DATA 02 DataNet: Data Conservancy (UCLA-DC).

Team Members
- Christine Borgman, Faculty, PI*
- David Fearon, Graduate Student*
- Sharon Traweek, Faculty
- Laura Wynholds, Graduate Student

* Primary Contact

Overview
The Data Conservancy is a five-year project lead by Johns Hopkins University, funded by NSF’s DataNet initiative by the Office of Cyberinfrastructure. A primary goal of DC is development of a repository that will archive data and facilitate collaborative access for a range of sciences. Its design will be based upon social science research of data curation and collaboration practices for science communities who are likely users of the DC repository. UCLA’s researchers, administered through CENS, will conduct research on practices and data curation requirements for astronomy and astrophysics. The outcomes of their research will contribute to the design of the Data Conservancy system architecture.

Approach
UCLA Data Conservancy (UCLA-DC) is coordinating with social science teams at The Center for Informatics Research in Science and Scholarship (CIRSS) at University of Illinois, Urbana-Champaign, and the National Center for Atmospheric Research (NCAR), who are studying data practices and needs for a range of science communities expected to contribute their data to the Data Conservancy and who will be users of DC’s products and services. UCLA-DC is studying scientific data practices and data curation requirements for the fields of astronomy and astrophysics. We are examining initially the Sloan Digital Sky Survey (SDSS), and its relation to two subsequent sky survey projects: the Large Synoptic Survey Telescope (LSST) and the Pan-STARRS project. Also examined are the Infrared Processing and Analysis Center (IPaC), the International Virtual Observatory Alliance (IVOA) and Space Telescope Science Institute Archives (STScI) for relevant issues in data practices and metadata standards for astronomical objects and digital archives. Additionally, we are conducting a series of interviews with astronomers and astrophysicists, including university faculty, postdoctoral and graduate students, and data center staff. Core questions for astronomers are focused on data sources, usage, and preservation. Investigators will determine which forms of data are used, which are selected for sharing and curation and which are discarded, how they are curated, and expected future uses of curated data. In addition to interviews, we are conducting a range of qualitative methods to address these goals including analysis of documented history of projects, observations of labs and workstations, and social network analysis of the involvement of key participants over time.

System(s) Description and/or Experiments
One central activity of the project will be the building and analysis of a relational database of documentation on project sites. The database will track project history, funding, personnel, timelines, and relationships among projects. The database will also incorporate for analysis data generated by our other methods. The database is built on an SQL network platform with a user interface operating in FileMaker Server.

Accomplishments
Specific accomplishments during the reporting period:
- Hired GSR David Fearon for a full-time post-doctoral staff position for the project year 2 with Jim Gray Gift funding from Microsoft Research
- Acquired and began analysis of the Sloan Digital Sky Survey listserv archive held by the Data Conservancy project.
- UCLA GSR Laura Wynholds met with the Data Conservancy CIRSS project partners at UIUC, before iConference at Urbana-Champaign. Contributed to collaborative partnership between UIUC and UCLA.
- Completed 25 interviews, most with faculty astronomers in California, including senior astronomers on prominent projects. These complete the "round-one" interviews, 32 total.
- Completed initial round-one coding and analysis of interviews.
- Round-one Analysis products included two conference posters, a white paper for meetings with Data Conservancy project partners, and a conference paper currently in review.
Co-PI Sharon Traweek spent much of August at the KEK High Energy Accelerator Research Organization in Tsukuba, Japan, engaging with several astrophysicists and gathering background on data curation in Japan.

Held two half-day meetings with astronomer Alyssa Goodman, discussing astronomy data curation, publication data linking.

Future Directions
We are beginning our second round of interviews in spring 2011, continuing through summer, primarily with Southern California astronomy/astrophysics institutions, will continue. Interviews will focus in particular on the preservation of recently published data, linking of data to publications, and integration of archived sources for data-intensive research. Concurrently, we will be continuing our document and listserv archive about SDSS development and other project sites, and extending integration of our project databases with our interview and ethnographic data, and social network analysis capacities. We will be coordinating activities with CIRSS, NCAR, MBL, and Microsoft partners.

All empirical results, use cases, and accounts of scientists’ data practices and conceptualizations of data will be provided to the Data Conservancy Data Practices and Data Concepts groups with regular working sessions for joint interpretation. The UCLA-DC team will also collaborate with DC researchers at CIRSS, NCAR, and MBL to integrate related findings, sharing results and interpretations of findings through the project wiki, conference calls, and scheduled face-to-face meetings. The overall analysis will produce a taxonomy of data practices and data attributes for assessment of curation needs and to facilitate curation activities.