MANAGING THE DATA DELUGE: UNDERSTANDING SCIENTISTS' NEED FOR DATA CURATION SERVICES

JEANINE M. SCARAMOZZINO, COLLEGE OF SCIENCE AND MATHEMATICS LIBRARIAN
MARISA RAMIREZ, DIGITAL REPOSITORY LIBRARIAN
KAREN MCGAUGHEY, ASSISTANT PROFESSOR OF STATISTICS
CALIFORNIA POLYTECHNIC STATE UNIVERSITY - SAN LUIS OBISPO

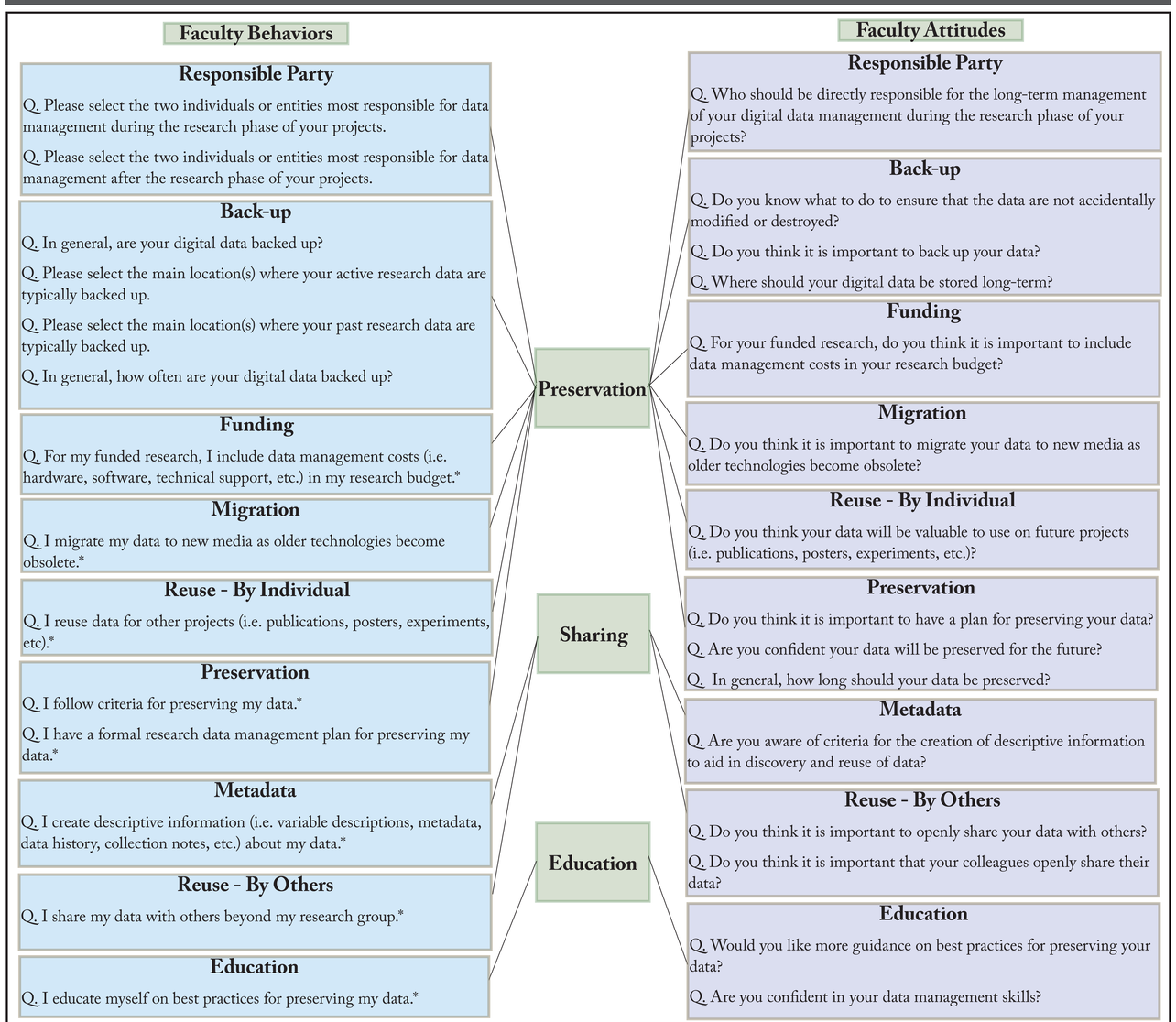

Figure 1: Survey questions employed to determine faculty behaviors and attitudes as they relate to data preservation, data sharing, and data curation educational needs.

***Semantic differential responses (e.g. always, frequently, occasionally, rarely, never).**

SUMMARY

Data curation is defined as "…the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education…activities [that] enable data discovery and retrieval, maintain its quality, add value, and provide for re-use over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation."[1]

While library research on data curation is active and ongoing in the humanities and social sciences, the research regarding data curation within the sciences is in its infancy. The lack of knowledge about data creation, management, and reuse has a direct impact on librarianship, library services, and library users, as libraries are now being asked to provide services to archive data created at their universities. What are the data curation needs on campus, and what services are libraries and librarians willing and able to provide to meet these needs?

Information gathered from a survey distributed to Cal Poly State University San Luis Obispo science and mathematics faculty will help provide insight into the awareness of science researchers about data curation issues and their needs for data curation services and education regarding maintenance and management of data.

1 Data Curation. Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. 16 Mar. 2010. <http://www.lis.illinois.edu/academics/programs/ms/datacuration>

BACKGROUND

- Data curation -- the active and ongoing management of data created by scientific endeavors -- is the next big collection development and faculty education issue to face librarians.
- Academic publishers are beginning to require authors to submit datasets in concert with their completed manuscripts. Many funding agencies also require faculty to include a data management plan as a component of grant applications.
- Anecdotal evidence suggests that Cal Poly faculty do not actively curate their datasets after the course of their research and some do so poorly during the research process. Based on these patterns, we believe faculty will need to be better informed on how to care for their data to sufficiently satisfy publisher and funding agency requirements.
- While many well-funded "R1" academic institutions (e.g. MIT, Purdue, Cornell) are actively engaged in curation of large datasets, little attention has been paid to the handling of smaller datasets or to the development of educational components to support faculty data curation practices.
- Because many academic libraries have an institutional repository infrastructure in place to support the acquisition and delivery of published information, libraries are uniquely poised to provide technical support and education on metadata creation, retention and migration of datasets.

PROJECT OBJECTIVES

The aim of this research is to:

- Understand scientists' current data management activities.
- Assess faculty awareness of data curation issues.
- Identify gaps in scientists' understanding of best practices for maintenance and management of data.
- Identify educational opportunities to enhance researchers' data management practices.
- Develop reusable educational modules to educate faculty on data curation best practices.

METHODOLOGY

- Research questions:
  - What are the data curation needs of Cal Poly scientists?
  - What is the level of awareness of data curation best practices among Cal Poly faculty?
  - What services can libraries and librarians provide to meet these needs?
- A survey (Fig. 1) was developed to evaluate current faculty data preservation behaviors, attitudes and awareness of best practices (i.e. do faculty know what they should be doing with data and do they do what they should be doing).
- The draft survey was pilot-tested and revised based on feedback from a representative cross-section of science and math faculty.
- Survey invitations were emailed to 330 Cal Poly State University, San Luis Obispo, College of Science and Mathematics (COSAM) faculty April 7th with the support of the COSAM dean and department chairs.
- Survey responses are being collected on a secure website from April 7th-April 19th.

ANALYSIS AND NEXT STEPS

- Survey responses will be analyzed to determine faculty data curation behaviors and faculty attitudes.
- Responses will also reveal the level of awareness of data management best practices among faculty.
- Questions on current data management practices are generally paired with questions on faculty attitudes in order to highlight mismatches which may indicate a need for data curation services (Fig. 1).
- If severe mismatches are found, it may indicate that faculty need help optimizing their data curation practices. The library can then consider appropriate strategies to provide such help.
- Based on the current awareness level, educational initiatives can be developed by the library to inform faculty of data curation issues and strategies.

SELECTED READING

- Berman, Francine. (2008). "Got Data? A Guide to Data Preservation in the Information Age." Communications of the ACM 51(12):50-56.
- Gabridge, Tracy. (2009). "The Last Mile: The Liaison Role in Curating Science and Engineering Research Data." Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC 265:15-21.
- National Academy of Sciences, National Academy of Engineering, and Institue of Medicine of the National Academies. (2009). Ensuring the Intergrity Accessiblity, and Stewardship of Research Data in the Digital Age. Washington, DC: National Academies Press.
- Pryor, Graham and Donnelly, Martin. (2009). "Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career." International Journal of Digital Curation 4(2):159-170.
- Witt, Michael. (2008). "Institutional Repositories and Research Data Curation in a Distributed Environment." Library Trends 57(2):191-201.

ACKNOWLEDGEMENTS

- We would like to thank Anna Gold, Associate Library Services Dean, Michael Miller, Library Services Dean, Phil Bailey, College of Science and Mathematics Dean, and the COSAM Department Chairs and faculty for their support of this research.
- We would also like to thank the COSAM faculty that provided valuable feedback on the draft survey.

CAL POLY LIBRARY SERVICES
ROBERT E. KENNEDY LIBRARY

Download this poster: http://digitalcommons.calpoly.edu/lib_fac/56