Title
Making CENS Data Useful to Scientists, Engineers, Students

Permalink
https://escholarship.org/uc/item/4t8446n1

Authors
Jillian Wallis
Christine Borgman
Noel Enyedy
et al.

Publication Date
2005

Making CENS Data Useful to Scientists, Engineers, & Students

Jillian Wallis, Christine Borgman, Noel Enyedy, & Staša Milojević
CENS & CENSei Data Management

CENS Data Management

Our Goals & Objectives for CENS

1. To make CENS a leader in the *selection* and *implementation* of standards and technologies to support *multidisciplinary data* and *sensor network portals*.
2. To make CENS data *accessible* to a variety of audiences and maintain the *integrity* of those data for the *future research* and *educational efforts* of CENS participants.
3. To *collect, organize, and preserve scientific data* so that they are useful to multiple users having *different levels of subject expertise* and *different uses* for the data.

Requirements for Data to be Useful

1. **There exists some scientific data** - What counts as scientific data within the context of sensor networks?
2. **There exists some scientist willing to share his or her data** - What are the conditions that support sharing?
3. **There exists some scientist with the authority to share his or her data** - Who holds intellectual property rights to CENS data?
4. **There exists some method of sharing scientific data** - Is private access to data or a public data repository the best means to the sharing end?

Ensuring sensor data will be useful to the multiple current and future communities through understanding the data management needs and practices of these communities.

1. Scientific Data Collected by Sensors

• Data is a blanket term that includes any number of data types and formats.

• For example, data can include raw numerical data right off the sensors, images and video as in the NIMS node, audio samples, published graphs and charts, and everything in between.

• In order to avoid the "Data Deluge" inherent to regular sampling activities there needs to be some metric for selecting data that is worth preserving for future use.

2.a. Data Sharing in the Sciences

• Data sharing depends on trust, incentives, infrastructural support, institutional policies, and the utility of data collected by others to support the scientist's own research.

• From our preliminary research we have found that data sharing in the sciences varies by discipline.

• For example: Seismologists value the preservation and sharing of historical and global data and thus have a rich system of data repositories, policies, metadata and transmission standards, which all support this need.

2.b. Sharing Data with Science Students

• Sharing data with science students is a special case of scientific data sharing.

• Not only must the data be available to the students, but it is not accessible to the students without the implicit knowledge and methodology of scientists.

• Metadata to capture the implicit methodologies and practices of CENS scientists will support this end.

• We will be using data sharing with students as a test case for CENS data sharing on a larger scale.
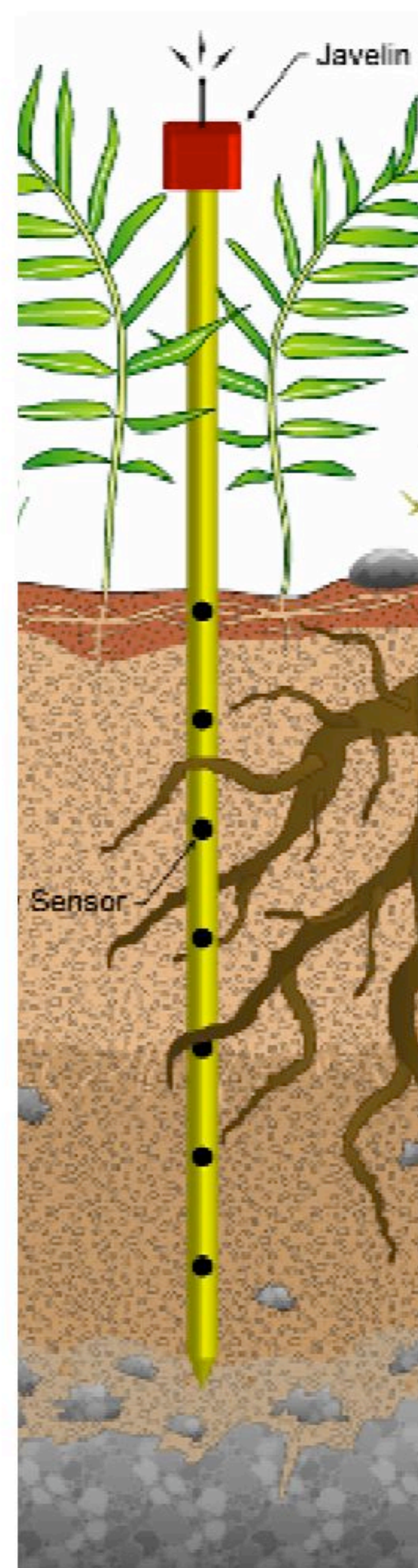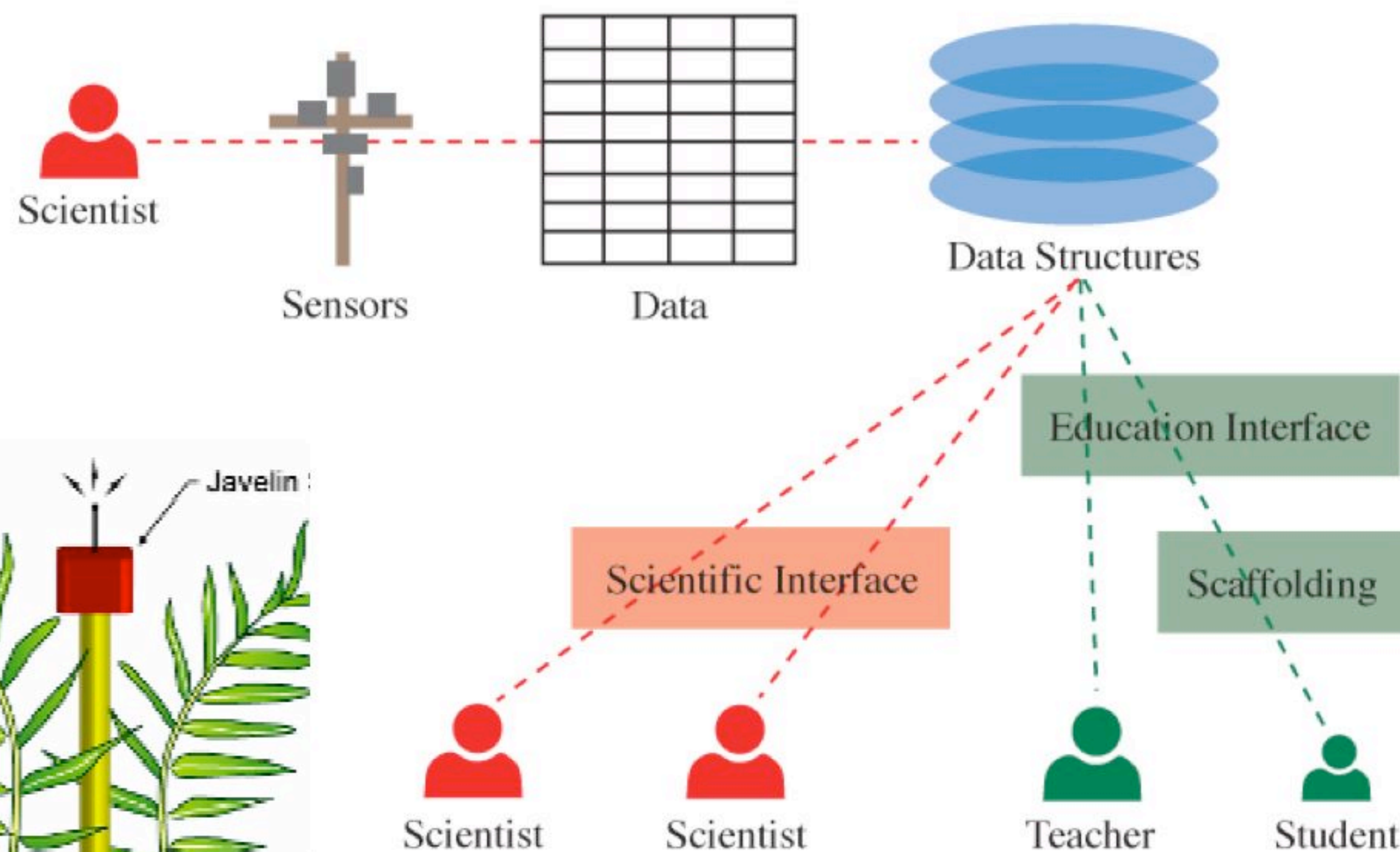
3. Scientific Data Authorship

• Data authorship and intellectual property rights will eventually determine who has access to CENS data.

• The concept of data authorship is affected by many factors, such as patenting, corporate interest, protection of publishing rights, effort that has gone into the collection or clean up of data, etc.

• From our preliminary research we have found that the claim of authorship can be proportional to the amount of effort expended during collection.

4. Methods of Scientific Data Sharing

• The sharing of scientific data can be encouraged through both social and technological systems.

• Social systems include: data validation, researcher trust and collaboration, citation and recognition for shared data, and data policies that advocate sharing.

• Technological systems include: data repositories, data portals, data organization and retrieval systems, metadata and transmission standards.

• Ideally a cyberinfrastructure to support sharing would include a mixture of these systems to that end.

Cyberinfrastructure to Support Sharing

• Data creation relies on a scientist to place a sensor on the CENS network, or trigger a sampling event.

• Sensor data is captured by the sensor and moves into some data structure.

• Other scientists can then access this data through some portal which allows them to retrieve and manipulate applicable data.

• Through the education portal interface built by CENSei, the teachers can access the data they need for lessons.

• Science students go through an additional layer of scaffolding that gives the students a context for learning scientific methods, which can be slowly stripped away as the student progress.

Our Next Steps:

• Understanding the current data needs and practices of communities of CENS scientists, through interviews and an analysis of data and requirements.

• From our preliminary research we have uncovered some of the needs and practices of CENS scientists, but we need to find out if they are pervasive throughout or localized phenomena.

• Work closely with select CENS researchers to develop metadata structures to support their research, specifically projects such as the Ballona Wetlands, Palmdale, and Point Mugu.

• Inventory of data and metadata standards in selected domains.

• Case study of science application for educational module.

• Develop data models for science application and education module.

• Identify policy issues in data sharing and data management.

Javelin or pylon sensors in use at the Ballona Wetlands and Point Mugu to track soil contaminant transport.