drive storage. Copies of the electronic data will be preserved off-site at Rice University's storage facility. Original laboratory notebooks will be secured by the PI in his campus office or laboratory. If requested, access to the data will be provided via contact with the PI. Data will, in principle, be available for access and sharing as soon as is reasonably possible, and not longer than two years after the acquisition of the data. The data will be preserved for at least three years beyond the award period, as required by NSF guidelines.

This project will not involve the acquisition of either animal or human subjects data.

We do not anticipate that there will be any significant intellectual property issues involved with the acquisition of the data. In the event that discoveries or inventions are made in direct connection with this data, access to the data will be granted upon request once appropriate invention disclosures and/or provisional patent filings are made.

The data acquired and preserved in the context of this proposal will be further governed by Rice University's policies pertaining to intellectual property, record retention, and data management.

Sample Data Management Plan: Biosciences

The data generated through the work described in this proposal will consist of [*nucleotide and protein sequences, graphs, three-dimensional structures, reconstructed images, photographs, hand-recorded observations, biological expression signatures, movies, etc.*] in both raw and curated forms, and will include relevant statistical analyses. These data will consist primarily of [*describe specific type of data being collected in the proposal, e.g., “arrays of gene expression data from cells grown in control and test conditions as described in the body of the proposal”*]. These data will be collected using the instruments and methods described in the proposal and will include datasets generated from commonly accepted data acquisition software, with essential metadata presented as headers in the relevant electronic files, or included along with the indexed laboratory notebook narrative.

Records of results will be labeled and will be stored as either hard copy or as digitized images. [*If needed, add a statement such as, “In some cases, observations of [organism or cell] behavior, viability, or phenotype will be retained in hand or electronic notation that will be dated and labeled in dated laboratory notebooks.*]

These data will provide an experimental look at the [*describe project briefly, e.g., “responses of cells to experimentally manipulated changes in gene expression*]. As such, they will be of interest to the [*in this case, genomics*] community, as well as the [*any other community, e.g., pharmacology*] community.

All members of the investigative team with access to data will receive instruction in the Responsible Conduct of Research (RCR); this instruction will include CITI training and additional training onsite. Original data notebooks will be retained in a secure location in the PI's laboratory, with electronic data backed up on Rice University's off-site storage servers whenever the nature of the data makes such archiving possible. If requested, data will be made available for sharing to qualified parties by the PI, so long as such a request does not compromise intellectual property interests, interfere with publication, invade subject privacy, betray confidentiality, or precede data curation. Data that are shared will include standards and notations needed to interpret the data, following commonly accepted practices in the field. Data will be available for access and sharing as soon as is reasonably possible, normally no longer than two years after its acquisition.

Data will be retained for at least three years beyond the award period, as required by NSF guidelines. In the event that discoveries or inventions are made in direct connection with this data, access will be granted upon request once appropriate invention disclosures and/or provisional patent filings are made. Key data relevant to the discovery will be preserved until all issues of intellectual property are resolved.

Insert one of the following statements, modified as necessary:

- A) The data to be acquired in the proposed project will not involve human subjects.
- B) The data to be acquired in the proposed project will include human subjects data that require Institutional Review Board approval. As detailed in the human subjects section of the proposal, all rules and regulations related to privacy (i.e., HIPPA) will be observed with specific regard to collected data.

If data are de-identified: To preserve confidentiality, each subject will be assigned an arbitrary code that will be associated with the data, and only data stripped of all potential identifiers will be stored in the collected and curated data sets. One file that contains the links to subject names and identifiers will be kept in a password-controlled file that will be accessible only to the subject coordinator. De-identified electronic data will be stored on external hard drives and DVDs, and copies of de-identified data will be preserved off site on the Rice University server. *[If needed, expand on this to cover details specific to the proposed project.]*

If data include identifiers: It is recognized that potential identifiers may exist in the datasets. All hard copies of data that contain such identifiers will be kept in a locked area of the laboratory or office, with key access available only to the PI and subject coordinator. All electronic data will be encrypted and stored in password-protected files that will be accessible only by the PI and the subject coordinator. All members of the Rice information technology community who may have access to the digitized storage files will receive appropriate HIPPA training from CITI.

- C) The data to be acquired in the proposed project will include data collected from experimental animals, as described in the institutional animal care and use section of the proposal. All such data will be collected only under IACUC-approved protocols that will be preserved with approval numbers and signatures by the PI. Observations of animal behavior, viability, or phenotype will be preserved as described above.

The data acquired and preserved in the context of this proposal will be further governed by Rice University's policies pertaining to intellectual property, record retention, and data management.

Sample Data Management Plan: Social and Behavioral Sciences

The proposed project will include human subjects data consisting of background demographic information (including age, gender, education, and handedness), reaction times and error rates from computer-based language and working memory tasks, and structural and functional neuroimaging data from MRI scans, with the functional data collected during the performance of the computer-based tasks.

The functional neuroimaging data will be used to determine the brain regions subserving language processing and to determine their overlap with brain regions supporting working memory. The behavioral data will be used to verify that the findings from the computer-based tasks replicate previous findings regarding the conditions affecting the difficulty of sentence comprehension and the capacity of working memory. The demographic data will be needed for published reports to convey the characteristics of the subject population.

The demographic data will be collected from a questionnaire administered by the PI and student/postdoctoral researchers associated with this project, and will be entered into electronic spreadsheets. The data from the computer-based tasks will consist of tab-delimited output from the programs running the tasks. Raw structural and functional magnetic resonance data will be obtained in DICOM format and transferred on DVD or external hard drives to Rice University. DICOM files will be converted to NiFTI format for compatibility with various neuroimaging analysis packages. Analyzed neuroimaging data will be stored in the AFNI file format.

Since these data will be from human subjects, approval for human subjects research will be obtained through the Rice Institutional Review Board. As detailed in the human subjects section of the proposal, because of confidentiality issues, each subject will be assigned an arbitrary code, which will be associated with the demographic, behavioral, and neuroimaging data. One file that contains the correspondence between subject names and codes will be kept in an encrypted password-controlled file accessible only to the PI and subject coordinator. Personal information (name, date of birth, etc.) will be permanently removed from raw neuroimaging data with various DICOM viewers (e.g., OsiriX) prior to data analysis.

The de-identified electronic data will be preserved on DVDs and external hard drives. Copies of these data will also be preserved offsite at Rice University's storage facility. Completed questionnaires (as well as representative blank questionnaires) and human subject consent forms will be stored in a locked area of the PI's laboratory accessible only to the PI and subject coordinator.

If requested, access to the de-identified data will be provided by contacting the PI. Data will in principle be available for access and sharing as soon as is reasonably possible, normally not longer than one year after publication of the data. The data will be preserved for at least three years beyond the award period, as required by NSF guidelines.

We do not anticipate that significant intellectual property issues involved with these data will arise. However, in the event that discoveries or inventions are made in direct connection with these data, access to the data will be granted upon request once appropriate invention disclosures and/or provisional patent filings are made.

The data acquired and preserved in the context of this proposal will be further governed by Rice University's policies pertaining to intellectual property, record retention, and data management.

Useful Links

NSF *Award and Administration Guide* and the relevant section on data management plans:

- http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4

NSF's FAQs about data management plans:

- <http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp>

NSF directorate-specific guidance documents regarding data management plans:

- <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>

Blog post offering insights about data management plans gleaned from various guidance documents:

- <http://researchremix.wordpress.com/2010/10/08/revised-nsf/>

Step-by-step guide for formulating a data management plan:

- http://www.dcc.ac.uk/sites/default/files/documents/templates/DMP_checklist.pdf
Note: This guide is based on UK governmental rules and guidelines.

MIT library guide to data management planning:

- <http://libraries.mit.edu/guides/subjects/data-management/>

University of Minnesota library guide to data management planning:

- <http://www.lib.umn.edu/datamanagement>