

UC Merced Library Strategic Agenda to Support Data-Intensive Research

The UC Merced Library participated in the 2012-13 ARL/DLF/Duraspace E-Science Institute, as part of a cohort of twenty-one institutions, to develop a strategic agenda for the library in support of data-intensive research on campus. The Library conducted interviews with key campus stakeholders to assess campus directions and needs. This document presents findings on the institutional context and proposed next steps.

Team members:

Donald Barclay, Interim University Librarian
Susan Borda, Digital Curation Librarian
Susan Carter, Director of Research Development Services, Office of Research
Emily Lin, Head of Digital Assets, Library
Samuel Traina, Vice Chancellor for Research

I. Background

Opening its doors in 2005, the University of California is the tenth campus in the UC system and the first U.S. research university of the twenty-first century. For 2012-2013 the campus has about 5,760 students and 153 ladder-rank faculty. In eight to ten years, the campus expects to grow to 10,000 students and 350 ladder-rank faculty. The campus has no academic departments; instead, there are three schools (Engineering; Natural Sciences; and Social Sciences, Humanities and the Arts) and a strong commitment to interdisciplinary studies. There are currently two organized research units on campus that serve as nexuses for interdisciplinary research: the Sierra Nevada Research Institute and the newly established Health Sciences Research Institute.

The current Chancellor and all three deans have been in place for less than two years; a new Provost arrived in December 2013; the campus is beginning a search for a permanent CIO and has an interim University Librarian. The University of California's historical emphasis on shared governance means that the campus has a strong Academic Senate with considerable power to shape campus programs and policies. The campus leadership has embarked on a strategic focusing initiative to direct faculty hiring in selected areas of excellence, with the aim of attaining "Research High" status in the next eight to ten years. There is a push to increase graduate enrollment to a minimum of 10-20% and to improve the student to faculty ratio. UC Merced has a strong STEM emphasis in both research and teaching and expects to maintain that emphasis of 50% or more students majoring in STEM. As with all institutions of public higher education in California, UC Merced has been constrained by shrinking state support for higher education.

The interviews the library conducted with various campus leaders and faculty show that leadership and infrastructure are wanted for managing data, and the deans and VC of Research see a role for the library as a “core facility” for research. There is a deep distrust of campus IT, a unit that has shown no interest in supporting research computing under interim leadership. There is already strong recognition on the campus that the Library places emphasis on “digital” and there is support for the library to serve as a clearinghouse for data. A significant proportion of the faculty are junior faculty who may not yet have complete training on or experience with the issues surrounding research data.

Strengths:

- Library has strong reputation of customer services and user orientation and has the moral support of administration to take the lead on data management
- Library has good relationship with the Office of Research; has conducted a survey of campus data management needs and submitted white paper to VC of Research and Provost to establish baselines awareness of needs
- New digital curation librarian brings skills and focus on developing data management services

Weaknesses

- Library lacks breadth and depth of technical expertise: visualization, GIS, programming, ontologies, data analysis
- Library lacks tools and resources for management: data clean-up, processing, description; could not just create a repository
- Library lacks well-established channels for external communication and marketing.
- Library has a broad mission without new resources: FTE, funding, infrastructure, space.

Opportunities

- New Provost/EVC comes from the NSF (Asst. Director for Engineering directorate), is well-versed in data management requirements, and brings emphasis on Big Data
- The VCR supports the library taking the lead on data management for the campus
- Several Deans/Directors alluded to potential possibilities and projects for library involvement:
 - Dean of Engineering sees a role for library to deliver top-notch data management plans
 - Deans of Natural Sciences, SSHA see a role for library to serve as data repository and “middleware” (providing data services)
 - HSRI director sees value in collecting longitudinal data about Central Valley population and the library having key role in collecting and disseminating such a database

- SNRI director would like to see data services that would free up faculty time

Threats

- Campus IT strategy is not yet defined and it is difficult to know the scope of consolidation, impact and roles/responsibilities across boundaries.
- The campus lacks critical infrastructure and faces continued limited resources or continued squeeze of funding.
- Without carrot/stick from funding agencies, researchers would not even bother with archiving data or prioritize data management, or buy into the library role.
- Faculty are overwhelmed with electronic communications. The library needs to invest considerable effort in individual follow-up and diversified communication channels, and this will further stretch limited resources.

II. Potential Opportunities that Align with Strategic Priorities

Data Curation

Mike Dawson (Natural Sciences) reached out to the library for support of his NSF-funded Marine Lakes Ecosystems research project, based on prior working relationship with the library. The library is embarking on a pilot to provide data management services to support this project by identifying a method to facilitate sharing and dissemination of data sets, advising on best practices for file organization and metadata, and supporting final archiving.

The library is seeking out other pilot opportunities as a means of developing its services and building expertise and stature. These should tie directly to the campus foci/strategic directions to garner ongoing support. Strategic opportunities include discussions with directors of the Health Sciences Research Institute and the Sierra Nevada Research Institute about:

- Developing a data collection with Health Sciences Research Institute.
- Supporting data management plans for the SNRI Critical Zone Observatory.

There appears to be an emerging interest in digital humanities, which the library should continue to investigate. The library should strengthen communication with Humanities and World Cultures faculty and promote its digital expertise and resources.

Education and Outreach

The Office of Research and the library are planning to collaborate on training faculty regarding data management policies, planning and practices, including integrating the library into Office of Research orientation for new faculty as well as the faculty grant-writing institute and Research Week programs.

The library and the Office of Research will pursue co-sponsoring a “Big Data” institute, with incentives for faculty attendance such as mini-grants or competitive spots. (Example: <http://orci.research.umich.edu/cidays/>)

The library is going to do more direct meeting with faculty (especially new faculty) to discover their needs and inform them of services and facilities available to them locally, via the UC system, and nationally.

The University Librarian will seek to advocate and communicate about strategic agenda within the Academic Senate.

Facilities

Library will seek a role in advising how the campus will support research computing. The library is planning to acquire storage space in the cloud that we will use to store small data sets produced by faculty research.

The library will communicate research infrastructure needs and costs to the Provost, the VC for Budget and Planning, other campus decision-makers. The library needs to reach out to campus IT and establish a good relationship with the new CIO.

Library will continue to position itself to take advantage of system-wide and national/international resources for digital curation. The library will pursue opportunities to collaborate with other UC campuses and California Digital Library.

Public Dissemination

The library and Office of Research will explore the aggregation of research publications, data sets, grants information, faculty profiles via a research management system.

III. Risk Assessment

If the library does not align itself with the campus research agenda, it will lose relevance and not be viewed as a core research facility. Researchers will continue to expend institutional and extramural resources to support individual projects or needs without considering aggregate or long-term costs.

The provost and campus administration do not want to commit to decisions without a new CIO in place, and it will take time to develop a vision and strategy for campus information technology. There is risk that IT/computing resources will continue to be decentralized, expenses will be redundant, systems will lack interoperability, and data will be lost.

There is risk that pilot projects do not succeed or are not seen as significant. Faculty may not see the significance or need for library services, or funding agencies may change direction and not emphasize or place resources in support of data management.

Unexpected technological developments or unanticipated solutions may make current approaches obsolete. The institution may underestimate the costs of building research infrastructure and capacity. The institution may wait for or depend on grant funding and lose opportunities.

There is no institutional umbrella strategy for data security and no follow-through on policies.

IV. Organizational Implications

Staff capacity

The library does not have full data curation expertise. The library needs to identify what expertise it really needs and expand our capacity through partnerships.

- Inventory skills, identify gaps and prioritize needs
- Develop “in-service” training for librarians, staff
- Pursue opportunity for bringing on board postdoctoral fellows, data scientist

Tools and resources

The library needs to approach the issue in terms of “why invest in data curation?” and explain the return on investment. The library needs to develop a business plan for long-term sustainability of services.

Identify analytics and other tools for workflow, metadata, visualization, etc. as well as opportunities to partner in development for these tools.

Institutional partners

The library needs to communicate with the new EVC/Provost to clarify what the campus process for strategic priorities will entail and how the library’s strategic agenda will align.

The library needs to educate the Interim VC for Budget and Planning on data curation needs and research computing infrastructure needs so that they are included in resource and campus planning.

The library will work with directors of research institutes to identify pilots. Other institutional contacts to pursue include Anand Gadre, Director of the Stem Cell Foundry, who reports to Office of Research. The library should expand and strengthen its range of external contacts to include regional institutions such as CSU Fresno (strong biochemistry department), CSU Stanislaus, and peer institutions.

V. Next steps

- Meet with VC Research in early January to review budget proposal and strategic agenda, including specific collaborations with Office of Research:

faculty “data institute,” research management system (e.g. Symplectic demo), potential Research Hub pilot

- Create a visual representation for strategic agenda.
- Determine date, initial plan for Research Week program on data management; schedule with UCSD and UCB by mid January.
- Determine program plans for Grant Writing Institute sponsored by Research Development Services in the Office of Research by February.
- Schedule a meeting with Tom Peterson for mid January with Sam Traina and ESI team.
- UL will present strategic agenda to Academic Senate for input.
- Survey faculty affiliated with HSRI and SNRI (through executive directors) to identify vital datasets and opportunities to provide data management services. Individuals also include
 - SSHA:
 - Will Shadish (Quant)
 - Holley Moyes (tools for visualization)
 - Mario SiFuentes (oral histories)
 - Ruth Mostern (SPARC)
 - Ana Song (anonymized)
 - Matt Hibbing (voting patterns/behavior)
 - Chris Kello, Mike Spivey (perception data on eye tracking)
 - SoE:
 - Lilian Davila
 - Gerardo Diaz
 - Tom Harmon
 - Florin Rusu
 - NS:
 - Teamrat Ghezzehei
 - Linda Hirst
- Work with AVC Research (Motton) on opportunities for graduate student training, specifically QSB course; research compliance.
- Write up article on ESI experience for *Panorama*, campus communication.