

# **UCLA**

## **Papers**

### **Title**

Adding Context to Content: The CENS Deployment Center

### **Permalink**

<https://escholarship.org/uc/item/5br1t8tj>

### **Authors**

Mayernik, Matthew

Wallis, J C

Borgman, C L

et al.

### **Publication Date**

2007-09-26

UCLA Graduate School of  
Education & Information Studies  
*Publications*  
(University of California, Los Angeles)

---

*Year 2007*

*Paper asist07*

---

Addng Context to Content: The CENS  
Deployment Center

Matthew S. Mayernik \*

Jillian C. Wallis †

C L. Borgman ‡

Alberto Pepe \*\*

\*University of California, Los Angeles

†Center for Embedded Networked Sensing

‡University of California, Los Angeles

\*\*UCLA

This paper is posted at the eScholarship Repository, University of California.

<http://repositories.cdlib.org/gseis/pubs/asist07>

Copyright ©2007 by the authors.

# Addng Context to Content: The CENS Deployment Center

## **Abstract**

Scientists and engineers working with embedded networked sensing systems in the environmental sciences are acquiring data at unprecedented rates. Scientific data do not emerge from a vacuum. There is considerable contextual information that surrounds the process of data acquisition that is critical to interpret and analyze data. Current techniques for data sharing involve considerable manual effort to prepare, describe, and transfer this contextual information along with the data itself. This paper reports on a study of the UCLA-based Center for Embedded Networked Sensing (CENS), an interdisciplinary NSF research center that supports collaborations to develop and implement innovative wireless sensor networks. We report here on the development of the CENS Deployment Center, a database for CENS deployment information. The goals of the CENSDC are to facilitate better deployment organization, and to provide a central location for key information that describes the context of data capture. We also describe our plans within CENS to use a tri-partite approach to capture and share sensor data resources using technical recommendations currently being developed within the context of the Open Archive Initiative for Object Reuse and Exchange (OAI-ORE).

## **Adding Context to Content: The CENS Deployment Center**

**Matthew S. Mayernik**

Dept of Information Studies  
Graduate School of Education & Information  
Studies, UCLA  
310-206-0029

[mattmavernik@ucla.edu](mailto:mattmavernik@ucla.edu)

**Jillian C. Wallis**

Center for Embedded Networked Sensing, UCLA  
310-206-0029

[jwallisi@ucla.edu](mailto:jwallisi@ucla.edu)

**Christine L. Borgman**

Dept of Information Studies  
Graduate School of Education & Information  
Studies, UCLA  
310-825-6164

[borgman@gseis.ucla.edu](mailto:borgman@gseis.ucla.edu)

**Alberto Pepe**

Dept of Information Studies  
Graduate School of Education & Information  
Studies, UCLA  
310-206-0029

[a pepe@ucla.edu](mailto:a pepe@ucla.edu)

**Scientists and engineers working with embedded networked sensing systems in the environmental sciences are acquiring data at unprecedented rates. Scientific data do not emerge from a vacuum. There is considerable contextual information that surrounds the process of data acquisition that is critical to interpret and analyze data. Current techniques for data sharing involve considerable manual effort to prepare, describe, and transfer this contextual information along with the data itself. This paper reports on a study of the UCLA-based Center for Embedded Networked Sensing (CENS), an interdisciplinary NSF research center that supports collaborations to develop and implement innovative wireless sensor networks. We report here on the development of the CENS Deployment Center, a database for CENS deployment information. The goals of the CENSDC are to facilitate better deployment organization, and to provide a central location for key information that describes the context of data capture. We also describe our plans within CENS to use a tri-partite approach to capture and share sensor data resources using technical recommendations currently being developed within the context of the Open Archive Initiative for Object Reuse and Exchange (OAI-ORE).**

### **Introduction**

Modern science is distinguished by the extent to which its practices rely on the generation, dissemination, and analysis of data. These practices are themselves distinguished both by the massive scale of data production and by the global dispersion of data resources. The need to develop and deploy an integrated framework for data management is keenly felt by scientists who generate massive quantities of data via wireless sensor networks (Akyildiz 2002; Culler & Hong 2004; Elson & Estrin 2004; Pottie & Kaiser 2000, 2006). These are

systems of sensors that are embedded in the environment of the phenomena on which data are sought, and connected via communication networks so that data from numerous locations can be collated and analyzed either within the network or external to it. In habitat ecology, for example, scientists use multiple types of sensors to observe phenomena, each at different sampling rates. Sensors vary widely in type and capability. Some sensors capture data continuously or at discrete intervals for indefinite periods of time; some sensors are activated only when triggered by an event (e.g., the movement of an animal into the field of vision of a camera), requiring Bayesian statistical models. Scientific data such as these are difficult to use, reuse, or interpret without adequate understanding of the context within which they were collected. Consequently, scientists need assistance in identifying and selecting data that are useful in individual contexts, and preserving and curating data that are of future value, whether to the originators or to others.

Traditionally, contextual information surrounding data acquisition has been recorded and preserved in lab or field notebooks, and later included in varying degrees of completeness in research publications. The interdisciplinary nature of the wireless sensing community complicates the task of capturing contextual information about data, as different disciplinary cultures do not always match processes and terminologies. In addition, standards and expectations for documenting the data collection effort vary across disciplines. As the scale of data generation and of collaborations increases, new technological approaches to capturing data context are required. Our research to date within the Center for Embedded Networked Sensing (CENS), a National Science Foundation Science and Technology Center established in 2002 [<http://www.cens.ucla.edu/>], explores the data practices of scientists working with wireless sensing systems, including the difficulties they encounter in scaling their current practices to the increasing amounts of data being generated (Borgman, Wallis, & Enyedy 2006, in press; Borgman, et al. 2007). Data sharing in this community usually occurs through direct interpersonal interaction, with data being cleaned, described, and contextualized prior to, and as a part of, the data transfer. This process is currently performed manually, and is typically quite time consuming. Data digital libraries can mitigate this time expenditure by providing direct access to large and disparate data sets, eliminating the need for direct communication between the individual who desires a particular data set and the data owner. But in order for data stored in digital libraries to be useful and trustworthy for a general audience, there must be metadata associated with the data that describes the people, instrumentation, and techniques that were involved in the data capture process.

### **The CENS Deployment Center**

CENS supports multi-disciplinary collaborations to develop and implement innovative wireless sensor networks. New sensor technology is being deployed in a wide variety of settings, including habitat ecology, marine microbiology, environmental contaminant transport, seismology, and urban sensing. The CENS Deployment Center (CENSDC) is a multi-purpose web-tool and database service to support CENS research. As the database is populated, the CENSDC will provide access to past, current, and future CENS deployments in a centralized web location, searchable by structured metadata such as the deployment location, technologies used, and participating students, staff, and faculty. Among the key goals are to indoctrinate inexperienced individuals, streamline deployments, and increase remote staff efficiency, helping to reduce deployment errors and facilitate knowledge transfer throughout CENS. It will provide a cross-linkable source of contextual information for CENS data, facilitating access to important documentation that may otherwise only reach a limited audience.

The backbone of the CENSDC is a searchable relational database of deployment information. Researchers have the ability to create and publish a pre-deployment plan through an easy-to-use web interface. Following a deployment, researchers will be able to generate self-contained deployment reports based on participant evaluations. The CENSDC planning services will help ensure higher data quality through better preparation and smoother deployments. The report generating system will compile feedback from the deployment team to create a report illustrating the types of problems that may arise in the field. These deployment reports will be searchable by a number of facets, including experiment type, technology, location, people, or any combination of elements, by any user who wants to understand how a deployment ran, the context for the collection of a specific dataset, or to find similar deployments to one being planned in the future.

The system consists of a back-end SQL database that uses PHP to generate web pages dynamically.

Information is submitted to the database through a series of forms, enabling users to contribute important contextual deployment information usually confined to field notebooks. Users have the capability to create simple or complex deployment records as they see fit. Deployment specifics, such as task and equipment lists, can be entered into the system at an arbitrary level of granularity. Detailed task lists may not be necessary for routine procedures, but for new or difficult processes a high level of task list detail provides a valuable source of documentation for future use. Similarly, specific equipment details are not necessary for generic gear such as ladders or wire cutters, but detailed information illustrating which notes and sensors were used is critical in interpreting data generated during that deployment.

Several XML-based standards and protocols exist for this diverse community, but none of them are stable or widely adopted. The lack of established data standards contributes to the entrenchment of ad hoc management techniques and minimal documentation. The structure most relevant to describing wireless sensing systems is the Sensor Modeling Language (SensorML), supported by the Open Geospatial Consortium, which describes sensor network equipment and relationships. SensorML is complemented by the Observations and Modeling (O&M) language to express ecology data captured by the sensor network. We are using SensorML, which is in the final stages of being accepted as a formal standard (Botts 2006), to derive an ontology to enable users to input specific sensor information, such as manufacturer, model, serial numbers, and calibration information into the CENSDC in a standardized format.

Ease of use has been another key consideration in our design decisions. To simplify the process of setting up a deployment plan, which can be quite extensive, and because CENS research groups often deploy similar sensor systems repeatedly in the same locations, the CENSDC includes a “Make Like Deployment” function, which creates a new record for an upcoming deployment and automatically populates it with values from a pre-existing deployment record. The user can then simply edit the new plan as necessary to accommodate any differences in deployment plans between the past deployment and the upcoming deployment. This function significantly reduces the time and energy necessary to create a deployment plan, as no redundant data entry is necessary.

Finally, the CENSDC allows a user to add a deployment report link to any web accessible data sets that resulted from that deployment. These data sets may be held on shared file servers, local lab or individual computers, or in SensorBase.org, a dynamic repository for CENS data (Chang, et al. 2006).

### **Current CENSDC Work and Future Directions**

The CENSDC project proposal was presented in September of 2006, and development began in January of 2007. Functional requirements for the system were extracted from studies of CENS scientific data practices (Borgman, Wallis, & Enyedy 2006, in press; Borgman, et al. 2007), supplemented by discussions with CENS researchers. Those requirements led to the design of the database structures, web interfaces, and other functionalities of the system. Upon completion of the first iteration, the system was tested on selected deployments. The system will be tested more widely on a variety of deployments during the summer of 2007. Further discussion of system design and results of evaluation will be reported at the November ASIST conference. The CENSDC, as a source for contextual deployment information, will be an excellent training tool for students and other individuals new to field research. It will be a valuable resource for researchers both current and future, aiding in efficient generation, interpretation, and analysis of sensor data.

### **Tri-partite approach to contextualizing scientific data**

Within CENS, we are taking a tri-partite approach to capturing and sharing contextual information about sensor data. As described in this paper, the CENSDC focuses on deployment information, creating deployment plans and reports on a deployment-by-deployment basis. Data generated during CENS deployments are being captured in SensorBase.org. The current CENS publication database, containing bibliographic information and limited full-text access to CENS publications, is now being migrated to the eScholarship Repository, California Digital Library's OAI-PMH-compliant scholarly archive. A key future direction of work will be to integrate the three systems to provide a flow of research documentation from the contextual deployment information, to the data generated from those deployments, to the publications

stemming from those data, capturing the entire data life cycle and preserving the intellectual chain of scholarly communication and knowledge generation (Pepe, et al. 2007).

In order to knit the three systems of digital resources, we plan to use and extend the technical recommendations currently being developed within the context of the Open Archive Initiative for Object Reuse and Exchange (OAI-ORE) (Van de Sompel, et al. 2006). Our framework for interoperability will consist of a layer (a light triple-store digital repository) in which inter-related data entities residing in their own system will be aggregated into compound objects, i.e. surrogate representations of their constituents. In our framework, as data are put into SensorBase.org, they will be linked to the corresponding deployment report in the CENSDC. Data are stored in SensorBase.org on a “project” basis, with each “project” corresponding to a single sensor deployment in the CENSDC. The links will be ensured through the use of persistent unique identifiers, in the form of URL’s, for each deployment report and SensorBase.org project. Ultimately, data sets in SensorBase.org and deployment reports in the CENSDC will be able to be linked to research papers held in the eScholarship Repository. Such linkage among heterogeneous data will be embodied by a digital compound object that will be both harvestable, via standard OAI services, and discoverable, via popular web crawlers and search engines.

### Acknowledgements

CENS is funded by National Science Foundation Cooperative Agreement #CCR-0120778, Deborah L. Estrin, UCLA, Principal Investigator; Christine L. Borgman is a co-Principal Investigator. SensorBase.org research in CENS is led by Mark Hansen and Nathan Yau. Interoperability research to integrate SensorBase.org, CENSDC, and the CENS bibliographic database is led by Alberto Pepe, with funding from Microsoft Technical Computing Initiative. Development support for the CENSDC provided by Margo Reveil, Christina Patterson, James McEvoy, and Andrew Tai of UCLA Academic Technology Services.

### References

- Akyildiz, I.F., et al. (2002). Wireless sensor networks: A survey. *Computer Networks*, **38**: p. 393-422.
- Borgman, C.L., J.C. Wallis, and N. Enyedy. (in press). Little Science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*.
- Borgman, C.L., J.C. Wallis, and N. Enyedy. (2006). Building digital libraries for scientific data: An exploratory study of data practices in habitat ecology. in *10th European Conference on Digital Libraries*. Alicante, Spain J. Gonzalo, et al. LINC 4172 Berlin: Springer: p. 170-183.
- Borgman, C.L., et al. (2007). Drowning in Data: Digital Library Architecture to Support Scientists’ Use of Embedded Sensor Networks. in *JCDL '07: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*. Vancouver, BC. Association for Computing Machinery.
- Botts, M. (2006). *Sensor Modeling Language (SensorML) Status*. Retrieved from <http://stromboli.nsstc.uah.edu/SensorML/status.html> on 24 May 2007.
- Chang, K., et al. (2006). SensorBase.org - A Centralized Repository to Slog Sensor Network Data. in *Proceedings of the International Conf. on Distributed Networks (DCOSS)/EAWMS*.
- Culler, D.E. and W. Hong. (2004). Wireless sensor networks. *Communications of the ACM*, **47**(6): p. 30-33.
- Elson, J. and D. Estrin. (2004). Sensor networks: A bridge to the physical world, in *Wireless sensor networks*, C.S. Raghavendra, K.M. Sivalingam, and T.F. Znati, Editors, Kluwer Academic: Boston.
- Pepe, A., et al. (2007). Knitting a fabric of sensor data and literature. in *Information Processing in Sensor Networks*. Cambridge, MA. Association for Computing Machinery/IEEE.

Pottie, G.J. and W.J. Kaiser. (2000). Wireless integrated network sensors. *Communications of the ACM*, **43**(5): p. 51-58.

Pottie, G.J. and W.J. Kaiser. (2006). *Principles of embedded networked systems design*. Cambridge, England: Cambridge University Press.

Van de Sompel, H., et al. (2006). An Interoperable Fabric for Scholarly Value Chains. *D-Lib Magazine*, **12**(10).