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

BACKGROUND

• Data curation -- the active and ongoing management of data created by scientists -- is an emerging field.

• Academic publishers are beginning to require authors to submit datasets alongside manuscripts.

• Many funding agencies are also requiring data management plans as a component of grant applications.

• Anecdotal evidence suggests that Cal Poly faculty do not actively care for their data once the research phase of their projects is complete.

• Because many academic libraries have an institutional repository within their collection development and faculty education missions, it is important for libraries to determine faculty data curation needs.

The draft survey was pilot-tested and revised based on feedback from a representative cross-section of science and math faculty.

While library research on data curation is active and ongoing in the humanities and some 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 librarian roles. 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.

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 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


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.

ACKNOWLEDGEMENTS

• We would like to thank Ariana Gold, Associate Library Services Dean, Michael Miller, Library Services Dean, Phil Bailey, College of Science and Mathematics Dean, and the COSAM Department Chair and faculty for their support of this research.

• We would also like to thank the COSAM faculty who provided valuable feedback on the draft survey.

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).

<table>
<thead>
<tr>
<th>Faculty Behaviors</th>
<th>Faculty Attitudes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responsible Party</strong></td>
<td><strong>Responsible Party</strong></td>
</tr>
<tr>
<td>Q. Do you think it is important to include data management costs in your research budget?</td>
<td>Q. Do you think it is important to share your data with others?</td>
</tr>
<tr>
<td>Q. Do you think it is important to make your data widely available in the future?</td>
<td>Q. Do you think it is important to ensure that the data is not accidentally modified or destroyed?</td>
</tr>
<tr>
<td>Q. Do you think it is important to back up your data?</td>
<td>Q. Do you think it is important to keep your data secure?</td>
</tr>
<tr>
<td>Q. Do you think it is important to keep your data confidential?</td>
<td>Q. Do you think it is important to ensure that the data is not accidentally modified or destroyed?</td>
</tr>
<tr>
<td>Q. Do you think it is important to include data management costs in your research budget?</td>
<td>Q. Do you think it is important to share your data with others?</td>
</tr>
<tr>
<td>Q. Do you think it is important to make your data widely available in the future?</td>
<td>Q. Do you think it is important to ensure that the data is not accidentally modified or destroyed?</td>
</tr>
<tr>
<td>Q. Do you think it is important to back up your data?</td>
<td>Q. Do you think it is important to keep your data secure?</td>
</tr>
<tr>
<td>Q. Do you think it is important to keep your data confidential?</td>
<td>Q. Do you think it is important to ensure that the data is not accidentally modified or destroyed?</td>
</tr>
</tbody>
</table>

Survey responses are being collected on a secure website from April 7th - April 19th.

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 its use over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation.”

While library research on data curation is active and ongoing in the humanities and some 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 librarian roles. 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.

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 its use over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation.”

While library research on data curation is active and ongoing in the humanities and some 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 librarian roles. 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.

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 its use over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation.”

While library research on data curation is active and ongoing in the humanities and some 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 librarian roles. 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.

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 its use over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation.”

While library research on data curation is active and ongoing in the humanities and some 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 librarian roles. 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.

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 its use over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation.”

While library research on data curation is active and ongoing in the humanities and some 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 librarian roles. 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.