Disciplinary Perceptions of Data and Data Management Practices

Pamela L Shaw¹ MSLIS, MS and Cunera M Buys² JD, MS, MSLIS

Galter Health Sciences Library¹, Northwestern University Clinical and Translational Sciences Institute (NUCATS)², Seeley G. Mudd Library², Feinberg School of Medicine¹, Northwestern University¹ & ²

Brief Background Highlights

- NSF requires data management plans since 2011 [1]
- NIH has a data sharing policy since 2003 [2]
- White House OSTP released memoranda calling for data sharing plan requirements for awards from federal agencies with more than $100 million R & D expenditures [3,4]
- Published study shows differences in data sharing or reuse among different academic disciplines [5]
- Academic researcher dissatisfaction with long-term data storage, funding, infrastructure and instruction in data management [5]
- Libraries becoming involved in research data management [6]
Methods

- Survey conducted by the Northwestern University E-Science Working Group (ESWG), a collection of representatives from Northwestern libraries, research computing, information technologies and the Office of Sponsored Research
- Survey questions adapted from survey materials shared by other academic institutions
- Survey constructed in Qualtrics, consisting of 21 questions in the following categories:
  - Researcher demographics
  - Types and size of data
  - Data storage
  - Data retention
  - Data sharing
  - Data management planning
  - Training or assistance needed
- Survey was distributed via email link to over 12,000 Northwestern faculty, staff and students across all academic schools and disciplines
- Survey results analyzed with Atlas.ti text analysis package and with Qualtrics built-in cross-reference functionality
- Link to the completed report with appendices: https://www.library.northwestern.edu/sites/www.library.northwestern.edu/files/images/5-27-14%2FFinal%2520Report%25202012.psedit.pdf

831 responses; 788 completed the survey (response rate of 6.4%)

Representatives from 159 different departments responded to the survey, with the department of chemistry the largest respondent
The numbers in the "Don't know" category of this chart suggest that the individuals who may spend the most time generating and working with data (research staff and graduate students) may not have as clear a concept of the cumulative data storage needs for projects.

**Future Data Storage Size Needs by Status**

- Tenure-track faculty
- Non-tenure-track faculty
- Post-doctorate
- Graduate student
- Staff
- Other (please specify)

**How long are data stored?**

- Many researchers prefer to retain raw data and published data indefinitely, citing the value of raw data for new analyses, longitudinal studies or replication of results.
- The time span of 5-10 years was also commonly selected, often cited as required by NIH funding mandates.
• 54.6% of respondents indicated that they share data by personal choice, others share because of funder requirements or recommendations.
• Those who do not plan to share their data cite protection of human subject privacy, protection of their intellectual property rights and beliefs that others would not be interested in their data.
Disciplinary Perceptions of Data & Data Services

Respondents from the schools of education and management requested information on accessing sources of data more than any other schools. Responses from humanities were low and many humanities researchers did not perceive their output as "data," despite the fact that text files compose the largest category of data collected among all researchers.

Schools most likely to have data management plans were schools of education (69%), journalism (60%) and medicine (59%). Schools least likely to have DMPs were law (0%) and management (23%).

Limitations of Survey Design

- Department designation
  - Respondents were given a text box to supply department name, instead of a list of options to select from. This made coding the responses for correlation statistics by department virtually impossible with our limited staff resources. As a result, our plan of "disciplinary" analysis was not easily implemented.
- Contact follow up
  - Several respondents indicated that they would be willing to be contacted for a follow up interview, but only 2 provided contact email addresses.
- Term usage
  - The survey asked about data needs for each new "grant." Not all respondents had grants, so the word "project" would have been a better choice.
- Missing questions
  - Questions about sensitive data were left out of the survey, but would have provided valuable information on practices in managing data privacy.
Barriers to Good Data Practices

- Poor storage choices
  - Users are storing data on potentially unstable computer hard drives
  - Lack of choice in long-term storage options at the university level, and user misunderstanding of current university storage options for long term use
  - NUIT is investigating options for long-term storage
- Lack of organization
  - Text analysis indicated that "frustration" and "disorganization" were a common theme in text responses from all user communities
  - This is an opportunity for libraries and IT to provide training and consultation in file naming, metadata and file version control practices
- Lack of knowledge
  - Respondents at all levels indicated interest in library and NUIT guidance on all levels of data management, from federal requirements to local storage and analysis services.

References

Additional Resources


• Read, K.B (July 2015). Starting the data conversation: informing data services at an academic health sciences library. J Med Lib Assoc 103(3): 131–135. DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005