

UCLA

Posters

Title

CENS Data Sharing Practices (KNO 3)

Permalink

<https://escholarship.org/uc/item/2jh9b5k7>

Authors

Matthew Mayernik

Alberto Pepe

Jillian C. Wallis

et al.

Publication Date

2006

CENS Data Sharing Practices

Matthew Mayernik, Alberto Pepe, Jillian C. Wallis, & Christine L. Borgman
CENS Data Management

Introduction

Data sharing vs. data volume

- At CENS we are *committed to data sharing*
The commitment is not spelled out clearly for all CENS researchers as to the *How* or *When* the data sharing should take place.
- **Data sharing practices vary greatly**
Practices remain very *localized*, varying by research group and requiring a lengthy exchange between the data provider and the data requester.
- **With existing practices data sharing will become a hardship**
As the volume of data and data requests increase, the burden on the data provider will become unwieldy for CENS researchers.

Systems & standards for data sharing

- **A system for data sharing**
We propose the leveraging existing data practices and work flows to assist both data providers and data consumers, through the use of *metadata*.
- **Interoperability and standards**
Add *accessibility*, *interoperability*, and other functionality through the implementation of appropriate *standards*, such as OAI harvesting protocols.
- **Data sharing policy**
Develop a more *robust data sharing policy* to encourage data sharing and system use.

Problem Description: Our current data sharing practices will not scale to meet future need

- **Data sharing requires a significant investment of resources on the part of the data provider**
 - See results below
- **Data volume is quickly becoming unwieldy**
 - The majority of CENS researchers are at the point where they have more data than can use, or foresee hitting this point in the near future
 - The majority of CENS researchers are finding they have less time to properly mark up data for their own use or use by others
- **CENS data potentially has a very broad audience of users**
 - See Fig. 1 below
- **Data sharing practices do not scale well to new volumes**
 - As CENS gains more exposure, demand for CENS data will also increase
 - As data volume increase these researchers will have less time to engage in the data sharing exchange
 - As the ratio of publications to data decreases, the publication will no longer serve as an adequate pointer or context provider for data

Proposed Solution: Leveraging current data management practices and standards to alleviate scalability challenges in data sharing

Data Practices Interview Study

- **Interview study method**
 - Interview study based on the results of a pilot interview study
 - Interview subjects included 22 researchers across the CENS community
 - Sample included application science researchers and technological science researchers, from the faculty, student, and staff populations
 - Interview questions ranged from data characteristics, sharing, policy, and architecture
- **Interview study results**
 - Sharing practices differ between and within research groups
 - Occasionally the group will have a dictated format and path for sharing data (e.g. lab servers)
 - More often researchers share in whichever way is the most convenient
 - Data will be shared in different ways by the same researcher depending on the document size
 - Summary of sharing within the CENS community
 - Shared server space -- public and password protected (e.g. JR DMS, CENSweb, etc)
 - Use of email to send URL, FTP location, or files
 - Trading data in Excel or tab-delimited and comma-delimited text files on jump drives
 - Hard copies
 - Summary of sharing with researchers external to CENS
 - Requests for data are usually issued in response to a published paper, which provide contextual information about the data and the data collection effort for the data consumer
 - Data needs to be cleaned, annotated and otherwise processed before being given to a requestor
 - Trust in shared data comes from trusting the person who collected the data and their instruments
 - Sharing is almost always mediated by a human being
 - Typical data sharing exchange (fig. 2)
 - Data provider publishes results from data
 - Requestor reads about data through publication
 - Requestor contacts data providers
 - Requestor supplies their potential use for data and data format needs
 - Data provider filters and formats data to the requestor's needs
 - Data provider send data or data pointer to requestor

System Recommendations

- **Recommendations**
 - Develop a system for supporting data access by researchers and consumers alike
 - Fit into existing work flows
 - Leverage existing data practices for use by others
 - Capture metadata used by researchers to describe their own data
 - Require metadata about data for storage in SensorBase or other CENS repository
 - Give researchers better access to their own data through metadata
 - Use metadata as possible hooks for data consumers to exploit
 - Adopt XML metadata standards for interoperability and discovery
 - Open Archives Initiative (OAI) harvesting protocol
 - SensorML - Sensor Markup Language
 - EML - Environmental Markup Language
 - Provide alternate documentation for data
 - Metadata and deployment reports -- see CENS Deployment Center Portfolio Proposal
 - Makes SensorBase data more accessible to potential consumers
 - Automates the data sharing exchange and requires less human-mediation
 - Fulfills the requirements for ensuring trust through provision of information
 - Develop a reasonable data sharing policy
 - Set up reasonable guidelines for how and when sharing can occur
 - Fit with system outlined above to encourage use

