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

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.

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.

Results
Table 1 (top) shows the frequencies of five data management mechanisms among funded and unfunded NSF proposals. 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.

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 NSFs’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 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.


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 NSFs’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 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.