

2.7 Research Testbeds and Data Practices (DATA)

These projects are both testbeds and empirical research; the common theme is collaboration, both local and distributed. We are exploring multiple aspects of cyberinfrastructure requirements for embedded sensor networks and other types of multi-disciplinary science and technology research. Funding sources include NSF, Microsoft, CENS, and the University of California.

Monitoring, Modeling, and Memory is a distributed collaboration with partners at the University of Michigan, University of Pittsburgh, and Georgetown University. We are comparing data practices across multiple distributed sci-tech collaborations, including CENS, Long Term Ecological Research Network (LTER), and the Earth System Modeling Framework. This is largely a social science collaboration of investigators studying science. We are in year three of three.

The Data Conservancy, which is a very large distributed collaboration based at Johns Hopkins University, addresses the curation of scientific observations. It is a partnership of scientists, social scientists, and technologists to build systems and to study data curation in multiple fields. Our focus is on astronomy, which is outside the realm of CENS per se, but draws upon the theory and methods applied to other studies of CENS data practices. We are in year two of five.

Object Reuse and Exchange is a technology project that builds upon our earlier work in CENS data practices, extending it to deploy new technical standards to represent and to link research objects. We partnered with leads in ORE development at the Los Alamos National Laboratory. This project is in year 4 of 4. Funding for this research came primarily from Microsoft Research.

The CENS Deployment Center captures knowledge of CENS research activities and makes those records reusable for future deployments by the same and other CENS teams. The eScholarship project captures CENS publications. The Data Discovery Library and the CENS Annual Report System both were outcomes of these projects, which began with a seed grant from CENS funding. While CENS DC per se is ending, the follow-on projects continue, and a grant proposal is in progress to transfer stewardship of the CENS data and publications to the UCLA library.

The mobile computing project, also funded by Microsoft, developed a scientific data collection module for the Microsoft phone platform. While technology development ended when Microsoft made a major platform shift that was not forward compatible, we gathered useful empirical data on microblogging for scientific applications.

The multiple CENS testbeds are exemplars of the collaborative cyberinfrastructure built by CENS teams. These testbeds yield empirical data and are available to other teams to extend their work.

Theses and Dissertations

MMM, Deployment Center, and eScholarship: These three projects together are the basis for the dissertations of Matthew Mayernik, who will complete his degree in June, 2011, and Jillian Wallis, who defended her proposal in March, 2011, and plans to finish in June, 2012.

Object Reuse and Exchange: Alberto Pepe, who completed his dissertation in June, 2010, extended the ORE project with additional study of social networks in CENS.

Mobile scientific data collection: Research done jointly by Mayernik and Pepe contributed to both of their dissertations.

Data Conservancy: David Fearon's master's degree project is based on his work with astronomy data from this project.

Awards and Prizes

In this reporting year, our team has received multiple awards:

- Matthew Mayernik: Best Dissertation Proposal Award, UCLA Department of Information Studies
- Alberto Pepe: Best Dissertation Award, American Society for Information Science and Technology
- Jillian Wallis: Team lead, Winner of Student Design Competition, American Society for Information Science and Technology
- Laura Wynholds: Inaugural Best Student Paper Award, International Digital Curation Conference