Data Management Plans for Successful NIH and NSF Grants

Presented by: Michael Lesiecki, PhD

Provided by:

Principal Investigators Association™
Dear Professional:

Thank you for ordering this webinar from Principal Investigators Association Library. We developed this series of topic–specific publications so our clients could have practical, how-to guidance addressing some of their most critical challenges all in one place, without searching far and wide for information on their managerial and funding activities.

In addition to the numerous special reports that make up the Library, we offer a weekly e-Alert, and a year-long series of audio conferences – devoted to helping you improve performance and spend more time doing what you love.... the research.

Our goal is to be the world’s leading source of real-world, results-orientation information in all fields of science. Our unique approach -- delivering targeted guidance, case studies, success strategies, and best practices -- has earned us a reputation for depth, usefulness, and high-value information as well as a loyal group of researchers who rely on that information to help them with their administrative and funding duties. We’re glad you’ve joined them, and invite you to review all of our products and services at www.principalinvestigators.org.

We are always on the look-out for interesting topics, researcher needs, and ways we can be of service to you. If you have a success story you’d like to share with your colleagues, please do not hesitate to contact us. We’d be thrilled to hear from you, and look forward to serving you and your organization with the best advice and information available in the months and years to come.

Best Regards,

Leslie Norins
Founder
Principal Investigators Association
9990 Coconut Road, Suite 316
Bonita Springs, FL 34135
info@principalinvestigators.org
Data Management Plans For Successful NIH and NSF Grants

Michael Lesiecki, Executive Director
Maricopa Advanced Technology Education Center
An NSF Funded Center of Excellence in the Advanced Technological Education Program
June 29, 2011

About the Presenter

- I have had the privilege to be a PI and Reviewer for both NIH and NSF grants since 1989.
- My technical background is chemical physics with a lot of focus on applications of laser technology in science and medicine.

Objectives

- Knowing what counts as data
- Understanding the agencies’ expectations and requirements for the data management plan
- Following policies for access and sharing
- How to determine the requirements specific to the Directorate, Office, Division, Program, or Institute.
- Funding the plan and finding model data management plans and process
There Are Many Sources of Information on This Topic

- University of Idaho, http://www.uidaho.edu/research/fundingagencies/proposal/nsfdatamanagementplan
- University of California, http://www.cdlib.org/services/uc3/datamanagement/funding.html

Data

- Knowing what counts

![Image of data center](http://example.com/data-center.jpg)

Courtesy Pittsburgh Supercomputing Center

What Counts as Data? Examples

- **Observations**: atmospheric profiles, image collections
- **Experiments**: NMR, short pulse waveforms
- **Simulations**: Computational Fluid Dynamics, T-allele polymorphism
- **Compilations**: data sets in the Institute for Polar Ecology of Kiel University and to make these data sets available for input into OBIS
Data Generated by Models

- Novel Laboratory Model Reveals Clues to How Blood Starts Clotting; Nicole Rager Fuller - NSF

What Are the Agencies Expectations?

Experimental Computing cluster at the university of Illinois
Zina Deretsky, National Science Foundation

NIH

- "The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers. ...Investigators submitting an NIH application seeking $500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible."
**NSF**

- "Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing."
- Plan is a two page attachment to the proposal

---

**What Does the Plan Look Like? (NSF Key Elements)**

- A two page supplementary document
- Costs are allowed
- Available in a reasonable length of time
- Not preliminary data
- Follow confidentiality and Human Studies regulations

---

**Saving Data**

- Biological
  - In the freezer
- Hard copies, physical specimens
  - In the file cabinet
- Databases
  - In the cloud
  - In a digital library repository
Grant Proposal Guide (GPG) Says
The Plan Includes
1. Types of data
2. Applicable standards for data and metadata
3. Policies for access and sharing
4. Policies for re-use, distribution and derivatization
5. Plans for archiving

This Might Not Be The Best Plan For Archiving

There Are Solutions for Sharing, Distribution and Archiving
- NSDL is the nation's online portal for education and research on learning in Science, Technology, Engineering, and Mathematics

Incorporated Research Institutions for Seismology
More Solutions and Examples

- The Dataverse Network software, http://thedata.org/, is an open-source, digital library system for management, dissemination, exchange, and citation of virtual collections (dataverses) of quantitative data.

How I Did It

1. Contacted University of Wisconsin, Madison, Internet Scout project
2. They have developed an open source platform, CWIS, Collection Workflow Integration System, http://scout.wisc.edu/Projects/CWIS/
3. Installed CWIS on our College Server

And Then

- Tried to understand their metadata tagging system based on the Dublin Core standards
  - User defined special vocabulary
How I Did It continued

- CWIS was specifically created to help build collections of Science, Technology, Engineering, and Math (STEM) resources and connect them into NSF’s National Science Digital Library, but can be (and is being) used for a wide variety of other purposes.

More How I Did It

- To “create” the resource, a pretty large excel sheet in this case, I had to give it a unique url on our server
- Now in CWIS, I describe what it is, add descriptive search terms, add metatags and “submit”
- Anyone coming to our site or Internet Scout can search, access and download the file and description
  - Google too

What I Achieved

- I complied by identifying standards, schema, and metatags
- I provided a policy for use
  - Open access through a simple login
- I insured long-term integrity and archiving through the dual server and our institution
And More

- I connected to the larger world of the National Science Digital Library

Synoptic Optical Long-term Investigations of the Sun Kevin Schramm, NSO/AURA/NSF

One More Example

- OBIS allows users to search marine species datasets from all of the world's oceans

More Data Repositories

- MIT has an excellent site on these, http://libraries.mit.edu/guides/subjects/data-management/publishing.html#icpsr
Individual Directorates

- Link to specific directorate requirements: http://www.nsf.gov/bfa/dias/policy/dmp.jsp

Remember the Grant Proposal Guide (GPG) Says the Plan Includes

1. Types of data
2. Applicable standards for data and metadata
3. Policies for access and sharing
4. Policies for re-use, distribution and derivatization
5. Plans for archiving

We are good!

The NIH Perspective

- Final Research Data
  - This does not mean summary statistics or tables; rather, it means the data on which summary statistics and tables are based.
  - Applies to clinical research, surveys and other types; including human and pre-clinical studies
NIH Expectations

- For most studies the “final research data” will be a computerized data set.
- If an application describes a data sharing plan there is a rather distinct expectation that this plan will be enacted.

More Constraints

- Human Studies and Privacy
  - Data should be
    - Redacted to strip all identifiers
    - Minimize risks of unauthorized disclosure
  - If you are proposing clinical trials think about study design, informed consent and the structure of the resulting dataset
    - There may be HIPAA restrictions

Methods for Data Sharing

- Under the PI
- Using a data Archive
- Using a data Enclave
- Mixed mode

- **Enclave**: a controlled secure environment for analysis of restricted data resources
Data-Sharing Agreements

- Imposes appropriate limitations on users

ECological Speciation, Marcus Kronforst, Harvard University

Proprietary Data

- For SBIR projects allows non-disclosure for a period of up to four years
- Industry co-funding scenario:
  - The need to invoke provisions of a sponsored research agreement may delay the release of data
  - State this in the plan

Provisions for Archiving (UConn ref.)

- How long should your data be kept?
- Have you identified a repository or archive?
- How will the data be prepared for long term?
- Will funding be required for preservation?
- What is the long term data location for preservation and backup?
- How will documentation and curation responsibilities be transferred from one entity to another?
Access

- Ethical issues or privacy concerns?
- Confidentiality?
- Embargoed for IP? (Intellectual Property)
- Who?

  Tahiti Sea Level Expedition, 
  ECORD/IODP

More Access

- Any reason to restrict use?
- Any reason to license
- Do you wish to release under Creative Commons?
  http://creativecommons.org/choose/zero/

Cost

- NIH recognizes that it takes time and money to prepare data for sharing. Thus, applicants can request funds for data sharing and archiving in their grant application.

  Sequencing the corn genome, 
  Nicole Rager Fuller, NSF
Review Considerations

- Reviewers will not factor in a Data Plan into a scientific merit or priority score.
- At the NSF, reviewers judge its adequacy
- At the NIH, staff members are the judge

After All This at the NIH

- Looking for a brief paragraph following the Research Plan section
  - Not included in page count
- May appear in:
  - Budget and Justification
  - Background and Significance
  - Human Subjects Section

NIH Examples


Tip of An Atomic Force Microscope
University of Illinois
Summary: Data Plan
(see UConn reference)
1. the **types** of data to be authored;
2. the **standards** that would be applied for format, metadata content, etc.;
3. **provisions** for archiving and preservation;
4. access policies and provisions for re-use of the data; and
5. plans for eventual transition or termination of the data collection in the long term future.

Questions

- **Frequent**
  - If my proposed research does not generate data, do I need a data management plan?
  - I may not be able to find one of these collective databases to provide access. What can I do?
  - How long should I plan on maintaining my data collection?