DEVELOPING AND IMPLEMENTING DATA SERVICES AT AN UNDERGRADUATE INSTITUTION

Jeanine M. Scaramozzino and Russell A. White
Robert E. Kennedy Library
Cal Poly State University
1 Grand Avenue
San Luis Obispo, CA 93407-0600

Abstract:
California Polytechnic State University (Cal Poly) is a primarily undergraduate, comprehensive polytechnic university with roughly 19,000 students. The Cal Poly Data Services program currently provides basic support for faculty and students in data curation, data management, GIS, and scholarly communication. Whereas data service programs at most research universities focus on institutional data management, the program at this undergraduate institution is focused on developing a data literacy curriculum through instruction and peer-to-peer reference. The physical home for Data Services, the Data Studio, supports the virtual world of data literacy. The development and implementation of Cal Poly’s Data Services program and the lessons learned by its staff are discussed to help other undergraduate universities to begin or refine the development of library data services.

Keywords: Geographic information systems, electronic information resources – management, library outreach programs, data literacy.

Introduction and Background
California Polytechnic State University (Cal Poly), founded in 1901, is part of the 23-campus California State University System. It is an undergraduate and master’s granting university with approximately 19,000 students, 2,000 staff, and 1,300 faculty. The focus is on undergraduate education and it offers 64 bachelor’s majors, 68 minors, 31 master’s programs, and 14 credentials (California Polytechnic State University 2012). One direction to take when developing and implementing data services at an undergraduate institution is to move beyond institutional research data services and data curation toward a focus on data literacy. Using this lens and the “learn by doing” philosophy of the campus undergraduate curriculum helped Cal Poly refocus and define the role data play in the larger organization. The Data Services mission at Cal Poly’s Robert E. Kennedy Library is to support students and faculty in finding and using data for scholarship, teaching, and learning. This supports the larger Data Services vision to provide the resources and services to equip students with data literacy skills and collaborative opportunities to prepare them as 21st century citizens.

The Library is part of Information Services (IS), which also contains Information Technology Services (ITS) and the Center for Teaching, Learning and Technology (CTLT). Data is a common interest within this organization. ITS maintains servers, campus data backups, and supports Big Data needs. CTLT provides faculty support and instructional workshops. During the past few years the Library’s data services have evolved from solely providing basic GIS support toward a broader definition including data access, analysis, use, visualization, management, sharing and curation support, and the formalization of the Data Services Unit. In conjunction with Library and Campus ITS, Data Services is defining, employing, and maintaining the necessary hardware, software, and services for data discovery, access, use,
dissemination, and storage within the Library, on campus, and in the community. This work contributes to the foundation for data literacy development.

The following is a discussion of what is being done, what needs to be known to move forward, and the challenges and opportunities at Cal Poly. The process can be relevant or applicable to other institutions at any stage of program evolution. Understanding the development and implementation of the Cal Poly program may help begin or improve other libraries’ data services programs.

**Scoping Data Services**

It is important to remember that the functions and activities of data services mirror those of broader library services. Collections, outreach, reference and instruction are part of managing journal articles, books, subscriptions, or data. Most of these general library functions are reflected in “data services” with a few additions like data management and scholarly communication. It is important to understand that rather than attempting to develop data services separately, Cal Poly instead incorporated them into the broader library services already in place. For traditional library services, reference, instruction, collections, etc., there are similar foundations that are translatable across disciplines. For example, finding an article for engineering is somewhat similar to finding an article in agriculture, and using the article is also similar in that it is read, ingested, and its content communicated. It is not quite the same in the data world; there is no “EBSCO for data.”

A challenge in scoping data services is to consider the overall size of the task. Data Services as envisioned is concerned with both front-end and backend services. This in itself is beyond the traditional realm of subject librarians, who may be involved on the backend with collections development and various planning and coordination activities etc., but who are not typically involved with administering technical infrastructure, software licensing, etc. There are both front-facing and backend Data Services at Cal Poly. Front-facing services include the Data Studio (computer lab/physical space), outreach, instruction, and reference. Backend services include data set storage, software and licensing, collections, and strategic planning. The Cal Poly student body consists of over 95% undergraduates. Figure 1 and Table 1 illustrate the proportion of students within the five colleges and the focus on front-facing services more than on backend. Cal Poly has been expanding front-facing services, but in order to effectively expand these front-facing services it is necessary to expand the backend services in support, while recognizing domain-specific data areas and needs.
Figure 1. Front-facing and Backend Service Levels Over Various Subject Areas.

<table>
<thead>
<tr>
<th>College</th>
<th>Undergrad</th>
<th>Post Bac./Grad</th>
<th>Total</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture, Food &amp; Env. Sciences</td>
<td>3,644</td>
<td>84</td>
<td>3,728</td>
<td>20%</td>
</tr>
<tr>
<td>Architecture &amp; Env. Design</td>
<td>1,453</td>
<td>80</td>
<td>1,533</td>
<td>8%</td>
</tr>
<tr>
<td>Business</td>
<td>2,297</td>
<td>37</td>
<td>2,334</td>
<td>12%</td>
</tr>
<tr>
<td>Engineering</td>
<td>5,007</td>
<td>404</td>
<td>5,411</td>
<td>29%</td>
</tr>
<tr>
<td>Liberal Arts</td>
<td>2,623</td>
<td>108</td>
<td>2,731</td>
<td>15%</td>
</tr>
<tr>
<td>Science &amp; Mathematics</td>
<td>2,575</td>
<td>273</td>
<td>2,848</td>
<td>15%</td>
</tr>
<tr>
<td>Others</td>
<td>81</td>
<td>13</td>
<td>94</td>
<td>1%</td>
</tr>
<tr>
<td>Total</td>
<td>17,680</td>
<td>999</td>
<td>18,679</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1. Number Of Undergraduates and Graduates Per College and As Percentage Of Student Body.

On the front end, Data Services has been envisioned in line with the traditional library\subject specialist roles of reference and instruction, in addition to more involved programming and outreach (ex. The Data Studio Presents speaker and workshop series, Open House, Geography Awareness Week, GIS Day). The nature of reference in Data Services is also different; where traditional reference is primarily concerned with finding appropriate resources - which Data Services serves - it also deals with working through technical challenges in utilizing available resources. Further, consultation on methodological development for students and faculty seeking to apply data to a particular project or question falls under the scope of Data Services. Finally, in contrast to the traditional model of subject librarians, Data Services staff are tasked with providing resources to support the entire breadth of disciplines on campus. The breadth of the subject specific needs is daunting at Cal Poly. Understanding the technical data-related needs, space, resources, and literacy components that would benefit the wide range of subject and discipline-related activities, and the wide range of existing student technical abilities are still questions the Data Services program needs to address.
It is critical to understand who Data Services’ patrons and stakeholders are, what their data needs are, how those needs vary by discipline, how services can be developed to meet those needs, and in light of this, how the scope of data services can be refined to provide the greatest benefit to patrons. This is part of a campus conversation on data and quantitative literacy at Cal Poly. Data Services within the library is not going to answer this question independently, but it is a logical and significant resource in supporting campus-wide data and quantitative literacy efforts.

There is a lot that libraries in general are seeking in the area of data services. Properly aligning expectations and actual service capabilities in this new area is not necessarily a given. By reviewing other data services programs, libraries can help scope their capabilities for service. Cal Poly reviewed more than 30 institutions. There are elements of different services that appear similar: data and GIS lab spaces, peer assistants, speaker series events, technical workshops, etc. Implementing Data Services at Cal Poly requires understanding resources to find and develop a model that works. What has been learned so far from reviewing the data services landscape is that the defined breadth and depth of data services at other institutions is wide. There are one to more than 70 staff serving various disciplines at different levels of service. An important take-away is that effective data service is a program, not a position.

Self-assessment is required before planning a program. At Cal Poly four staff members have some level of data curation, data management, and GIS skills sets. Cal Poly does not have an ARL library and cannot serve the campus community in the same way or at the same level of staffing as those institutions. A review of reference statistics revealed that 10% of all recorded reference questions have to do with GIS or data. This indicates a need for internal workshops, campus training, and continued Data Services staff trainings. Opportunities presented themselves when the Library was integrated into IS, while opportunities were sought through trainings geared to the Office of Research and Economic Development. The biggest obstacle at this time is lack of resources and unsustainable growth. Due to this Cal Poly chose to provide data literacy support to faculty, graduate, and undergraduate students in lieu of supporting institutional data curation and Big Data.

Program Development Highlights 2009-2013
Since the program’s inception in 2009, there have been a number of changes and highlights that show where the Data Services program is now. In 2009 a survey was deployed at Cal Poly, “A study of faculty data curation behaviors and attitudes at a teaching-centered university” and the results were published in College & Research Libraries (C&RL), July 2012 (Scaramozzino et al. 2012). In 2010 a data management LibGuide was developed (http://libguides.calpoly.edu/data) and basic training was offered to campus collaborators, including the Office of Research and Economic Development. In 2011 another survey of faculty was deployed, “A study of faculty data service needs at a teaching-centered university” (manuscript in progress), to determine a baseline for scientists’ awareness and interest in library based infrastructure. In 2012 an area of the library was renovated and the Data Studio was created; a Data Services Working Group was convened resulting in the Data Services Program - First Year Program Report and the writing of Cal Poly’s Digital Scholarship Program - A Proposal; and the first Annual Data Studio Open House was held.

The introduction to the campus of the Library’s ideas about data services included having Peter Wiley Booth (the major Data Studio renovation donor), the President, Provost, and CIO cutting the ribbon to the Data Studio during the first Annual Open House. The event also included a showcase of student and faculty work. New York Times Technology Editor Quentin Hardy’s plenary evolved into The Data Studio
Presents, a lecture and workshop series on topics regarding GIS, data management, digital humanities, and more (see http://libguides.calpoly.edu/datastudio for additional information).

In 2013 a full-time Numeric and Spatial Data Specialist was hired and conducted an environmental scan of more than 30 institutions and universities with data services plus the other 22 CSU campuses. The scope of services and what can be accomplished by libraries can be informed by the examples available – there are individual spatial and numeric specialists providing basic access to social science data all the way to a 70 person campus-wide educational and research ITS groups. The Data Studio Presents events are now expanding to include STEAM (Science, Technology, Engineering, Art, Math) focused events. The Data Studio as place continues to be a space for students to experiment with technology and to collaborate.

Next Steps and Conclusion
Data Literacy Instruction
There are a lot of data out there, free and accessible to anyone, and there are general strategies for finding data. Students and faculty don’t “need” the library to gain access to these datasets in the same way as journals that are purchased through an institutional subscription or physical books are housed in the library. However, the data are not well organized, at least not to the same degree that library collections are organized. Data literacy instruction requires the ability to create information, use the information, analyze the data, share data, manage data, and collaborate with others regarding the data.

Many students are generally comfortable finding, reading, and ingesting information from a journal article but the same is not usually true for raw datasets. Students often have limited experience dealing with datasets they have created. This experience is not necessarily transferrable to finding other datasets and dealing with others’ data. Learning to find, view, manipulate, and perform analysis on datasets is a skillset like any other and takes time to develop. It takes practice, work, instruction, and curiosity. It takes multiple passes over the same ground. In short, it is a great deal more than “access” to enable students to work well with data.

Ideally the purpose of Data Services in the long term is to provide a platform for cultivating inquiry. Data literacy will be part of a larger “data sandbox” effort to help students communicate through writing, visualization, and more. The idea is that students should be met and helped whatever they are on the data literacy continuum. Are they learning to use a calculator, exploring Excel™, learning scripting, or creating web services? This will be accomplished in a comprehensive “Learning Lab” staffed in collaboration with the University Writing & Rhetoric Center, Statistics Department, and Graphics Communication Department and their peer assistants. The hope is that each student regardless of field will become an educated consumer, producer, and communicator of data, which requires learning to find and use data effectively. Cal Poly is currently defining the necessary skill sets needed for the data literate undergraduate student.

This goal is bigger than the library but the Library can have a large part to play. Common resources in the Library include datasets, software, computing, visualization/sharing/collaboration spaces and some quantitative expertise. Making these resources available to all students is a major step forward. The idea is supporting “data for non-majors” (e.g., non-statistics majors). The use of data is not a traditional strength of libraries; this is largely new territory, so it’s a learning process. It includes bringing in people who can build connections (e.g. a service\education-minded-data specialist) and engaging new campus partners to add their expertise.
Strategic Planning
Data Services has begun preparing a 2014 strategic planning effort to engage Library and campus constituents in defining how the program can best meet identified needs. Through the formalization of a vision statement and the development of a strategic plan, staff will continue to build support and working relationships within the Library, IS, institutional and community stakeholders (e.g. GIS working group, Big Data interests, campus researchers) to more clearly define the roles Data Services may play in supporting student, staff and faculty success. This will also allow for additional advancement and development opportunities.

Final Words
How institutions decide to define their roles in supporting this process at their institutions will vary greatly. Self-assessment is critical to how an institution’s data services develop. The major reason for the success of Cal Poly’s program has been administrative buy-in and support. Making sure that there are appropriate resources and support is critical.

References