

2.8 Statistics and Data Practices (SDP)

Statistics and data practices have transitioned from relatively “offline” to “critical real time” components of CENS systems and practices. Our developments in this past year are in the context of mobile and social computing systems and designing for data quality, use and reuse.

Mobile devices and social computing

Three projects employ mobile phone technology for wireless sensing applications in very different ways. *your.flowingdata* is a participatory sensing application that can be used for health monitoring and a variety of social activities (Figure 6). What’s Invasive is a participatory sensing application for citizen science that enables individuals to take geotagged images of invasive species and contribute them to a shared database for tracking and control of these plants. The Mobile Scientific Data Collection project is an application to assist scientific teams in collecting data in the field more cleanly, quickly, and simply than with traditional notebook devices. *your.flowingdata* is built directly on the Twitter platform, aggregating tweets from individual users into a database, and providing visualization tools. The Mobile Scientific Data Collection project also is based on micro-blogging, although not the Twitter platform, and enables scientists in the field to document their data and activities with short statements.

Capturing cleaner data

Several projects aim to improve the data capture and management process. The Anomaly Detection project enables scientists to identify patterns in sensing data in real time, thereby allowing them to modify their methods in the field and to correct for errors in faulty sensors. The CENS Deployment Center captures information about deployment activities, ranging from planning and equipment to documentation on the successes and failures of those activities. The social and technical knowledge gathered in the CENS DC provides context for data collected in the field, thus improving the reliability of current and future deployments. Similarly, the Mobile Scientific Data Collection project allows the real time capture of documentation on data and activities.



Figure 6. *your.flowing data* website.

Data use and reuse

By capturing cleaner data earlier in the scientific and technical life cycle, CENS data become more usable by the teams that produced them and more amenable to reuse by others. The Monitoring, Modeling, and Memory project is studying the entire life cycle of CENS data, as part of a four-university partnership. CENS data activities are being compared to several other large, distributed, collaborative scientific projects. The results will provide guidance on best practices to each of the participating centers. CENS publications, along with posters, working papers, and other scholarly products, are being deposited or recorded in the eScholarship repository of the University of California. CENS is now one of the largest of the eScholarship sub-repositories. Our scholarly products are now exposed to the

larger online community, and readily discoverable by search engines. Upon the request of the 2009 NSF Site Visiting Team to expose our datasets as well, we launched the Data Discovery Library project. The Data Discovery Library is being constructed as a layer on the CENS Deployment Center, which also leverages that project. We streamlined the workflow of the CENS Annual Report process by requiring CENS participants to deposit publications and datasets, thus capturing them for reuse.

Data modeling

Modeling of sensing data, is critical at multiple scales. At the most micro scale, the Unblinking project models environmental phenomena with latent geometric structure. At an intermediate scale, the Data Conservancy project is modeling scholarly products of, by, and about the astronomy community in an effort to design systems for long-term data curation. The most encompassing data modeling project is the Object Reuse and Exchange effort, whose aim is to model relationships among all extant types of scholarly artifacts. The ORE project is part of a larger international standards activity, part of the Semantic Web, to aid in the linking and discovery of scholarly products. CENS is among the first to test the ORE on scientific data and publications.

In sum, the Statistics and Data Practices area of CENS has matured in productive and integrative ways to study phenomena throughout the scientific life cycle.