November 12, 2014

TO: Harrison Dekker (Library Data Lab) - Chair
  Jim Church (Economics)
  Lynne Grigsby (Systems)
  Trina Pundurs (Catalog & Metadata Services)
  Lydia Peterson (Licensing)
  Susan Powell (Physical Sciences)
  Hilary Schiraldi (Business)
  Elliott Smith (Life Sciences)

FR: Jean McKenzie
RE: Working Group to Develop a Program to Acquire Datasets

Context
The Library established a Data Lab in 2008 to respond to requests for numeric and spatial data resources, mostly from social science disciplines. Some resources have been purchased each year, funded by a small data fund or by selector funds when possible. Demand for more numeric and spatial data sets from all parts of the world has outstripped the capacity of those funding sources and disciplines may be unevenly served for a variety of reasons. Numeric and spatial data, and possibly other data resources, are increasingly required for the complex research questions that faculty and students are asking.

In addition, multifarious conversations and activities regarding data are swirling among funding agencies, information organizations, and academic libraries. At our local level, librarians provide access and training for researchers using Data Lab resources; a group of librarians working with Social Sciences D-Lab faculty and staff has produced a focused report on restricted use data; and there is a new campuswide “Big Data” faculty research group housed in the Library. Librarians are already asked for advice on how to find, use, publish, store, re-use, and preserve data and some have expressed a need for professional development and training opportunities to better answer data questions.

To respond to the demand across all disciplines, as well as provide development opportunities, the AUL for Collections will establish an initial central pool of funds from which selectors can apply for data purchases.

Goals of the Program for Acquiring Unrestricted Data
• Respond to the needs of campus researchers by identifying and acquiring core commercially available numeric and spatial datasets
• Create an innovative new campuswide program for library – enhance the offerings of the Data Lab
• Provide new learning opportunities for library personnel (new roles in learning about data, instructing on use, possibly becoming more closely involved in research process – consultation on other aspects of data literacy)
• Develop researcher relationships, for example, by asking them to propose and evaluate data sets to purchase and return later to talk about how the data were to do new research and reach new insights

Charge
The AUL for Collections is charging a Working Group with developing this program, including criteria for eligibility and developing a process and timeline for calls for proposals. The final report should answer the questions below and identify anything that would stop the program from being implemented relatively quickly.

1. Set the parameters of the program:
   a. Assume that we are including both numeric and non-numeric datasets that are commercially available.
   b. Identify how different kinds of data might require different kinds of procedures and services.
   c. Discuss the difference between real-time computing and not.
   d. Establish size limits for datasets that can be acquired and stored locally.
   e. Discuss how this is not a program for faculty-created data.
   f. Provide definitions for terms that may be new to librarians and library staff and users.

2) Environmental scan of other research libraries that have programs for acquiring and supporting data and describe the elements that are most appealing.

3) Outline options and recommendations for the best practice in the following areas of a data purchase program, considering that The Library might choose to participate in some of these areas for some products and not for others:
   a. Eligibility
      i. What are the criteria for eligibility? Who can make proposals? (from librarian liaisons only and/or direct from users?)
   b. Proposal content
      i. Any parameters on kinds of data that will be considered?
      ii. Who can use the data (preference for data sets that can be used by all UCB faculty, students, and staff? Or not?)
      iii. Require or request any cost sharing?
      iv. Etc.
   c. Licensing
      i. What are the crucial elements of a license for data?
ii. Would we consider sharing licensing terms with faculty for whom we will not do the acquisition?

d. Acquisition
   i. What would we pay for, and what is outside our mission/scope?
   ii. How much are we willing to pay for data for what size audience? If we were to establish a process for receiving proposals, what size annual budget would make this viable?
   iii. When would we be purchasing the data directly and when might we secure access via memberships?
   iv. When would we pay more for real-time computing?
   v. Are different procedures needed in Ordering, and ERU for these products?
   vi. How are data sets received and where are they stored?

e. Systems support
   i. Are different procedures needed in Systems, if ingesting content is required? Is the Materials Acquired Digitally (MAD) process useful/adequate here?
   ii. What implications might there be when buying access to data that is “in the cloud”? When supporting real-time computing?
   iii. How would data be backed up?
   iv. What options do we have for preservation?

f. Discoverability and Cataloging
   i. Are different procedures needed in Cataloging for these materials?
   ii. Other ways to make these materials discoverable? Search widget?

g. Tools needed to use the data; what should the library buy, and what should the library provide training for?

h. Publicity
   i. How to broadly publicize the data purchasing program?
   ii. How to publicize the Library’s ongoing services regarding data resources?

Timeline
Phased implementation? What’s the low hanging fruit? What can we purchase this year? What elements of a data program would be possible this year? What should be considered phase 2?

Plan to have a preliminary report to the AUL for Collections by February 1, 2015.

---

1. “Report of the Restricted Use Data Task Force.” Distributed to library staff September 5, 2014. Members of this group are working with Social Sciences D-Lab to develop a program to for restricted use data (RUD). When completed, this program may be extended to cover the purchase of RUD.