An Analysis of Data Management Plans in University of Illinois NSF Grant Proposals

Alexandra L. C. Krogman, William H. Mischo, Mary C. Schlembach, and Megan N. O'Donnell
Grainger Engineering Library Information Center, University Library, University of Illinois at Urbana Champaign

Introduction

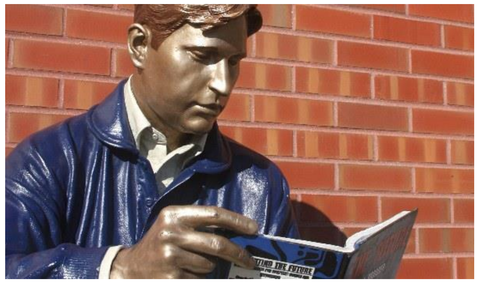
The University of Illinois at Urbana-Champaign (UIUC) Library analyzed 1,260 Data Management Plans (DMPs) from grant proposals submitted to the National Science Foundation (NSF) from July 2011 through November 2013. A team of Library staff members assigned each DMP controlled vocabulary terms that summarized its proposed data storage and sharing mechanisms. The Library constructed a database composed of the proposal's title, PI's name, PI's department, PI's college, NSF grant number, funded status, and assigned vocabulary terms. NSF funded a total of 298 of the 1,260 proposals as of May 2014. There was no significant statistical differences in the proposed data storage or sharing mechanisms between the funded and unfunded NSF proposals. However, there was significantly higher campus institutional repository use and disciplinary repository or cloud storage use in proposals submitted to NSF after October 2012.

Purpose

How do Illinois researchers address data storage and sharing?

Do funded NSF proposals employ common data management practices?

Answering these questions allows the Library to participate in a dialog with University of Illinois administrators about research data services and to develop resources that are campus-wide and can be used by University of Illinois researchers to manage their data.



Methodology

A total of **1,260 NSF grant proposals** submitted to NSF between **July 2011 and November 2013** by **Illinois researchers** were analyzed.

A team of staff reviewed each proposal's DMP in order to develop **controlled vocabulary terms** that would be assigned to each DMP. These terms addressed questions about data storage and sharing.

Graduate assistants then read through the DMPs of each proposal and assigned appropriate terms based on content.

Term	Definition
PI Server	Computers, servers, hard drives, or workstations that the PIs (and/or their staff) use to store project data
PI Website	Websites usually edited or ran by the PI or a group that they belong to
Analog	Physical records not including specimens, samples, or artifacts Example: Lab notebooks
Optical Disc	DVDs, CDs, or Blu-ray discs
Department	When a department is mentioned as providing a storage or hosting resource Example: Departmental backup service
Campus	Services located on or provided by the Illinois campus Example: IDEALS
Remote	Services and sites not located on the Illinois campus Example: Governmental repository services
Disciplinary	Disciplinary data repositories Example: GenBank
Cloud	Storage services that use cloud technology Example: Amazon Cloud
Publication	Traditional scholarly outputs Example: papers and presentations
Specimens	Physical specimens, samples, or artifacts Example: DNA samples
No Data	When the research will produce no data products Example: Theoretical studies
Not specified	When the text of the DMP was not specific enough to record many details
Template	When the researchers utilized the Grainger Engineering Library DMP template

Results

Chart 1 (right) : Percentage of funded NSF grant proposals with controlled vocabulary terms assigned to them

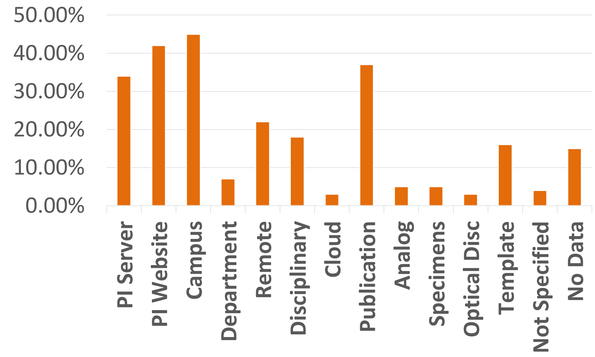


Chart 2 (left) : Controlled vocabulary terms assigned to DMPs of funded and unfunded NSF grant proposals

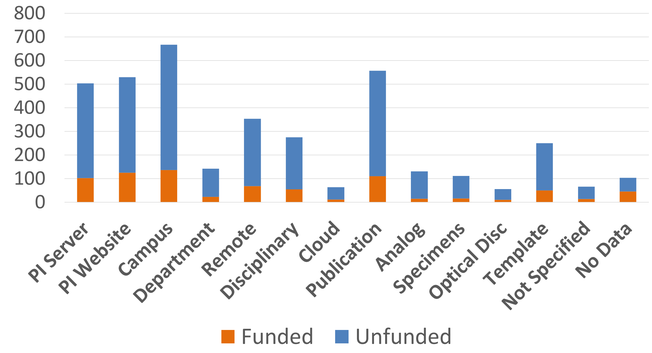


Table 1 (top) shows the frequencies of five data management mechanisms among funded and unfunded NSF proposals. The chi-square values do not indicate that significant differences exist.

Table 2 (bottom) shows the frequencies of two data management mechanisms among NSF proposals submitted prior to October 2012 and proposals submitted after October 2012. The chi-square values indicate that proposals submitted after October 2012 specify use of the University's institutional repository and disciplinary repositories or cloud storage services at a higher frequency.

Type of proposed storage mechanism	Funded	Unfunded	Chi-Square Value
PI Server / PI Website	183	569	0.7
Illinois institutional repository	62	197	0.02
Campus storage services	139	474	0.74
Departmental Server	24	102	1.67
Disciplinary / Cloud Services	67	241	0.85

Type of proposed storage mechanism	Before October 2012	After October 2012	Chi-Square Value
Illinois institutional repository	108	166	4.59
Disciplinary / Cloud Services	121	182	4.33

Discussion & Conclusion

- 556 DMPs mentioned publications as a method of data dissemination. This high number may be partially due to the vagueness of the NSF DMP guidelines, but it could also be a side effect of NSF's focus on sharing processed data and a PI's natural tendency to associate processed data with publications.
- There were no significant differences between the data storage and sharing mechanisms proposed in funded and in unfunded proposals. This indicates that researchers are just entering the DMP lifecycle and that communities of practice and best practices have yet to emerge.
- Proposals submitted after October 2012 specified use of the Illinois institutional repository and disciplinary repositories or cloud storage services at a higher frequency. Researchers may be responding to the Library's educational and assistance efforts with respect to data management.
- **Data management is an institutional-wide issue requiring collaborative working relationships between multiple stakeholders. It is critical that campuses and other institutions awarded NSF grants either develop or access key infrastructure services that will give researchers enhanced data management capabilities and provide mechanisms for compliance with federal grant requirements and mandates.**

Mischo, William H., Mary C. Schlembach, and Megan N. O'Donnell. 2014. "An Analysis of Data Management Plans in University of Illinois National Science Foundation Grant Proposals." Journal of eScience Librarianship 3(1): e1060. http://dx.doi.org/10.7191/jeslib.2014.1060