SDP 02 Towards a Virtual Organization for Data Cyberinfrastructure

SDP 02.1 People

- Principal Investigator: Christine Borgman
- Faculty: Christine Borgman (UCLA, PI), Geoffrey Bowker (Santa Clara University, Co-PI), Thomas Finholt (University of Michigan, Co-PI)
- Researchers: Katie Vann (SCU)
- Graduate Students: Jillian Wallis (UCLA), David Fearon (UCLA) Mouly Kumaraswamy (UMich)
- Undergraduates: Victor Quintanar-Zilinskas (SCU)

SDP 02.2 Overview

Our research work is primarily directed towards the design and development of tools and to allow efficient and durable management of data coming from embedded sensor networks. CENS, being an inter-disciplinary and multi-institutional center of innovation, is a convenient testbed to develop an integrated data management model as it fully encompasses the richness, heterogeneity and quantitative extent of current sensor network observations and their conceptual fragmentation across different academic groups and departments. Our research group is especially concerned with scientists’ and engineers’ data practices and their implications for science and education. Most of the group’s work is devoted to research on data sharing, usability, preservation, accessibility and interoperability.

As part of our on-going research agenda to understand CENS as a collaboratory and the cyberinfrastructure necessary to support such a collaboratory we have joined with colleagues from University of Michigan, Santa Clara University, and Georgetown University to develop a comparative study of CENS and other collaboratories. Our long-term goal is to establish a CI Virtual Observatory for the study of data, data analysis, and visualization. Our short-term goal (in the time period of a one-year Small Grant for Exploratory Research (SGER)) is to conduct the background research to develop such a virtual organization, to identify the most effective models for a virtual organization, and to identify the research questions about data, data analysis, and visualization that must be addressed to construct a viable cyberinfrastructure. The promised outcomes of our work are:

- To produce fundamental social science about infrastructure development for virtual organizations and for data;
- To create a functioning virtual organization with a specific set of goals to study data practice and data policy;
- To identify methods for formative evaluation of CI efforts on data, data analysis, and visualization that will be enriched by continual comparisons between projects.

SDP 02.3 Approach

Our virtual organization began with the writing of the SGER proposal, and has been extended in two convergent directions. First, we have conducted the initial stages of the project (literature review and development of methods and instruments) via distributed collaboration technologies that link the three university sites (UCLA, University of Michigan, Santa Clara University). Our use of technologies such as CTOOLS (a University of Michigan implementation of SAKAI), DimDim (web conferencing tools), audio conference services, videoconferencing (iChat and Skype), Google Docs, and Basecamp (a low-cost commercial site for sharing files, sending messages, whiteboard space) have enabled us to be participant-observers in our own collaboratory. The content of these distributed meetings is divided between administrative tasks and shared reading and discussion of relevant research. While each of these tools has strengths and weaknesses, Basecamp and audioconferencing are the most robust. We will continue to work with these tools as they mature. Recently, Thomson-Reuters made us a gift of Endnote licenses for use in testing its robustness for distributed collaboration. We have now settled into our chosen products and use them regularly for communications and to add items to the communal store.
The second direction is to extend our virtual organization through larger and longer grants to carry this research beyond the SGER award. We used the same distributed technologies noted above to develop these collaborations, adding more investigators from the same universities. Two of the proposals we submitted in early 2008 received funding. These are (1) HSD AOC – Monitoring, Modeling & Memory: Dynamics of Data and Knowledge in Scientific Cyberinfrastructures (Paul N. Edwards, UM, PI; Co-Pls Borgman, UCLA; Bowker, SCU; Thomas Finholt, UM; Steven Jackson, UM: David Ribes, Georgetown; Susan Leigh Star, SCU) and (2) VOSS Mapping Uptake of VO Technologies (Finholt, UM, PI; Co-Pls Jackson, UM; Ribes, Georgetown). These grant writing activities, and subsequently the awards, lead to the establishment of regular meetings between collaborating sites, both in-person and virtual. We held a 2-day in-person meeting of the investigators and graduate students of these three projects in December at Santa Clara University (with 2 participants joining by videoconference).

SDP 02.4 System(s) Description and/or Experiments
Part of the work stipulated in this research project was to administer a small program of research on data practices across multiple cyberinfrastructure projects. Part of our collaborative effort has been the development of a shared research instrument, with questions of interest to all of the investigators, and which capture a baseline of data to motivate future rounds of interviews. This instrument also can serve as exploratory research for the longer-term grants to follow.

Once our instrument was agreed upon, all institutions applied for and received full IRB approval. IRB approvals were received in November and December across the three participating universities. While awaiting approval of our instruments, we developed a shared sampling method based on centrality measures of participants in our research sites. Since we received our administrative approval we have contacted our sampled participants and begun interviews. Of the 40 total interviews to be conducted across the 3 sites, the majority of interviews have been conducted. We are now in the midst of data transcription and analysis.

SDP 02.5 Accomplishments
- Developed a cross-institutional research instrument and method
- Developed a sampling method based on author centrality measures and applied it to our research populations
- Obtained IRB approval of our study at all sites
- Invited participants and collected interview data
- Developed a shared method for data analysis
- Initiated data analysis

SDP 02.6 Future Directions
Over the next year we need to complete data collection, transcription, and analysis. Additionally we need to demonstrate the viability of this method for use within a cascading research organization that would allow future research to scale to meet the needs of both our growing virtual organization and the growing cyberinfrastructure projects we are studying.

SDP 02.7 External Research Partnerships
Anita Borg Institute (current)

This project is funded by NSF award #OCI-0750529, "Towards a Virtual Organization for Data Cyberinfrastructure". Christine L. Borgman UCLA, PI; Geoffrey Bowker, Santa Clara University, Co-PI; Thomas Finholt, University of Michigan, Co-PI. Start date: 2/1/08; End date: 1/31/09.