

# LA-SIGMA

Louisiana Alliance for Simulation-Guided Materials Applications

## Data Management

Lawrence R. Pratt and Joel E. Tohline  
Tulane University and LSU



## Data Management Plan (from NSF Grant Proposal Guide)

Plans for data management and sharing of the products of research. **Proposals must include a supplementary document of no more than two pages labeled "Data Management Plan"**. This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (see below), and may include:

1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
2. the 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);
3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. policies and provisions for re-use, re-distribution, and the production of derivatives; and
5. plans for archiving data, samples, and other research products, and for preservation of access to them.

Data management requirements and plans specific to the Directorate, Office, Division, Program, or other NSF unit, relevant to a proposal are available at: <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>. If guidance specific to the program is not available, then the requirements established in this section apply. **Simultaneously submitted collaborative proposals and proposals that include subawards are a single unified project and should include only one supplemental combined Data Management Plan, regardless of the number of non-lead collaborative proposals or subawards included.** Fastlane will not permit submission of a proposal that is missing a Data Management Plan (separate place to upload). Proposals for supplementary support to an existing award are not required to include a Data Management Plan. A valid Data Management Plan may include only the statement that no detailed plan is needed, as long as the statement is accompanied by a clear justification. Proposers who feel that the plan cannot fit within the supplement limit of two pages may use part of the 15-page Project Description for additional data management information. Proposers are advised that the Data Management Plan may not be used to circumvent the 15-page Project Description limitation. The Data Management Plan will be reviewed as an integral part of the proposal, coming under Intellectual Merit or Broader Impacts or both, as appropriate for the scientific community of relevance.

### Dissemination and Sharing of Research Results

**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. Privileged or

# Requirements:



- May vary with Directorate/Program
- "... may include only the statement that no detailed plan is needed ..."
- "... 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.

# NSF Data Management Plan



- Relevant NSF URLs
  - Requirements: <http://www.nsf.gov/eng/general/dmp.jsp>
  - FAQs: <http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp>
- CCT strategy (draft)
  - [http://www.cct.lsu.edu/NSF\\_DataManagement\\_Plan](http://www.cct.lsu.edu/NSF_DataManagement_Plan)
- Tulane policy statement (not specific to computational sciences)
  - <http://tulane.edu/asvpr/ora/upload/NSF-Data-Management-Plan2.pdf>



# CCT Strategy (draft)

## Scope:

- Every proposal submitted to NSF must have a two-page (maximum) Data Management plan.

## Ambiguity:

- No clear instructions for implementation
- Does not include information about types of data that must be included
- Data preservation longevity requirements are not clear

## CCT's Data Management Plan for Immediate Use:

- CCT has set up a web page [[www.cct.lsu.edu/NSF\\_DataManagement\\_Plan](http://www.cct.lsu.edu/NSF_DataManagement_Plan)] containing information about the procedures we follow at CCT for Data Management related to NSF projects and proposals. This URL can be referenced in the Data Management Plan section of every proposal. Information contained on this web page may also serve as a boilerplate for the Data Management section of any individual proposal.
- This web page can be updated regularly to reflect changes in the CCT's or LSU's Data Management strategies so that PIs can retrieve the latest version of the document.

# CCT Strategy (draft)



## Short Term Solution:

- Offer a module on the CCT's Subversion (SVN) repository for each submitted proposal that the PI and his/her team can use as a Data Management Repository for their project. Access to any module can be restricted on a per folder basis to everyone, anyone, to the world or a group. Instructions to do so will be added on the CCT's subversion help files. A link to the particular project/proposal module can be included in the Data Management portion of the proposal.
- Data that is relevant to each funded project – as defined by the PI in the project proposal's Data Management Plan – can be stored throughout the life of the project in an appropriate folder within the project's SVN module. As files are updated, SVN will handle versioning automatically. All SVN modules will be automatically and routinely backed up and archived by CCT's IT staff.
- Recognized problem: handling large data files (e.g., > 30 MB binary) through SVN will likely become tedious.

# CCT Strategy (draft)



## Long Term/Permanent Solution:

➤ We propose to implement a Data Repository with Metadata and Search Engine capabilities that can handle growing information and files sizes, along with multiple levels of access control. At the present time, it is unclear what scale of such a system will be required to meet NSF's requirements. We expect that, over time, more detailed instructions from NSF will clarify this situation. Before such a system is implemented, a committee with faculty and technical staff needs to formulate a plan and select a data management/sharing system that can be customized to serve the broad needs of CCT faculty/researchers.

# CCT Strategy (draft)



## Progress over the past year:

- Participated in the 2011 RDLM Princeton Workshop (see accompanying slide) to understand how other research-intensive universities are tackling this issue
  - have involved both LSU’s information technology services (Dee Childs) and LSU’s libraries (Will Hire & Brian Melancon)
- Also in tune with strategies being developed/implemented ...
  - among Internet2 leadership
  - at San Diego Supercomputing Center (SDSC) using OpenStack (<http://cloud.sdsc.edu/hp/about.php>)
- LONI proposes to dedicate a modest amount (few hundred terabytes) of “tier 1” storage toward addressing NSF’s data management mandate
- We’re supportive of the decision made by LSU’s ITS to purchase the Isilon (hardware) “tier 2” storage solution from EMC (<http://www.isilon.com/>), presently expandable up to 16 petabytes.
- Not yet clear ...
  - How very long-term, archival (e.g., tape) storage will be handled at LSU or across LONI
  - What searchable scheme for data retrieval will be adopted

# 2011 RDLM Princeton Workshop

<http://rds.columbia.edu/rdlm>



## Research Data - Storage & Management at the Louisiana State University

***Position paper for the Research Data Lifecycle Management Workshop,  
July 18 - 20, 2011***

Ravi Paruchuri, Joel Tohline, Faye Phillips, Melody Childs, Sam White, William Armstrong, Will Hires, Brian Melancon - Louisiana State University, Baton Rouge LA

[http://www.columbia.edu/~rb2568/rdlm/Paruchuri\\_LSU\\_RDLM2011.pdf](http://www.columbia.edu/~rb2568/rdlm/Paruchuri_LSU_RDLM2011.pdf)





Louisiana  
EPSCoR