

# UC San Diego

## Scripps Institution of Oceanography Technical Report

### Title

Ocean Informatics Monograph. Ocean Informatics Initiative: an Ethnographic Study (2002-2006). Part 2: Appendices

### Permalink

<https://escholarship.org/uc/item/1np8b208>

### Authors

Millerand, Florence  
Baker, Karen S

### Publication Date

2011-06-01

# OCEAN INFORMATICS MONOGRAPH

## Ocean Informatics Initiative: an Ethnographic Study (2002-2006) Part 2: Appendices

Florence Millerand<sup>1</sup> and Karen S. Baker<sup>2</sup>

<sup>1</sup>Department of Social and Public Communication, University of Quebec, Montreal (UQAM)

<sup>2</sup>Scripps Institution of Oceanography, University of California, San Diego (UCSD)



This is an ethnographic study of the development of Ocean Informatics,  
an initiative to grow digital infrastructure that addresses  
field-oriented, scientific data and information needs.

Scripps Institution of Oceanography Technical Report  
June 2011

# Table of Contents

## Appendices

<b>1</b>	<b>Appendix: Integrative Oceanography Division (IOD) Web page</b>	<b>3</b>
<b>2</b>	<b>Appendix: UC Marine BioOptics</b>	<b>5</b>
<b>3</b>	<b>Appendix: Research Publications about the Ocean Informatics initiative</b>	<b>8</b>
<b>4</b>	<b>SIO Requests for Action</b>	<b>14</b>
4.1	<i>Appendix: SIO Time Capsule and Long-Term Data</i>	14
4.2	<i>Appendix: Response to SIO Director Search Request</i>	15
<b>5</b>	<b>Appendix: Ocean Informatics Reading Groups</b>	<b>16</b>
<b>6</b>	<b>Appendix: Ethnographic Research</b>	<b>24</b>
6.1	<i>Appendix: What do you mean by 'social'? An imaginary dialogue between a social scientist and an ecological scientist</i>	24
6.2	<i>Appendix: Ethnographic Field Hand-outs</i>	27
6.3	<i>Appendix: Interview questions</i>	30
6.4	<i>Appendix: Memos on Grounded Theory</i>	31
6.5	<i>Appendix: Steps of Grounded Theory Analysis</i>	36
6.6	<i>Appendix: Response - Steps of Data Management</i>	38
6.7	<i>Appendix: Memo on Conceptualizing Categories</i>	39
6.8	<i>Appendix: Memo on Ocean Informatics Definitions</i>	41
6.9	<i>Appendix: Memo on Ocean Informatics as a 'community of practice' (CoP)</i>	42
<b>7</b>	<b>Appendix: Qualitative Analysis Software</b>	<b>44</b>
7.1	<i>Appendix: References on the differences between various qualitative analysis software</i>	44
7.2	<i>Appendix: Semato in Review</i>	45
7.2.1	<i>Appendix: Memo on use of Semato for project</i>	45
7.2.2	<i>Appendix: Semato – A software for quantitative and qualitative analyses (home page)</i>	49
7.2.3	<i>Appendix: Description of Semato main features and their use in qualitative data analysis</i>	50
7.2.4	<i>Appendix: Semato Glossary</i>	53
7.2.5	<i>Appendix: Transcription format for Semato</i>	54
<b>8</b>	<b>Appendix: Paper on Role of Information Management</b>	<b>55</b>
<b>9</b>	<b>Appendix: Event Logger: Summer Project Proposed Paper/Poster</b>	<b>62</b>
<b>10</b>	<b>Appendix: Ocean Informatics Posters: Technical and Conceptual</b>	<b>68</b>
<b>11</b>	<b>Appendix: Ocean Informatics Event Gallery</b>	<b>117</b>

## **Abstract**

The report presents an initial monograph on Ocean Informatics (OI), an information infrastructure initiative in the ocean science community. Using ethnographic methods, we observed and analyzed the development of the OI Initiative based at Scripps Institution of Oceanography over a period of 4 years (2002-2006). The focus of the report is the formation of an information environment that provides information management and information systems design expertise focusing on biological and ecological oceanography in particular. OI is specifically framed as conducive to support of scientific data practices, data curation, design practices, and information managers' professional development when our understanding of these elements is under development amidst an era of transitions relating to digital data production and access. The effort aims to address short-term needs for information management while formulating and planning for the growth of infrastructure over the long-term. As an interdisciplinary initiative that spans multiple organizational units, its development is framed by a keystone relationship with the scientific environments with which it partners and within which it is embedded. It began as an oceanographic site in the Long-Term Ecological Research program (LTER) and subsequently partnered with the California Cooperative Fisheries Investigations (CALCOFI) as well. In bringing new attitudes and insights relating to living systems, the ecological perspective may also have significant ramifications in considering digital configurations. The OI Initiative highlights the envisioning of infrastructure efforts as having local, situated elements and how such efforts contribute to science today. The report captures the views of the diverse participants associated with the Initiative, thus providing a living portrait of Ocean Informatics whose development continues today. The report is in two parts with appendices appearing in a separate volume as Part 2.

# Appendices

## 1 Appendix: Integrative Oceanography Division (IOD) Web page

The Integrative Oceanography Division web page text was inspired and developed by Ocean Informatics participants.

8650 Discovery Way . La Jolla, CA . 92093 08

### Integrative Oceanography Division

SCRIPPS INSTITUTION OF OCEANOGRAPHY, UNIVERSITY OF CALIFORNIA, SAN DIEGO

- Home
- About Us
- People
- Business Services
- Research
- Outreach
- Links of Interest
- Data Zoo
- Courses
- Infrastructure

#### ABOUT US THE INTEGRATIVE OCEANOGRAPHY DIVISION

The organizing principle of the Integrative Oceanography Division (IOD) resides in a shared commitment to collaborative, interdisciplinary science. "Integrative" denotes our philosophy that multiple approaches are important in creating a better understanding of the ocean system. We take great pride in our membership of physical, biological, chemical and geological oceanographers, climate and information scientists, data and information managers, engineers, technicians, education specialist, administrative and management professionals. IOD distinguishes itself as scientific home for researchers working at the boundaries of traditional academic spheres, and generating growing programs in the integration of research with informatics as well as with education and public outreach. From pelagic to benthic ecology, shoreline to open ocean dynamics, data to information systems, our research encompasses field work, laboratory experimentation and computer modeling to acquire, integrate, synthesize and understand diverse data sets to elucidate the underlying dynamics of complex, multidimensional ocean systems.

#### SIO/IOD

#### Resources

- [IOD Phone List](#)
- [UCSD Directory](#)
- [BLINK](#)
- [Scripps Website](#)
- [UCSD Website](#)
- [Deadline & Event Calendar](#)
- [Other Scripps Business C...](#)
- [Online Resources](#)

UCSD Official web page of the University of California, San Diego

## COMPUTATIONAL INFRASTRUCTURE SERVICES (CIS) THE INTEGRATIVE OCEANOGRAPHY DIVISION

---

### ABOUT US

IOD's physical computational infrastructure consists of a multi-platform array of servers, providing email, file, and other network services. The infrastructure team maintains the integrity of these servers and provides user support for both network and local resources.

---

### CONTACTS

#### Computational Infrastructure Services Team

[Jerry Wanetick](#) (Director)

[Nate Huffnagle](#)

[Charles Baker](#)

CIS Help Desk email - [help@coast.ucsd.edu](mailto:help@coast.ucsd.edu)

CIS Help Desk Phone - **858-246-0555** (x60555 on campus)

#### UCSD Contacts

---



Official web page of the University of California, San Diego

## 2 Appendix: UC Marine BioOptics

Overview from 1999 SIO report about the University California Marine BioOptics cross-campus group.

### UC Marine Bio-Optics Group

*RAYMOND C. SMITH*  
*KAREN S. BAKER*

University of California Marine Bio-Optics (UCMBO) is a new IMR multi-campus group formed in 1981 with centers at the University of California at San Diego co-ordinated by Karen S. Baker and at the University of California at Santa Barbara directed by Professor Raymond C. Smith. This group is concerned with investigating and understanding the role of radiation in natural waters in order to quantitatively describe and predictively model the marine photoenvironment and the corresponding bio-optical ocean properties. In pursuit of these goals, instruments are designed, measurements are made at sea and analysis and computer modeling are carried out. Since the changing marine light field is directly related to and is an influencing factor for physical, chemical and biological water properties, the UCMBO investigations are multi-disciplinary. For instance, a bio-optical model relating dissolved and suspended biogenous material in ocean waters to the corresponding optical properties has been developed and continues to be improved. Multispectral satellite imagery and multiplatform sampling strategies (ships, satellites, buoys and aircraft) are used to examine the distribution and variance of phytoplankton biomass. The relation of this biomass to regional productivity and the mesoscale ocean phenomena influencing phytoplankton distributions and productivity are research objectives of the group.

This group is composed of approximately ten people including five UCSB graduate students. Recent projects have initiated strong collaborations with other groups including the IMR Food Chain Research Group at SIO, the Rosensteel School of Marine and Atmospheric Science Group located at the University of Miami, the Warm Core Ring Group with field work out of Woods Hole Oceanographic Institution and the Optical Dynamics Experiment Group organized through the University of Oregon. Interaction continues with the National Aeronautic and Space Administration Laboratories at Pasadena, California (Jet Propulsion Laboratory) and at Virginia (Wallops Space Flight Center) as well as with the Environmental Protection Agency Laboratory at Athens, Georgia.

Major areas of investigation for the 1979-1984 time period include (1) Southern California Bight studies; (2) warm core rings project; (3) optical dynamics experiment; (4) bioluminescence and optical variability study; (5) ultraviolet radiation and suspended sediment studies.

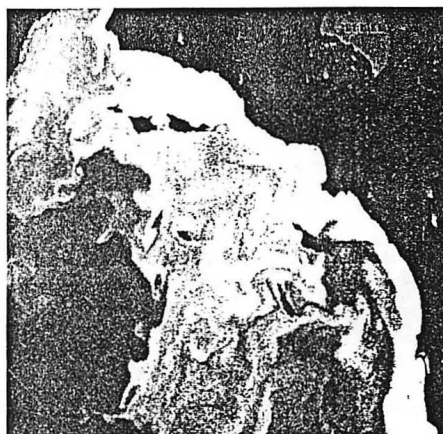


Figure 38. A calibrated image (Nimbus 7 satellite, Coastal Zone Color Scanner instrument) of the Southern California Bight region on 6 March 1979. The land and cloud areas are masked black. In the water, the lighter the area, the higher the chlorophyll on a scale from  $.01 \text{ mg/chl/m}^3$  (black) to  $5 \text{ mg/chl/m}^3$  (white).

**Southern California Bight Studies.** An investigation of the distribution and variance of phytoplankton over a full range of space and time scales has been undertaken in order to obtain a more basic understanding of mesoscale biological processes in productive coastal waters. Complementary ship and satellite (Nimbus-7 satellite Coastal Zone Color Scanner instrument) bio-optical data from the Southern California Bight Region has been obtained for many time periods. An initial chlorophyll time series has been published giving a quantitative assessment of chlorophyll and its variance in these waters and the group is currently

working to provide a quantitative time series of chlorophyll distribution and of sea surface temperature for a longer time period. It has been found that in a period of less than two weeks the integrated chlorophyll in the Bight changed by more than a factor of two. In collaboration with D. Au (National Marine Fisheries Service) and P. Dustan (College of Charleston) we are also investigating the possibility of using sea-surface temperature and chlorophyll as determined by ship and satellite as habitat descriptors related to the distribution of marine mammals. This work has been sponsored by the National Aeronautics and Space Administration.

The California Space Institute sponsored a multi-campus investigation of an integrated remote sensing program particularly for the study of California's coastal zone. It also sponsored the collaborative work between R. W. Eppley and UCMBO.

The California Sea Grant College Program has furthered UCMBO studies of the phytoplankton dynamics in eutrophic coastal water by sponsoring a pilot program to examine problems related to the database management of disparate data sets obtained from multiplatform sampling strategies. In addition, this program has provided two Sea Grant Traineeships for the training of graduate students in the relatively new skills of ocean remote sensing.

**Warm Core Rings (WCR).** UCMBO participation in the WCR project which involves more than 25 principal investigators at thirteen different institutions has been sponsored by the National Aeronautic and Space Administration. A warm core ring forms when a meander separates to the north of the Gulf Stream to form an anticyclonic vortex. This project conducted an interdisciplinary study of the structure and the dynamics of Gulf Stream warm core rings by carrying out a series of multi-ship cruises to follow the evolution of a ring. Rings range in size from 30 to 150km in diameter and exist in depth to hundreds of meters. They exist for a period ranging from months to over a year and travel as a body in a general southwesterly direction from latitudes of approximately 40 degrees north to as far south as 28 degrees while maintaining their own internal clockwise rotation.



Figure 39. Nimbus 7 image of the Atlantic coast on 25 April 1982. The land is masked. In the lower right corner, the Gulf Stream is visible. A meander in the stream is forming. A detached Warm Core Ring is visible in the lower center. Both the Gulf Stream and the Warm Core Ring have lower chlorophyll values than the surrounding waters. Image processed at RSMAS, University of Miami.

The primary objective of the UCMBO participation has been to determine the bio-optical properties and the distribution and variance of phytoplankton biomass of a warm core ring. The spatial and temporal variability of these properties, as well as related hydrodynamic properties, throughout the evolution of several rings have been observed through the use of a newly designed state-of-the-art Bio-Optical Profiling System (BOPS) instrument during participation in five WCR cruises in 1981-1982. The analysis of over 100 days of along-track and on-station ship data in concert with contemporaneous data from NASA P3 aircraft flying an Airborne Oceanographic Lidar system and with color and temperature satellite data is providing an understanding of the mesoscale processes influencing primary production and the distribution and variance of phytoplankton biomass in WCR's and their environs as well as demonstrating the effectiveness of multi-platform sampling strategies. Specific bio-optical data has been obtained for the investigation of the radiometric sensitivity of the Coastal Zone Color Scanner satellite instrument which will aid in the calibration of and the algorithm development for that instrument.



### Optical Dynamics Experiment (ODEX).

ODEX, sponsored by the Office of Naval Research, has as its goals the development and testing of oceanic models in the upper mixed layer linking physical, biological and optical properties with respect to physical forcing functions such as winds and storms. Complementary color and temperature satellite images were captured to provide an overview to the 1982 field program.

The first major at sea data collection segment of ODEX occurred in October and November of 1982 in the North East Pacific (141°N, 35°W) with scientists aboard the Naval Post Graduate School's R/V Acania and the Scripps' R/P Floating Instrument Platform (FLIP). Optical, physical, biological, and chemical oceanic parameters were sampled intensively over a grid of stations in order to delineate synoptically the structure of an apparent instability in the subtropical ocean front. Nearer shore sampling was carried out in order to be able to contrast the oceanographic structure at the site with the waters of the California Current and the transition waters in between. Further field programs near the 35°W latitude line continue in order to investigate seasonal variability in this region.

**Bioluminescence and Optical Variability in the Sea (Biowatt).** This is a recently initiated study with the goal of identifying causal links between the variability in light attenuation

and light production in the ocean. The issues addressed include behavioral relations among macrozooplankton and micronekton, to the dynamics of absorbancy and scattering populations, to the physical dynamics of the upper layers of the ocean. Field work includes an Atlantic cruise in 1985.

### Ultraviolet Radiation and Suspended Sediment Studies.

This work, funded by the United States Environmental Protection Agency, consists of empirical as well as theoretical studies. Aquatic photochemical and photobiological processes depend upon both the amount and the spectral composition of solar radiation penetrating to depth in natural waters and are often found to be particularly effected by the ultraviolet region of the spectrum. A predictive model of the spectral radiant energy of natural waters (with particular emphasis in the UV portion of the spectrum) in terms of their biogenous components as well as their suspended sediments has been developed. This model allows the assessment of photochemical processes (e.g., photolysis rates) that account for the transformation of pollutants in aquatic environments. Laboratory studies using a submersible ultraviolet spectroradiometer and computer studies using Monte Carlo modeling techniques have helped in our quantitative modeling of these photoprocesses.

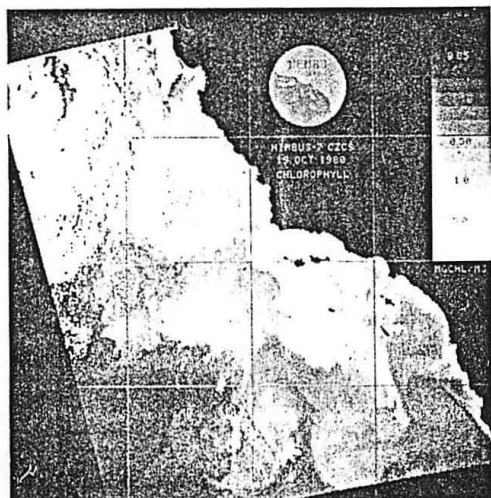


Figure 40. Satellite image from 19 October 1980 of the ocean off the Southern California coast where land is black and clouds are white. The optical data from the Nimbus 7 satellite Coastal Zone Color Scanner instrument has been atmospherically corrected and absolutely calibrated to produce an image of chlorophyll-like pigments.

### 3 Appendix: Research Publications about the Ocean Informatics initiative

#### Technical Reports

- Baker, K.S. (2005). *Informatics and the Environmental Sciences*. Scripps Institution of Oceanography (SIO) Technical Report Series, University of California San Diego. <http://escholarship.org/uc/item/0179n650>
- Millerand, F. and K.S. Baker (2011). Ocean Informatics Monograph (2002-2006). Technical Report Series. University of California San Diego.
- Baker, K.S., M.Kortz, and J.Conners (2011). *DataZoo, an Information System*. Scripps Institution of Oceanography (SIO) Technical Report Series, University of California San Diego.
- Donovan, J.M., and K.S.Baker (2011). *The Shape of Information Management: Fostering Collaboration across Data, Science, and Technology in a Design Studio*. Scripps Institution of Oceanography (SIO) Technical Report Series, University of California San Diego.

#### Pre-Ocean Informatics Technical Reports

- Karasti, H., K. Baker, and G.C.Bowker (2003). Proceedings of the Computer Supported Scientific Collaboration Workshop (CSSC), Eighth European Conference on Computer Supported Cooperative Work (ECSCW), Helsinki, Finland, 14 September 2003. <http://escholarship.org/uc/item/31m1m9qc>
- Baker, K.S., and H.Karasti (2004). *The Long-Term Information Management Trajectory: Working to Support Data, Science and Technology*. Scripps Institution of Oceanography (SIO) Technical Report Series, University of California San Diego. <http://escholarship.org/uc/item/7d64x0bd>

#### Papers and Proceedings

- Karasti, H., and K. S. Baker (2004). Infrastructuring for the long-term: ecological information management. *Proceedings of the Hawai'i International Conference on System Sciences (HICSS) 2004*, 5-8 January, Big Island, Hawaii, IEEE, New Brunswick, NJ, 2004.
- Jackson, S. J., and K. S.Baker (2004). Ecological Design, Collaborative Care, and Ocean Informatics. *Proceedings of the Participatory Design Conference, PDC-04, Vol 2*.
- Baker, K.S., S.J. Jackson, and J.R. Wanetick, 2005. Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment in *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS) 2005*, 3-6 January, Big Island, Hawaii, pp. 1-10, IEEE, New Brunswick, NJ.
- Baker, K. S., and F. Millerand (2007). Scientific Information Infrastructure Design: Interdependent Provinces and Knowledge Environments. *Proceedings of the American Society for Information Systems and Technology Conference*. October 18-25, Milwaukee, Wisconsin.
- Baker, K.S. and F.Millerand (2007). Articulation Work Supporting Information Infrastructure Design: Coordination, Categorization, and Assessment in Practice in *Proceedings of the 40th Hawaii International Conference on System Sciences*.

- Baker, K.S. and K.I.Stocks (2007). Building Environmental Information Systems: Myths and Interdisciplinary Lessons in *Proceedings of the 40th Hawaii International Conference on System Sciences*.
- Baker, K. S., and C. L. Chandler (2008). Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics. *Deep Sea Research Part II*, 55(18/19): 2132-2142.
- Karasti, H. and K.S. Baker (2008). Community Design: Growing One's Own Information Infrastructure in *Proceedings of the Participatory Design Conference*. 30Sep-04Oct, 2008, Bloomington, IN.
- Baker, K.S. and L.Yarmey (2009). Data Stewardship: Environmental Data Curation and a Web-of-Repositories. *International Journal of Digital Curation* 4(2):12-27.
- Baker, K.S. and F.Millerand (2010) Infrastructuring Ecology: Challenges in Achieving Data Sharing. In *Collaboration in the New Life Sciences*. J.Parker, N.Vermeulen, and B.Penders (eds), Ashgate, Surrey, England: p. 111-138.

#### Posters about Ocean Informatics

Included are posters about the Ocean Informatics Initiative and Information Management. The abstracts with numbers as identifiers are given in Appendix 10. Posters are online: <http://oceaninformatics.ucsd.edu/media-gallery/?id=1>

10. Title: LTER Growing Information Infrastructure: Data Lifecycles and Subcycles  
 Author(s): Karen Baker, Florence Millerand, Lynn Yarmey  
 Date: 2009-09-14
20. Title: INTEROP Scientific Infrastructure Design: Information Environments and Knowledge Provinces  
 Author(s): Karen Baker, Florence Millerand  
 Date: 2007-10-19
24. Title: LTER Environmental Data Management: Infrastructure Studies Insights  
 Author(s): Florence Millerand and Karen Baker  
 Date: 2007-08-02
26. Title: LTER: Long Term Informatics  
 Author(s): KBaker, CChandler, AGold, FMillerand, JWanetick  
 Date: 2007-08-02
28. Title: LTER: Research in Infrastructure Studies: Social & Organizational Perspectives on Ecological Data Management  
 Author(s): Florence Millerand and Karen Baker  
 Date: 2006-09-20
33. Title: Initiating the Data Dialogue: 2005 CalCOFI Conference Interactive Poster  
 Author(s): Karen Baker  
 Date: 2005-12-06

34. Title: CalCOFI Data Management: Overview and Reflection  
Author(s): Karen Baker, Karen Stocks  
Date: 2005-12-05

Pre-Ocean Informatics Publications

- Baker, K. S. (1996). Development of Palmer Long-Term Ecological Research Information Management in *Proceedings of Eco-Informa Workshop, Global Networks for Environmental Information, 4-7 November 1996, Lake Buena Vista, FL*, pp. 725-730.
- Baker, K. S. (1998). Palmer LTER information management in *Data and information management in the ecological sciences: a resource guide (Proceedings of workshop, held at University of New Mexico, Albuquerque, NM, 8-9 August, 1997)*, pp. 105-110, 1998.
- Baker, K. S., B. J. Benson, D. L. Henshaw, D. Blodgett, J. H. Porter and S. G. Stafford (2000). Evolution of a multisite network information system: the LTER information management paradigm. *BioScience*, 50 (11), 963-978.

Other Information Management Articles

The follow are articles from Databits, the LTER Information Management Committee Newsletter (<http://databits.lternet.edu>).

***Author, Title, Newsletter Issue: Category***

- Baker, K.S., Palmer Field Work, 91Summer: News Bit
- Baker, K.S., Software Tips, 92Fall: News Bit
- Baker, K.S., Palmer Field Work, 92Spring: News Bit
- Baker, K.S., Palmer Field Work, 92Fall: News Bit
- Baker, K.S., Palmer Field Work, 93Summer: News Bit
- Baker, K.S. , Palmer Field Work, 94Spring: News Bit
- Baker, K.S., Technical Training, 99Fall: News Bit
- Baker, K.S., Site Survey/Education Outreach/Good Read/Ecologist in the News, 99Fall: News
- Baker, K.S. and M. White, LTER Newsletter Databits New Design, 99Spring: Feature
- Baker, K. and J.Brunnt , Site Information Manager-Network Office Exchanges, 99Spring: Feature
- Baker, K. and J.Brunnt, Database Design Tools, 99Spring: Feature
- Baker, K.S., Electronic Multi-Authoring, 99Spring: Feature
- Baker, K., LTER Site Description Directory Update, 00Spring: News
- Baker, K., Information Manager Guide, 00Spring: FAQ
- Baker, K., Online Computing Dictionary, 00Fall:, FAQ
- Sheldon, W., Evolution of a Multisite Network Information System, 01Spring: Good Read Review
- Baker, K., Moving Toward Network Identity, 01Fall: Feature
- Baker, K., Ecology Through Time, 01Fall: Good Read
- Baker, K., Biodiversity Data Diversity, 01Fall: Good Read
- Baker, K., Managing Scientific Metadata, 02Spring: Good Read
- Vernet, M. and K.Baker, Is it Time to Bury the Ecosystem Concept?, 02Spring: Good Read
- Baker, K., SCI2002 Conference: Ecoinformatic Challenges at International Conference, 02Fall: News
- Baker, K., Ecological Vignettes: A History of the Ecosystem Concept in Ecology, 02Fall: Good Read

Baker, K. and H. Karasti, Whirlwind Tour of Collaborative Practice, 03Spring: , Commentary

Baker, K., Information Ecology, 03Spring: Good Read

Baker, K., The Invisible Present, 03Spring: Good Read

Baker, K., BioScience January 2003 Special Issue LTER, 03Fall: Good Read

Baker, K., Steps Towards an Ecology of Infrasctructure, 03Fall: Good Read

Baker, K.S., S.R.Haber, and M.White, Postnuke Portal Software: Community, Content, and Collaborative Management System, 04Spring: Feature

Baker, K., Data Grids, Collections, and Bricks, 04Spring: Good Read

Jackson, S., The Dry and the Wet, 04Spring: Good Read

Baker, K., J.Wanetick, and S.Haber, The Cognitive Style of Powerpoint, 04Fall: Good Read

Campbell, C., Infrastructuring for the Long-term: Ecological Information Management, 04Fall: Good Read Review

Baker, K., Data at Work: Supporting Sharing in Science and Engineering, 04Fall: Good Read

Baker, K., L.Yarmey, L.Powell, and W.Sheldon, Designing a Dictionary Process: Site and Community Dictionaries, 05Spring: Feature

Baker, K., Atkins Report on CyberInfrastructure, 05Spring: Good Read

Baker, K.S., Revolutionizing Science and Engineering through Cyberinfrastructur, 05Spring:Good Read

Millerand, F., Building the Virtual State: IT and Institutional Change, 05Spring: Good Read

O'Brien, M., Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment, 05Spring: Good Read Review

Millerand, F., K.Baker, B.Benson, and M.Jones, Lessons Learned from EML about the Community Process of Standard Implementation, 05Fall: Feature

Haber, S. and K.Baker, Web Communication Strategies in a Collaborative Environment: Lessons Learned, 05Fall: Feature

Ribes, D., Incorporating Semantics in Scientific Workflow Authoring, 05Fall: Good Read

Yarmey, L. and K.Baker, The Meaning of Everything, 05Fall: Good Read

Baker, K., From Databases to Dataspaces: Opening up Data Processes, 06Spring: Good Read

Haber, S., Designing Interfaces, 06Spring: Good Read

Haber, S., Design Patterns: Elements of Reusable Object-Oriented Software, 06Spring: Good Read

Baker, K., L.Yarmey, S.Haber, F.Millerand, and M.Servilla, Creating Information Infrastructure through Community Dictionary Processes, 06Spring: Feature

Kortz, M., File Sharing Options: Elements of a Collaborative Infrastructure, 06Spring: Feature

Baker, K., Governance Working Group Proposes Updates to LTER By Laws, 06Spring: News

Yarmey, L., The Importance of Intertwining, 06Spring: Good Read

Baker, K., D.Pennington, and J.Porter, Multiple Approaches to Semantic Issues: Vocabularies, Dictionaries and Ontologies, 06Spring: Feature

Yarmey, L., Ocean Informatics Matlab Working Group, Mirroring the LTER Community Approach, 06Fall: Feature

Kortz, M., Three Challenges in Supporting Shared Workspaces, 06Fall: Feature

Baker, K., Scientific Meetings: Rigor, Relevance, and Variety, 06Fall: Editorial

Millerand, F., NSF Workshop: History and Theory of Infrastructure. Lessons for New Scientific Cyberinfrastructures, 06Fall: News

Baker, K., Metadata: Implementation of an International Framework, 06Fall: Good Read

Gragson, T., Data Curation in E-Science, 06Fall: Good Read Review

Conners, J., Database Storage Model Considerations: XML and Relational Database Approaches  
07Spring: Feature

Baker, K., J.Wanetick, N.Huffnagle, and M.Kortz, Information Infrastructure: Transitioning  
Directory Services, 07Spring: Feature

Haber, S., A Web Developer's View of the Research World and the Entertainment Industry,  
07Spring: Feature

Kaplan, N., C.Gries, K.Baker, D.Henshaw, T.Valentine, and J.V.Castle, Information  
Management Committee: GIS, Technology, and Changing Organizational Structures,  
07Spring: News

Millerand, F., On-going research collaboration-interopability, 07Spring: News

Baker, K., Computer Systems Development: History, Organization and Implementation,  
07Spring: Good Read

Grabner, S., Information Ecology: Open System Environment for Data, Memories and Knowing,  
07Spring: Good Read Review

Baker, K. and J.Campbell, What is the rationale for publishing DataBits twice a year?, 07Spring:  
FAQ

Kortz, M., Web-Based Data Visualization With JPGraph, 07Fall: Tools

Conners, J., YUI: An Open-source JavaScript Library, 07Fall: Tools

Baker, K.S. and R.Thombley, Place, Location, and Geographic Conventions, 07Fall: Good Read

Yarmey, L., Figuring on Insight through an Insightful Figure, 07Fall: Good Read Review

Baker, K., Professional Learning Opportunities: Conferences, Meetings, and Mindsets, 07Fall:  
Feature

Conners, J. and M.Kortz, Developing and Using APIs in System Design, 08Spring: Feature

Baker, K. and S. Grabner, Big Science and Local Meetings, 08Spring: Commentary

Yarmey, L., Preservation Metadata: Another Chapter in the Metadata Story, 08Spring:  
Commentary

Yarmey, L., Data Quality: Yet Another Chapter in the Metadata Story, 08Spring: Commentary

Baker, K.S., Cyberinfrastructure Primer, 08Spring: Good Read

San Gil, I., Digital Data Practices and the Long Term Ecological Research Program,  
08Spring:Good Read Review

Baker, K., Whirlwind Tour of Digital Curation in the UK, 08Fall: Commentary

Kortz, M., Getting Started with Web Services, 08Fall: Feature

Simmons, B. and J.Conners, Telling the Story Behind the Photos, 08Fall: Feature

Yarmey, L., Clutter is Failure of Design, 08Fall: Commentary

Conners, J., MySQL Workbench: A Visual Database Design Tool, 08Fall:Tools

Baker, K., Disputed Definitions, 08Fall: Good Read

Kaplan, N., Enabling Long-Term Oceanographic Research, 08Fall: Good Read Review

Palfner, S., Cyberinfrastructure Travels: Sharing & Shaping Time, Space and Data,  
09Spring:Feature

Petersen, R.I., Representing Geographic Features, 09Spring: Feature

Yarmey, L., Vocabulary Development as a Tool for Community-building, 09Spring: Feature

Baker, K., Pacific Coast Zooplankton Working Group: Data and Information Infrastructur,  
09Spring: News

Kortz, M., Data at Work: Supporting Sharing in Science and Engineering, 09Spring: Good Reads

Baker, K. and M.Bietz, Informatics and the Electronic Geophysical Year, 09Spring: Good Read

Yarmey, L., Continuing Education Options for Information Managers, 09Fall: Commentary

Wiley, S., Firebug: Web Customizing To Fit Your Needs, 09Fall: Tools

Kaplan, N. and K.Baker, Experiences from an Information Management Cross-Site Visit, 09Fall: Feature

Conners, J., Matplotlib: An Open Source Python 2-D Plotting Library, 09Fall: Tools

Kortz, M., LTER Unit Registry: Products and Processes, 09Fall: News Bits

Baker, K., Identifying Best Practice and Skill for Workforce Development in Data Curation, 09Fall: Good Read

Yarmey, L., An Introduction to the Panton Principles for Open Data in Science, 10Spring: Feature

Baker, K. and J.Wanetick, SIO Ocean Informatics Update: Growing Infrastructure in Support of Scientific Research, 10Spring: Feature

Baker, K.S., Information Manager Extraordinary Teleconferences: An ET Moment, 10Spring: News Bits

Henshaw, D., Webs of users and developers in the development process of a technical standard, 10Spring: Good Read Review

Baker, K., Note on Category Formation, 10Fall: Feature

Conners, J., Addressing Scaling Associated with Data Access, 10Fall: Feature

Baker, K., N.Kaplan, and E.Melendez-Colom, IMC Governance Working Group: Developing a Terms of Reference, 10Fall: Feature

Yarmey, L., Transitions and Comparisons, 10Fall: Feature

Kortz, M., Enactment and the Unit Registry, 10Fall: Feature

Baker, K. and E.Melendez-Colom, Evolution of Collaboration in Ecology, 10Fall: Good Read

Baker, K.S. and N.Kaplan, Network Identity: 2009 All-Site Milestone and Governance Issues, 11Spring, Feature

Baker, K. and M. Kortz, LTER Information management: Continuing Education and Site Change, 11Spring, Feature

Baker, K., Collaborative, cross-disciplinary learning and co-emergent innovation in eScience teams, 11Spring, Good Read

Baker, K., A Special Issue of Science on Data, 11Spring, Good Read

Haber, S., Technical Roles: Am I In IT?, 11Spring, Commentary

Conners, J., Notes on Design. 11Spring, Commentary

Donovan, J., Making Space for Information Management, 11Spring, Feature

Kortz, M., Review: The PersonnelDB Design and Development Workshop, 11Spring, Feature

Baker, K., Information Management, Data Repositories and Data Curation. 11Spring, Commentary

Baker, K., Wordle: Application for Generating Text Visualization. 11Spring, Good Tools.

## 4 SIO Requests for Action

### 4.1 Appendix: SIO Time Capsule and Long-Term Data

Subject: SIO Centennial Time Capsule and Long-Term Data  
Date: Mon, 16 Jun 2003 11:14:58 -0700  
From: Karen Baker <karen@guardian.ices.ucsb.edu>  
To: Kevin Hardy <khardy@ucsd.edu>  
CC: ckennel@ucsd.edu, tcollins@ucsd.edu, lshaffer@ucsd.edu, evenrick@ucsd.edu, kbaker@ucsd.edu

Dear SIO Centennial Organizers,

The call for contributions to a Scripps time capsule is thought provoking:

- > Subject: Re: [Ancient Mariners] Ancient Mariner eNews. May 2, 2003
- > 7. Time Capsule contents suggestions needed, due 01 August 2003.
- > On the Friday of our Centennial, two time capsules will start a journey
- > for Scripps 50 and 100 years hence. .The big question is: What . do
- > you think we should send along to our academic descendants? Send
- > an e-mail with subject "Time Capsule" to Kevin Hardy <khardy@ucsd.edu>.

A time capsule reaches back into an institution's past, displays the institution's present and reaches forward toward its own future. Historical data and interpretation of those data interact in distinctly different ways with future scientists. Persistent relevance is a hallmark of the long-term data sets and time series needed to understand environmental change. Scripps is renowned for such seminal work and irreplaceable data, e.g. C D Keeling's atmospheric CO<sub>2</sub> data, SIO pier time series, the CalCOFI and Santa Barbara Channel data sets. Including some of Scripps's noteworthy data in the time capsule would not only represent one of the Institution's most salient contributions to contemporary science, but would also preserve those data through the ensuing century.

Preparation of data for long-term preservation is a critical though often overlooked and underestimated task. In addition, providing access to the data is as important as providing the data themselves. If the time capsule recipients no longer have the means to convert a DVD to their contemporary presentation mode, the data will effectively be lost. Therefore, a "modern" presentation of data could be coupled with a visual presentation on a durable medium, such as paper. Perhaps the Keeling Curve warrants a Rosetta stone and an accompanying contemporary volume of contentious discussions surrounding the Kyoto Protocol.

We suggest an invitation be issued to the SIO community for contributions of selected long-term datasets, along with their stories, to be included into the time capsule.

-Karen Baker, Jerry Wanetick, Dawn Rawls  
SIO Integrative Oceanographic Division



## 4.2 Appendix: Response to SIO Director Search Request

### SIO Director Search 2006

A request was made to the faculty and staff prior to the search for a new director of SIO in 2006. Below is the response sent by an Ocean Informatics participant influenced by the notions of infrastructure, sociotechnical, and long-term.

#### **Comment**

In a new director I would look for an awareness and commitment to developing and supporting new interdisciplinary data practices, information interfaces, and learning environments; someone who knows the difference between information science and information technology and can create a balance that bridges to the scholarship of information studies and information systems design. As we face the challenges of developing new approaches to both long-term local endeavors and connectivity to global collaborative programs, data and knowledge management - frequently lumped under the cyberinfrastructure banner today - are traditionally underdeveloped, narrowly defined, and organizationally unrecognized.

#### **Specific Recommendations**

1. In the SIO search process seek an individual with an openness, sensitivity, and/or understanding of information stewardship and information infrastructure as part of their vision for contemporary scientific work.
2. Consider informatics, information management, and data stewardship as additional 'alternative' categories.

-Karen Baker

Palmer Station and California Current Ecosystem Information Manager  
Long-Term Ecological Research Program (PAL, CCE LTER)  
Ocean Informatics Initiative, Integrative Oceanography Division

## 5 Appendix: Ocean Informatics Reading Groups

OI held Reading Groups from 2003 to 2010. The purpose of a reading group is to foster conceptual development, create mental frameworks, and broaden perspectives through shared readings. Reading group characteristics include meeting regularly over time to stimulate dialogue, generate shared experiences, and build common vocabulary.”

### 1. Summer Informatics Reading Group 2010

#### 15 Jul 2010

KS Baker and FMillerand, 2010. Infrastructuring Ecology: challenges in achieving data sharing. In *Collaboration in the New Life Sciences*. J.Parker, N.Vermeulen, and B.Penders (eds). [http://interoperability.ucsd.edu/docs/10BakerMillerand\\_infrastructuringEcology.pdf](http://interoperability.ucsd.edu/docs/10BakerMillerand_infrastructuringEcology.pdf)

#### 23 Jul 2010

GC Fox and DGannon, D., 2006. Special Issue: Workflow in grid systems. *Concurrency and Computation: Practice and Experience*, 18(10), 1009-1019. <http://dx.doi.org/10.1002/cpe.1019>

TOinn, MGreenwood, MAddis, MNAIpdemir, JFerris., KGlover et al., 2006. Taverna: lessons in creating a workflow environment for the life sciences. *Concurrency and Computation: Practice and Experience*, 18(10), 1067-1100. <http://dx.doi.org/10.1002/cpe.993>

#### 30 Jul 2010

Char Booth, In The Library with the Lead Pipe, July 21, 2010. <http://www.inthelibrarywiththeleadpipe.org/2010/librarians-as-shapeshifting-at-the-periphery/>

John Graybeal, The Good Enough System, July 11, 2010. <http://marinemetadata.org/blogs/graybeal/the-good-enough-data-system>

Chris Rusbridge, Semantic Web of Linked Data, July 24, 2009. <http://digitalcuration.blogspot.com/2009/07/semantic-web-of-linked-data-for.html>

#### 06 Aug 2010

IHacking, 1983. The creation of phenomena. (Chapter 13) In *Representing and Intervening: Introductory Topics in the History of Natural Science*. Cambridge University Press, Cambridge. p220-232.

HSBecker, 1986. Telling about society. (Chapter 7). In *Doing things together: selected papers*. Northwestern University Press. Evanston, Illinois. P121-135.

#### 13 Aug 2010

LManovich and JDouglas, 2009. *Visualizing Change*.

<http://lab.softwarestudies.com/2008/09/cultural-analytics.html>

LManovich, Software Takes Command, 2008. In There is Only Software.  
<http://lab.softwarestudies.com/2008/11/softbook.html> )

LManovich, 2007. Databases as a symbolic form. In New Media.  
<http://con.sagepub.com/content/5/2/80.full.pdf+html>

### **20 Aug 2010**

RTomako, 2004. How I Explained REST to My Wife; <http://tomayko.com/writings/rest-to-my-wife>

AMiles, 2009. REST-not-so-easy? Data-Sharing Networks and the Atom Publishing Protocol. <http://alimanfoo.wordpress.com/2009/12/15/rest-not-so-easy-data-sharing-networks-and-the-atom-publishing-protocol/>

DHinchcliffe, 2008. What Is WOA? It's The Future of Service-Oriented Architecture (SOA)  
<http://hinchcliffe.org/archive/2008/02/27/16617.aspx>

### **27 Aug 2010**

EAronova, KBaker, and NOreskes, 2010. From the International Geophysical Year to the International Biological Program: Big Science and Big Data in Biology, 1957-present. *Historical Studies in the Natural Sciences* 40(2): 183-224.  
[http://interoperability.ucsd.edu/docs/10AronovaBakerOreskes\\_HNS.pdf](http://interoperability.ucsd.edu/docs/10AronovaBakerOreskes_HNS.pdf)

### **3 Sep 2010**

Latour, B. (1992). Where are the Missing Masses? *Sociology of a Few Mundane Artefacts*. In W. Bijker and J. Law (Eds.) *Shaping Technology, Building Society: Studies in Sociotechnical Change*. Cambridge, Mass, MIT Press: 225-258.  
[http://spiral-ulg.be/cours/STS\\_09-10/Lectures/11-03\\_Séance%20D/LATOURE%20%281992%29\\_The%20Missing%20Masses.pdf](http://spiral-ulg.be/cours/STS_09-10/Lectures/11-03_Séance%20D/LATOURE%20%281992%29_The%20Missing%20Masses.pdf)

## **2. Summer Informatics Reading Group 2009**

Memorable Quote: “The thing about this group is it’s a technology group that doesn’t think technology is the answer.”

### **18 Jun 2009**

JBirnholtz and MBietz, 2003. Data at Work: Supporting Sharing in Science and Engineering. *Proceedings of the ACM Conference on Supporting Group Work*, Sanibel Island, FL, November 9 – 12, 2003.

### **30 Jun 2009**

DNBaker, WKPeterson, and PFox, 2008. Informatics and the 2007-2008 Electronic Geophysical Year. *EOS* 89(48): 485-500.

### **14 July 2009**

JGrudin, 1988. Why CSCW Applications Fail: Problems in the Design and Evaluation of Organizational Interfaces. Proceedings of the 1988 ACM conference on Computer-Supported Cooperative Work: 85-93, ACM Press New York, NY, USA.

### **04 Aug 2009**

BRZeeberg, JRiss, DWKane, KJBussey, EUchio, WMLinehan, JCBarrett, JNWeinstein, 2004. BMC Bioinformatics 5:80.

SVeretnik, JLFink, PEBourne, 2008. Computational Biology Resources Lack Persistence and Usability. PLoS Computational Biology 4(7). <http://www.ploscompbiol.org>

### **09 September 2009**

Lee, C., Dourish, P., and Mark, G. 2006. The Human Infrastructure of Cyberinfrastructure. Proc. ACM Conf. Computer-Supported Cooperative Work CSCW 2006 (Banff, Alberta), 483-492. [//www.dourish.com/publications/2006/cscw2006-cyberinfrastructure.pdf](http://www.dourish.com/publications/2006/cscw2006-cyberinfrastructure.pdf)

## **3. Ocean Informatics Reading Group 2005**

This group re-emerged recently in response to the recognition of the benefits of integrating, reflecting, exploring, articulating, and dialoguing (iREAD!) on new perspectives enabled by contemporary information science and technology. The emerging plan is to meet monthly and to identify strategic design teams or working groups to pursue topics of immediate interest to the community. Occasional guest authors of papers will be invited.

### **March 14, 2005 - Semantics of the Web**

Tim Berners-Lee, JHendler and OLassila, 2001. The Semantic Web. Scientific American 0501:35-43.

### **6 April 2005 - Marine Metadata**

RLRiali, FMarincioni, and FLLLightsom, 2004. Content Metadata Standards for marine Science. USGS Report 2004. <http://pubs.usgs.gov/of/2004/1002/images/pdf/site.pdf>

Related Links:

MRIB System: <http://mrib.usgs.gov/>

MRIB Metadata: <http://mrib.usgs.gov/meta/>

MRIB Controlled Vocabulary: [http://mrib.usgs.gov/controlled\\_vocabulary/](http://mrib.usgs.gov/controlled_vocabulary/)

### **11 May 2005 – Ontologies: A Learning Trajectory**

DRibes and GCBowker, submitted. A Learning Trajectory for Ontology Building (2009). Between meaning and machine: learning to represent the knowledge of communities. Information and Organization 19(4):199-217. [http://interoperability.ucsd.edu/docs/09RibesBowker\\_Inf&Org.pdf](http://interoperability.ucsd.edu/docs/09RibesBowker_Inf&Org.pdf)

### **21 June 2005 – Information Ecologies**

Thomas Davenport, 1997. Information and Its Discontents: An Introduction. Chapter 1 in Information Ecology, Oxford University Press.

### **26 July 2005 – Information Exchange**

PCornellian, JGallagher, TSgouros, 2003. OpenDAP: Accessing Data in a distributed, heterogeneous environment. Data Science Journal 2: 159-169.

PCornellian, 2005. What Is a Data System, Anyway? Educause Review, March/April 2005, p.10-11.

## **4. Information Studies Reading Group 2004-2005**

This group began Fall Quarter 2004 as a collaborative learning mechanism for the Comparative Interoperability Project. Readings explore sociotechnical and human dimensions of information systems, data and information management. Participation includes Interoperability project participants, UCSD and SIO staff as well as students. Guest authors of papers will be invited occasionally.

### **18 Nov 2004, 6-8pm**

JFountain, 2001. Build the Virtual State: Information Technology and Institutional Change [chapters 1 (p3-17), 2 (p.18-30), 7(p.107-128), 11(p193-206) + footnotes]

Two factors of immediate interest: the distinction between objective technology and enacted technology as well as the collection of interesting case studies. Although the focus is on organizations (projects with goals/products) within institutional cultures (governments with processes/rules), the organization-institution distinction becomes less distinct in university settings where the project can be the organization within the university institution while simultaneously the university can be the organization within the NSF institution. Pertinent to those working with national computational centers, there is a lack of application of the objective-enacted distinction at the time of software development. Because the author tended to lump technical with objective and social with enacted, we were prompted into a lively discussion of how the technical was relevant in the enacted phase and the social in the objective phase.

### **09 Dec 2004, 6-8pm**

Atkins Report, 2003. Revolutionizing Science and Engineering Through Cyberinfrastructure. NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure ([http://www.communitytechnology.org/nsf\\_ci\\_report](http://www.communitytechnology.org/nsf_ci_report)). -Foster, Kesselman,

Tuecke: The Anatomy of the Grid, 2001  
(<http://www.globus.org/research/papers/anatomy.pdf>)

Internet Computing and the Emerging Grid, 2000  
(<http://www.nature.com/nature/webmatters/grid/grid.html>) The Atkins report three chapters total just over 100 pages and is not particularly dense. To direct our efforts read Appendix A and C but skip/skim the other appendices.] So what is this new beast "cyberinfrastructure"? It's related to the grid-eScience for which we

have two overview papers for background.

**13 Jan 2005, 6-8pm**

MCallon and BLatour, 1981. Unscrewing the Big Leviathan: how actors macrostructure reality and how sociologists help them to do so. *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies*. K.Knorr-Cetina and A.V.Cicourel. Boston, Mass, Routledge

S.S. Strum and B. Latour, 1987. Redefining the social link: from baboons to humans. *Social Science Information* 26(4):783-802

B. Latour, 1992. Where are the Missing Masses? The Sociology of a Few Mundane Artifacts. In *Shaping Technology, Building Society: Studies in Sociotechnical Change*. edited by Weibe E. Bijker and John Law, 225-258. Cambridge, MA: MIT Press

Moving from grand theories of society (structure) and the ethnomethodological highlight of everyday work of construction and negotiations to actor network theory with the origins and manifestations of power emerging from a blend of structure and process, the Latour readings bring focus to 'the experts' and to extending our language resources to include social (meaning to associate), macroactors, leaky black boxes, translation, negotiation, technomorphism, and obligatory passage points (OPP). These papers present Latour's sociotechnical ponderings on the concept of macro-actors, performative/negotiated social arenas, and prescriptive elements. So do we understand what part such perspectives, roles, and the 'distribution of competence' play in social science in general and in our work in particular? And can we see where technology (or technological artifacts) contribute to the shaping of society as the process of simplification occurs, the taking of the complex to the complicated in order to make it durable?

**10 Feb 2005, 6-8pm**

SLStar, Power, technology and the phenomenology of conventions: On being allergic to onions, in *A Sociology of Monsters: Essays on Power, Technology and Domination*

Star, 1989, Institutional Ecology, "Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39

Abbate, J. (1999). *Inventing the Internet*. Cambridge, MA, MIT Press Introduction (6p) and Chp5: The Internet in the International Standards Arena (34p)

Other/support readings:

Star, 1990, The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving, in *Distributed Artificial Intelligence*, vol2, Morgan Kaufman Publishers, Inc

Star and Bowker, 2002, How to Infrastructure, in *Handbook of New Media*, LA Lievrouw and S Livingstone (eds), London, Sage Publications

Star and Strauss, 1999, *Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible*

Work, Computer Supported Cooperative Work 8:9-30

Abbate, Inventing the Internet Chp1: White Heat and Cold War: The Origins and Meanings of Packet Switching Chp2: Building the ARPANET: Challenges and Strategies Chp3: The Most Neglected Element: Users Transform the ARPANET Chp4: From ARPANET to Internet Chp5: The Internet in the International Standards Arena Chp6: Propularizing the Internet

**10 Mar 2005, 6-8pm**

TAFinholt, 2004. Collaboratories. In Annual Review of Information Science and Technology. E B. Cronin (ed)

GM Olson and JS Olson, 2000. Distance Matters Human-Computer Interaction 15:137-178

SS Hale, AHMigliarese, MP Bradley, TJBelton, LDcooper, MTframes, CAFriel, LMHarwell, REKing, WKMichener, DTNicolson, BGPeterjohn, 2003; Managing Troubled Data: Coastal Data Parnerships Smooth Data Integration. Environmental Monitoring and Assessment, Kluwer Academic Publishers 81: 133-148

**14 April 2005, 6-8pm; guest Susan Sim**

Author Susan Sim will join us for a phone conference discussion.

Technical: SSim, SEasterbrook and RHolt, 2003. Using Benchmarking to Advance Research: A Challenge to Software Engineering in Proceedings, 25<sup>th</sup> International Conference on Software Engineering, Portland, Oregon, May, 2003

Social: JO'Connell, 1993. Metrology: The Creation of Universality by the Circulation of Particulars Social Studies of Science 23(1): 129-173. (pdf)

**19 May 2005, 6-8pm; guest David Obstfeld**

KEWeick, KMSutcliffe and DObstfeld, in press. Organizing and the Process of Sensemaking - Projects and Routines: Toward A More Concrete Specification of the Exploration-Exploitation Perspective. in preparation. Author David Obstfeld will join us for a phone conference discussion.

K.E.Weick, K.M.Sutcliffe, and D.Obstfeld, Organizing and the process of sensemaking. Organization Science, 16(4), 409-421

**28 June 2005, 6-8pm**

Amit Sheth, Changing Focus on Interoperability in Information Systems: from System, Syntax, Structure to Semantics in Interoperating Geographic Information Systems. Goodchild, Egenhofer, Fegreas, Kottman (1999)

**5. Science Studies Technology Reading Group 2003-2004**

This group met as an offshoot of the UCSD Science Studies Program under the guidance of Geoffrey Bowker (past Chair of the UCSD Communication Department and current Director of the Center for Science, Technology and Society at Santa Clara University).

Memorable Quote: A 'perspicacious object' is something that illustrates, provokes, challenges or otherwise opens up the question of technology in interesting ways.

**08 Oct 2003**

Italo Calvino - Six Memos for the Next Millennium, 1987

**05 Nov 2003**

Participants - Perspicacious Objects

Leo Marx, 1994. The Idea of Technology and Post Modern Pessimism. In Does Technology Drive History. MSmith and LMarx, eds, pp237-257.

**04 Dec 2003**

David de Leon - Building Thought Into Things, 1999

Ernest Boesch - The Sound of the Violin

**21 Jan 2004**

Dava Sobel - Galileo's Daughter, 1999 (Chp 4, 5, & 28)

M.G.Winkler and A.Van Helden, 1992. Representing the Heavens: Galileo and Visual Astronomy. Isis 83(2): 195-217.

**19 Feb 2004**

Bruno Latour - Paris, Invisible City; <http://www.bruno-latour.fr/virtual/index.html>

**18 Mar 2004**

Donald MacKenzie- Mechanizing Proof: Computing, Risk and Trust, 2003

Chp 2: Boardwalks Across the Tar Pits

Chp 9: Conclusion: Logics, Machines and Trust

**20 May 2004**

Kevin Warwick- March of the Machines, 1997

Chapter 1: In the Year 2050

Chapter 8: The Reading Robots -- An Overture

Chapter 9: Our Robots Today

Chapter 10: What Next With the Robots?

Chapter 11: A Fantastic Future?

**6. Ocean Informatics Reading Group Discussion 2003**

Discussion began about forming a group drawing on the SIO Integrative Oceanography Division interdisciplinarity and the Long-Term Ecological Research Program information management "community-of-practice". The reading group purpose is to discuss articles about data, information



management, information systems, and informatics topics with the explicit aim of exploring and learning to communicate about the concepts of information management, the Ocean Informatics Initiative, and an information environment.

December 2003. Take Back the Net. PC Magazine.

<http://www.pcmag.com/article2/0,4149,1400257,00.asp>

## 6 Appendix: Ethnographic Research

### 6.1 Appendix: What do you mean by ‘social’? An imaginary dialogue between a social scientist and an ecological scientist

What do you mean by ‘social’? An imaginary dialogue between a social scientist and an ecological scientist

By Florence Millerand, February 2006

*The Ecological Scientist (ES)*: I’ve heard that you’re involved in LTER as a social scientist. I’m curious, what kind of research do you do exactly?

*The Social Scientist (SS)*: As a social scientist, I’m generally interested in the study of social behavior and social arrangements. I’m particularly interested by the study of scientific communities (like the LTER community) and technological development, as results of human and social activities.

Currently, I work on a project about technologies that enable the exchange of data among scientists (we call cyberinfrastructures or large-scale information infrastructures). We compare three scientific communities, each tries to resolve the challenge of data sharing in their own way. In doing so, we do not merely concentrate on the chosen technical solution but also consider the often overlooked but crucial social and organizational dimensions of such technological projects.

*ES* : Sounds interesting. But I think I didn’t get all the concepts. Could you be more specific about what you look at when you say ‘social and organizational dimensions of technological projects’? And, I’d like to add, what’s the scientific purpose of your research?

*SS* : Let me begin first with a answer to your last question: what do we do this research for? The scope of my research consists of providing a better understanding of the *organizational complexity* of scientific cyberinfrastructure projects. One possible outcome is a better understanding of the changes and challenges associated with the development of these large-scale information infrastructures. We hope our findings will be useful to both communities we study and other communities with similar technological projects.

*ES*: It’s good if it’s useful ☺. So when you talk about organization, for me organization refers to structure, shelves, classification... and social refers to groups (and parties ☺). For instance in ecology, ‘social’ species are the ones that are organized in colonies as opposed to disseminated species, spread individually...

SS: Well, you're right, that's what these words can mean, but the way we use them is different. What we understand under organization is human association, for example groups of researchers or structured communities.

As for social, I agree, it's a fuzzy term that can mean many things. Actually even if it's a crucial category in social sciences, its meaning is difficult to capture because it may not be directly observable and visible. Basically, what we understand by social is human and group interactions and the results thereof. For example, science is a social production. It implies informal as well formal interactions between individuals, groups, and institutions.

So that's what we study, these interactions and ties.

ES: So how is social different from organizational?

SS: Good point. They overlap but they remain distinct concepts. For example, when you implement a new communication technology in a research institution, for example e-mail, you may look at the consequences at the organizational level (what may affect the hierarchy, allocated resources, and policies) and at the social level (how the researchers' identities, working practices, and relationships are changed).

ES: I think I get it, so the organizational is like the structure, the formal stuff, whereas the social is more human and informal?

SS: Yes kind of. Let's say that the organizational is what is the most visible, and the social the often invisible or unexpressed.

ES: Why is it important to study something invisible and unexpressed?

SS: Well, because of our background as social scientists, we know that this invisible stuff is critically important, for instance when we try to understand the impact of technological change. Let me give you an example. When an assembly line is introduced into a factory, the work efficiency may increase. But at the same time, the informal communication and relationships that were important sources of motivation for the workers in the former shop may be lost. A new technology always comes with the new working practices that fit it, and with the organizational structure that supports the whole. In this case, informal communication didn't fit the new organization of working practices associated with the assembly line. As a consequence, it may increase workers discontent that may translate to loss of productivity, strikes, and so on.

ES: Ah interesting.

SS: From this example, we draw that this technical thing, the assembly line, is in fact sociotechnical: it consists of technical as well as social components that are intertwined. In our research about cyberinfrastructures, we assumed that these technological projects were also sociotechnical in nature. The interoperability strategies that are put in place by the different communities require the simultaneous mobilization of community, technical and organizational resources. Because all these components are tied together, I talk about *configurations* of communities, technologies and organizations.

*ES* : You say ‘configuration’ because the communities, technologies, and organizations you talk about may be organized differently from a cyberinfrastructure project to another?

*SS*: Right. The three communities we study choose different approaches to achieve data interoperability, each of which is a configuration of a specific technology with specific communities in specific organizational arrangements.

*ES*: Understood. So why is your work so important?

*SS*: Well, the problem is: most of the time, these cyberinfrastructure projects are considered as technical issues (what would be the most suitable technology to achieve data interoperability) while in fact, such large-scale projects imply also important underestimated challenges at the social and organizational levels which may result in significant delays, costs, or frustrations.

*ES*: I think I get it. So, since I’m involved with LTER, and since we have adopted the EML specification as our metadata standard, I was wondering, what do you think of it?

*SS*: What we’ve noticed is that this strategy implies that the LTER researchers describe their datasets using EML, which represents a significant investment (of time, resources...) without an immediate benefit. In this case, the technology might seem good but implies a big burden upon the shoulders of your information managers. And what to say about the difficulty of convincing the researchers that the required investment is worth making since nothing in the current system rewards them in the short run?

Another problem in the long run is that the standard itself needs maintenance over the years. This requires a certain level of expertise, skilled people whose time has to be dedicated to work on the standard sustainability. But, the experts who have developed the standard are now working on other projects, and LTER information managers have enough work enacting the standard across the network. Who is going to take this job and the responsibility that goes with it? The organizational challenges are far reaching and might impact the very existence of the infrastructure in the long term.

Further, putting the standard into practice may imply some changes in the way scientific data are recorded and managed at the labs and research stations inside the community. When can a special measurement unit that has been created at a specific research station be acknowledged as a special unit or recognized as an LTER unit?

To conclude, the enactment of the EML standard in LTER and more broadly in the ecological research community comes with the transformation of the daily practices and organization of ecological science.

I hope this is helpful to understand what we’re doing.

*ES*: It was helpful, thank you. So, what is the link between social and party again ☺?

## 6.2 Appendix: Ethnographic Field Hand-outs

March 2007

### Ethnographic Fieldwork and Infrastructure Studies

Florence Millerand and Karen Baker

*Project: Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study*

Project Web Page: <http://interoperability.ucsd.edu>

NSF Program(s): Scientific Testbeds; Human Social Dynamics: Agents of Change

NSF URL: <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0433369>

**We ask: How are data going to be made available...and widely usable? How will infrastructure and information systems be built to support data sharing?**

**Our goal: Comparative analysis with the goal of identifying features and facets of collaborative communities working on issues of interoperability.**

This project explores the centrality of collaborative, interdisciplinary work in building information infrastructure. As new scientific infrastructure is emerging, a central question being posed is how to share data across time and across distributed organizational and social contexts. This issue is particularly important since some of the great political questions of our day, such as understanding climate and developing a sustainable relationship with our environment, depend upon the ability to federate data across organizational and disciplinary contexts. There have been a wealth of suggestions for technical fixes for this pressing concern, but there has been little study - and no comparative study - of the organizational and social dimensions of differing data handling and integration strategies.

As contemporary scientific questions increase in scope, conceptual and methodological frameworks must also broaden. Our project brings together a collaborative interdisciplinary team to address jointly selected contemporary cyberinfrastructure issues focusing on local practices and technology use that supports long-term scientific endeavors. We are looking simultaneously at the interdependent technical, organizational, and social processes involved in informatics and information system design including classification strategies, organizational structures, and ways of working as well as participant roles and responsibilities.

Through comparative study of three scientific communities - GEON, LTER, and Ocean Informatics - we seek to develop a grounded understanding of the complexities involved in producing and sustaining a shared scientific information infrastructure. Our methods draw from qualitative research - and include grounded theory, action research, design and sociotechnical analysis as well as systems and information science approaches. We conduct ethnographic analysis on documents and interviews; we use collaborative design in order to consider and facilitate interfaces with and between data, technology, and participants. Through design and articulation work such as community dialogue and mutual learning, we focus on building awareness of configurations and ramifications of technology use in today's scientific data handling arena.

Our work blends research and application, stretching from theory to enactment. While conducting infrastructure research, **we are sensitizing informatics, environmental science, and science studies communities to the need to consider in partnership the social and organizational dimensions of local work practices together with the technological.**

Project References:

KSBaker, DRibes, FMillerand, GCBowker, 2005. Interoperability Strategies for Scientific Cyberinfrastructure: Research and Practice. Proceedings of the American Society for Information Systems and Technology  
[http://interoperability.ucsd.edu/docs/05ASIST\\_CIP\\_wbox.pdf](http://interoperability.ucsd.edu/docs/05ASIST_CIP_wbox.pdf)

August 2006

**Ethnographic Fieldwork and Design Studies**  
Karen Baker, Brian Lindseth, and Florence Millerand

*Project: Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study*

Project Web Page: <http://interoperability.ucsd.edu>

NSF Program(s): Scientific Testbeds; Human Social Dynamics: Agents of Change

NSF URL: <http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0433369>

**We ask: How are data going to be made available...and widely usable? How will infrastructure and information systems be built to support data sharing?**

This project explores the centrality of collaborative, interdisciplinary work in building cyberinfrastructure. As new scientific infrastructure is emerging, a central question being posed is how to share data across time and across distributed organizational and social contexts. This issue is particularly important since some of the great political questions of our day, such as understanding climate and developing a sustainable relationship with our environment, pivot on the ability to federate data across organizational and disciplinary contexts. There have been a wealth of suggestions for technical fixes for this pressing concern, but there has been little study - and no comparative study - of the organizational and social dimensions of differing data handling and integration strategies.

As contemporary scientific questions increase in scope, conceptual and methodological frameworks must also broaden. Our project brings together a collaborative interdisciplinary team to pose and address jointly contemporary cyberinfrastructure issues including local practices and technology use that supports long-term scientific endeavors. We are looking simultaneously at the interdependent technical, organizational, and social processes including information system design, organizational structure, ways of working, participant roles and responsibilities.

Through comparative study of three differing scientific communities: GEON, LTER, and Ocean Informatics, we seek to develop a grounded understanding of the complexities involved in producing and sustaining a shared scientific information infrastructure. Our methods draw from qualitative research - and include: grounded theory, action research, design and sociotechnical analysis, as well as systems and information science approaches. We conduct ethnographic analysis jointly on documents and interviews; we use collaborative design in order to facilitate interfaces with and between data and technology. Through design and articulation work including community dialogue and mutual learning, we focus on building awareness of configurations and ramifications of technology in today's scientific data handling.

Our work blends research and application, reaching from theory to enactment. **While conducting infrastructure research, we are sensitizing informatics, environmental science, and science studies communities to the need to consider in partnership the social and organizational dimensions of local work practices together with the technological.**

Project References:

KSBaker, DRibes, FMillerand, GCBowker, 2005. Interoperability Strategies for Scientific Cyberinfrastructure: Research and Practice. Proceedings of the American Society for Information Systems and Technology. [http://interoperability.ucsd.edu/docs/05ASIST\\_CIP\\_wbox.pdf](http://interoperability.ucsd.edu/docs/05ASIST_CIP_wbox.pdf)

July 2004

## **Introduction to the Ocean Informatics Project**

Karen S. Baker, Steven Jackson, Jerry R. Wanetick

In recent decades, changes in the nature and practice of ocean science have driven (and in some cases, been driven by) parallel shifts in the information technology and computational landscapes. But there have been as yet few scholarly attempts to explore the intersection of these two worlds, and fewer still examining these dynamics in concrete organizational settings. The Ocean Informatics Project, based in the Integrative Oceanography Division (IOD) at the Scripps Institution of Oceanography (SIO), joins ocean, information, and social scientists in a collaborative effort to design adaptive and scalable information systems suitable for supporting the diverse work worlds of integrative ocean science. The success of larger-scale collaborative efforts depends upon facilitation of communication and data handling as well as upon a supportive technical, organizational, and social infrastructure.

Research during the initial phase of the project draws heavily on interviews and participant observation conducted with members from across IOD and related communities, ranging from PIs to information managers, graduate researchers and administrators. Such ethnographic field work places emphasis on identifying past and current work practices, with a particular focus on shifting patterns of data collection, use, sharing, and storage. Subsequent phases of the project will draw on initial findings and employ participatory, collaborative design methods developed in the social sciences to support the growth of locally-appropriate innovation strategies responsive to changes in the real-world data practices of ocean science.

---

Karen Baker is an information manager at SIO at UCSD working with the Long-Term Ecological Research Program's Palmer Station site and the California Coastal Ecosystem site. Her research has ranged from bio-optical oceanography to informatics, and more recently from cooperative scientific work to collaborative systems design. She can be reached at [kbaker@ucsd.edu](mailto:kbaker@ucsd.edu), tel: 858-534-2350.

Steven Jackson is a doctoral candidate in the Department of Communication and the Science Studies Program at UCSD, and a doctoral fellow at Harvard University's National Center for Digital Government. His dissertation research explores the work of computer modeling and water policy in the American Southwest and U.S.-Mexico border region. He can be reached at [sjjackson@ucsd.edu](mailto:sjjackson@ucsd.edu), tel: 858-452-8056.

Jerry Wanetick, director of the IOD Computational Center at SIO, provides support for long-term physical oceanographic studies through information system administration, data management, and cyberinfrastructure design. His work ranges from design and field deployment of data acquisition systems including remote sensing and wireless to a current focus on creating a working environment for information systems development. He can be reached at [jwanetick@ucsd.edu](mailto:jwanetick@ucsd.edu), tel: 858-534-7999.

Team members Charleen Johnson (Logistics and Transcription), Dawn Rawls (Science Editor and Informal Education), Beth Simmons (Education and Outreach) and Shaun Haber (Computational Infrastructure and Collaborative Tools) add breadth and depth to Ocean Informatics endeavors.

### 6.3 Appendix: Interview questions

Interview Outline (for Data/Information managers)

JUNE 15, 2008

1. What is your job here at Scripps?
2. What roles does that imply?
3. How long have you been working at Scripps?
4. How has your job evolved during that time?
5. What is your relationship with the scientists?
6. What is your role with data? How do you interact with it?
7. How would you define interoperability?
8. According to you, what could be described as infrastructure in this project?
9. Can you define Ocean Informatics?
10. What are the objectives associated with OI?
11. What are the main challenges OI needs to face?



## 6.4 Appendix: Memos on Grounded Theory

Two memos summarize and reflect on the grounded theory approach.

### **Grounded theory: analytical process and comparison with other similar approaches By Claude Arsenault**

#### What is grounded theory?

Grounded theory is a qualitative analysis method. Its goal is to build theories concerning the structure and constancy of social phenomenon which have rarely been studied before. It is based on the comparison of different instances of a certain phenomenon. It does not try to achieve description of the phenomenon.

Its particularity lies in the fact that it tries to provide a valid account of the social phenomenon by using legitimate systematic analysis methods and a well chosen data sample. The broadness of the sample will not, however, allow the researcher to verify his theory. The main goal of grounded theory is to create or suggest a new theory, which explains that, throughout the sampling, the researcher will look for theoretical sufficiency rather than empirical sufficiency. Theoretical sufficiency can be defined as the integration, in the theory which is being built, of all the instances of the social phenomenon. Using this method, the sampling will stop when the new incidents found do not permit the researcher to identify different incidents.

Grounded theory is based on two main methodological principles. The first one, which originated in American pragmatism, states that a phenomenon must be grounded in its context to be studied properly. It therefore justifies the use of *in situ* observation as to make changes and processes more obvious. The second principle came from phenomenological philosophy. According to this principle, in order to build the theory, the researcher must put aside all pre-existing literature and data relative to the phenomenon he is studying. The concepts and hypotheses must be built as the research progresses.

The process of grounded theory is systematized and clearly defined. Its main steps are included in the codification, which is the core of continuous comparative analysis. Codification itself can be divided in three levels, between which the research will constantly be alternating. Each level brings a higher integration and greater delimitation.

Another key concept of grounded theory lies in theoretical sampling. The sampling in this type of research is based on the research question. Its goal is to be representative conceptually rather than statistically. It needs to reflect all the possible instances of the phenomenon under study. Theoretical sampling following the same general levels as codification and stops when theoretical sufficiency is achieved.

**Reference:** Lapierre, A. (1997). La théorisation ancrée (grounded theory): démarche analytique et comparaison avec d'autres approches apparentées. In J. Poupart, Deslauriers, Groulx, Laperrière, Mayer & Pires (Eds.), La recherche qualitative. Enjeux épistémologiques et méthodologiques (pp. 309-340). Boucherville, QC Gaétan Morin Editeur.

**Grounded Theory Method**  
**By Brian Lindseth**  
**060520**

**Definition:**

Grounded theory represents one type of qualitative analysis with underlying assumptions regarding development of theory usually emerging concurrently with field research and based on data (experiential, participative, documents). Analytic operations involve data collecting, theoretical coding and memoing. The method aims for rigor through systematicity of coding, iterative development of categories from data-analysis work, and comparative analysis. Grounded theory is a label and a literature with key concepts including theoretical sampling and theoretical selection or sensitivity as distinct from empirical selection. Need for explicitness in methods, object of study, and level of abstraction (descriptive to general theory) eventually adopted.

**Related concepts:** qualitative analysis, theoretical sampling, theoretical coding, theoretical sensitivity, constant comparative method

**References:**

1. Strauss, Anselm L. Qualitative analysis for social scientists  
Cambridge, Cambridge University Press, 1987.
2. Strauss, Anselm L. and Juliet Corbin. Grounded Theory Methodology: An Overview. In Norman Denzin and Yvonna Lincoln (Eds.) Handbook of Qualitative Research. p273-285, Thousand Oaks, CA, Sage Publications, 1994.
3. Glaser, Barney G.  
Title Theoretical sensitivity in Advances in the methodology of grounded. Chap 3.  
Publisher Mill Valley, CA, Sociology Press, 1978.
4. Historical  
1967 Glaser & Strauss, Discov of Grounded Theory: Strategies for Qual Research  
1990: Strauss & Corbin, Basics of Qualitative Research  
1997: Strauss & Corbin, Grounded Theory in Practice

**Grounded Theory**

**General**

Grounded theory specifies an approach in which the analyst remains 'close' to the social world that is the object of investigation. Theory emerges out of an analyst's engagement with the world being studied in a process of continuing interaction and revision. Founded by Anselm Strauss and Barney Glaser, grounded theory is defined against the kind of speculative theorizing characteristic of mid century sociologists such as Talcott Parsons. In 1965 the two published Awareness of Dying in 1965, and followed up with The Discovery of Grounded Theory in 1967.

### **Core concepts**

Strauss and Glaser describe the process of constructing grounded theory as occupying three steps--**data collection, coding and memoing**. The analyst **observes** a social world of interest. Then, he or she can start to identify distinctions that are important in this world. Looking at sick patients, an analyst could distinguish between machinery that enters the patients' bodies and machinery that does not. Then, these distinctions can be **dimensionalized**. In other words, sub distinctions that are important in the world under investigation can be identified. In addition to observations, analysts can also draw from the experiences we as social actors bring to the investigation (the fact that we can be said to know things that the people observed might also know).

**Coding** refers to how data is conceptualized and can refer to the work of coming up with categories in which to fit the data or the relationships between categories. Tentative conclusions and questions about the relations between categories or dimensions can be treated as hypotheses to be held up against further observation and experience. In the coding of data--the ways in which an observed phenomenon fits into a category--care must be taken to make sure the categories and the coding of observations and interviews is tightly linked with the observations and interviews in their context.

Theories emerge as memos based on the coding are examined in relation to each other and to the process of coding and observation. As they are being constructed, attention should be paid to what distinctions and relationships are more important than others or which ones will lie at the core of a theory. These core categories are ones that play a central role in integrating various facets of a theory. When the links and relationships are examined, held up against observations and experience, they can provide **conceptual density** to theories. Once theories emerge--or are constructed--they should be verified, held up once again to observations and experience. Once further observation or analysis won't add anything to the theory, it has reached a point of **saturation**.

It is perhaps important to note that--while Strauss and Glaser identify the stuff of research as three phases, observation, coding and memoing--these phases are very much intertwined in practice as a researcher is constantly holding emerging theories against observations.

**Theoretical sampling** is another core concept for grounded theory. It refers to the ways in which further observation or data collection can be guided by emerging theory and the insights that can be gained by comparing observations gained in different samples or among different populations. It seems as though theoretical sampling could provide an approach to the use of **comparison** (and the logic behind the selection for potential sites).

### **In relation to..**

Grounded theory inherits emphases on action and the problematic situation from pragmatism and its emphasis on qualitative methods (and much else) from the 'Chicago School' of sociology.

**Robert Park** is seen as one of the founders of the 'Chicago School' of American sociology that prevailed at the University of Chicago roughly from 1920s through the 1950s. This tradition is often associated with the use of field observation and interviews to grasp the views of actors embedded in the social world under investigation.

**Everett Hughes** was a student of Park's. I know very little about Hughes other than that his emphasis on work came to be influential to Howard Becker.

A student of Hughes at the University of Chicago, **Howard Becker** is often viewed as the founder of labelling theory in his work Outsiders: Studies in the Sociology of Deviance published in 1963. Having worked as a jazz musician through graduate school, Becker researched the world of jazz musicians, investigating topics such as marijuana smoking among musicians by interviewing musicians. His later work, Art Worlds (1982), provides a compelling and accessible account of the worlds in which artists work. His attention to the material world which the artists operate and in which art is located and travels (and is constrained) is interesting when considered next to recent emphases in science studies on materiality and the environment. Becker provides insight into his methodology and emphasis on clear writing in Tricks of the Trade (1998) and Writing for the Social Scientists: How to Start and Finish Your Thesis, Book, or Article (1986).

The emphasis on close engagement with the social world being studied can also be seen in the work of scholars such as **Herbert Blumer**. Strauss was student of Blumer's at the University of Chicago in the 1940s. Himself a student of George Herbert Mead, Blumer is often associated with 'symbolic interactionism.' Here there is an emphasis on the individual in (a specific) context--in in his or her 'natural world'--and the importance of meaning. Meaning is considered to be a source of action and emerges out of interaction. This view problematizes the traditional notion that meaning operates as an attribute inscribed in objects, independent of any interaction with people (as in the Kantian noumenal realm). There is also an emphasis here on the importance of interpretation that emerges in something like an internal dialogue. While the meaning of objects comes from people, objects can resist our conceptions or the meanings we might assign to them.

Grounded theory seems to share interesting similarities to themes in **actor network theory**. Here an emphasis on following the actors resembles the emphasis, in grounded theory, on maintaining a close relationship with the data and the milieu out of which it can become 'data.' Here there is an emphasis on tracing the links by which a heterogeneous set of actors and objects can hold together as a network. The network is a kind of accomplishment here as the links must be continually enacted for the network to 'hold.' The approach of picking an object and 'following it' as different people touch it and as it becomes embedded in the practice of different social worlds describes one way of trying to trace the links in a milieu of interest.

It might be interesting to investigate the similarities and differences between grounded theory and approaches such as **ethnomethodology**. Often associated with the name of **Harold Garfinkle**, ethnomethodology could be considered to be a related approach. It is qualitative and emphasizes a close engagement with the social world under investigation. Here the emphasis on the embeddedness of the investigator in the world under investigation seems to resemble the rejection, in grounded theory, of speculative theorizing that presents itself as scientific, removed from its object as something made of a different substance. Further, Latour seems to have been influenced by some of these kinds of qualitative approaches. Also, I would like to reread some of Leigh Star's work to see grounded theory in action.



## 6.5 Appendix: Steps of Grounded Theory Analysis

A table synthesizes the 6 steps of grounded theory analysis from coding to theorizing.

### Steps of Grounded Theory Analysis

By Claude Arsenault

From: Paillé, P. (1994). L'analyse par théorisation ancrée. *Cahiers de recherche sociologique*, 23, 147-181.

In English

STEP	SUB-STEPS	DESCRIPTION	QUESTIONS	OBJECTIVE(S)	RESULT(S)
CODIFICATION		Identify, name, summarize, thematize each and every line	What is there here?	Identify what is being said in the interview	Codified corpus
		what is being said in the text on which the analysis is being done	What is it? What is it about?	without repeating the actual text	Codes
CATEGORIZATION	1. Creation of a list of already existing categories	Bring the analysis to a conceptual level by naming the phenomena and events which come out of the data in a richer and more encompassing way	What is happening here?	Bring the analysis to a conceptual level	Categories Analytical memos
	2. Second reading (more conceptual)		What is it? Which phenomenon am I facing?	Put the phenomenon in perspective	
	3. Construction and consolidation of the categories				
	3.1. Definition	Specify what the category refers to	What do you mean by...?		
	3.2. Identification of the properties	Establish the category's attributes	What is the category composed by? What are its attributes?		
	3.3. Specification of social conditions	Determine which elements are essential for the category to be applied			
	3.4. Identification of their different forms	Follow the evolution of many dimensions of the phenomenon			
LINKAGE	1. Write a list of the categories	Compare the categories of the analysis		Find links between the categories	Graphic representation
	2. Examine the categories with the questions	3 possible approaches:	Is what I have here linked with what is there?	Improve the analysis	
	2.1. Graphic representation	- empirical approach - speculative approach - theoretical approach	How and in what is it linked?	Evolve from description to explanation	
INTEGRATION		Establish precisely the study's subject	What is the main problem? In general, which phenomenon am I facing? In definitive, what will my study be about?	Determine the precise objet of the analysis	Precise object
ESTABLISHMENT OF A PATTERN		Reproduce as reliably as possible the organization of the	What is the type of this phenomenon?	Identify the important attributes of the phenomenon,	Model
		structural and functional relations characterizing a	What are the attributes of this phenomenon?	its usual steps, its strong moments, its consequences	
		phenomenon, an event of a system	What precedes this phenomenon? What are the consequences of this phenomenon? Which processes are involved in this phenomenon?	on different levels, etc.	
THEORIZATION	3 strategies:	Reinforce the theory and weaken the diverging explanations		Consolidate the theory	
	- theoretical sampling	Sample the different manifestations of a phenomenon		Understand its variation	
	- verification of theoretical implications	Indicate the implications logically following from the theory Confront the explanation of the	If... then?	Make sure the data supports the hypotheses	

In French

ÉTAPE	ÉTAPES	DESCRIPTION	QUESTIONS	OBJECTIF(S)	PRODUIT
CODIFICATION		"Dégager, relever, nommer, résumer, thématiser,	Qu'est-ce qu'il y a ici?	"Dégager le témoignage livré lors de l'entrevue en évitant de répéter le verbatim"	verbatim codifié
		presque ligne par ligne, le propos développé à l'intérieur du corps sur lequel porte l'analyse"	Qu'est-ce que c'est?		codes
			De quoi est-il question?		
CATÉGORISATION	1. Dresser la liste des catégories déjà formées	"Porter l'analyse à un niveau conceptuel en nommant de manière plus riche et englobante les phénomènes et événements qui se dégagent des données"	Qu'est-ce qui se passe ici?	Porter l'analyse à un niveau conceptuel	Catégories
	2. Nouvelle lecture plus conceptuelle		De quoi s'agit-il?	Mettre le phénomène étudié en perspective	Mémos analytiques
			Je suis en face de quel phénomène?		
	3. Construction et consolidation des catégories				
	3.1. Définir	Spécifier ce à quoi la catégorie renvoie	Qu'entends-tu par...?		
	3.2. Dégager les propriétés	Déterminer les caractéristiques de la catégorie	De quoi la catégorie est-elle composée? Quels sont ses attributs?		
	3.3. Spécifier les conditions sociales	Déterminer ce qui doit être présent pour que la catégorie s'applique			
3.4. Identifier leurs diverses formes	Suivre l'évolution de plusieurs dimensions du phénomène				
MISE EN RELATION	1. Dresser une liste des catégories	Comparer des catégories de l'analyse		Trouver les liens entre les catégories	Schématisation
	2. Examiner les catégories à l'aide des questions	3 approches possibles:	Ce que j'ai ici est-il lié avec ce que j'ai là?	Raffiner l'analyse	
	2.1. Schématisation	- approche empirique - approche spéculative - approche théorique	En quoi et comment est-ce lié?	Passer de la description à l'explication	
INTÉGRATION		Déterminer précisément l'objet d'étude	Quel est le problème principal?	Délimiter l'objet précis de l'analyse	Objet d'étude précis
			Je suis en face de quel phénomène en général?		
			Mon étude porte en définitive sur quoi?		
MODÉLISATION		"Reproduire le plus fidèlement l'organisation des relations	De quel type de phénomène s'agit-il?	"Dégager les caractéristiques importantes du phénomène, son déroulement	Modèle
		structurelles et fonctionnelles caractérisant un phénomène,	Quelles sont les propriétés du phénomène?	habituel, les moments forts de son existence, ses conséquences à divers	
		un événement ou un système"	Quels sont les antécédents du phénomène? Quelles sont les conséquences du phénomène? Quels sont les processus en jeu au niveau du phénomène?	niveaux, etc. "	
THÉORISATION	3 stratégies:	Renforcer la théorie et affaiblir les explications divergentes		Consolider la théorie	
	- échantillonnage théorique	Échantillonner les diverses manifestations d'un phénomène		Cerner la variation	
	- vérification des implications théoriques	Indiquer les implications découlant logiquement de la théorie	Si... alors?	Vérifier si les données soutiennent les hypothèses émises	
	- induction analytique	Confronter l'explication du phénomène aux "cas négatifs"			

## 6.6 Appendix: Response - Steps of Data Management

The memo 'Steps of Grounded Theory Analysis' inspired a memo on steps of Data Management.

### Steps of Data Management and Steps of Grounded Theory

By Karen Baker

13 January 2008

The basis of grounded theory is well thought out. When summarized as steps, the steps appear pertinent to the fields of data and information management as well as to grounded theory. Because the steps have been developed at a meta or conceptual level, it may be possible to use them to generate awareness of the 'Steps of Data Management'.

1. issue framing
2. sampling design
3. data collection
4. codification
5. categorization
6. linkage
7. integration
8. establishment of a pattern
9. theorization

The first outcome of such an exercise is the realization that there are missing first steps. The new first step I have struggled to articulate over the years within the realm of information management: the framing or bounding of the issue at hand. I have identified this step as the point at which the context of subsequent work is constrained for a scientist. This early constraint creates barriers to comprehending subsequent data handling issues in the data workflow. In order to make this understanding visible, I have labeled this comprehension capacity as a 'readiness factor'. I wonder whether Grounded Theory would benefit from including an 'issue framing' step, and what would be the ramifications of this step-making process?

Steps 2 and 3 involve sampling design and data collection. These influence data description and analysis significantly. When listed, they become 'part of grounded theory'. A great deal of reparation work seems to result when the list is considered in a narrower, fragmented or atomized approach without steps 1-3. Information Systems Journals encourage submissions today to include discussion of sampling and collection specifically. In data reuse, metadata requires gathering of information about these topics so that the data can stand independently outside the immediate location and collection activity.

A second outcome seems to be that steps 2 and 3 are frequently absent so unavailable to innovation in making or analyzing a collection. This absence, in turn, makes difficult, if not impossible to perform the later steps of linkage, integration and establishment of a pattern. One may hypothesize that Steps 1 through 3 are missing because their focus is the scientific experiment/observation theory, logistics of sampling, sample collection and capturing of the data. This focus is one of data management rather than on the data record and its organization, that is, on information management.

\*1 This memo draws on the steps of grounded theory table prepared by Claude Arsenault that summarizes an article by Pierre Paillé "la théorisation ancrée, résumant un peu les étapes".



## 6.7 Appendix: Memo on Conceptualizing Categories

### Qualitative analysis by conceptualizing categories

By Claude Arsenault

From: Paillé, P., Mucchielli, A. (2003). L'analyse qualitative à l'aide des catégories conceptualisantes. In *L'analyse qualitative en sciences humaines et sociales*. Paris: Armand Colin, p.147-179.

### What is the analysis by categories?

The analysis by categories is a qualitative analysis method. It permits the researcher to conceptualize and theorize while the analysis is going on, unlike any other method. This implies that during the annotation of the corpus, the researcher writes down category names in the margins. Throughout the process, those categories will be worked on, merged, divided or refined. They will be the core of the analysis and of the research report.

Paillé identifies three types of analytic work that are essential to the analysis by categories. First, a work of *analytical description*: the first categories to be created will simply name the phenomenon, making the immediate significations present in the corpus more obvious without adding an analytical dimension to them. When this step is completed and the corpus is well annotated, the researcher will continue with *interpretative deduction*. At this point, he will start to create significations, by either one of two methods. The first method is to use theoretical references, which help to situate the texts in a broader context. The second method implies using *theorizing induction*. This term refers to the construction by the researcher of his own categories and to their identification with precise terms and unique expressions. For example, Paillé mentions Auziol's "double communication".

The category is at the center of this type of analysis. It represents a set of condensed significations, meaning that its definition takes into account every aspect of the phenomenon it defines. It can be applied to every type of research material used for a certain project, no matter their nature or the nature of the phenomenon they describe. The elaboration of the category takes place in two fundamental steps, clarifying the category and validating it. To ensure that the category is clarified successfully, Paillé suggests to try and define the category, to specify its properties and to identify its existence conditions.

When talking about the internal and external validity of a research using analysis by categories, it is important to understand that the classic concept of external validity does not apply here, since what makes such a process so valuable is its unique character and its tight link with the theoretical background of the researcher.

### What is the difference between analysis by themes and analysis by categories?

Contrary to a theme, a category designates directly a phenomenon. It goes beyond the descriptive nature of the theme by going over the simple content designation. It represents the creation of significations, and it is a fundamental constitutive element of the analysis and of the theorisation which will follow.

In the analysis by themes, the use of the category does not have the same level of importance. The concept of category itself is different. In theme analysis, the term category designs a rubric, whereas in category analysis, the category is directly involved in the analysis.

The gap which exists in theme analysis between the definition of the categories and their analysis is erased in category analysis by the role of the categories in the analysis.

### **What are the main strengths and weaknesses of the analysis by categories?**

The analysis by categories is conceptually different from the other types of qualitative analysis. Before engaging in an analysis of this type, it is essential to understand what separates it from content analysis or theme analysis. The concept of category as seen in this type of analysis is not instinctive. It is important to grasp all its significations before proceeding to category analysis.

The categories can apply to any research material, but it is obvious that some categories will be at different levels, and that some will be more dense than others. We must see this variability as a strength of category analysis, because it permits the researcher to work with significations, slowly building the analysis. Each element thus has its place.

Two main difficulties might arise while using category analysis. First of all, the researcher has to be very careful not to paste in his analysis interpretations coming from anterior work or the literature. The interpretation must here rest on the construction of categories unique to the corpus. Similarly, in order for the categories to define entirely and perfectly the studied phenomenon, they must be unique to them, and not borrowed from other works.

**Reference:** Paillé, P., Mucchielli, A. (2003). L'analyse qualitative à l'aide des catégories conceptualisantes. In *L'analyse qualitative en sciences humaines et sociales*. Paris: Armand Colin, p.147-179.

## 6.8 Appendix: Memo on Ocean Informatics Definitions

A targeted analysis on ethnographic materials (interviews) prompted the writing of a memo on Ocean Informatics various definitions by participants.

MEMO - O.I. DEFINITION - FROM TARGETED ANALYSIS

JANUARY 25, 2008

CLAUDE ARSENAULT

Through the analysis of four interviews held for the Ocean Informatics monograph project, different visions and definitions of Ocean Informatics emerged. The core notion at the heart of each of these definitions seems to be interoperability – whether from a technical standpoint or a social one. Every individual seems to have a unique perception and conception of Ocean Informatics’s nature and of its role, centered on different dimensions of the project: communication, infrastructure, or data.

IM1<sup>1</sup>, for instance, describes in her interview Ocean Informatics as being a **communication** mechanism between people. To her, OI is “the bubble which allows communication to happen”. The physical infrastructure and data interoperability are means which help the participants attain communication. To describe this, the expression “social interoperability” seems the most relevant. This concept refers both to the human aspect of Lynn’s definition and to interoperability, seen in a communicative perspective.

IM2 has a completely different approach. He defines OI as an “effort to bring together a lot of disparate areas of both data and expertise”. His description focuses on the integration of the data from a **technical standpoint**. Contrary to IM1, he sees communication as a way to enable the technique. He agrees that channels of communication and a shared vocabulary must be created in order to achieve interoperability, but sees them as means rather than the goal to achieve. IM2 also mentions the iterative nature of OI’s development and its particular culture of acute conceptualization and identification.

IM3 and IM4 both center their OI definition on **data**. To IM3, Ocean Informatics is the physical infrastructure which allows the scientists and the data managers to format and integrate the data. OI is a way to put data together and easily access it by achieving interoperability. What it provides ultimately is an easy access to information. To differentiate IM3 from IM4, we can say the former insists on **data use** whilst the latter concentrates on **data formatting** and the whole process behind obtaining integrated data. IM3 defines OI as a “common data management practice”, and IM4 as data format rules.

---

<sup>1</sup> To ensure confidentiality, Original names have been replaced by letters.

## 6.9 Appendix: Memo on Ocean Informatics as a ‘community of practice’ (CoP)

A discussion on a shared practice tying together members of the OI team prompted the writing of a memo on Ocean Informatics as a ‘community of practice’.

### **MEMO: Is Ocean Informatics a community of practice? Notes from discussion on Wenger’s text on CoP**

**By Florence Millerand**

From: Wenger, Etienne. ONLINE. Communities of practice: a brief introduction. Web site: <http://www.ewenger.com/theory/index.htm>.

Could we define OI as a community of practice? Wenger provides 3 characteristics for a CoP: the domain, the community, and the practice. We apply these characteristics to Ocean informatics.

(a)The domain: What would be the shared domain of interest for OI members?

“The domain: A community of practice is not merely a club of friends or a network of connections between people. It has an identity defined by a shared domain of interest. Membership therefore implies a commitment to the domain, and therefore a shared competence that distinguishes members from other people. (You could belong to the same network as someone and never know it.) The domain is not necessarily something recognized as "expertise" outside the community. A youth gang may have developed all sorts of ways of dealing with their domain: surviving on the street and maintaining some kind of identity they can live with. They value their collective competence and learn from each other, even though few people outside the group may value or even recognize their expertise.” (Wenger, online)

The shared domain of interest that grounds OI identity is *informatics* defined as the design and organization of data, systems and practices as applied to oceanography (OR: is *informatics* applied to oceanography i.e. system design for oceanographic data organization, use and preservation.

(b)The community: What makes OI a community and not merely a group of people? What ties together OI members?

“The community: In pursuing their interest in their domain, members engage in joint activities and discussions, help each other, and share information. They build relationships that enable them to learn from each other. A website in itself is not a community of practice. Having the same job or the same title does not make for a community of practice unless members interact and learn together. The claims processors in a large insurance company or students in American high schools may have much in common, yet unless they interact and learn together, they do not form a community of practice. But members of a community of practice do not necessarily work together on a daily basis. The Impressionists, for instance, used to meet in cafes and studios to discuss the style of painting they were inventing together. These interactions were essential to making them a community of practice even though they often painted alone.” (Wenger, online)

Examples of OI activities, OI communication and collaboration mechanisms (\*think of formal as well as informal interactions):

-activities: work at the design table; interactions with STS scholars and others scholars; OI reading groups; expectations that we take time to draw in theory (FM: not sure it fits here, maybe in domain?);

-events: OI luncheon

-tools: blog, web site, shared infrastructure e.g. datazoo

-?

(c)The practice: What do OI members share as a common practice? What defines them as practitioners?

“The practice: A community of practice is not merely a community of interest--people who like certain kinds of movies, for instance. Members of a community of practice are practitioners. They develop a shared repertoire of resources: experiences, stories, tools, ways of addressing recurring problems—in short a shared practice. This takes time and sustained interaction. A good conversation with a stranger on an airplane may give you all sorts of interesting insights, but it does not in itself make for a community of practice. The development of a shared practice may be more or less self-conscious. The "windshield wipers" engineers at an auto manufacturer make a concerted effort to collect and document the tricks and lessons they have learned into a knowledge base. By contrast, nurses who meet regularly for lunch in a hospital cafeteria may not realize that their lunch discussions are one of their main sources of knowledge about how to care for patients. Still, in the course of all these conversations, they have developed a set of stories and cases that have become a shared repertoire for their practice”. (Wenger, online)

As practitioners, OI members share a common practice comprised of:

-experiences: in data and system design, in application development for information systems;

-stories and cases: in interfacing (partnering) with oceanographers...

-ways of addressing recurring problems: through common design approaches, e.g. participatory design approaches...

-common knowledge: about ecological data, information system design...

## 7 Appendix: Qualitative Analysis Software

### 7.1 Appendix: References on the differences between various qualitative analysis software

#### References on the differences between various qualitative analysis software

By Claude Arsenault

Lejeune, C. (s.d.). Des outils libres, *Sociologie qualitative et analyse de contenu*. Consulté le 19 juillet 2007 au <http://analyses.ishs.ulg.ac.be/logiciels/opencaqdas.html>.

This website briefly describes and compares some free software which can be used in qualitative analysis. The author also states the main advantages of using free software in social sciences.

Brugidou, M., Escoffier, C., Folch, H., et al. (2000). Les facteurs de choix et d'utilisation de logiciels d'Analyse de Données Textuelles, *Lexicometrica*. Consulté le 19 juillet 2007 au <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2000/pdf/04/04.pdf>.

This article describes a research which presents criteria on how to choose an appropriate software tool in qualitative research. It also has the particularity of presenting this from the point of view of the user.

Tesch, R. (1990). *Qualitative Research, Analysis Types and Software Tools*. New York : Falmer Press.

This book is a comparison of analysis types in qualitative research. It also reviews software tools, but since it was written over 15 years ago, that review is not completely accurate today.

Barry, C. (1998). Choosing Qualitative Data Analysis Software: Atlas/ti and Nudist Compared, *Sociological Research Online*, vol.3, no. 3. Consulté le 19 juillet 2007 au <http://www.socresonline.org.uk/3/3/4.html>.

This article compares two broadly used software tools in Qualitative Analysis, Atlas/ti and Nudist, by conceptualizing their differences concerning the structural design of the software and the complexity of the research project.

Kelle, U. (1997). Theory Building in Qualitative Research and Computer Programs for the Management of Textual Data, *Sociological Research Online*, vol. 2, no. 2. Consulté le 19 juillet 2007 au <http://www.socresonline.org.uk/2/2/1.html>.

This article weighs the pros and cons of using a computer-assisted method in qualitative research, and links this method to certain methodological approaches. It also issues warnings about possible methodological confusion due to increasingly powerful software tools.

## 7.2 Appendix: Semato in Review

### 7.2.1 Appendix: Memo on use of Semato for project

#### **On Using Qualitative Analysis Software: The Case of *Sémato* By Florence Millerand**

Using software for doing qualitative analysis of ethnographic material is a common practice in social sciences. Many are available<sup>2</sup>, yet choosing the right tool is not an easy task, and debates on the usefulness of software versus on doing the analysis manually are still relevant today (see: Wanlin, 2007). Having discovered a new tool developed at University of Québec in Montréal, called *Sémato* <http://semato.uqam.ca/>, we started to experiment with it. We trained a research assistant, Claude Arsenault in 2006, to use the software and carry on the analysis. Unfortunately, the software proved to be unsuitable for our research project given our heterogeneous ethnographic material and the openness of our research questioning. We detail our findings below.

#### *A qualitative data analysis software based on semantic analysis*

*Sémato* is a semantic analysis tool for text documents. As a qualitative data analysis software, it is different from most software such as AtlasTI or nVivo, because of its linguistic technology that provides a semantic assistance for categorization (coding) and for text mining on the corpus. It can create themes by linking words which belong together (as an example, the theme “live” would also include occurrences of the words “alive”, “inhabit”, “know”, “life” and “living”). Of course, there are semantic complications to take into account. For example, “live” may also mean “to demonstrate” or “demo real-time”. One of its features, the GTH (“génération de thèmes”, which literally stands for theme creation), automatically creates themes (or codes), making it easier to see quickly what a certain interview is about. The same feature can also be used on a whole corpus, giving a good idea of which themes are present in more than one interview. After those themes are suggested by the software, the researcher is free to merge them, divide them or refine them. Also, the software produces tables and graphs based on the themes.

It is important to remember that even though the software suggests themes, it does not understand the significations behind the words. For instance, when creating the theme “knowledge”, the software adds all the occurrences of “know”, “knowledge” and the other words in this lexical field. The software cannot see the difference between a sentence which is related to knowledge and a sentence which ends with “you know” (since the phrase may be very frequent in interviews, as a way to make sure your interlocutor is following, this may pose some difficulties). As a result, the “knowledge” theme will not be accurate, and would require sorting out relevant phrase and exclusion of non relevant ones.

---

<sup>2</sup> The web site Content-Analysis.de presents a well-stocked and updated list of softwares for qualitative analysis : <http://www.content-analysis.de/software/qualitative-analysis>.

### ***An online tool for all platforms suitable for cooperative work***

Sémato is an online tool, which means it can be accessed using a web browser from any computer and any platforms, allowing cooperative work from diverse platforms. This software, although limited to an interface that is only in French, can analyze documents in English as well, since its lexical database includes both English and French. The vocabulary chosen by the creators is unique to the software (See: Semato Glossary), which can be unsettling at times.

The software requires the texts to be formatted in a certain way, which may require an extensive amount of work. For instance, questions and answers must be on separate lines and beginning of each question and answer must be identified by the same word or expression followed by a hyphen (“ – “), after which the actual text can start (see: Memo on transcription format for Semato). When formatted, the text can be imported into the software, and then indexed. It is strongly recommended to avoid all spellings mistakes in the documents, since misspellings can alter the software’s capacity to recognize the words and therefore cause problems with the automatic theme generation.

Once the themes are finalized, the software is able to produce interesting tables and graphs, showing similitude and dissimilarity between participant interviews for instance.

Sémato has all the features of traditional qualitative analysis software applications, including the option to attach analytical notes, to manually create themes and to link different documents. It is very user-friendly and can be mastered fairly quickly, using the online tutorial.

### ***The wrong tool for our research project***

The research project sought to build a monograph of the Ocean Informatics initiative, i.e. a detailed, descriptive, and exhaustive study (as possibly it can be) of the initiative. As a monographic research, it aims at highlighting general features of a phenomena, undertaking or entity (such an organization) from a case study – in this case involving the complexity of information infrastructure development for scientific communities from the Ocean Informatics initiative case.

We began the interview data analysis in an exploratory manner, having a set of loosely tied research questions rather than a well defined and circumscribed research problem. We first aimed at providing descriptive accounts on, for instance, what the participants think of the Initiative, how they define it, how they see their roles in it, with some more specific questions related to infrastructure or interoperability (what it meant to them), and to their jobs (their professional trajectory, their role with data, etc.). The semato software promised to be very useful, specifically with regard to automatic theme generation that would allow data exploration and thus potential theme discovery.

Unfortunately, the software proved to be unsuitable for our research project, mainly due to the high level of heterogeneity of our ethnographic material. The software is more suitable for homogeneous data such as data obtained with well structured interviews with identical questions



and/or a well-defined set of themes where the problem is well formulated (as shown in the example used in the software tutorial). In contrast, our corpus contained a mix of semi-structured and open-ended interviews, and our research problem was far from being well circumscribed. Besides, interviews were conducted with different categories of actors, thus leading to a vast array of themes and different questions order.

Why the Semato tool didn't work well for us can be summarized by the following points:

- Text formatting was time consuming: an extensive amount of work (approximately three hours per interview) was required to format and correct each of the 17 interviews which were to be analyzed.
- Vocabulary specifics was a barrier: Even if the software is user-friendly and the online tutorial easier and useful to learn, appropriating Semato specific vocabulary took some time, and continued to be a barrier between team members who have learned the software and those who have not.
- Application was unsuitable for heterogeneous data: Depending on the corpus analyzed, automated theme creation could result in a list of a few themes (in projects where interviews content is homogeneous and addresses a limited number of topics) to hundreds of themes (in projects where themes are heterogeneous and disparate), thus limiting strongly the relevance of this feature in the latter case. Yet, the software provides a "best themes" option that selects the 25 themes which are the most precise semantically and the most significant considering the corpus. When we ran the automatic theme creation feature using this option, 21 themes out of 25 included the word "data" – as we could have expected, and providing no new or unexpected theme. In other words, the automatic theme generation didn't provide any new theme that we haven't anticipated. This could be interpreted as a good match between our research intuitions pre-analysis and the analysis results, or as a limit of the tool in terms of theme generation.

Important, additional factors other than those related to the software need to be stressed. These factors relate to the research project management more generally:

- Lack of resources (time and expertise): with our list of automatic created themes, some more work was required to refine them until they would represent semantic units, and thus potential new categories. This kind of work cannot be automated, e.g. carried out by the software, but has to be done by the research team with the research questioning and conceptual framework in mind. Our research assistant, who was the person trained on the software, left the team just before this phase had started, leaving the task unfinished. The supervisors had delegated the task of working with Semato and neither was in a position to make the investment required to appropriate the software. It represented too much of a burden for the other team members, and the interface in French only limited communication about the analysis. We decided to pursue the analysis manually, using a compiled file with all interviews (allowing for text search).

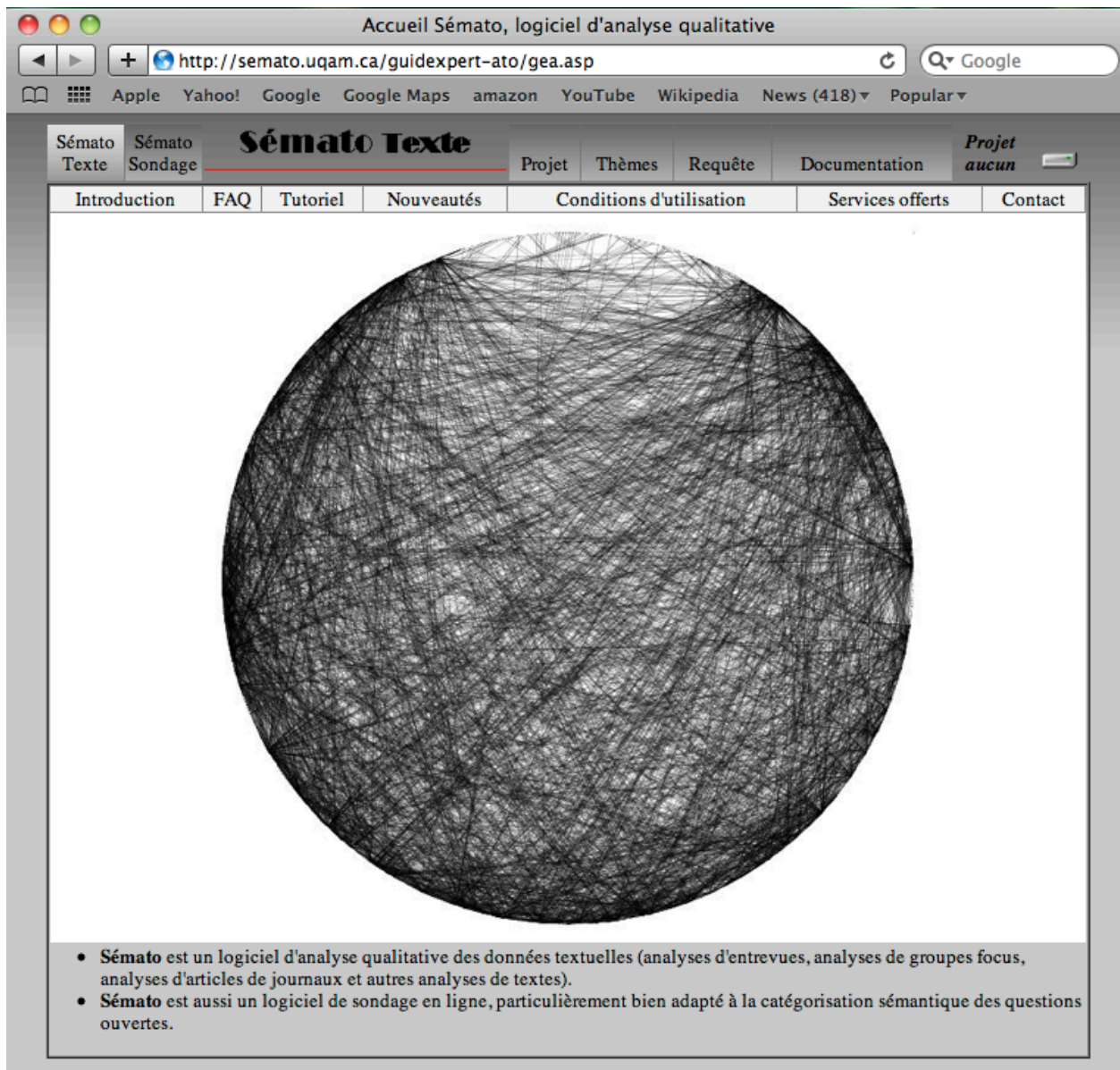
- Lack of success may be attributed to naivete in terms of time and planning in addition to the typical enthusiasm for use of available local tools that seem to hold many possibilities. There was an overestimation of student capacity to use the tool. From a management perspective, we made the mistake of lumping together skill acquisition with technology (ability to investigate and lay out options) with the capacity for synthesis and decision-making.

### ***Conclusion:***

As happens often with qualitative data analysis software, our Sémato experiment revealed an unacceptable balance with a disproportionate effort between investments made (in terms of resources) and gains obtained. The more tangible benefits of this experiment are the many texts that have been produced in the course of the project that improved our understanding of both the software and its limitations, of our methods, and of our research analysis. We wrote a series of memos on the software itself (the main features of Semato (Appendix 8.8.3), Semato glossary (Appendix 8.8.4), Semato transcription format (Appendix 8.8.5), References on the differences between various qualitative analysis software (Appendix 8.7.1)), on our methods (Grounded theory (Appendix 8.6.4), Steps of Grounded Theory Analysis (Appendix 8.6.5), a Response to Conceptualizing Categories for Information Management (Appendix 8.6.6)), and on various analysis elements (memo on Ocean Informatics as a community of practice (Appendix 8.6.9), memo on Ocean Informatics Definitions (Appendix 8.6.8)). The writing of these memos was prompted by our analysis strategy based on grounded theory – where memo writing is a key process in analysis (see: Strauss and Corbin, 1998) – and also because of our need to formalize and circulate texts to improve learning and mutual understanding between us. The memos helped a lot in circulating and sharing understanding about the analysis process while carried on with the software, and continued to be useful afterwards while pursuing the analysis manually.

Though somewhat familiar with the claims and realities of using the qualitative analysis software packages NVivo and Atlas TI, we began work with Semato with a great deal of enthusiasm. Not only was it a locally developed application that worked on diverse platforms, those supporting it were responsive to our inquiries. In retrospect, a lesson learned is the value of identifying a few representative interviews to serve as a limited set of materials to be prepared for use as a test case. Developing materials and queries to be run as a pilot study would allow investigation of both expected basic functionality as well as new, advanced features.

## 7.2.2 Appendix: Semato – A software for quantitative and qualitative analyses (home page)



The screenshot shows the homepage of the Semato software. The browser window title is "Accueil Sémato, logiciel d'analyse qualitative". The address bar shows the URL "http://semato.uqam.ca/guidexpert-ato/gea.asp". The page features a navigation menu with the following items: "Sémato Texte", "Sémato Sondage", "Sémato Texte" (highlighted), "Projet", "Thèmes", "Requête", "Documentation", and "Projet aucun". Below the navigation menu is a secondary menu with "Introduction", "FAQ", "Tutoriel", "Nouveautés", "Conditions d'utilisation", "Services offerts", and "Contact". The main content area displays a large, circular network graph with a dense web of black lines connecting numerous nodes. Below the graph, there is a list of bullet points describing the software's capabilities.

- **Sémato** est un logiciel d'analyse qualitative des données textuelles (analyses d'entrevues, analyses de groupes focus, analyses d'articles de journaux et autres analyses de textes).
- **Sémato** est aussi un logiciel de sondage en ligne, particulièrement bien adapté à la catégorisation sémantique des questions ouvertes.

### 7.2.3 Appendix: Description of Sémato main features and their use in qualitative data analysis

#### **Memo: Sémato : Description of the main features and their use in qualitative data analysis By Claude Arsenault, October 2007**

##### “Analyse Express” (Express Analysis):

- provides descriptive statistics for the whole corpus or a selected section
- the statistics are mainly about the project categories (as defined by the researcher) and the main themes (found by the software)
- should be used after the transfer of documents into Sémato
- the command produces 4 tables:
  - the first table shows the occurrence of the project categories and the 100 best GTH generated themes, per document
  - the second table shows the occurrence of the project categories and the 100 best GTH generated themes, per text
  - the third table presents descriptive statistics concerning the project categories, per document
  - the fourth table presents the frequency of the themes in each document, calculated in percentages and with the chi-square
    - the results can be sorted according to the %em column (it evaluates the variation between a result and the average) in order to see which theme is associated more closely with a certain document, as monitored by the chi-square
- at the end of the results page, there is a chi-square independence test, evaluating the resemblances or differences between the documents in regard to the GTH generated themes.
- using the results of this command, the GTH generated themes can be modified, merged or deleted
- if themes have already been created before the Express Analysis, the command will use those rather than to generate new ones using the GTH.

##### “Lexique Express” (Express Glossary)

- builds conceptual alphabetical and frequencial indexes
- helps to go through the data easily and fast by organizing it
- can be built for the whole corpus, a single document or a project category. Using the “mode ET”, some of those options can be combined (very useful to analyze only the texts corresponding to the answers and not the questions, for instance).
- creates two files: an alphabetical glossary and one classified by frequency
- the particularity of this glossary is that it not only includes words, but also expressions (eg: *database analysis* or *ocean informatics*)
- by clicking on a word or expression, the according binding page opens

##### GTH

- automated generation of themes by the software
- separates the themes it creates in three files, depending on whether they represent objects (what is talked about), actions (what is done), or qualities (qualifications)

- the generated themes should be considered as semantic gathering propositions, which can be modified or refined manually afterwards
- clicking on a theme will accept it and include it in the project's theme list
- automatically produced after the indexation of the texts
- can also be commanded manually to use its more advanced functions: custom GTH

#### “GTH sur mesure” (Custom GTH)

- GTH can be customized in three ways

##### 1. **Definition of the domain**

- selects only a section of the corpus, either according to:
  - a project category
  - a document

##### 2. **Definition of the number of themes**

- the software will select the best themes (25, 50, 75 or 100)
- those themes will automatically be added to the theme list and not to the GTH list
- this option favors themes with multiple “synapsies” and are thus more precise semantically
- interesting to use when beginning a project before building themes manually

##### 3. **“GTH orientée”: oriented GTH.**

- allows the researcher to know whether a certain theme differentiates a document with regards to a project category using the chi-square
- can also identify themes which do not differentiate a document, but are generalized to the whole corpus
  - must be used with the selection of the best themes.
- the themes provided by the custom GTH will be identified as such by the presence of two letters in their name
- can be useful when combined with “introjection”, which helps to refine a theme by using a custom GTH for a section of the corpus delimited by a manual theme. The results of the custom GTH can be used to enrich the manual theme.

Manual theme > Custom GTH > Ingredients for the theme > AST

#### “Assistant Scripteur de themes (AST)” ( Assistant Theme Writer)

- helps to define automatically generated themes by analyzing their ingredients and suggesting new ingredients semantically associated
- can also be used to create a theme if the researcher indicates ingredients to start with
- on the results page, the name of all the ingredients are links which lead to a new page where the ingredients' context is shown
- there are two types of ingredients:
  - **Ingredients A:** general ingredients which were used to start with and the elements of their semantic field
  - **Ingredients B:** expressions which contain a certain ingredient A
- selecting the ingredients will include them in the theme
- a group of words can be used as an ingredient

- The AST should be used a few times in a row. Each time, the ingredients which were selected the time before become basic ingredients

#### “Page d’arrimage thématique” & “mémors analytiques” (Thematical Binding Page & Analytical Memos)

- a thematical binding page will be opened whenever there is an interlink on a text or a word (for instance in the AST)
- a number is assigned to every text of the corpus and another one to every sentence of each text when the document is indexed, those numbers will be used for the thematical binding
- the thematical binding page presents the texts, the sentences which compose them and the themes or memos linked to each sentence or text.
- themes and memos can be added or deleted from a text or sentence using this feature

#### “Requêtes” (Requests)

- 2 types: tracking and analysis
  - **Tracking**: allows searching the corpus to find texts with specific characteristics
  - **Analysis**: presents a few options for preliminary analysis on the corpus

#### “Réseaux de similitude” (Similitude Networks)

- type of analysis request
- analyzes the distance between the textual gathering units (created by the project categories) in the corpus
- looks for the resemblance between each textual unit pair, the pairs being set by the researcher when he selects the project category he wants the analysis to be based on
- can be built for different linguistic levels (depending on the use of “synapsies”, lemmas, semantical fields or selected themes)
- it is recommended to try using different levels in order to discover different types of similitudes
- the researcher must decide whether he wants to use the frequencies in the sentences or in the texts

#### “Requêtes de repérage” (Tracking requests)

- 3 types:
  - **Project categories**: searches the corpus to find texts according to a parameter defined by the researcher
  - **Themes**: searches the corpus for texts related to a certain theme
  - **Text search**: by typing in an expression or a sentence (in French), the researcher will obtain the most similar occurrences in the corpus
    - to get more significant results, the function “with semantical field” can be used

## 7.2.4 Appendix: Semato Glossary

Sémato: Glossary of the main terms used by the software October 2007			
French Word or Expression	English Equivalent	Definition	Semato Feature
Arrimage	Binding	To create a link between a sentence or a text and an analytical element	Analysis features
Arrimage thématique	Thematical binding	To apply a theme to a sentence or a text	Thematical Binding Pages
AST (Assistant-scripteur de thème)	Assistant Theme Writer	Analyzes the ingredients of themes created by the GTH and suggests new semantically associated ingredients	AST
Bouton cible	Target link	Link which, when next to a text, leads to the texts that are most similar to it. when next to a group of themes, it commands a tracking request for those themes.	Analysis features
Catégorie de projet	Project category	Categories determined by the researcher, which define each document	Analysis and Requests features
Configuration focus	Key configuration	Configuration of a similitude networks which presents the most sub-networks	Similitude Networks
Contexte d'un ingrédient	Context of an ingredient	Broader expression which includes a certain ingredient	Themes
Document	Document	A text file (an interview)	All
GTH (Génération de thèmes)	Automated Theme Generation	Automated generation of themes by the software	AST
Indexation	Indexation	Broad analysis of the corpus by the software, required to be able to use the documents with the software. also produces the automatic GTH.	Indexation
Ingrédient	Ingredient	Word, expression or theme which composes a theme	Themes
Introjection	"Introjection"	To find elements for a theme using custom GTH	Custom GTH
Lemme	Lemma	General form of a word (for instance, the infinitive for a verb)	
Mode ET	AND mode	Allows the researcher to select more than one option when defining a command	Most commands
Phrase	Sentence	Section of a text	All
Poids d'un réseau	Weight of a network	Formula used to find the key configuration of a network. It uses the level of resemblance between the elements of a network and the number of elements in the said network.	Similitude Networks
Poids d'une configuration	Weight of a configuration	Formula using the weights of the networks and the number of networks to establish the key configuration	Similitude Networks
Portrait catégoriel	Categorical description	Gives the composition of a certain network in regard with the selected project categories	Similitude Networks
Synapsie	"Synapsie"	Recurrent expression which has a unique meaning	
Synapsies multiples	Multiple "synapsies"	Recurrent expressions composed of more than one word and are more precise semantically	
Texte	Texte	A section of a document (in the case of an interview, each question and each answer would be a distinct text)	All

## 7.2.5 Appendix: Transcription format for Semato

### **Memo: Transcription format for SEMATO**

**By Claude Arsenault**

**August 8, 2007**

- Questions and answers must be on separate lines
- The beginning of each question and answer must be identified by the same word or expression (it can either be the name of the person speaking if there is only one interviewer and one interviewee, or any expression, as long as it is constant all through the interview).
- If there is more than one interviewer or interviewee, a distinction between both can be made by adding a second identification before the text, as long as it is after the original “-“
- The identification must be followed by a “ – “, after which the actual text can start.
- It is also important to try and avoid all spelling mistakes, since they can alter the software’s capacity to recognize the words.
- Avoid mentions of the time in the first part of the text (the identification of the speaker)

*Example with only one interviewer and one interviewee:*

Florence - What did you say, the person that .... who gets.

Lynn - Yea, gets or who gives them what they want.

Florence - OK. But you say get or give. Sorry ....

Lynn - I said get. Who gets them what they want.

Florence - OK. OK. Sorry. OK. And how do you think they see your role with the data.

*Example with two interviewers:*

Question - K - Hmm

Elizabeth - Ahh. I got into the field because I wanted to go to sea and there was a slot.

Question - S - Oh.

Elizabeth - But that's not what you wanted to know.

Question - S - Well that's a good reason. I mean I'd, every time I come down here I want to become an oceanographer.



## **8 Appendix: Paper on Role of Information Management**

Paper submitted, reviewed, revised but not accepted in 2008 for Environmental Information Management (EIM) Conference.

### **Thoughts on the Role of Information Management: A Local Example**

Karen Baker\*

\* Scripps Institution of Oceanography, University of California, San Diego, 92093-0218, USA, email: [kbaker@ucsd.edu](mailto:kbaker@ucsd.edu)

Keywords: informatics, information management, information mediation, information systems design, infrastructure, sociotechnical

#### **Abstract**

Information management faces many challenges in an era of ever expanding digital recordkeeping. In this paper a key information management challenge is considered: scaling of local site capacities in order to address calls for public data reuse but also to support new local data uses and reuses as well as new concerns with respect to sustainability. Information management arrangements at two LTER sites illustrate a changing role described by an information management strategy that augments previous efforts. Development of a strategy and accompanying goals contribute to design of a local information infrastructure.

#### **Introduction**

The term information management is in common use yet its definition and relationship to digital records are understudied. With scientific support over the last decade focusing on changes in technology use, data sharing, and cyberinfrastructure building, issues relating to local information mediation and management arise. The LTER research community shares an overarching community goal of long-term ecology that is grounded by local biome field studies and long-term datasets. This community is organized as a multi-level configuration of sites and network with an information manager at each site working in conjunction with an information technology team at the network office. In considering how information management conducted at sites changes over time, elements relevant to the vitality of local information management and to community networking may be identified.

#### **Background and Setting**

The LTER network makes visible aspects of information management and networking that are related to data curation and to interdisciplinary, collaborative work. The synergistic arrangement of site-based scientists and information managers charged with a) studying a local biome and 2) leading participation in network-based synthesis creates an organizational structure for negotiating a well-recognized tension in priorities between local study and global synthesis. This may be viewed as a set of problems to resolve, balances to arrange, and/or dichotomies from which to learn.

The development of an updated site information management strategy was triggered for two LTER marine sites by a series of events – site-based and network-based. Palmer Station (PAL) became part of the LTER network in 1991 with a field study area off the Western Antarctic Peninsula; California Current Ecosystem (CCE) LTER joined LTER in 2005 with a field site offshore of Southern California. Information management for the two sites is collocated in the Integrative Oceanography Division at Scripps Institution of Oceanography, University of California San Diego where the majority of CCE research participants reside. PAL investigators are geographically distributed with the lead institution originally located at UCSB and moving to Marine Biological Laboratory in Woods Hole in 2008. The information system of PAL and CCE developed in response to a noticeable absence of support for first access and then query or improved access to (small) but complex datasets that are key to local studies of the biome. In this case, local is defined as close to the source of data collection.

### **Discussion**

Strauss (1988) juxtaposed the notion of short-term care with that of long-term care. Both are needed – immediate and extended-horizon responses – and are intertwined in a complex site information management development trajectory. Once out of a setting that is logistically well-organized, highly-instrumented technically, and placed organizationally (i.e. that of a lab with lab-based data collections), data move to an open system influenced by local circumstances, human factors, and cultural forces. New types of infrastructure are required to continue the blend of existing and new needs into the longer-term trajectory of information management. This involves a scaling of individual efforts through development of a local information infrastructure that enables sites to be active nodes - reactive in terms of data collection priorities and proactive in terms of data description and exchange. For the two sites, this has involved developing a local environment that recognizes and supports 1) interdisciplinary partnering, 2) new types of roles for information mediation; and 3) growth of sociotechnical infrastructure.

Partnering with social scientists has provided PAL and CCE the opportunity to reflect on data and information management practices. This generated awareness and a deeper understanding of concepts such as information management, community, data practices, classification systems, informatics, and information infrastructure. We have researched communication through storytelling as a form of narrative where the story may be seen as a case of extreme metadata (Karasti et al, 2002). Ethnography is being used to capture the experience of information managers speaking about the role of information management as having multiple dimensions involving scientific services, data services, and technology work (Karasti and Baker, 2004). An understanding of all roles including those of interdisciplinary partners is being explored (Ribes and Baker, 2007; Karasti and Baker, 2004; Millerand and Baker, submitted). More recently, we have explored the ties of information management with data practices in terms of data stewardship (Karasti et al, 2006) and data curation (Karasti et al, 2007)

The LTER multilevel arrangement of sites and a network of sites has wide spread ramifications in terms of our conceptual model as well as our strategies. It represents a setting that on a continuing basis makes explicit the inherent dichotomies and enhances comparative study as well as creates an interesting environment that can attract and retain the expertise needed for contemporary informatics initiatives. For example, in order to comprehend the full trajectory of

site developments in the LTER Decade of Regionalization from 1990-2000, a period when the Internet became a factor in how information management played out, it is important to recall that site information managers worked on independent Network Information System (NIS) modules – site description, bibliography, and climate – as well as a framework for assembling them (Brunt, 1998; Baker et al, 2000). In practice, the LTER sites themselves are both training grounds – places that are teaching about informatics while doing local information management – and learning environments – places where experiences meld with theory to create a foundation for the applied work of informatics. A network with the role of information manager organizationally prescribed for each site presents a unique configuration that can highlight particular types of information mediation. The configuration also has implications for information system development as well as local adoption of community standards. Delays in the deployment of standards within the LTER have been interpreted as “revealing neither the capacity of resistance of the users (information managers plus scientists) facing enactment of a community standard nor the limits of the EML standard itself as a shared standard.” (Millerand and Bowker, in press). In practice, when information managers make visible and explicit the difficulties of enacting a standard, they accomplish a number of things critical to infrastructure design and community coordination by contributing to the elaboration of data processes.

Growth of local information management represents a contemporary strategy for designing information infrastructure that is integrative. Technical scaling frequently appears deceptively straight-forward but experience reveals a myriad of related sociotechnical factors addressed by design initiatives drawing on studies of language and categories, the theory of social sciences and informatics, and integrative activities with partners across multiple studies (e.g. Cherns, 1976; Bijken and Law, 1992; Fischer, 2002; Kling and Lamb, 1999; Mumford, 2003). The LTER is rich in experience with diverse scientific practices, data practices and collaborative practices. Folk definitions of experience state: “Experience is what you get when you don’t get what you were expecting” or “Experience is what you get when you don’t get what you want”. Our interdisciplinary team has considered the concepts of building/growing infrastructure (Bowker et al, in press), articulation work (Baker and Millerand, 2007a), standards-making (Millerand and Bowker, in press), knowledge-making in differing knowledge provinces (Baker and Millerand, 2007b), local information environments (Baker and Chandler, in press). In order to meet the need for data query and integration at the site, a technical choice was identified that involved moving from text data files to a relational database on the backend while sociotechnical concepts informed the information system design effort as part of a larger information infrastructure initiative.

### *Case Example*

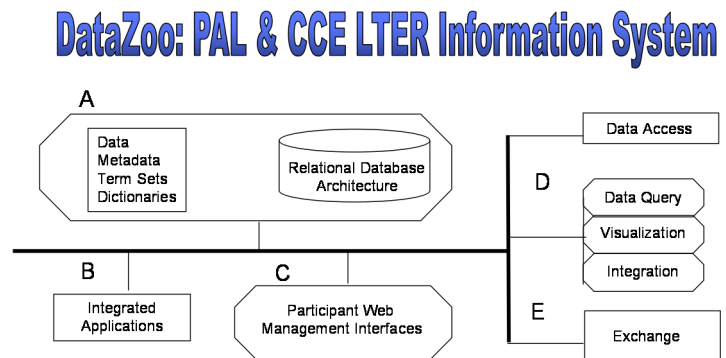
The development of information management at PAL & CCE provides an example of changes in information management that have occurred over time. An understanding of technical data management has broadened to information management with explicit sociotechnical dimensions. Ocean Informatics represents a conceptual framework for a team of information specialists from multiple projects. Drawing upon interdisciplinary partnerships, the role of information management expands, bringing together conceptual, organizational, and social elements along with the technical.

Informatics is an applied field that works at the intersection of a domain science, information sciences, and social sciences, that mediates interactions, interfaces, and interdependencies of information while facilitating informatics research. In the case example, the domain is marine science. Ocean Informatics (Baker, Jackson, and Wanetick, 2006) provides a local information management identity separate from any one project, a forum for cross-project discussions and emphasizes the applied nature of our work while drawing in information and design theory. Working with multiple projects provides a larger context for each project – offering up a window onto a larger set of circumstances that frequently foreshadows a future context or provides a reminder of past issues.

The fieldwork of PAL and CCE is organized around cruises and seasons. Each is considered a study, i.e. the January 2007 cruise or the 2007-2008 season with Spring-Summer Antarctic sampling from October to March. The original data system was designed with twin aims: access and simplicity. It used a hierarchical file structure that reflected the organization typically used in participant labs. The system architecture mimicked that of the NIS schema; it focused on individual support modules—personnel, bibliography, and dictionaries and research modules of data and metadata. The similarity of site and network models was not surprising at a time when sites and network office were gaining experience jointly with data practices, design, and networking. After a number of years of making individual data files available on a study-by-study basis, the Palmer investigators requested an information system redesign that would provide both data query and data integration.

The design of an updated information system was initiated in 2002 with a design approach growing over time to address new system architecture requirements: the ability to handle multiple projects, to facilitate data exchange, and to engage participants in new ways. The concurrence of events including development of social science partnerships, of an Ocean Informatics framework and of joint efforts with the CCE project spurred assembly of resources and an informatics team interested in the challenge of designing information systems and an information infrastructure to support scientific research as part of an enriched information management trajectory.

**Figure 1** shows a schematic of the Ocean Informatics information system, DataZoo (<http://oceaninformatics.ucsd.edu/datazoo>). Planned and recognized as being in a permanent state of redesign, DataZoo is a data and metadata repository system and publishing forum that includes a dataset catalog, personnel directory, and help system. Dictionaries and term sets play a key role in the architecture and use of the system. Site metadata takes into account local and community standards building upon the Ecological Metadata Language to include unit, attribute and column qualifier details. DataZoo is recently organized into three web-based functional units – data, resources, and management, the latter two designed specifically to



**Figure 1:** DataZoo information system schematic showing a service bus (thick lines) connecting multiple system elements: a) databases, b) independent applications, c) web enabled participant management interfaces, d) ‘data use’ functionality and e) data exchange.

engage the learner and the data contributor stretching the system reach to the lab or desktop so contributors are able to upload their own data - and reciprocally to update their data practices. This approach contributes to the transformation of the role of information management from one largely of locating, proofing, and ingesting data to one of mediating and collaborating, designing and analyzing.

Carrying out design in the midst of developing concepts, frameworks, and initiatives, the Ocean Informatics team found it valuable to revisit information management development using two coordination mechanisms. First, there is an effort to identify and articulate an information management strategy. Second, following the example of scientific research components that state their objectives succinctly, the process of capturing specific information management aims in a set of objectives was initiated.

### Concluding Remarks

In conclusion three aspects of an information management trajectory have been highlighted: a) partnering that draws in additional expertise in social and information sciences as well as across related projects in order to better understand the tasks of information management, b) roles of information mediation that bridge data practices and theories, and c) the growth of integrative infrastructure that supports information systems. This work contributes to the conceptual development of local information management using a case example to illustrate interdisciplinary partnering, information mediation, and the growth of local infrastructure. The trajectory concept provides a framework within which multiple factors are brought together. The distinction between technical and sociotechnical growth is illustrated, including the social and organizational along with the technical. The case example suggests two mechanisms for augmenting local information management: the development of an information management strategy and statement of local information management objectives.

Starting with the LTER ISSE figure that provides a community context for cyberinfrastructure, local information management for the case study may be represented by a modified model (Figure 2). Informatics is added as a research element in its own right and information infrastructure is added as an integrative substrate across all research activities. Six facets of information management are portrayed as supporting environmental & information action and awareness.

Implications of this work relate to the value of conceptualizing the role of local information management. The local information management perspective provides field experience that shapes and informs infrastructure building within a local scientific research team close to the data source but also within networks of partners. Perhaps the best way to end is to restate the question that represents a starting point for

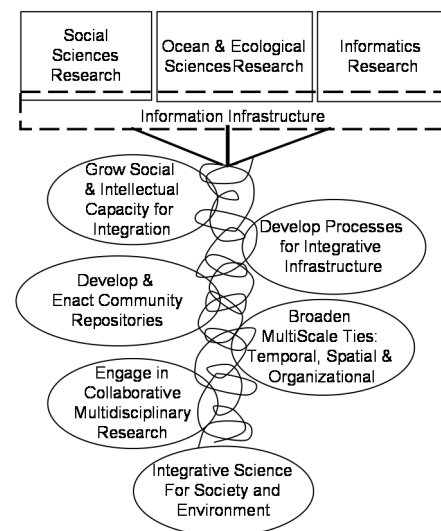


Figure 2. A local example of site science and social sciences partnering with informatics (inspired by LTER ISSE initiative brochure figure).

data stewardship and data curation research: What is the vision for local information management?

## Acknowledgements

Acknowledgement is given to the Ocean Informatics team – researchers, technicians, programmer/analysts, system administrators – contributors, designers, and users all. In particular, recognition is given to the information design work of Mason Kortz and James Conners; the analysis work of Lynn Yarmey; the systems administration of Jerry Wanetick and Nate Huffnagle; and the community support of PAL and CCE participants.

## References

- Baker, KS, BBenson, DLHenshaw, DBlodgett, JPorter and SGStafford, 2000. Evolution of a Multi-Site Network Information System: the LTER Information Management Paradigm, *BioScience* 50(11):963-983.
- Baker, KS, SJJackson, and JRWanetick, 2005. Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment. Proceedings of the 38<sup>th</sup> Hawaii International Conference on System Sciences. HICSS38, IEEE Computer Society, 2-6 January 2005, Big Island, Hawaii.
- Baker, KS and FMillerand, 2007a. Articulation work supporting information infrastructure design: coordination, categorization, and assessment in practice. Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07), January 03-06, 2007, IEEE Computer Society, Washington, DC, 242a.
- Baker, KS and FMillerand, 2007b. Scientific infrastructure design: information environments and knowledge provinces. Conference Proceedings of the 70th Annual Meeting of the American Society of Information Science and Technology (ASIS&T), 19-24 October, Milwaukee, WI, Vol. 44.
- Baker, KS and CChandler, in press. Oceanographic Data and the Ecology of Information. Deep Sea Research II. Series Palmer Long-Term Ecological Research.
- Bijken, WE and JLaw, 1992. Shaping Technology/Building Society: Studies in Sociotechnical Change. Cambridge, MA, MIT Press 205-225.
- Bowker, G, KBaker, FMillerand and DRibes, submitted. Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In International Handbook of Internet Research, JHunsinger, MAllen and LKlasrup (eds), Springer.
- Brunt, JW, 1998. The LTER network information system: A framework for ecological information management. Pgs 435-440 in Aguirre-Bravo, C, Franco CR, eds. North American Science Symposium: Toward a Unified Framework for Inventorying and Monitoring Forest Ecosystem Resources; 2-6 Nov 1998; Guadalajara, Mexico. Fort Collins (CO): US Department of Agriculture, Forest Service, Rocky Mountain Research Station. Proceedings RMRS-P-12.
- Cherns, A., (1976), The principles of sociotechnical design, *Human Relations*, Vol. 29, No. 8, pp. 783-792.
- Fischer, 2002. Beyond "Couch Potatoes": From Consumers to Designers and Active Contributors. *First Monday* 7(12).[http://www.firstmonday.org/Issues/issue7\\_12/fischer/](http://www.firstmonday.org/Issues/issue7_12/fischer/)
- Karasti, H, KSBaker and GCBowker, 2002. Ecological Storytelling and Collaborative Scientific Activities, *SIGGROUP Bulletin* 23(2):29-30.
- Karasti, H, KSBaker, and GCBowker, 2003. ECSCW 2003 Computer Supported Scientific Collaboration (CSSC) Workshop. *SIGGROUP Bulletin* 24(2): 6-13.
- Karasti, H and KBaker, 2004. Infrastructuring for the Long-Term: Ecological Information Management. Hawaii International Conference for System Science Proceedings. of the 37<sup>th</sup> Hawaii International Conference on System Sciences. HICSS38, IEEE Computer Society, January 2004, Big Island, Hawaii.
- Karasti H, KSBaker and KSchleidt, 2007. Digital Data Practices and the Long Term Ecological Research Program, Proceedings of the Digital Curation Conference, DCC2007, Washington, D.C.
- Karasti, H, KSBaker, and EHalkola, 2006. Enriching the notion of data curation in e-science: data managing and information infrastructuring in the Long Term Ecological Research (LTER) network. *Computer Supported Cooperative Work* 15, 321-358.
- Kling, R. & R Lamb, 1999. IT and Org Change in Digital Economies: A SocioTechnical Approach. *Computers and Society* 29(3): 17-25.

- Millerand, F and KSBaker, in press. Who are the users? From a single user to a web of users in the design of a working standard. Information Systems Journal (ISJ); Special Issue: User - the Great Unknown of Systems Development: Reasons, forms, challenges, experiences and intellectual contributions of user involvement. Eds: JIivari, HIsomäki, and SPekkola
- Millerand, F and GCBowker, in press. Metadata Standard: Trajectories and Enactment in the Life of an Ontology. In Formalizing Practices: Reckoning with Standards, Numbers and Models in Science and Everyday Life. M.Lampland and S.L.Star (eds). Cornell University Press.
- Mumford E., (2003) Redesigning Human Systems. Idea Group Inc.
- National Research Council, 2007. Environmental data management at NOAA. The National Academies Press, Washington, D.C.
- Ribes, D and KSBaker, 2007. Modes of Social Science Engagement. In Proceedings of the Third Communities and Technologies Conference, CSteinfeld, BTPentland, MAckerman, and NContractor (Eds). Michigan State University, Springer, p107-130.
- Strauss, A, 1988. The Articulation of Project Work: An Organizational Process. The Sociological Quarterly, 29(2), 163-178.

## 9 Appendix: Event Logger: Summer Project Proposed Paper/Poster

The event logger was the focus of a summer project of Brian Lindseth in 2006. The account of the logger was prepared in a format for CHI (Computer Human Interfaces) but was not accepted for presentation.

---

### Collaborative Design of an Oceanographic Event Logger

**Brian Lindseth**  
Department of Sociology  
University of California,  
San Diego  
La Jolla, CA 92093  
brianlindseth@gmail.com

**Karen S. Baker**  
Scripps Institution of Oceanography  
University of California,  
San Diego  
La Jolla, CA 92093-0218  
kbaker@ucsd.edu



---

Copyright is held by the author/owner(s).  
CHI 2007, April 28 - May 3, 2007, San Jose, USA  
ACM 1-xxxxxxxxxxxxxxxxxxx.

#### Abstract

We report on a design project that is in-progress. Iterative prototyping of a new field instrument, used to link data collecting practices and data collection preservation, also addresses issues of time, space and categorization that influence subsequent data integration. The development of an event logger system is unfolding as an oceanographic field research case study and provides an opportunity to consider collaborative design as one aspect of local information infrastructure building.

#### Keywords

Infrastructure, iterative design, invisible work, oceanographic research

#### ACM Classification Keywords

H.5.2 User Interfaces: User-Centered Design, Prototyping; H.5.3 Group and Organization Interfaces: Computer-supported Cooperative work; E.2 Data Storage Representation: Object Representation, Design

#### Introduction

We provide an account of an event logger system deployed recently on a series of oceanographic research cruises. Shipboard activities are carried out by investigators with different disciplinary and institutional



Quote #1. A CalCOFI participant discusses the dilemma of data that does not fit into existing categories: "CalCOFI's always been focused on stations, so anything between the stations has been ancillary or extra. And it's not—it's not reported by CalCOFI or stored in their databases for public consumption. You know, it's all on a DVD somewhere or a CD somewhere, but it's not easy to get at".

Another goal of the project has been to implement a reflexive design process—one open to incorporating insights from the design literature—broadly conceived—as well as from the successes and failures encountered in the field.

### Design: Setting, System, and Process

Our object of study is an event logger system designed to provide for the possibility of both participant dialogue and interoperable data across organizations. Therefore, while much attention will be paid to the event logger system itself, the evolution of the design process is also an object of analysis. Some assumptions underlie our study: the recognition of the importance of existing practices, the value of a design approach that is both iterative and collaborative, and recognition of the long term implications of infrastructure building processes (Bowker and Star, 2002; Baker et al, 2005). Methods for our work include use of design, participatory action research, ethnographic methods, and an interdisciplinary research team as collaborative mechanisms supporting data integration.

Webs of practices and technologies of each organization—the data management conventions, the actors, skills and routines linking measurement strategies and on board data collection together with database and reporting technologies into a functioning process—could be considered as local infrastructures. They provide a relatively invisible, relational framework along which the measurements from a cruise travel as data into a database to be delivered on a website. Such frameworks embed classification systems that both include and exclude data (Bowker and Star, 1999).

One way we can see classifications in play is in the ways in which space is measured. For both sets of researchers, the categories are stations in contrast to the more granular geolocation elements of latitude and longitude. And, as a CalCOFI participant noted (see Quote # 1), if a measurement does not find an easy home in the categories of a given infrastructure, it can

fall between the cracks. *Extra work is needed to find the data that do not fit easily into existing categories of the local infrastructure* (Walsh and Ungson 1991; Bowker 2006).

The event logger system provides a unique identifier in the form of an event number for each measurement as well as a common measure of space and time. It creates a bridge along which oceanographic data and data collections can travel between organizational infrastructures. Designed as a local, 'ground up' solution (and not an overarching or generalized solution), the event logger could be described as a dedicated gateway (David and Bunn, 1988; Egyedi, 2001; Edwards, forthcoming), linking infrastructures that are, themselves more stabilized and resistant to change (Hanseth et al, 1996).

In emphasizing a design approach that is both iterative and collaborative, we are building on previous work that focuses on the distinction between designers, users, and support personnel (Clement 1993; Bratteteig 2003; Grudin 1993; Kanstrup 2005). Here, the openness towards collaboration finds theoretical resonance in efforts to broaden the role of designers.

Kanstrup and Bertelson (2006), for example, examines the role that IT support workers play in designing the systems they support. Suchman highlights the 'artful integration' work that goes on in everyday practice (2002, 1994). This literature seeks to extend our notions of who can count as a designer. This move brings visibility to the work traditionally seen as being less significant – or not seen because it is perceived as less significant (Star and Strauss, 1999; Suchman 1995, 2000). In addition, work on the distributed

*Quote #2. Incorporating diverse voices over time, a lead designer retrospectively observed: "the development's been incremental, doing robustness checks along the way."*

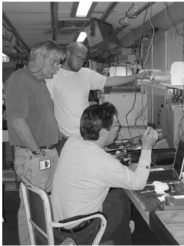


Figure 3. Working to install the event logger on the ship.

nature of cognition brings visibility to processes, tools, and features of the environment in which cognition, more narrowly conceived, takes place (Hutchins 1995; Hollan et al 2000).

The event logger describes an on board computer technology and related processes. An event log is produced as an output, a list of each measurement that has taken place on board the cruise. It is a sequential, digital record in a standard format that can be accessed by all cruise participants. Together, the event logger and the event log were designed to span the flow of data - from measurements in the field to a shared data repository. The standard time and location allows for translation of measurements across organizational boundaries and a unique identifier (the event number) allow commensurability to be enacted at the level of the database, establishing relations across research organizations and ultimately enabling queriability. In this context, the event logger was designed as a way of distributing data management work (Hollan et al, 2000; Turner et al, 2006). Instead of relegating the work of cleaning up the data to post cruise data users, the work of "keeping the data clean" - organizing and relating the data - begins at the source of the data itself in measurements on board cruises. In extending post-cruise reconciliation efforts into the cruise, the event logger functions to distribute this work over time and a wider set of actors (Hutchins 1995, Hollan et al, 2000, Halverson & Ackerman, 2002).

From their beginnings, the event logger and the event log can be construed as boundary objects that span organizations collecting the data (Star and Ruhleder, 1994; Halverson and Ackerman, 2002). As the system has evolved, it has gained momentum (Hughes 1987)

and become more embedded (Egyedi, 2001) or stabilized (Latour 1987) in the infrastructures of each organization. The event logger is in the process of becoming by being a tool in use (Norman, 1988; Suchman, 1987). It has become part of the environment in which measurements start to become data; it serves as a prompt for communication and an architecture component linking the organizations' infrastructures. This shift has not proceeded without incident.

With early iterations uncertainty stemmed from changing ships and therefore the system's environment. At times, it was not known whether the system could be positioned on the bridge or placed in a ship laboratory, whether connection to the ship's network would be wireless or by cable, and/or whether the system could be networked into the ship's GPS system or would need to determine position independent of the ship's systems. With each cruise, the system design evolved to absorb the effects of variation in environment. Incorporating contingencies made the system more configurable. Thus using simple text configuration files to replace hard coded options in the system allowed support technicians to change the configuration of the event logger on the fly. The system was made more mutable, less rigid as explained by one of the designers.

#### FINDINGS AND FUTURE DIRECTIONS

The primary finding of the design project has been the recognition of the form iterative design has taken and the value of an open design process. This openness has led to recognition of the importance of configurability as the event logger system has been deployed in different environments. With the open collaborative stance, each

In this context, it has seemed natural to adopt a loosely **iterative design process**.

Quote #3. A lead designer discusses flexibility: "the configuration file [was] flexible so you have lab events, you have bridge events, you have marine mammal events.... And every computer running the event log can have their own selections to choose from. And they essentially just share the position file or data and the event number gets updated."



Figure 4. Iterative design process in the design studio.

iteration has allowed for the possibility of extending the circle of designers. After each iteration, the definition of the event logger unfolds. In a sense, one could say that this open stance allowed the variability in the system's environment to have a voice as sources of action in the design process (Latour 1987). Each cruise has represented an opportunity to learn from snags and successes that were encountered. *In incorporating multiple voices, the event logger had become an actor in its own right.*

From a design perspective, these opportunities can be seen as windows into articulating much of the invisible work (Star and Strauss, 1999; Suchman 1995; Clement 1993) that is required for the success of the event logger system. On a more fundamental level, we could say that this support work is necessary for the event logger to be an event logger instead a collection of components. While uncertainties and breakdowns can bring attention to tools, instead of the processes in which they are used, they can also be viewed as opportunities to gain valuable feedback in the design process (Suchman, 2000).

In the more concrete case of the event logger as the openness to collaboration has led to many valuable insights. The primary insight has involved the recognition of the importance of flexibility in bridging local infrastructures—of building the event logger system in a way that it can be easily configurable to meet the multiple demands of actors approaching it (or the cruises) from different organizational milieus.

#### Acknowledgements

*This work is supported by the Comparative Interoperability Project (CIP NSF/SBE/SES #04-33369)*

*and benefits from associations with the Ocean Informatics, LTER, & CalCOFI (NSF/OPP #02-17282, NSF/OCE #04-17616, NSF/OCE #03-40839) and conversations with Florence Millerand.*

#### Citations

- Baker, K.S., Jackson, S.J., and Wanetick, J.R. Strategies supporting heterogeneous data and interdisciplinary collaboration: towards an ocean informatics environment. *Proceedings of the 38<sup>th</sup> Hawaii International Conference on System Sciences*. HICSS38, IEEE Computer Society (2005).
- Bowker, G.C. *Memory Practices*. MIT Press, Cambridge, MA, 2006.
- Bowker, G.C. and Star, S.L. *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA, 1999.
- Bowker, G.C. and Star, S.L. How to Infrastructure. In L. Lievrouw and S. Livingstone, eds, *Handbook of the New Media*. London. Sage, 2002.
- Bowker, G.C. Biodiversity Datadiversity. *Social Studies of Science* 30, 5 (2000), 643-683.
- Bratteteig, T. Making change: dealing with relations between design and use. PhD Thesis, University of Oslo, Norway, 2003.
- Clement, A. Looking for the designers: Transforming the 'invisible' infrastructure of computerised office work. *AI & Society* 7, 4 (1993), 323-344.
- David, P. A. and Bunn, J.A. The Economics of Gateway Technologies and Network Evolution. *Information Economics and Policy* 3 (1988), 165-2002.
- Edwards, P. *The World in a Box*, forthcoming.

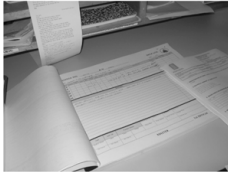


Figure 5. A handwritten paper bridge log.

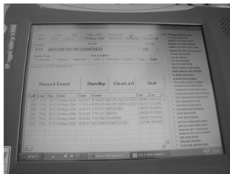


Figure 6. The digital event logger.

Egyedi, T.M. Infrastructure Flexibility created by Standardized Gateways: The Cases of XML and the ISO Container. *Knowledge, Technology and Policy* 14, 3 (2001), 41-54.

Grudin, J. Interface: An Evolving Concept. *Communications of the ACM* 36, 4, April (1993)

Halverson, C. A. and Ackerman, M.S. 'Yeah, the Rush ain't here yet - Take a Break': Creation and use of an artifact as organizational memory. *Proceedings of HICCS 03* (2002)

Hanseth, O, E. Monteiro, M. Hatling. Developing Information Infrastructure: The Tension Between Standardization and Flexibility. *Science Technology, ollen J., E. Hutchins, D. Kirsh. Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. ACM Transactions on Human Computer Interaction* 7, 2, June (2000), 174-196.

Hughes, T. P. The Evolution of Large Technological Systems. Pp 51-82 in *The Social Construction of Technological Systems*. Edited by Wiebe Bijker, Thomas Hughes and Trevor Pinch. MIT Press, 1987.

Hutchins, E. *Cognition in the Wild*. MIT Press, 1995.

Kanstrup, A.M. Local design; volume 1: an inquiry into work practices of local IT-supporters. PhD Thesis, Aalborg University, 2005.

Kanstrup A.M. and Bertelson, P. Participatory IT Support. PDC 2006 - *Proceedings of the 9<sup>th</sup> Participatory Design Conference* (2006), 87-94.

Latour, B. *Science in action*. Harvard Univ. Press, 1987.

Luhmann, N. *Observations on Modernity*. Translated by William Whobrey. Stanford University Press, 1998.

Norman, D.A. *The Design of Everyday Things*. MIT Press, Cambridge, MA, 1988.

Star, S.L. and Ruhleder. K. Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. *ACM Computer Supported Cooperative Work*, Chapel Hill, 1994.

Star, S.L. and Strauss, A. Layers of silence, arenas of voice: the ecology of visible and invisible work. *Computer Supported Cooperative Work* 8 (1999), 9-30.

Suchman, L.A. Working Relations of Technology Production and Use. *Computer Supported Cooperative Work*, 2 (1994), 21-39.

Suchman, L.A. *Plans and Situated Actions*. Cambridge University Press, Cambridge, 1987.

Suchman L. Making Work Visible. *Communications of the ACM*. 38, 9, September (1995).

Suchman, L. Organizing alignment: a case of bridge-building. *Organization* 17, 2 (2000), 311-327.

Suchman, L. Practice Based Design of Information Systems: Notes from the Hyperdeveloped World. *The Information Society*, 18 (2002), 139-144.

Walsh, J. P. and Ungson, G.R. *Organizational Memory*. *The Academy of Management Rev* 16, 1 (1991), 57-91.



Figure 1. An ocean research vessel – a sea going laboratory equipped with an event logger.

**The question** driving the current review is, what are the lessons we have learned in the process of deploying this technology - in a variety of shipboard environments - that contribute to bridging differing organizational

One of the **primary goals** of this project has been to provide for the possibility of interoperable data while respecting differences in data management practices that serve the

backgrounds and with a variety of research interests brought together by a shared interest in describing and understanding the marine system off the California coast. Perhaps unsurprisingly, the data collected by different scientific teams and particularly across organizational boundaries require extensive work to be rendered commensurable. Bowker (2000, 2006) describes precisely this kind of work in efforts to and discipline. Seemingly small differences in how data is gathered in the field become progressively difficult to reconcile as the ship data returns to land and is carried by investigators into their labs. As cruise experiences fade into stories exchanged in conversational encounters, such discrepancies become progressively more difficult to resolve as they become integral to subsequent data handling practices.

Problems that might appear to be mundane—determining the sequence in which measurements are taken—are key to bringing measurements into a shared data repository. Yet, the time stamp associated with shipboard measurements is subject to varying configurations - from ships' dual global positioning satellite systems (GPS) to differing arrangements in GPS receiving equipment as well as to a bevy of unsynchronized shipboard computer clocks. So time is an issue.

Another issue is space--the position at which measurements are taken. Investigators are from different institutions where each organization has a distinct way of considering and recording a measurement. Established stations with unique names mark the location where measurements are to be taken. On the face of it reconciling two station-naming

conventions appears straightforward, simply determine the latitude and longitude for a given station or set of stations in order to provide an accurate representation of location. Still needed, however, are the mechanisms that tie an objective location to the subjective work done by stations. Here stations represent a subjective notion that encompasses field experiences and evokes memories. Further, they are easily pronounced and tied to human recall. And yet, there is the issue of how measurements taken *near* a given station are to be lumped into or related to the category of that station. These differences could be characterized as different frames of reference or could be considered as a gap between metrologies (Latour 1987) embedded in each organization.

These kinds of issues are not limited to the gaps between organizations. They can be found between groups of researchers operating within the same organization as well. In the current case the gap between physical and biological oceanography finds a local correlate in the gap between researchers affiliated with Scripps Institution of Oceanography at the University of California, San Diego (SIO/UCSD) and the National Ocean and Atmospheric Administration's Southwest Fisheries Science Center (NOAA/SWFSC), a fisheries management laboratory. Though both groups of researchers in question comprise jointly the California Cooperative Oceanic Fisheries Investigations (or CalCOFI) program, the organizational gap between SIO and NOAA lines up with the larger divide between physical and biological oceanography and the measurement strategies associated with each. Organizational conventions appear, at the same time, as conventions in data management *and* metrology.

# 10 Appendix: Ocean Informatics Posters: Technical and Conceptual

Posters are summarized in the table below. Brief descriptions follow the table, grouped into three sections: A) about Ocean Informatics, B) by Ocean Informatics: Conceptual, and C) by Ocean Informatics: Technical. Posters are online: <http://oceaninformatics.ucsd.edu/media-gallery/?id=1>. Posters are given unique identifiers in sequential order by date.

Date	Title	Description	Creator
2003-09-18	Palmer LTER: Information Flow and Management	Organizational repositories are needed today to address the needs of scientific information management. Given the social aspects of information, building useful information systems requires multi-faceted infrastructure.	Karen Baker, Anna Gold, Frank Sudholt, Helena Karasti, Geoffrey Bowker
2005-12-05	CalCOFI Data Management: Overview and Reflection	A CalCOFI White Paper (2005) provides an overview of the current state of data and its management within the California Cooperative Ocean Fisheries Investigations (CalCOFI) program.	Karen Baker, Karen Stocks
2005-12-06	Initiating the Data Dialogue: 2005 CalCOFI Conference Interactive Poster	The interactions surrounding the 2005 CalCOFI Data Management poster are captured through photographs of updates and additions made to the poster during the poster session of the annual conference.	Karen Baker
2006-04-01	Ocean Informatics: Conceptual Framework for Marine Science Information Management	The work of Ocean Informatics is represented at the union of oceanography, information science and social science domains. Participants range from data and information managers to technical specialists, archivists, scientific researchers, educators.	Karen Baker, Jerry Wanetick, Shaun Haber, Lynn Yarmey, Mason Kortz, Florence Millerand, Jesse Powell, Jim Wilkinson, Robert Thornbley, Julie Thomas, Beth Simmons
2006-09-20	CCE LTER: Information Management (2004-2006)	The California Current Ecosystem information management efforts were launched with inquiries into existing data practices. This was followed by design, development and deployment of elements of an information infrastructure.	Karen Baker, Lynn Yarmey, Mason Kortz, Jerome Wanetick
2006-09-20	LTER: Research in Infrastructure Studies: Social & Organizational Perspectives on Ecological Data Management	In the mist of major changes in ecological data collecting, managing and sharing, an interdisciplinary team of information, ecological, and social scientists has been brought together at LTER PAL and CCE sites to facilitate the growth from site-based to larger-scale efforts.	Florence Millerand and Karen Baker
2006-09-20	Palmer LTER: Design of a Queriable Ocean Information System	Field data, originating with domain understandings and practices that shape sampling and collection, has informed development of the PAL LTER information system. In becoming digitally preserved, data capture may in turn be influenced by information system work.	Karen Baker and Shaun Haber
2006-09-20	LTER IN Articulation Work: Developing Community Web Recommendations	Over the past two years, the Web Site Design Recommendations Working Group developed recommendations for web sites in response to challenges of first generation LTER web sites. They worked to align a set of social, technical and organizational elements.*	Nicole Kaplan, Karen Baker, Barbara Benson, John Campbell, Corinna Gries, James Laude, Jasmina McCann, Eda Melendez, James Wilkinson, Karen Baker
2006-12-04	CalCOFI: An Oceanographic Event Logger	Local data management, informed by field sampling and data use, supports community coordination at the interface of data collection and data curation. An oceanographic event logger recently deployed on a series of research cruises extends data management to the field.	Mason Kortz, Lynn Yarmey, James Conners, Karen Baker
2007-08-02	LTER: Data Integration in the Decade of Synthesis	As data availability, findability, and even queriability become more ubiquitous, the need to make sense of data from multiple, disparate sources increases. Data integration and data synthesis allow extension of the scope of data beyond local use.	Florence Millerand and Karen Baker
2007-08-02	LTER Environmental Data Management: Infrastructure Studies Insights	In the mist of major changes in ecological data collecting, managing and sharing, an interdisciplinary team of information, ecological, and social scientists has been brought together at LTER PAL and CCE sites to facilitate the growth of site-based information management.	Karen Baker, Cyndy Chandler, Anna Gold, Florence Millerand, Jerry Wanetick
2007-08-02	LTER: Long Term Informatics	With the information age as one of the many ramifications of the Internet, our understandings, cultures, and communities are undergoing change.	Jerry Wanetick, Karen Baker, Nate Huffnagle, Lynn Yarmey, Mason Kortz, James Conners
2007-09-17	CCE LTER Information Infrastructure	Information Infrastructure is an arrangement of computational systems, an iTeam, information systems and partnerships associated with a core interest in informatics.	Karen Baker, Mason Kortz, James Conners, Jerry Wanetick
2007-09-17	Ocean Informatics Information System: One Element of an Information Infrastructure	Focus is on an Information system for managing data - DataZoo 2.0 -at the heart of a configuration of computational systems, an iTeam, informatics work, and a complex set of partnerships.	Lynn Yarmey, Karen Baker
2007-09-17	A working Standard: Augmenting the Ecological Metadata Language	Metadata standards are an integral and necessary part of data sharing as they provide a structure and format to allow comparisons of data context.	Karen Baker, Florence Millerand
2007-10-19	INTEROP Scientific Infrastructure Design: Information Environments and Knowledge Provinces	Conceptual models and design processes shape the practice of information infrastructure building in the sciences. We consider two distinct perspectives: (1) a cyber view of disintermediation where information technology enables data flow from the 'field'.	Mason Kortz, James Conners, Karen Baker
2007-11-17	CalCOFI & Ocean Informatics DataZoo: A Multi-Project Data Publishing System	The DataZoo information system is a hub in the Ocean Informatics learning environment that creates a central forum for data exchange, collaborative design, and community building. It is a central repository for data and metadata of member projects.	James Wilinon, Karen Baker, Rich Charter
2007-11-17	CalCOFI Data Management: Developing Community Standards	CalCOFI represents a partnership of multiple agencies conducting quarterly joint oceanographic cruises, CalCOFI field team members work as a cohesive cross-agency unit to accomplish the cruise goals.	Lynn Yarmey, Karen Baker, James Conners
2007-11-17	CalCOFI Local Metadata: Augmenting the Ecological Metadata Language	Metadata is an integral and necessary part of data sharing; the enactment of a metadata standard not only guides the creation of local metadata documents but is also a link between local and broader communities.	Mason Kortz, James Conners, Karen Baker
2008-08-10	LTER Abstracting Functionality and Access: Facilitating Data System Manageability and Site Coordination	As the functionality of site data systems increases, frequently so does the complexity. Organizing system functionality through distinct layers of abstraction, from low-level system access to high-level user access, is key to maintaining a manageable set of systems.	Kbaker Baker, Nicole Kaplan, Inigo San Gil, Margaret O'Brien, Florence Millerand
2008-08-10	LTER Information Managers: A Community of Practice	Communities of Practice are groups of people who share a concern or a passion for something they do and who want to learn more about how they do it. Such a community is more than a group of people having the same job or a network of connections between people.	Lynn Yarmey, Karen Baker
2008-08-10	LTER Information Infrastructure: Emergent Roles, Responsibilities and Practices	Human activities together with technical elements and collective practices are core elements for growing local infrastructure as well as for bridging with other communities and networks. Site information management activities create a shared data curation opportunity.	Karen Baker, Beth Simmons, Ryan Rykaczewski, Alison Cawood, Peter Davison, Moira Decima, Melissa Garren, Andrew King, Andrew Taylor, Jesse Powell, Melissa Soldevilla, Mike Stukel
2008-08-10	Scientific Communication and Information Infrastructure	Scientific communication is central to collaborative scientific endeavors. A shared information infrastructure facilitates communication and collaboration. Digital information infrastructure occurs in multiple forms. The poster presents examples from CCE LTER.	Karen Baker, Mason Kortz, James Conners, Lynn Yarmey
2008-11-17	CalCOFI Biological Data Management	An information system designed for working with multiple oceanographic biological data collections is presented. DataZoo is an extensible system that supports data discovery, access, query, and exchange for data such as the CalCOFI integrated biological data.	Karen Baker, Lynn Yarmey, James Conners
2009-05-13	CCE LTER: An Oceanographic Eventlogger as One Part of an Information Environment	The CCE LTER initiated at SIO in 2004 enabled launch of "Ocean Informatics", a new approach to design of information infrastructure in support of interdisciplinary science. CCE works synergistically with Palmer Station LTER and with California Cooperative Fisheries Investigations.	Lynn Yarmey, Karen Baker
2009-09-14	LTER: A Web of Repositories	The movement and exchange of data are frequently described using a 'flow' or a 'pipeline' model. We differentiate a unidirectional data 'flow' from an alternative model, a web-of-repositories. A web-of-repositories is a federation of diverse nodes.	Robert Petersen, Sean Wiley, Nicole Kaplan, Eda Melendez, Karen Baker
2009-09-14	LTER Information Management History Database (HistoryDB)	Organizational history requires a facility to manage, archive and present event details as well as narratives that provide perspective to the events. While events form a historical thread, storied narratives weave these threads together into a retrospect	James Conners, Mason Kortz, Lynn Yarmey, Karen Baker
2009-09-14	PAL & CCE LTER: A Site-Based Information Architecture	Designing infrastructure to support the management of diverse data presents unique challenges for each site. Described here is the current information system architecture, as well as targeted architectural features, implemented by the Ocean Informatics team.	Karen Baker, Florence Millerand, Lynn Yarmey
2009-09-14	LTER Growing Information Infrastructure: Data Lifecycles and Subcycles	Information infrastructure, a vital aspect to many contemporary scientific investigations, is in transition. A lifecycle model for digital datatogether with plans for standards provide a framework for site-based information management	Mason Kortz, Lynn Yarmey, James Conners, Todd Ackerman, Karen Baker
2009-09-14	LTER Unit Working Group Projects: Dictionary and Registry	Units of measurement are a fundamental element of scientific discourse and data integration. The LTER Unit Working Group has developed two initiatives to promote consistent use of units throughout the network including the Unit Dictionary.	Karen Baker, Mason Kortz, Ed Weber, Rich Charter, Susie Jacobson, Sam McClatchie, Bill Watson, Tony Koslow
2009-12-07	CalCOFI Toward Integrated Data: Web Access to CalCOFI Ichthyoplankton Data	IchthyoDB ( <a href="http://oceaninformatics.ucsd.edu/ichthyoplankton">http://oceaninformatics.ucsd.edu/ichthyoplankton</a> ) is a queriable web application that provides data about abundance of fish eggs and larvae sampled as part of the CalCOFI program.	







3

Poster date: 2005-12-06

Title: Initiating the Data Dialogue: 2005 CalCOFI Conference Interactive Poster

Description: The interactions surrounding the 2005 CalCOFI Data Management poster are captured through photographs of updates and additions made to the poster during the poster session of the annual conference.

Authors: Karen Baker

### Initiating the Data Dialogue: 2005 CalCOFI Conference Interactive Poster



2005-12-06

Poster date: 2006-04-01

Title: Ocean Informatics: Conceptual Framework for Marine Science Information Management

Description: The work of Ocean Informatics is represented. Participants range from data and information managers to technical specialists, archivists, scientific researchers, educators.

Authors: Karen Baker, Jerry Wanetick, Shaun Haber, Lynn Jarney, Mason Kortz, Florence Millerand, Jesse Powell, Jim Wilkinson, Robert Thombley, Julie Thomas, Beth Simmons

# Ocean Informatics

**Conceptual Framework for Marine Science Information Management**  
Promoting Resource and Repository Infrastructure  
organizational • technical • social  
Towards a Shared Data-space

**Towards an Ocean Informatics Learning Environment**

The work of Ocean Informatics is represented at the union of oceanography, information science and social science domains. Participants range from data and information managers to technical specialists, archivists, scientific researchers, educators, as well as those working in science and infrastructure studies.

Ocean Informatics is a community of practice emerging to meet the challenges of articulating requirements and designing for complex systems of information that support heterogeneous data collections and diverse local practices. Engaging with issues of interdisciplinary communication and collaboration, the focus is on developing strategies from the design process to mindful variety, from shared language to boundary objects. In fostering a multi-dimensional sociotechnical awareness, the focus is to design a thick infrastructure that enables interoperability and facilitates collaborative science and scientists.

This is not a temporary state of affairs; the design of an information infrastructure is an ongoing process. Our products are both immediate and long-term: a deeper understanding of and engagement with information flows and learning environments

**A New Scientific Communication Regime:**  
**Joint Database Design Teams**  
*Heterogeneous Data Types:* shipboard underway, event logging, survey data, autonomous & manned shore stations, moorings, meteorological & glider data  
**Data Interoperability:** data defined technically, semantically, politically, and within organizations in such a way that like types can be identified for comparison and combination; entails community-accepted conventions and standards, syntactic and semantic agreements.

Data Management Elements	Ocean Informatics Environment System Components
<b>Data Sets</b> • Accessible data • Local metadata	<b>Repository Systems</b> • Homogeneous data • Metadata standards
<b>Physical Infrastructure</b> • Computer data storage (local) • Software (local-external) • Support (by project) • Product flows • Technology-oriented	<b>Information Infrastructure</b> • Division level with local access • Software (open-external) • Support (by organization) • Product flows • User-oriented
<b>Outreach</b> • Temporal effort • Short-term projects • Interdisciplinary • Out-of-disciplinary communication • Little DM	<b>Outreach &amp; Involvement/Teaching/Learning</b> • Embedded • Long-term strategies • Interdisciplinary • Out-of-disciplinary communication • Shared DM • Design processes • Iterative assessment

**A New Scientific Information Regime:**  
**Cyberinfrastructure Endeavors**  
*Federated Data Flows:* centers, mediators, applications, designers, services, communication mechanisms, transport, organizational structure, and information systems support  
*Community Building Mechanisms:* workshops, standards, best practices, reading groups, MMI, SIO Webheads

**What is Informatics?**  
Informatics is concerned with the applications of structure and organization to recorded information and the associated human activities. The term is used simultaneously today in a variety of ways, emphasizing applications of information technology, representing natural or human systems, and exploring multifaceted sociotechnical issues. Informatics draws on the information and social sciences in considering conceptual, organizational, technological and communication issues.

**What is Interoperability?**  
Interoperability is the seamless exchange of data and information across diverse networks, disciplines, and applications that enables data accessibility and information integration. Achieving data interoperability is understood here as a process that aims at information accessibility and sharing in order to enable reuse by various people, across diverging disciplines and across long periods of time.

**What are Information Infrastructure Elements?**

**OI Community Work**  
Shared Dictionaries  
User and Attribute Registries  
Best, Better and Good Enough Practices  
Database Schemas  
PODC, OGC, & OGC/MD Standards  
LTER BML Metadata  
Collaborative Design Methods

**OI Technical Work**  
Open Source Architecture  
Electronic Query Definition  
Digital Data Loggers  
Database Schemas  
Ecoinformation BML Metadata  
Negotiating organizational-community arrangements for the technical

**OI Collaborative Mechanisms**  
Strategic Design Teams  
Themed Working Groups  
Social Information Partnerships  
OI Blog and Development Forum  
Data Center & Shared Storage  
Reading Groups and Articulation Work  
IM Conferences and Newsletters  
Collaborative Exercises  
Implementation of Collaborative Field Practices

**AMF and the OOI Website: Integrated Cyberinfrastructure**  
Traditional Technical View

**AMF Website: Sociotechnical Cyberinfrastructure**  
Long-term Sociotechnical View

**References**  
KBaker, KJahoon, JWanetick, 2005 System Science Proceedings Towards an Ocean Informatics Environment, HIC3505  
HKarnal and KBaker, Informatics for the Long-Term: Biological Information Management. Hawaii International Conference for System Science Proceedings, HIC3504, IEEE Computer Society, 2004  
HKarnal, KBaker and OCBowser, Biological Storytelling and Collaborative Scientific Activities, BIOORCUUP Bulletin 28(2): 29-30, 2002.  
DRibes, KBaker, FMillerand, and GBowser, Comparative Interoperability Project: Configurations of Community, Technology, Organization. Joint Conference on Digital Libraries, 2005. Available at <http://dlm.interoperability.usst.edu>

**Timeline of Key Events:**

- 1903: A self-organizing scientific and technical community-of-practice with a common interest in informatics, design, data, earth science, and the long-term
- 1949: UCS-D
- 1959: LTER
- 1981: CCS DataZoo
- 1991: PAL LTER
- 2001: IOD
- 2002: Social Informatics
- 2003: Shared Infrastructure: iOcean
- 2004: Oceans Observing Systems
- 2004: OI Design Studio
- 2004: QAR TOD
- 2004: <MMI />
- 2004: OCE LTER
- OI Exchanges
- Data Zoo 2.0

**Authors:** Karen S. Baker, Jerry F. Wanetick, Shaun F. Haber, Lynn Jarney, Mason Kortz, Florence Millerand, Jesse Powell, Jim Wilkinson, Robert Thombley, Julie Thomas, Beth Simmons

2006-04-01




6

Poster date: 2006-09-20


# Title: LTER: Research in Infrastructure Studies: Social & Organizational Perspectives on Ecological Data Management

Description: In the midst of major changes in ecological data collecting, managing and sharing, an interdisciplinary team of information, ecological, and social scientists has been brought together at LTER PAL and CCE sites to facilitate the growth from site-based to larger-scale efforts.

Authors: Florence Millerand and Karen S. Baker



**Research in Infrastructure Studies:  
Social & Organizational Perspectives on Ecological Data Management**  
Florence Millerand and Karen S. Baker  
Université du Québec à Montréal, University of California San Diego, Scripps Institution of Oceanography



---

### A. Abstract

In the midst of major changes in ecological data collecting, managing and sharing, an interdisciplinary team of information, ecological, and social scientists has been brought together at LTER PAL and CCE sites to facilitate the growth from site-based to larger-scale federating infrastructures. While research endeavors traditionally focus either on the technical or the social aspects of information systems design, this project addresses simultaneously the technical, social, and organizational dimensions of the development, usage, and maintenance of large-scale information infrastructure in ecological science. Such dimensions include design methodology, change mechanisms and interdisciplinary collaboration as well as participant engagement, articulation processes, and data stewardship.

We present the project (Comparative Interoperability Study), the research area (Infrastructure Studies), and findings from a case study on the design, development and implementation processes of the Ecological Metadata Language standard in the LTER community. A conceptual framework based on the notion of enactment from organization theory is presented to broaden the understanding of large scale information infrastructure deployment. Initiated a decade ago, Infrastructure Studies appears to be a new and promising research area for the digital needs emerging in the natural sciences.

### What is Infrastructure Studies?

Infrastructure Studies is a social science research area that focuses on the study of *infrastructure* – water pipes and e-grids, community regulations and network standards, natural resources and computational resources, national parks and scientific data repositories – from a social, organizational and technical point of view. Infrastructure Studies draws from cross-disciplinary literatures including Large Sociotechnical Systems, Science and Technology Studies, Computer Supported Cooperative Work, Participatory Design, Information Science, and Social Informatics.

While the LTER program is engaged in the development of a cyberinfrastructure that will support and facilitate synthetic science across regional and network scales, Infrastructure Studies helps by providing new understandings of *technological dynamics* (from metadata standards to ontologies), *organizational concerns* (from site to networks) and *social dimensions* (from local to global).

### C. Case Study: Ecological Metadata Language (EML)


The Ecological Metadata Language has been adopted by the LTER community in 2001. Delays in its local implementation at some of the sites raise important issues in terms of EML functionalities as well as the state of community semantics and of LTER site 'readiness' for structured metadata production methods.

A Community Process working group was developed as a *communication mechanism* between EML developers and LTER information managers to help in articulating both perspectives. The diagram (below left) prompted the beginning of a *dialogue* through the mediation of social scientists from the Comparative Interoperability Project, that ended as a *'Lesson Learned' paper* with the 4 co-conveners as co-authors (Millerand et al., 2005) and a new understanding of community standards development with design and enactment as explicitly integral elements (below right).

### What is a Working Standard?

A working EML standard implies a maintaining of itself as a community-wide structure as well as an adaptation of itself (to fit diverse local needs and existing information systems) and of information managers' pre-existing work practices (to comply with new structured metadata production processes).

The viability of a working standard depends upon negotiated changes over time. That is, EML could not operate as a *standard per se* within the LTER community without ongoing mutual adjustments.



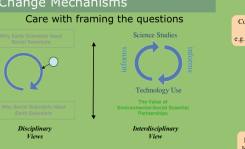
---

### B. Design Methodology & Change Mechanisms

#### Partnership with Social Sciences

Human Sciences	Earth Sciences	Information Sciences	Science Studies
<ul style="list-style-type: none"> <li>Anthropology</li> <li>Communication Studies</li> <li>Computer Science</li> <li>Design</li> <li>Education</li> <li>Engineering</li> <li>Environmental Studies</li> <li>Health, Behavior, and Society</li> <li>History</li> <li>Information Systems</li> <li>Law</li> <li>Library Science</li> <li>Management Science</li> <li>Media Studies</li> <li>Political Science</li> <li>Psychology</li> <li>Sociology</li> <li>Technology</li> <li>Urban Studies</li> <li>Visual Arts</li> <li>Work</li> </ul>	<ul style="list-style-type: none"> <li>Archaeology</li> <li>Biology</li> <li>Chemistry</li> <li>Ecology</li> <li>Environmental Science</li> <li>Geography</li> <li>Geology</li> <li>Hydrology</li> <li>Marine Biology</li> <li>Mathematics</li> <li>Physics</li> <li>Plant Biology</li> <li>Soil Science</li> <li>Statistics</li> <li>Systematics</li> <li>Zoology</li> </ul>	<ul style="list-style-type: none"> <li>Artificial Intelligence</li> <li>Computer Science</li> <li>Database Systems</li> <li>Design</li> <li>Human-Computer Interaction</li> <li>Information Systems</li> <li>Information Technology</li> <li>Library Science</li> <li>Management Science</li> <li>Media Studies</li> <li>Political Science</li> <li>Psychology</li> <li>Sociology</li> <li>Technology</li> <li>Urban Studies</li> <li>Visual Arts</li> <li>Work</li> </ul>	<ul style="list-style-type: none"> <li>Anthropology</li> <li>Communication Studies</li> <li>Computer Science</li> <li>Design</li> <li>Education</li> <li>Engineering</li> <li>Environmental Studies</li> <li>Health, Behavior, and Society</li> <li>History</li> <li>Information Systems</li> <li>Law</li> <li>Library Science</li> <li>Management Science</li> <li>Media Studies</li> <li>Political Science</li> <li>Psychology</li> <li>Sociology</li> <li>Technology</li> <li>Urban Studies</li> <li>Visual Arts</li> <li>Work</li> </ul>

#### Care with framing the questions



### CIP Project: Comparative Interoperability Study

The Comparative Interoperability Study participants are investigating three earth science research projects – GEON, LTER, OI – having distinct scientific information infrastructures and adopting different approaches to data interoperability. We aim at understanding the social and organizational dimensions of IIR, considering diverse forms of community engagement, and working with communities using an action research approach. For this empirical sociotechnical study we make use of cross-case analysis, ethnographic tools, and participatory mechanisms.

**GEON** is a cyberinfrastructure for the US geo-science aimed at providing scientific data and metadata sharing services to a broad range of disciplines to ensure a more integrated picture of earth processes (Keller, 2003).

**LTER** is a federated network of home sites with an information infrastructure for ecological sciences that aims at enabling interdisciplinary collaborative and processing data for the long-term (Hobbes et al., 2003).

**OI** is a research initiative for the ocean sciences based at US Scripps Institution of Oceanography that aims at providing a set of resources including shared scientific data and a design environment for testing, tool sharing and participatory design (Baker et al., 2005).

### D. Findings

#### Enactment and Organization Theory

In organization theory, the term *enactment* refers to the notion that when people act they bring structures and events into existence and set them in action. Thus, organizations are better understood as 'enacted environments' where individuals and organizations are constantly in the process of self-formation (Weick, 1979).

Applied to technology design and development, we distinguish between 'objective' versus 'enacted' technologies (Fountain, 2005), where an objective technology is a set of hardware and software components versus an enacted technology is the technology as perceived, designed or used in a specific context. For example, a metadata standard is both a generic standard for the wide range of environmental sciences and a specific tool to enhance metadata production at a local laboratory.

- From this perspective, implementing EML means more than doing 'implementative' work, it involves articulation work and enactment work.

### What is Articulation Work?

Articulation work is work – often invisible or unnoticed work. Articulation work refers to the interlarding of parts or the alignment of work elements, often involving a range of naming, categorizing, planning, coordinating, negotiating activities. Four distinct dimensions of this work are:

- Sensitizing to ramifications of names, categories, and dichotomies as well as communication, technology, community
- Juggling, rebalancing and refactoring in the present
- Aligning with legacy elements
- Aligning over time

### What is Enactment?

Enactment involves articulation work. Introducing the enactment concept informs our understanding of community processes such as developing working standards.

### What is one type of enactment that YOU are working on?

---

### What are Qualitative Methods?

Refers to the qualitative description of human social phenomena based on fieldwork to gain understanding of human arrangements and actions. It often includes ethnographic techniques and may include grounded theory as well as other conceptual frameworks such as life narratives or geographic in-depth interviewing. It typically involves selected small numbers of samples.

- Emphasizes the study of the phenomena in action (i.e. practice), e.g. a particular organization or group doing data management or infrastructure building.
- May include grounded theory with underlying assumptions regarding development of theory usually emerging concurrently with field research and based on data (empirical, participatory, documents) and the processes of data collection, coding, and meaning.
- Includes ethnography (from the Greek ethnos = nation and grapho = writing), using ethnographic techniques such as participant observation and may involve interviews, collaborative design, and document analysis.

### Our Research Contribution to the Grand Challenge is identifying and developing community mechanisms and strategies for 'process-building'

---

### Acknowledgements:

Thanks are given to LTER and Ocean Informatics community members especially the Infrastructure Management Working Group for their efforts in collaboration. Methods, clarity, and new conceptual frameworks.

### Community Communications:

Baker, K.S., Millerand, F. (2006). *Guiding and Articulating Community Standards for the Social Organization of Information Systems*. LTER Network Newsletter – Spring 2006

Baker, K.S., Millerand, F. (2005). *From Site to Community: Process of Standard Development in LTER Infrastructure Studies*. LTER Network Newsletter – Spring 2005

### References:

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F. (2005). *From Site to Community: Process of Standard Development in LTER Infrastructure Studies*. LTER Network Newsletter – Spring 2005.

Baker, K.S. and Millerand, F. (2007). *Articulation Work: Supporting Information Infrastructure Design, Construction, Organization, and Assessment in Practice*. Proceedings of the 2007 Conference on Design Research (CHI/CHI Extended Abstracts), San Jose, CA, April 28-30, 2007.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

Baker, K.S., Millerand, F., O'Brien, J. (2005). *Interoperability Strategies for Scientific Organizations: Research and Practice*. Proceedings of the 2005 International Conference on Information Systems (ICIS), Boston, MA, October 1-5, 2005.

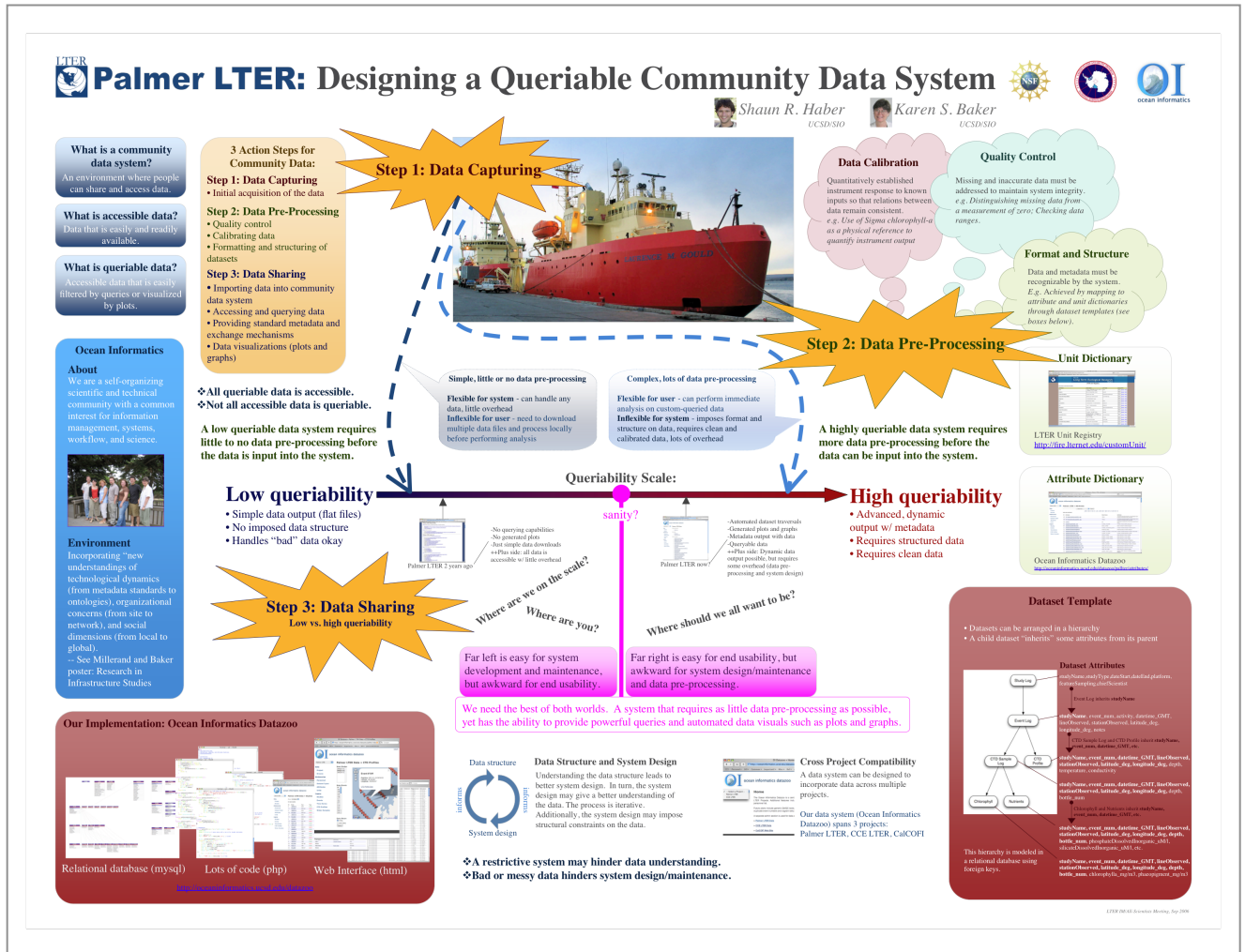
Baker, K.S., Millerand, F.,

Poster date: 2006-09-20

Title: Palmer LTER: Design of a Queriable Ocean Information System

Description: Field data, originating with domain understandings and practices that shape sampling and collection, has informed development of the PAL LTER information system. In becoming digitally preserved, data capture may in turn be influenced by information system work.

Authors: Karen Baker and Shaun Haber







Poster date: 2006-12-04


Title: CalCOFI: An Oceanographic Event Logger


Description: Local data management, informed by field sampling and data use, supports community coordination at the interface of data collection and data curation. An oceanographic event logger recently deployed on a series of research cruises extends data management to the field.

Authors: James Wilkinson, Karen Baker

# An Oceanographic Event Logger



A FIELD PERSPECTIVE

**James R. Wilkinson and Karen S. Baker**  
Scripps Institution of Oceanography, University of California San Diego



A DATA STEWARDSHIP PERSPECTIVE

### Field Practices

An oceanographic event logger, recently deployed on CalCOFI research cruises, extends data coordination into the data collection arena. The event logger system – consisting of networked PCs, communal event log, GPS coordinates and event indexing – promotes standard conventions & establishes relationships between diverse data efforts at the time of collection. The event logger addresses issues of time, space and categorization using standard vocabulary in order to assist subsequent data integration and exchange. It becomes one element of an information infrastructure that contributes to creating a common dataspace (Franklin et al, 2005), both conceptual and physical, that stretches from field to land and back again.

### Purpose

Project-specific data management must support data community coordination. By using the event logging system throughout the ship, all cruise participants, on any workstation, are able to:

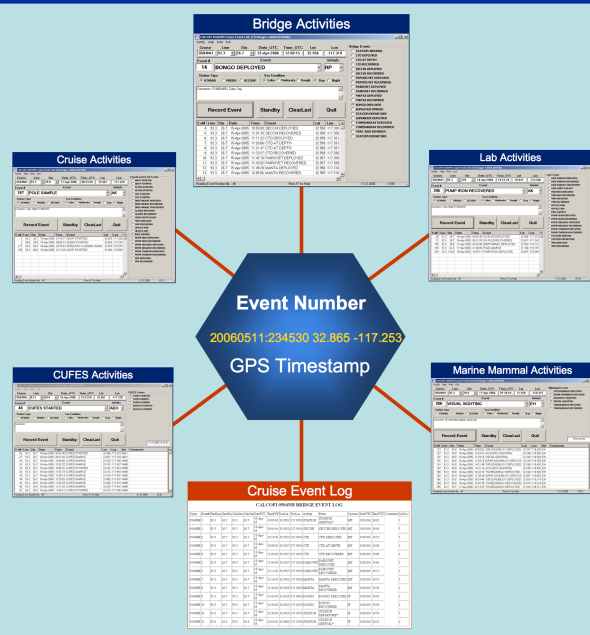
- ◆ contribute uniquely indexed events to the cruise event log.
- ◆ establish a common coordinate system to log all samples from shared activities, such as rosette sampling, using the event index and its ship-based GPS date, time, latitude & longitude.
- ◆ relationally link diverse data products from shared events post-cruise using the index as a relational database identifier.
- ◆ correlate continuous data, such that weather, acoustics, or sea-surface temperature measurement, to individual or grouped events using the common coordinate system.

### Methods

Installation of the event log software on individual, networked Windows workstations enables local, project-specific activities lists (main figure). Each workstation polls common GPS & event number files, incrementing the value when an event is recorded. A cruise log tabulates all logged events. Each workstation also generates a separate project-specific activities log. Current requirements:

- ◆ standalone: Windows pc with GPS
- ◆ networked: multiple Windows pcs with GPS on one and a mapped network directory

### Event Logger System = Event Logger(s) + Event Number + GPS => Event Log




Event Number  
20060511-234630 32.865 -117.253

GPS Timestamp

### Information Infrastructure

Digital practices from data collection to data preservation are supported by Information Infrastructure. Infrastructure-building draws on the fields of informatics, information systems, science and technology studies, library and information sciences as well as infrastructure studies. At SIO, an Ocean Informatics environment supports the design, development, deployment, and enactment of effective data practices as part of community infrastructure-building (Baker et al, 2006).

### Design and Use



Design theory and classification analysis are playing an active role in the iterative development of the CalCOFI Event Logger (Lindseth & Baker, 2006). Our methodology enables interplay of shared vocabulary-building and controlled vocabulary list-use. This facilitates a collective understanding of data and its organizing in addition to ensuring that data are well represented and integrated within the data community. Event Logger features include:

- ◆ Configuration files to create local flexibility required for varying ship platforms and group interests
- ◆ Authoritative lists of shared vocabulary established to accommodate changing vocabulary and maintainability as well as to contribute to sociotechnical process-building
- ◆ Event numbers enmeshed in field practices that travel from sea to land as a data index incorporated into the architecture of the information system

### Acknowledgements

We would like to thank the Ocean Informatics participants who are designing for the long-term, the field participants – ship and scientific staff – who are enacting the system, as well as the community participants who are co-constructing the system. In particular we recognize the work of Robert Thornberry, Sherril Devi, and Josiah Prewitt. This work has been supported by NOAA CalCOFI and NSF Long-Term Ecological Research and Human Social Dynamics Programs.

### References

- Baker, K., S.Jackson, and J.R.Warwick, 2005. Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment. Proceedings of the 38<sup>th</sup> Hawaii International Conference on System Sciences (HICSS), IEEE Computer Society, 2-4 January 2005, Big Island, Hawaii, 2005. Hawaii International
- Franklin, M., A.Haley, and D.Maler, 2005. From Dataspace: A New Abstraction for Information Management. SIGMOD Record 34(4):27-33.
- Lindseth, B. and K.Baker, 2006. Collaborative Design of an Oceanographic Event Logger. Proceedings of the Computer Human Interface Conference CHIE2007. (submitted).

CalCOFI Conference 2006

10

Poster date: 2007-08-02

Title: LTER: Data Integration in the Decade of Synthesis

Description: As data availability, findability, and even queriability become more ubiquitous, the need to make sense of data from multiple, disparate sources increases. Data integration and data synthesis allow extension of the scope of data beyond local use.

Authors: Mason Kortz, Lynn Yarmey, James Connors, Karen Baker

**What is Data Integration?**  
Data integration is a state of compatibility and comparability between data from varied and disparate sources. Data integration also refers to the process of achieving such a state.

**What is Data Synthesis?**  
Data synthesis refers to the comparison, manipulation, and recombination of integrated datasets to create new views into the data. Data synthesis creates new knowledge without the need for additional data gathering.

**What is DataZoo?**  
DataZoo is the data and metadata storage and publishing system currently in development by the SIO Ocean Informatics team. DataZoo combines a detailed metadata schema with a web interface to allow data providers to become part of the integration process. A second interface allows queries, ranging from simple retrieval of source data to plots of synthetic data products.

**Data Integration and Data Workflow**  
Data integration is not a single step in the data workflow; it is an ongoing iterative process that begins well before any data is gathered. Addressing integration issues early in the workflow prevents complications further down the line. However, some complications are not apparent until after they arise - in which case they inform the next generation of system design.

**Where is synthesis in the workflow?**  
Data integration forms the basis for reliable, meaningful synthesis. By doing integration work throughout the data workflow, data products are made available to the end user for quick, clean synthesis. Integration does not end with a synthetic data product. If the synthesis process is well documented in the metadata, the synthetic product can be ingested into the original (or another) integrated data system. The synthetic product represents new knowledge and is available for further, iterative synthesis, creating an even more refined information product.

**Data Integration and Metadata**  
Integration efforts, whether at the dataset, study, or organization level, must be documented in metadata, or data become ineffective when shared outside the original context. However, metadata itself must address the question of **semantic integration** - how can a metadata system enable integration with another system or user if they do not share a common language?

**DataZoo: The CCE/PAL LTER Metadata System**  
Users can produce synthetic data products by combining columns from multiple datasets. **Semantic restriction** via the unit and attribute dictionaries allows the system to automatically suggest compatible columns. Qualifiers are designed using **semantic suggestion**, allowing the user to make informed decisions about which column to combine, and how. Synthetic datasets can be shared or saved for later use.

**Data Integration and User**  
Integration enables both automated synthesis and user-controlled synthesis - services that can conflict with each other at the system design level.

**System Automation**  
• High initial time investment  
• Low per-use time investment  
• Creates standardized products  
• Most decisions made by system  
• Simple interface  
• Adding new features requires more back-end design  
• Can frustrate users who require specialized synthesis

**User Control**  
• Low initial time investment  
• High per-use time investment  
• Creates customized products  
• Most decisions made by user  
• Complex interface  
• Adding new features requires more front-end/interface design  
• Can overwhelm users with options and interfaces

**What's in the middle?**  
• An automated system with user-controlled overrides?  
• A user-controlled system with automated suggestions?  
• A hybrid system that remembers users' selections and uses them for future automation?

**Where are we on the scale? Where are you? Where do we all want to be?**

**Who is after you in the data integration process?**  
How much metadata do your data recipients want? What are their minimum standards for QA and QC? Will they be notified if the data and/or metadata are revised?

**Where are you in the data integration process?**  
• Individuals  
• Labs  
• Local Data Systems

**Who is before you in the data integration process?**  
Which elements of integration are data providers' responsibility, and which are yours? Is translation required between their dual metadata format and your own? Can they be engaged to participate in your design process?

**THE DATA ZOO**  
o d i o o i  
1 1 0 0 1 0

LTER MC Annual Meeting, San Jose, August 24, 2007




Poster date: 2007-08-02

Title: LTER Environmental Data Management: Infrastructure Studies Insights


Description: In the mist of major changes in ecological data collecting, managing and sharing, an interdisciplinary team of information, ecological, and social scientists has been brought together at LTER PAL and CCE sites to facilitate the growth of site-based information management.

Authors: Florence Millerand and Karen Baker



## Environmental Data Management: Infrastructure Studies Insights

Florence Millerand and Karen S. Baker  
 Université du Québec à Montréal, University of California San Diego, Scripps Institution of Oceanography



The Grand Challenge  
for Informatics & Long-Term Environmental Science  
Information Infrastructure Building (IIB)

### A. Abstract

In the mist of major changes in ecological data collecting, managing and sharing, an interdisciplinary team of information, ecological, and social scientists has been brought together at LTER PAL and CCE sites to facilitate the growth from site-based to larger-scale federating infrastructures. While research endeavors traditionally focus either on the technical or the social aspects of information systems design, this project addresses simultaneously the technical, social, and organizational dimensions of the development, usage, and maintenance of large-scale information infrastructure in ecological science. Such dimensions include design methodology, change mechanisms and interdisciplinary collaboration as well as participant engagement, articulation processes, and data stewardship.

We present the project (Comparative Interoperability Study) the research area (Infrastructure Studies) and findings from a case study on the design, development and implementation processes of the Ecological Metadata Language standard in the LTER community. A conceptual framework based on the notion of enactment from organization theory is presented to broaden the understanding of large scale information infrastructure deployment. Initiated a decade ago, Infrastructure Studies appears to be a new and promising research area for the digital needs emerging in the natural sciences.

**What is Infrastructure Studies?**  
 Infrastructure Studies is a social science research area that focuses on the study of *infrastructure* – water pipes and digital pipes, fire departments and help desks, power grids and e-grids, community regulations and network standards, natural resources and computational resources, national parks and scientific data repositories – from a social, organizational and technical point of view. Infrastructure Studies draws from cross-disciplinary literatures including Large Sociotechnical Systems, Science and Technology Studies, Computer Supported Cooperative Work, Participatory Design, Information Science, and Social Informatics.

While the LTER program is engaged in the development of a cyberinfrastructure that will support and facilitate synthetic science across regional and network scales, Infrastructure Studies helps by providing new understandings of **technological dynamics** (from metadata standards to ontologies), **organizational concerns** (from site to network) and **social dimensions** (from local to global).

Information Infrastructure Building  
Projects & Processes

2001-2002 Project  
**Interdisciplinary  
Ethnography  
Participatory Design  
Information System Science**

2004-2007 Project  
**Comparative Interoperability Study  
Comparative Analysis**

### B. Design Methodology & Change Mechanisms

**Partnership with Social Sciences**

**Human Sciences**

Anthropology  
Cognitive Psychology  
Communication Studies  
Cultural Studies  
Ecology  
Ergonomics  
Epidemiology  
Evolutionary Psychology  
Health Services Research  
History  
Human Geography  
Humanities  
Law  
Linguistics  
Management  
Marketing  
Media Studies  
Political Science  
Psychology  
Sociology  
Sociotechnical Systems  
Technology Studies  
Urban Studies  
Visual Studies

**Earth Sciences**

Archaeology  
Biology  
Chemistry  
Ecology  
Earth System Science  
Environmental Science  
Environmental Systems  
Evolutionary Biology  
Geography  
Geology  
Geophysics  
Hydrology  
Marine Biology  
Microbiology  
Oceanography  
Plant Biology  
Soil Science  
Terrestrial Ecology  
Zoology

Care with framing the questions

**Design**

Design Methodology  
Design Science  
Design Theory  
Design Research  
Design Practice  
Design Education  
Design History  
Design Critique  
Design Evaluation  
Design Innovation  
Design Management  
Design Policy  
Design Regulation  
Design Standards  
Design Tools  
Design Work  
Design Writing

**Science Studies**

Science and Technology Studies  
Science Communication  
Science Education  
Science History  
Science Policy  
Science Practice  
Science Regulation  
Science Standards  
Science Tools  
Science Work  
Science Writing

Technology Use

**Disciplinary  
Flows**

Information Science  
Information Systems  
Information Technology  
Information Systems Science  
Information Systems Research  
Information Systems Studies  
Information Systems Theory  
Information Systems Work  
Information Systems Writing

**Interdisciplinary  
Flows**

Information Science  
Information Systems  
Information Technology  
Information Systems Science  
Information Systems Research  
Information Systems Studies  
Information Systems Theory  
Information Systems Work  
Information Systems Writing

CIP Project: Comparative Interoperability Study

The Comparative Interoperability Study participants are investigating three earth science research projects – GEON, LTER, IIB – having distinct scientific, information infrastructure and deployment different approaches to data interoperability. We aim at understanding the social and organizational dimensions of IIB, considering diverse forms of community engagement and working with communities using an action research approach. For the empirical sociotechnical study we make use of interview analysis, qualitative research techniques (including interviews and focus group work), and participatory mechanisms.

IIB (US) participants are providing scientific data and resources sharing services to a broad range of disciplines to create a more integrated picture of earth resources (Miller et al. 2005).

LTER is a cyberinfrastructure research effort at providing scientific data and resources sharing services to a broad range of disciplines to create a more integrated picture of earth resources (Miller et al. 2005).

IIB (US) participants are providing scientific data and resources sharing services to a broad range of disciplines to create a more integrated picture of earth resources (Miller et al. 2005).

**What is Ethnography?**  
 Ethnography (from the Greek ethnos = nation and graphos = writing)  
 ✓ Refers to the qualitative description of human social phenomena, based on fieldwork  
 ✓ Emphasizes the study of the phenomena in action (in practice), e.g. a particular organization or group doing data management or infrastructure building  
 ✓ Involves participant observation and document analysis as main ethnographic research techniques (Hammersley and Atkinson, 1995).


### C. Case Study: Ecological Metadata Language (EML)

The Ecological Metadata Language has been adopted by the LTER community in 2001. Delays in its local implementation at some of the sites raise important issues in terms of EML functionalities as well as the state of community semantics and of LTER site 'readiness' for structured metadata production methods.

A Community Process working group was developed as a communication mechanism between EML developers and LTER information managers to help in articulating both perspectives. The diagram (below left) prompted the beginning of a *dialogue* through the mediation of social scientists from the Comparative Interoperability Project, that ended as a *Lesson Learned paper* with the 4 co-authors as co-authors (Millerand et al., 2005) and a new understanding of community standards development with design and enactment as explicitly integral elements (below right).

**What is a Working Standard?**  
 A working EML standard implies a maintaining of itself as a community-wide structure as well as an adaptation of itself (to its diverse local needs and existing information systems) and of information managers' pre-existing work practices (to comply with new structured metadata production processes).

The viability of a working standard depends upon negotiated changes over time. That is, EML could not operate as a **standard per se** within the LTER community without ongoing mutual adjustments.



Envisioning the cycle of Information Infrastructure Development:

The EML Working Group

The Working Group co-members included members from the Comparative Interoperability Project and members from the ecological data management community: Karen Baker (PAC/LTER Infrastructure Manager), EML 'Designer' Interoperability CAPS Barbara Brown (UCSD/LTER Infrastructure Manager), LTER IM Co-ordinator Chae-Hye Cho (UCSD), and Mark Jones (NCEAS, EML Developer, Florence Millerand (Interoperability Flow, Graduate Researcher).

### D. Findings

**Enactment and Organization Theory**

In organization theory, the term *enactment* refers to the notion that when people act they bring structures and events into existence and set them in action. Thus, organizations are better understood as 'enacted environments' where individuals and organizations are constantly in the process of self-formation (Weick, 1995).

Applied to technology design and development, we distinguish between 'objective' versus 'enacted' technologies (Fountain, 2005), where an objective technology is a set of hardware and software components versus an enacted technology is the technology as perceived, designed or used in a specific context. For example, a metadata standard is both a generic standard for the wide range of environmental sciences and a specific tool to enhance metadata production at a local laboratory.

From this perspective, implementing EML means more than doing 'implementation' work, it involves articulation work and enactment work.

**What is Articulation Work?**  
 Articulation work is work – often invisible or unnoticed work. Articulation work refers to the interrelating of parts or the alignment of work elements, often involving a range of naming, categorizing, planning, coordinating, negotiating efforts. Four distinct dimensions of this work are:

- ✓ Sensitizing to ramifications of names, categories, and dichotomies as well as communication, technology, community
- ✓ Juggling, rebalancing and redefining in the present
- ✓ Aligning with legacy elements
- ✓ Aligning over time

**What is Enactment?**  
 Enactment involves articulation work. Introducing the enactment concept informs our understanding of community processes such as developing working standards.

What is one type of enactment that YOU are working on?

Our Research Contribution to the Grand Challenge is identifying and developing community mechanisms and strategies for "process-building"

Poster date: 2007-08-02

Title: LTER: Long Term Informatics

Description: With the information age as one of the many ramifications of the Internet, our understandings, cultures, and communities are undergoing change.

Authors: Karen Baker, Cyndy Chandler, Anna Gold, Florence Millerand, Jerry Wanetick

**Long Term Informatics**  
longterminformatics.org

Karen Baker<sup>1</sup>, Cyndy Chandler<sup>2</sup>, Anna Gold<sup>3</sup>, Florence Millerand<sup>4</sup>, Jerry Wanetick<sup>1</sup>  
<sup>1</sup> SIO: Scripps Institution of Oceanography, UCSD; <sup>2</sup> WHOI: Woods Hole Oceanographic Institution; <sup>3</sup> MIT: Massachusetts Institute of Technology; <sup>4</sup> UQAM: University of Quebec, Montreal

Long Term Informatics recognizes local informatics environments and fosters the relations between these efforts. We use informatics to mean an interdisciplinary field of study, at the intersection of information and social sciences, using technologies as applied to a specific domain. Informatics is shaped by social, technological, and organizational arrangements (Baker and Bowker, 2002). It is a discipline concerned with the theory, history, and organization of information while engaged in the practice of data and information management.

**network:**  
a) an association of individuals having a common interest, formed to provide mutual assistance, helpful information, or the like; b) a wide variety of systems of interconnected components

**Loose networks of networks -**  
local informatics environments and information infrastructure capacity-building efforts:

- Environmental Informatics**
  - Canopy Studies: <http://canopy.evergreen.edu/>
  - Ecolinformatics: <http://ecolinformatics.org>
  - Ocean Informatics: <http://oceaninformatics.org>
    - SIO/UCSD: <http://oceaninformatics.ucsd.edu>
    - WHOI: <http://www.whoi.edu/sites/informatics>
- Library Informatics**
  - Data Work: <http://web.mit.edu/dig>
- Social Informatics and Science & Technology Studies**
  - Science-Technology: <http://rkesi.indiana.edu/>
  - Environmental Sciences: <http://interoperability.ucsd.edu>

A network of local informatics efforts enables information exchange, shared expertise, and learning through comparative analysis. We network in order to learn from and with each other, sharing understanding, tools, and methods. Documenting and analyzing successes AND failures as lessons learned can help redirect ongoing work.

**Next Steps:**  
 1: Gather nodes, network partnerships, and links  
 2: Define a vision anchored locally and connected globally  
 3: Develop a Long Term Informatics Initiative

As with environmental data, local information infrastructures come in a variety of sizes and shapes. They are representative of whole systems ... *wherein it's systems throughout.*

\*\*\*  
self-organize  
add yourself here ...

LTER 06C Annual Meeting  
San Jose, August 1-4, 2007

13

Poster date: 2007-09-17

Title: CCE LTER Information Infrastructure

Description: Information Infrastructure is an arrangement of computational systems, an iTeam, information systems and partnerships associated with a core interest in informatics.

Authors: Jerry Wanetick, Karen Baker, Nate Huffnagle, Lynn Yarmey, Mason Kortz, James Conners



# CCE LTER



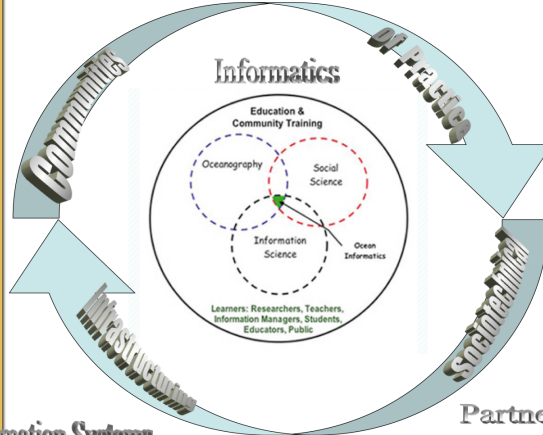
## Information Infrastructure

Jerry Wanetick, Karen Baker, Nate Huffnagle, Lynn Yarmy, Mason Kortz, James Conners  
Scripps Institution of Oceanography, University of California San Diego

### Computational Systems



### iTeam



### Information Systems

**DataZoo 2.0**  
<http://oceaninformatics.ucsd.edu/datazoo>

**SCCOOS**  
<http://sccoos.ucsd.edu/data>

**The Data Zoo**  
<http://zoo.ucsd.edu>

**CalCOFI**  
<http://www.calcofi.org>

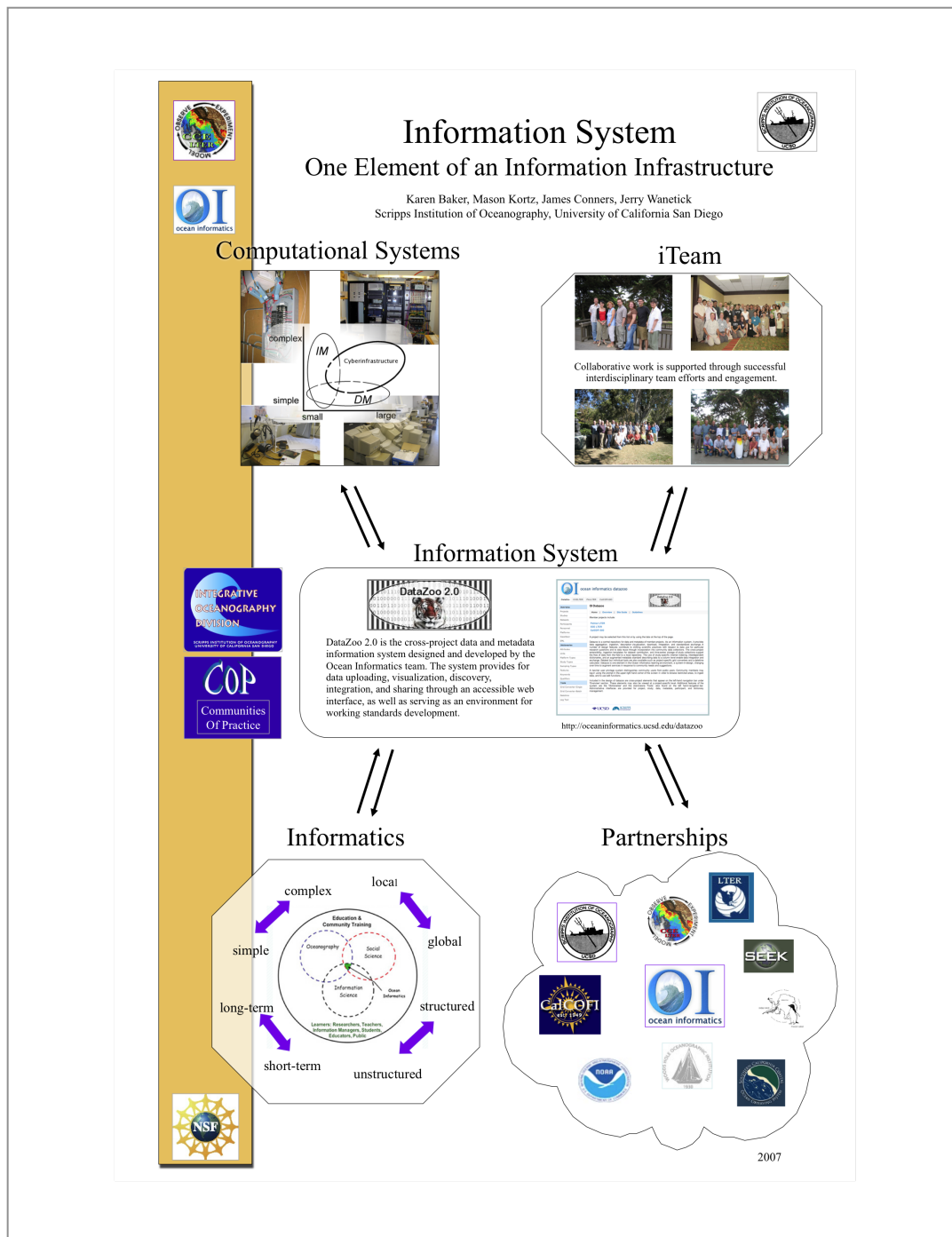
### Partnerships




Poster date: 2007-09-17

Title: Ocean Informatics Information System: One Element of an Information Infrastructure  
Description: Focus is on an Information system for managing data - DataZoo 2.0 -at the heart of a configuration of computational systems, an iTeam, informatics work, and a complex set of partnerships.


Authors: Karen Baker, Mason Kortz, James Connors, Jerry Wanetick





## A Working Standard: Augmenting the Ecological Metadata Language

Lynn Yarmey and Karen Baker



### Abstract

Metadata standards are an integral and necessary part of data sharing as they provide a structure and format to allow comparisons of data context. A full and complete metadata record is essential to understanding and using any dataset, as without the context of the data, values are meaningless. A metadata standard not only prepares for future dataset comparisons and integrations, but also prompts the user to consider of all parts of a complete metadata record, from descriptions of the field environment to detailed accounts of any and all analytical methods and quality control procedures performed. A standardized metadata format also allows for quick automated or visual comparisons of datasets and begins to lessen the impact from any workflow articulation differences. The Ecological Metadata Language (EML) is a standard with growing acceptance in the scientific realm, its strengths include attribute-level descriptions and a flexible architecture. In this poster, we discuss the adaptations and augmentations made to EML to better encapsulate the complexity inherent to our local datasets.

### Attributes

**EML attribute definition**

$$\text{EMLAttribute} = \text{definition} + \text{storageType} + \text{measurementScale} + \text{unit} + \text{method}$$

**Ocean Informatics Columns**

**Base attribute**

$$\text{BaseAttribute} = \text{definition} + \text{unit} + \text{measurementScale} + \text{storageType}$$

Those parts of the attribute that are common across unique but related attributes. Base attribute properties include a broad definition, the storageType, measurementScale and unit.

**Attribute Type**

The attribute type is the Unit Type for ratio and interval attributes, and "text," "code" or "datetime" for nominal, ordinal and datetime attributes. The attributeType is the broadest level of pre-ontological attribute categorization. (ex. length, massDensity, code)

**Attribute Categories**

The attribute categories were created as a readiness element for later ontological work and synthesis. There are two levels of granularity, a very broad discipline-based category (ex. biological vs. chemical) and a slightly more specific sub-category (ex. physical-atmosphere vs. physical-hydrographical).

**Attribute Qualifiers**

The qualifiers constitute the unique and detailed aspects of an attribute. The qualifiers provide a structure for comprehensive methods notes for collection, analysis, and processing steps as well as instrumentation details and further description of the attribute itself. Not only do the qualifiers begin to explore the black hole of the method field, they prompt data collectors to consider and record all aspects of the context of the data context, allowing the individuals unfamiliar with local practices to make their own decisions about appropriate comparisons and further analysis.

**Example - Two Approaches**

**Ocean Informatics column**

```

baseAttName: temperature
measScale: interval
storageType: float
unit: degC
type: temperature
category: Physical-Atmosphere
Qualifiers - sampleCollection
sampleRealm: air
samplingType: obsInst
dataSource: underwayMets
Qualifiers - dataAnalysis
calculationType: calculated
calculationDetail: averaged
calculationFrequencyIn: hertz
calculationInterval: 5-minute
Qualifiers - qualityControl
notes: data was visually reviewed
                    
```

**EML attribute**

```

name: average air temp
definition: air temperature from underway system
measScale: interval
storageType: float
unit: degC
                    
```

### Study-Attribute Link

**EML hierarchical structure**

```

graph TD
    Study --> Dataset
    Dataset --> Attribute
                    
```

The EML hierarchy narrowly defines the scope of allowed information and does not take into account the relationship between attributes and the study level.

**Ocean Informatics flexible structure**

```

graph TD
    Study --> Dataset
    Dataset --> DatasetNotes[Dataset Notes Table]
    DatasetNotes --> Attribute
    Attribute --> Qualifier
                    
```

The Ocean Informatics not only adds Qualifier-level description to the metadata structure, but links the attribute and its qualifiers back to the study.

**Examples**

- On one cruise, a problem with the CTD calibrations forced an extra processing step to correct the salinities, all other methods were the same.
- Partway through another cruise, a shortage of supplies meant half of the chlorophyll samples were filtered on a different type of filter paper.

**Working Standards**

A **standard** is established by some type of authoritative community process as a finished product that captures a realized general perspective that will endure... until there is a new standard or a new version of the existing standard.

A **working standard** is a strategy for initiating local convergence. It is more than a best practice but less established than a standard as it is expected to change with discovery through practice of more generalized perspective on the structure to be captured.

Millerand, F., & Baker, K.S. (forthcoming). Who are the Users? From a Single User to Webs of Users in the Design of a Working Standard Information Systems Journal.

Millerand, F., and G. C. Bowler. (forthcoming). Metadata Standard: Trajectories and Enactment in the Life of an Ontology in Formalizing Practices: Reckoning with Standards, Numbers and Models in Science and Everyday Life, edited by M. Lampland, and S. L. Star. Cornell University Press.

### Units

**EML unit definition**

$$\text{EMLUnit} = \text{unitType} + \text{abbreviation} + \text{parentSI} + \text{multiplierToSI} + \text{description}$$

**Ocean Informatics Units**


$$\text{EMLUnit} + \text{SourceSite} + \text{Scope}$$

"Well specified units are at the heart of measurement comparability. Today's data integration efforts are highlighting a range of unaddressed and unresolved unit definition issues involving syntactic and semantic ambiguities and conflicts. The community repository or "unit registry" presented here addresses some of these issues by sharing unit names, types, definitions, and forms while introducing the concept of site-working group-community-domain scope... This (registry) represents an LTER site-network collaborative design effort to meet community needs by creating a mechanism for locating units compliant with the EML standard, for bringing together local solutions, and for prompting cross-site discussion of units." - LTER Unit Registry Design Team (2005)

**Examples**

- micromolar (uM) - CCE
- microMolesPerLiter (uM/L) - PAL
- microMolesPerLiter (M) - GCE/FCE
- millimolesPerGram (mm/g) - EML 2.0.1



These units, taken from the Registry, illustrate the flexibility provided, and the need for further standardization work in the names, abbreviations, etc.



# Scientific Infrastructure Design: Information Environments and Knowledge Provinces

Karen Baker<sup>1</sup> and Florence Millerand<sup>2</sup>

1 SIO: Scripps Institution of Oceanography, UCSD; 2 UQAM: University of Quebec, Montreal

**The Grand Challenge  
for Informatics & Long-Term Environmental Science is  
Information Infrastructure Building**

Conceptual models and design processes shape the practice of information infrastructure building in the sciences (Atkins et al, 2008; Edwards et al, 2006). We argue that differences in conceptual models have critical implications for users and their working environments. With ‘cybersized’ views and pipeline arrangements receiving a lot of attention in current scientific endeavors, highlighting the multiplicity of knowledge provinces with their respective worldviews opens up understandings of design processes and knowledge work.

### Information Environment Models

Earth Science  
Data

→

Data  
Center

→

General User

**A. Pipeline Model: Disintermediated Perspective**  
Disintermediation for a global general user. A pipeline or disintermediation scenario is framed as an automated set of technical procedures designed for optimized dataflow from the field.

Earth Science  
Data

↔

Earth Science  
Collector-User

↔

Data  
Center

→

General User

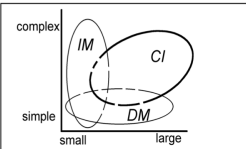
**B. Local Information Environment Model: Intermediated Perspective**  
Intermediation by the earth science data designer-collector-user teams constituting a local information environment supporting local science users as well as data delivery to a general end-user. A local information environment is a generative forum for dialogue about the data where participants create, support, and sustain communicative functions related to data and its flow.

### What is Cyberinfrastructure?

a) A solution  
b) A growth option  
c) One of multiple information infrastructure approaches

In case (a), cyberinfrastructure is seen as a general mechanism enabling global information flows, an upgrade to today’s independent centers. In this view, the aim is to align a single functional cyberinfrastructure. In (b), cyberinfrastructure is viewed as growing over time, gradually replacing earlier solutions. Claims of progress frequently attend these two cases. Case (c) highlights differences in arenas, suggesting the concept of knowledge provinces. Progress is active in this case as well, appearing with knowledge regions and frequently as integrative work at the interfaces of regions.

### What are Knowledge Provinces?



Knowledge provinces: a pluralistic view distinguishing interdependent work arenas including data management (DM), information management (IM) and cyberinfrastructure (CI). The concept of a plurality of knowledge provinces enables description of dynamic configurations with shifting boundaries and supports planning for a diversity of arrangements across the digital landscape. Attention to the growth of provinces is a strategy for changing how we think about generalizations with respect to knowledge-making and network federating.

### Ethnographic Studies

Ethnographic work in partnership with design and user communities within a variety of information environments is providing insight into the multiple dimensions of scientific work and relations to information infrastructure building (e.g. Star and Ruhleder 1996; Karasti and Baker, 2004; Baker et al, 2005; Lee et al, 2006; Ribes and Baker, 2007; Millerand and Bowker, forthcoming).

**Our Research Contribution to the Grand Challenge is identifying and developing conceptual approaches that open up the landscape for local AND global information infrastructure supporting a plurality of interrelated knowledge-making arrangements.**

**References**  
 Atkins, D. et al. (2008). NSF-AR Report: Revolutionizing science and engineering through cyberinfrastructure. NSF Report.  
 Baker, K. S., Ribes, D., Millerand, F. & Bowker, G.C. (2005). Disintermediation strategies scientific infrastructure: research and practice. AISST05.  
 Edwards, P.N., Jackson, S.J., Bowker, G.C., & Poyles, C. (2007). Understanding infrastructure: dynamics, tensions, and design. NSF Report of a Workshop: History and theory of infrastructure: lessons for new scientific cyberinfrastructures. <http://hdl.handle.net/2027.42/49353>.  
 Karasti, H. & Baker, K. S. (2004). Infrastructuring for the long-term: ecological information-management. HCSS37. IEEE Computer Society.  
 Lee, C.P., Dourish, P. & Mark, G. (2006). The Human infrastructure of cyberinfrastructure. CSCW06, Banff, Canada.  
 Millerand, F. & Baker, K.S. (forthcoming). Knowledge dynamics, tensions and enactment in the life of an ontology. In: S. L. Star & M. Linstead (Eds.), Formalizing practices: reworking with standards, numbers and models in science and everyday life.  
 Ribes, D. and Baker, K.S. (2007). Modes of Social Science Engagement. In Proceedings of the Third Communities and Technologies Conference, CSteinfield, BTPanband, Muckerman, and Nourbakhsh (Eds.). Morgan State University. Springer: 107-138.  
 Star, S.L. & Ruhleder, K. (1996). Steps Towards an ecology of infrastructure: design and access for large information spaces. Information Systems Research, 7(1), 111-134.

**Acknowledgement**  
 This work is supported by NSF NSD Grant IRI-03289 and the Social Sciences and Humanities Research Council of Canada Postdoctoral Fellowship #756-2003-0099. The work is in collaboration with LTER/NSF/OCSE site 1716 and NSF/OPTI 80-17202 and Ocean Informatics located at Scripps Institution of Oceanography.

AISST 2007, Milwaukee, Oct 19-24, 2007



85

Poster date: 2007-11-17

Title: CalCOFI & Ocean Informatics DataZoo: A Multi-Project Data Publishing System


Description: The DataZoo information system is a hub in the Ocean Informatics learning environment that creates a central forum for data exchange, collaborative design, and community building. It is a central repository for data and metadata of member projects.

Authors: Mason Kortz, James Conners, Karen Baker

## Ocean Informatics DataZoo: A Multi-Project Data Publishing System

Mason Kortz, James Conners, and Karen Baker - Scripps Institution of Oceanography, University of California San Diego



---

### Query Interface

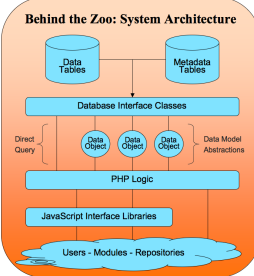
DataZoo's query interfaces allows users to find data easily and then retrieve data - as plots, tables, or text files - filtered by any column in the dataset. This means a user can view an entire dataset, or a subset associated with a particular cruise, time period, or location. This also allows a user to quickly find unusual values for any measurement in the dataset, a useful technique both for identifying interesting events and for quality control. Furthermore, the DataZoo interface can run queries over multiple datasets at once. This enables the user to integrate and compare data from multiple labs, ships, or even projects.

In addition to a tabular view of data, DataZoo's query interface can display dataset results as a line or scatter plot. Up to three variables in a dataset may be plotted. Combined with DataZoo's dataset integration feature, users are able to compare variables from multiple datasets as well. This feature allows users to quickly get an overview of the general shape and trends of the dataset, and is essential for quickly identifying interesting or anomalous data points. Plots can be saved after they are generated, allowing users to share particularly informative results with each other.

### Management Interface

DataZoo provides interfaces for uploading and describing datasets, as well as managing cruise information, personnel, and metadata dictionaries. A multi-tiered authorization system allows registered users to manage datasets on an individual, project, or global basis. Several management tasks are available. New datasets can be created and described, or existing datasets can be updated. Data can be uploaded in CSV format. Dictionary entries, such as attributes and units, can be added, edited, or removed. Cruise information may be recorded, as well as cruise participants. Data and metadata can even be published to outside repositories and archives with a push of a button.

### Behind the Zoo: System Architecture

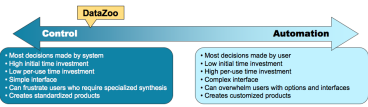


The diagram shows a layered architecture. At the top are 'Data Tables' and 'Metadata Tables'. Below these are 'Database Interface Classes'. A 'Direct Query' path leads to 'Data Objects'. 'Data Objects' are linked to 'Data Model Abstractions'. Below this is 'PHP Logic', followed by 'JavaScript Interface Libraries'. At the bottom are 'Users - Modules - Repositories'.

---

### Our Design Philosophy: Control vs. Automation

Presenting options to the user is a balancing act for any interface - too many options and the interface is overwhelming, too few and users cannot get what they want. In DataZoo, there is a strong element of user control. Automation is used to present suggested actions and prevent critical errors, but in general a more complex interface has been adopted to provide a more powerful system for the user.



The spectrum shows 'Control' on the left and 'Automation' on the right. Under 'Control' are: Most decisions made by system, High initial time investment, Low per-use time investment, Simple interface, Can frustrate users who require specialized syntaxes, Creating standardized products. Under 'Automation' are: Most decisions made by user, Low initial time investment, High per-use time investment, Complex interface, Can overwhelm users with options and interfaces, Creates customized products.

### What is DataZoo?

DataZoo is a multi-project data repository and publishing system currently in use by CalCOFI-SIO, LTER-CCE, and LTER-PAL. DataZoo provides a web-based interface to a database storing oceanographic data, metadata, and study information. DataZoo allows public users to search and retrieve data on a project, cruise, or dataset basis. Associated tools allow users to refine searches to a specific station or region, view multiple cruises as a timeseries, or compare data across projects, such as the CCF LTER project. Data providers can use the system to manage their own data and metadata through a user-friendly interface. DataZoo has been developed by the Ocean Informatics team collaboratively with community participants. It is an open source, open architecture using Apache, MySQL, PHP, JavaScript, and XML.

<http://oceaninformatics.ucsd.edu/datazoo>

### Our Design Philosophy: Owner-Managed Datasets


The goal of the DataZoo interface is to create a translation between a powerful and complex relational database architecture into a simple series of web forms that require little or no technical expertise. In addition to allowing information management personnel to work more efficiently, this enables data providers to edit the data and metadata for their own datasets. This direct access to the data system management ensures metadata can be provided by knowledgeable sources and prevents the information management component from becoming a bottleneck. It is critical to identify which tasks can benefit from direct owner management, and which tasks are best handled by the information management team.

---

### Data Description

A major goal of DataZoo is to provide well-described data. The DataZoo architecture provides a structure with required, suggested, and optional descriptors. These descriptors are drawn from dictionaries of terms, providing a common metadata language. Dictionaries are critical in the development of common vocabularies that prevent DataZoo from being overwhelmed by synonymous but inconsistent terms. This makes the browsing interface less cluttered for users and enhances the machine readability of the metadata for automatic transactions.

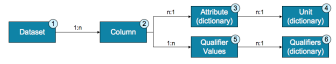
### Behind the Zoo: EML Standard



The DataZoo metadata schema is based upon an XML standard for ecological metadata exchange called the Ecological Metadata Language, or EML. Internally, DataZoo's schema is an augmented version of EML, with support for additional fields. Translation functions allow this schema to be delivered as XML fully compliant with the EML 2.0.1 specification. By enabling all datasets to be output as EML, DataZoo becomes part of a larger distributed system of repositories and archives. Translations from EML to other standards, such as PDK, further extend the interoperability of DataZoo.

### Our Design Philosophy: Qualifiers

Qualifiers allow data providers to annotate their datasets with an unrestricted vocabulary. This departure from the practice of controlled vocabularies was necessary to balance structure with flexibility. To provide a measure of consistency, qualifier terms used by previous users are presented as suggested, although not required, values.



The diagram shows a 'Dataset' containing 'Columns'. A 'Column' has an 'Attribute (dictionary)' and 'Qualifier Values'. A 'Qualifier' has 'Qualifier Values' and is linked to a 'Unit (dictionary)'. Metadata begins at the dataset level (1) with broad description, methods, and ownership information. Each dataset comprises many columns (2). Columns contain like metadata themselves, but each is described by a single attribute (3), drawn from an attribute dictionary. Attributes are modeled after the EML attributes (type and domain storage, representation, etc.). Qualifiers (4) are also associated with any number of qualifier values (5). Each qualifier value provides information about a specific qualifier (6). Qualifier values are entered freely using a suggestion interface to encourage consistency. A single qualifier may have multiple values for given columns, or no value.

---

### Behind the Zoo: The LTER Unit Registry

The first dictionary implemented for DataZoo was the unit dictionary, based off the LTER Unit Registry. The Registry is a database of unit definitions, including information on standards and conversion between units. Many ecological research sites submit locally defined units to the Registry and download units defined by other sites, creating a controlled vocabulary for cross-site interaction. DataZoo's unit dictionary is compatible with the Unit Registry and provides the source for CalCOFI, LTER-CCE, and LTER-PAL submissions. Future plans include building a new shared registry at the attribute level.

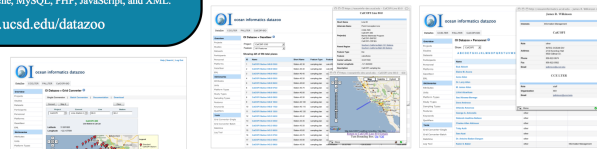
### Behind the Zoo: Code Libraries

DataZoo and its associated tools are built upon a foundation of common code. Some of this code is developed by the Ocean Informatics team to meet our specific needs. Other libraries are maintained and distributed by third parties, such as:

- Yahoo! User Interface API (JavaScript)
- Google Maps API (JavaScript)
- Google AJAXSLT Library (JavaScript)
- JGraph Graphing Libraries (PHP)

### Tools

DataZoo hosts a number of tools that have been created in either direct or indirect support of its main function as a data repository and publishing system. Outside tools have also been designed (or redesigned) to integrate with DataZoo.



The Gazetteer is a dictionary of locations important to the CalCOFI sampling plan. This information can be linked to via a PHP API to store project personnel, cruise participants, and dataset contributors.

The Ocean Informatics Personnel Database is an independent application that is integrated with DataZoo studies and datasets to provide a geographic context for the data and metadata stored in DataZoo.

---

2010-11-17

### Our Design Philosophy: Modular Design

DataZoo is a complex application that is continually developing. In order to add features in a timely fashion, DataZoo and other supporting applications have been designed in a modular fashion. Although the elements of DataZoo are integrated with each other, they are designed to operate independently as well. This modular design allows the logic of one DataZoo module to be upgraded without impacting the rest of the system, provided that the inputs and output to that module remain the same. Likewise, other Ocean Informatics applications that interact with DataZoo do so through abstracted layers, allowing the mechanisms on either side of the layer to change without disrupting any other integrated applications. These abstracted interfaces return results in a non-specialized format such as XML or as standard class objects. Modules interface through three types of abstraction:

- Programmatically accessible URLs, often used for AJAX queries
- PHP and JavaScript APIs
- PHP classes derived from a standardized template



18

Poster date: 2007-11-17

Title: CalCOFI Data Management: Developing Community Standards

Description: CalCOFI represents a partnership of multiple agencies conducting quarterly joint oceanographic cruises, CalCOFI field team members work as a cohesive cross-agency unit to accomplish the cruise goals.

Authors: James Wilkinson, Karen Baker, Rich Charter



## CalCOFI DATA MANAGEMENT: DEVELOPING COMMUNITY STANDARDS



**James Wilkinson<sup>1</sup>, Karen Baker<sup>1</sup>, and Richard Charter<sup>2</sup>**

<sup>1</sup> Scripps Institution of Oceanography, La Jolla, CA 92093 E-mail: jwilkinson@ucsd.edu, kbaker@ucsd.edu  
<sup>2</sup> NOAA Southwest Fisheries Science Center, La Jolla, CA 92037 E-mail: richard.charter@noaa.gov

### Introduction

CalCOFI represents a partnership of multiple agencies conducting quarterly joint oceanographic cruises. CalCOFI cruise participants work as a cohesive cross-agency unit to accomplish cruise objectives. Ancillary researchers frequently integrate their field measurements and sampling with the long-term core CalCOFI measurements and samples. Once a cruise concludes, however, this cohesive unit disperses; individuals return to their respective agencies and labs to process samples and analyze data. Each group uses legacy, lab or agency specific methods and software to generate data products in local formats. These diverse data processing methods, products, and storage formats create challenges for merging final datasets. Development and incorporation of shared data management practices and joint community standards enable data integration.

### CalCOFI-SIO Data Process



### Developing Data Integration Standards

**Typical data flow from field collection to publication**

- \*See: Water samples are collected, logged & analyzed. Net tows collected, logged, & preserved.
- \*Standard 1: all logs use joint standard formats with common station, event number & order occupied indexes.
- \*See: Preliminary data processing and quality control of individual samples types: salinity, nutrients, oxygen, chlorophyll.
- \*Standard 2: event logs, sample logs, & analytical output files are available on the network, all include common sample indices.
- \*Ashore: traditional data processing, merging of individual data types into a combined, local ascii format using in-house software. Preliminary data are merged, quality control protocols are applied. Final completion and data publication. CalCOFI Data Report as .txt, .pdf, and .html; contour plots, .EDL files (proprietary report format), all web accessible.
- \*Standard 3: Station & cast information, bottle, and CTD data are merged into a non-proprietary csv with common, queryable elements, standard formats, & labels.
- \*Ashore: Net tow data is processed with the flow meter calibrations and depth of tow, volume of water strained and other variables are calculated and added to the net tow dataset. The tows logs are then volume.
- \*Standard 4: A plankton volume report is generated with common elements, standard formats and labels.
- \*Ashore: The plankton samples are then sorted removing all fish eggs, larvae and squid paralarvae. The major species (genus, species, and larva) are identified and saved at the time of sorting. The remaining plankton sample goes to the SIO archives. The identified eggs and larvae are identified in the ichthyoplankton identification log. The fish eggs and larvae are then archived in the NMFS ichthyoplankton archives.
- \*Standard 5: An eggs and larvae dataset is produced in a standard format. An annual ichthyoplankton data report is produced in a standard format once all the cruises of the year have been identified.

### Establishing Shared Practices

Identifying and establishing common, queryable columns, such as *order occupied* and *event number*, and including them in final data products allows heterogeneous datasets to be related. In addition, standardizing data elements such as column headers, date-time specifications, spatial designations such as GPS decimal format are easy to implement with minimal impact on existing data production. Standard, linkable data elements allow ingestion into relational databases, applications, and other analytical tools such as Data Zoo using import templates.

**CalCOFI Standardization Strategies:**

- **Persistent vocabulary and formats** with defined standard data column labels
- **Date & position format conventions**  
example: YYYYMMDD HHMMSS.S UTC  
example: 32.5345,-117.2343
- **Standard Line Station grid designations**  
example: line 93.3, station 120.0
- **Order-occupied numbering** for sequential stations
- **Event numbers** as needed for distinguishing sta activities
- **Data distribution in non-proprietary format**  
example: CSV in addition to legacy IEH
- **Metadata** - definitions of measurements & equipment; translation tables for different unit attributes

### CalCOFI-SWFSC Data Process



### Cross-Project Data Interfacing

CalCOFI cruises generate multiple data formats such as station data, continuous meteorological, ADCP, & SCIMS data, surface & marine mammal visual observations and satellite recordings. Each research group has their own data publishing goals. It must be each group goal to generate a standard product with common indices for use by the data community. CalCOFI-SIO & CalCOFI-SWFSC are establishing a common vocabulary and standardizing final data formats to hydrographic, zooplankton, and ichthyoplankton data can be integrated.

**Data Interfacing Strategies:**

- **Establish a shared data product**  
Consider your final data and what you are able to share with the data community – some data processes take longer.
- **Develop a standard, persistent format**  
Cross-project partners can plan for a consistent data format and design ingestion mechanisms such as import templates.
- **Think collaboratively**  
The Ocean Informatics team is working together to automate the importing of CalCOFI data into DataZoo, a cross-project, web-based, information system.

### Shared Practices Begin in the Field

With quarterly cruises generating a persistent influx of data, the CalCOFI technical team must maintain an established routine to keep pace. Changes in procedure or protocol impact the expediency of the ongoing process. To minimize the impact of new data integration practices, the change process best begins at sea. Careful attention to sta activities & event logs create both a shared index and initiates a dialogue about organizational design.


2007-11-17

Poster date: 2007-11-17

Title: CalCOFI Local Metadata: Augmenting the Ecological Metadata Language

Description: Metadata is an integral and necessary part of data sharing; the enactment of a metadata standard not only guides the creation of local metadata documents but is also a link between local and broader communities.



Authors: Lynn Yarmey, Karen Baker, James Connors



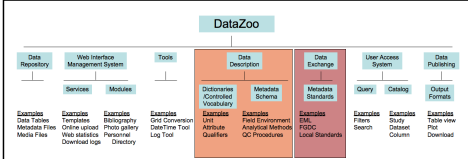
## Local Metadata: Augmenting the Ecological Metadata Language

Lynn Yarmey, Karen Baker and James Connors

Scripps Institution of Oceanography, University of California San Diego

**Abstract:** Metadata is an integral and necessary part of data sharing; the enactment of a metadata standard not only guides the creation of local metadata documents but is also a link between local and broader communities. A full metadata record, including but not limited to descriptions of the field environment, detailed accounts of analytical methods, and summaries of quality control procedures, is essential to the understanding and use of any dataset. Without the context of the data, measurement values are subject to misinterpretation and misuse. *Standardized* metadata functionally makes possible automated comparisons and visual presentation of datasets. In addition to establishing a local foundation for data sharing, a standard becomes an integrative bridge when developed in parallel with community and national standards.



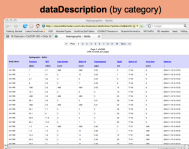
**DataZoo:** a web-delivered local information system consisting of a relational database for metadata and data, designed in collaboration with CalCOFI-SIO and CalCOFI-SWFSC, Palmer Station and California Current Ecosystem Long-Term Ecological Research (LTER) program participants. Design by the Ocean Informatics team supports multiple projects with a modular oriented architecture (MOA) featuring:

- a community data repository
- a data publishing system for data collectors
- a data access system for data users
- a set of web management interfaces
- a data description schema
- a data exchange capability
- a set of associated data tools

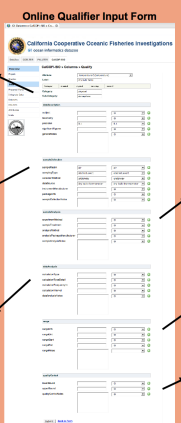
### Local Data Description DataZoo Metadata System

Metadata begins at the dataset level (1) with broad description, methods, and ownership information. Each dataset comprises many columns (2). Each column is described by a single attribute (3), drawn from an attribute dictionary. Attributes are modeled after the EML attribute module and contain storage, representation, and definition information. Each attribute contains a unit (4) from the unit dictionary, based on the LTER EML Unit Registry. Columns are also associated with any number of qualifier values (5). Each qualifier value provides information about a specific qualifier (6). Qualifier values are entered freely, using a suggestion interface to encourage consistency. A single qualifier may have multiple values for a given column, or no value.

**dataDescription (by category)**




**Online Qualifier Input Form**




**\*\*\* An Invitation \*\*\***

Do these qualifiers allow you to describe your data well?  
We invite your suggestions for additional qualifiers!

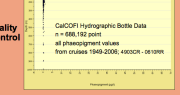
**sampleAnalysis**  
(includes any analysis, testing done on collected samples)




**sampleCollection**  
(includes whether the data is from a sample or a reading, any instrumentation, the earth realm of the sample, etc)



**range(s)**  
(includes age, flow rates, depth ranges, etc)




**quality Control**



### Community Data Exchange EML Ecological Community Metadata System

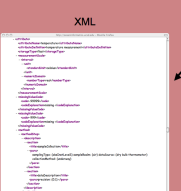
The Ecological Metadata Language (EML) provides a metadata specification with growing acceptance in environmental science communities. The DataZoo Information System application metadata schema, in being designed for optimized local data description and queryability, represents an expansion upon the granularity of expected EML data description.

**Adaptation to EML**



How we fit our locally optimized data descriptions (i.e. qualifiers) into EML.


**XML**



What metadata really looks like (how it is stored and sent).

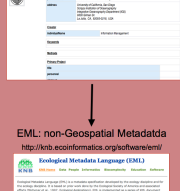
**Data Discovery Using EML**

<http://knb.econinformatics.org/>



Standardized EML allows for automated query by the larger community and public.

**EML translates to FGDC (Geospatial metadata standard)**



Standardized EML can be translated into other metadata standards.

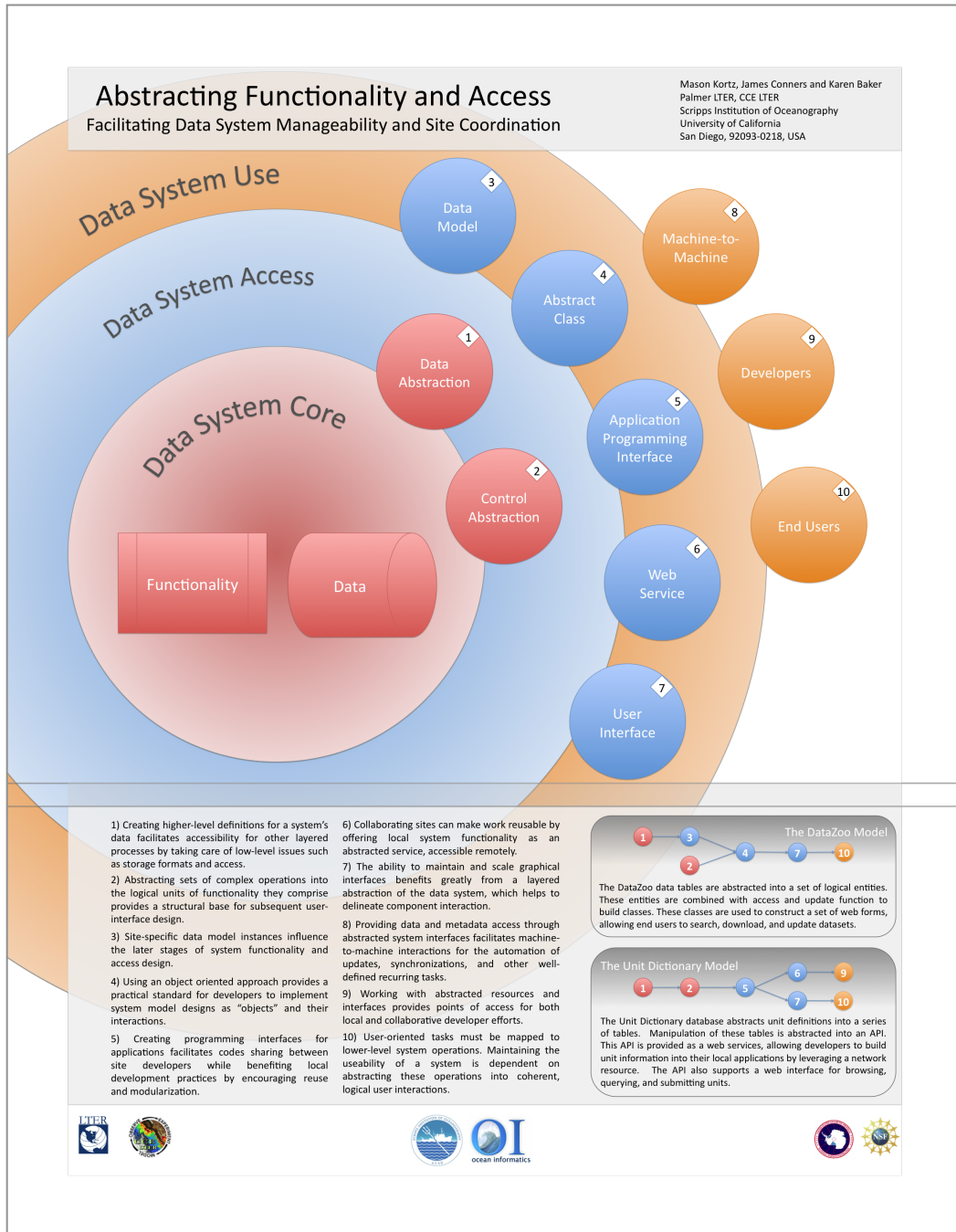
2007-11-17

Poster date: 2008-08-10

Title: LTER Abstracting Functionality and Access: Facilitating Data System Manageability and Site Coordination

Description: As the functionality of site data systems increases, frequently so does the complexity. Organizing system functionality through distinct layers of abstraction, from low-level system access to high-level user access, is key to maintaining a manageable set of systems.

Authors: Mason Kortz, James Connors, Karen Baker



Poster date: 2008-08-10

Title: LTER Information Managers: A Community of Practice

Description: Communities of Practice are groups of people who share a concern or a passion for something they do and who want to learn more about how they do it. Such a community is more than a group of people having the same job or a network of connections between people.

Authors: Karen Baker, Nicole Kaplan, Inigo San Gil, Margaret O'Brien, Florence Millerand



## LTER IMC Community of Practice: A Learning Environment



Karen Baker<sup>1</sup>, Nicole Kaplan<sup>2</sup>, Inigo San Gil<sup>3</sup>, Margaret O'Brien<sup>4</sup>, Florence Millerand<sup>1</sup>, Lynn Yarmey<sup>1</sup>  
<sup>1</sup>PAL & CCE LTER; <sup>2</sup>SGS LTER; <sup>3</sup>LNO; <sup>4</sup>SBC LTER

### What are communities of practice?

"Communities of practice (CoP) are formed by people who engage in a **process of collective learning** in a shared domain of human endeavor, e.g., a group of engineers working on similar problems, a band of artists seeking new forms of expression, a network of surgeons exploring novel techniques, a gathering of first-time managers helping each other cope..." (Wenger, 2008).

A community of practice is not merely a group of people having the same job or a network of connections between people. Communities of practice are groups of people who share a concern or a passion for something they **do** and **learn** how to do it better as they interact regularly.

### Three salient characteristics of CoP:

- 1. The domain:** the CoP identity is defined by a shared domain of interest. Membership implies a **commitment** to the domain, and therefore a **shared competence** that distinguishes members from other people.

*LTER Case*

Research domain: ecology, information science, computer science  
Practice domain: informatics  
Development domain: sociotechnical systems design, collaborative design  
Communication domain: communication studies, science & technology studies, infrastructure studies
- 2. The community:** Members build **relationships** that enable them to learn from each other through engaging in joint activities and discussions, and information sharing.

*LTER Case*

Relationships: working groups, LTER Information Management Committee, site-site, site-network, LTER Network Information System Advisory Committee  
Activities: annual LTER IMC meetings, best practices, collaborative design of modules & tools
- 3. The practice:** A community of practice is not merely a community of interest. Members are practitioners - they are **engaged in doing the work**.

*LTER Case*

Data gathering, data organizing, data describing, data preparing, quality control, data analysis, data synthesis, data exchange, data processing, IT evaluation, informatics research, technology development, assessment, informatics research, federation inquiry, community-building, remote sensing, site-network coordination

And more

Working groups, module development, prototyping, articulation, negotiation, knowledge mediation, standards-making, infrastructure-building, informal and formal communication facilitation

### Why our community of practice is important!

We have an organizational structure and a way of working that supports communications, a social organization for mentoring, a learning environment, a strategy for standards building, a mechanism for comparative analysis of experiences, a group identity...and more!

### With this foundation, how do we approach cyberinfrastructure and change?

**References**

Wenger, E. (2008). Communities of practice: a brief introduction. Online Guide: <http://www.ewenger.com/theory/index.htm>  
Karasil, H., Baker, K. (2004). Infrastructure for the Long-Term: Ecological Information Management. XXXX  
Wenger, E., McDermott, R., Snyder, W. (2002). Cultivating communities of practice: a guide to managing knowledge. Harvard Business school Press.  
Wenger, E. (2001). Supporting communities of practice: a survey of community-oriented technologies. <http://www.ewenger.com/tech>  
Wenger, E. (1998). Communities of practice: Learning, meaning, and identity. Cambridge: University Press.


EIMC 2008

Poster date: 2008-08-10

Title: LTER Information Infrastructure: Emergent Roles, Responsibilities and Practices


Description: Human activities together with technical elements and collective practices are core elements for growing local infrastructure as well as for bridging with other communities and networks. Site information management activities create a shared data curation opportunity.

Authors: Lynn Yarmey, Karen Baker



### Local Information Management and Information Infrastructure: Roles, Responsibilities and Practices

Lynn R. Yarmey and Karen S. Baker, Ocean Informatics, PAL and CCE LTER  
Scripps Institution of Oceanography, UCSD, 92093-0218, USA



**Abstract:** Human activities together with technical elements and collective practices are core elements for growing local long-term infrastructure as well as for bridging with other communities and networks. Site information management activities create a shared data curation experience where data curation refers to managing the capture, use and preservation of the data. Identifying local data activities opens up the complex set of arrangements that comprise site information management, including the variety of roles emerging to address mediation and collaboration. This articulation strategy may be seen as a preparatory step for conscientiously designing an effective data network.

Information management covers many diverse and overlapping realms, each realm has multiple roles and associated activities. Data curation is a link between the realms and is supported by the necessary foundation of both a well-maintained infrastructure and a complete information management strategy.

Information Management

**Roles we balance**

**Data Capture**

- Remote and Manual Sampling
- Informal Knowledge Sharing
- Traditional Sampling
- New Procedures
- Formal Training
- Event Logging

---

**Informatics Capture**

- Metadata Capture
- Data Transport
- Attribute Filtering
- Reformatting
- Data Centralizing
- Standardizing
- Organizing

---

**Use**

- Data Downloads
- Query and Filter
- Plot Creation
- Website Display
- Automated Data Dumps

---

**Preservation**

- Data Archival
- Papers
- Posters
- Presentations
- Press Releases
- Books (Scholarly, Children's)
- Informal Talks/Conversations

Curation Activities

Role	Activity	Responsibility
Research Planner	<b>Scope and System Design</b>	Formulate research question and determine scope and purpose. Methods and strategy development, quality assessment.
Acquirer	<b>Gather</b>	Acquisition (capture/record), quality assessment and calibration.
Scientific Assessor	<b>Scientific Assess/ Appraise</b>	Determine relevance to planned scientific research.
Informatics Assessor	<b>Informatics Assess/Appraise</b>	Consider relation to broader community efforts including standards-making.
Metadata Provider	<b>Describe</b>	Description (definition and environmental contextualization).
Receiver	<b>Ingest</b>	Ingestion to build collection (capture, translation, organization and registration).
Mediator	<b>Process</b>	Quality control.
User	<b>Use/Reuse</b>	Consider for original question (local user) or new questions and issues (remote user).
Analyzer	<b>Transform</b>	Create derived products or symbols.
Provider	<b>Deliver</b>	Enable public and/or other repository access through exchange or web access.

Infrastructure Management

**Dimensions we integrate**

**Social Infrastructure**

- Information Management Team
- Community Requirements
- Community Meetings/Events
- Best Practices/Protocols
- Metadata Standards
- Design Expertise
- Data/Quality Standards
- Collaborative Partnerships

---

**Organizational Infrastructure**



- IM Committee
- Network Office
- University Affiliation
- Internet/Email Services
- Computational Networking
- Data Lifecycle Support
- Data Repository
- Software Access Preservation
- Long-term System Maintenance

---

**Technical Infrastructure**

- Data Management Expertise
- Scientific Instrumentation
- Data Ingestion Forms
- Information System
- Hardware Updating
- System Administration
- Desktop Support
- Website Interface
- Security Enforcement
- Poster Printer

**References:**  
 Baker, K.S. and F. Millerand. Articulation work supporting information infrastructure design: coordination, categorization, and assessment in practice. Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS37), January 03-06, 2007. IEEE Computer Society, Washington, DC.  
 Baker, K.S. and L.R. Yarmey. Data Stewardship: Environmental Data Curation and a Web-of-Repositories. Proceedings of the 4th International Digital Curation Conference, December 1-3, 2008, Edinburgh, Scotland (submitted).  
 Karant, H. and K.S. Baker. Digital Data Practices and the Long Term Ecological Research Program Growing Global. International Journal of Digital Curation (in press).  
 Karant, H. and K. Baker. Infrastructure for the Long-Term: Ecological Information Management. Hawaii International Conference for System Science Proceedings, of the 37th Hawaii International Conference on System Sciences. HICSS38. IEEE Computer Society, January 2004. Big Island, Hawaii, 2004.  
 Jackson, S. J., Edwards, P. N., Bowker, G. C., and C.P. Knobel. Understanding Infrastructure: History, Heuristics, and Cyberinfrastructure Policy by First Monday, volume 12, number 6 (June 2007). URL: [http://firstmonday.org/issue/view/full\\_article.php?id=12\\_6\\_jackson/index.html](http://firstmonday.org/issue/view/full_article.php?id=12_6_jackson/index.html)





Poster date: 2008-08-10

Title: Scientific Communication and Information Infrastructure


Description: Scientific communication is central to collaborative scientific endeavors. A shared information infrastructure facilitates communication and collaboration. Digital information infrastructure occurs in multiple forms. The poster presents examples from CCE LTER.

Authors: Karen Baker, Beth Simmons, Ryan Rykaczewski, Alison Cawood, Peter Davison, Moira Decima, Melissa Garren, Andrew King, Andrew Taylor, Jesse Powell, Melissa Soldevilla, Mike Stukel



## Scientific Communication and Information Infrastructure

Karen Baker, Beth Simmons, Ryan Rykaczewski,  
Alison Cawood, Peter Davison, Moira Decima,  
Melissa Garren, Andrew King, Andrew Taylor,  
Jesse Powell, Melissa Soldevilla, Mike Stukel



CCE LTER

Scientific communication is central to collaborative scientific endeavors. A shared information infrastructure facilitates communication and collaboration. Digital information infrastructure occurs in multiple forms.

### DataZoo


DataZoo is an information system designed for publishing data of the CCE LTER Community. It is one part of the data curation process.

Participants: Andrew King, Moira Decima, Mike Stukel

URL: <http://oceaninformatics.ucsd.edu/datzoo/cce>  
URL: <http://cce.lternet.edu>

Data Publishing

### Children's Book



A Scientific Detective Story

Written by Mary Corbett and Beth Steiner  
Illustrated by Kristin Curran

Story Publishing

### Design Studio

The Ocean Informatics Design Studio provides a meeting place for students, staff, and researchers to consider data storage, access, and delivery together, jointly, as codesigners of an information system.

Participants: Ryan Rykaczewski, Jesse Powell, Moira Decima, Andrew King

Community Designing

### Picture Gallery


A picture helps capture the context of scientific measurement activities - whether they are pictures-of-the-day, a gallery of autotrophs or of marine mammals. Making the photos available in a shared infrastructure means they are available to the community and the public for browsing and as a need arises rather than upon request.

Participants: Andrew Taylor, Melissa Soldevilla

Multi-Media Publishing

### Referencing

Citations from Bibliography Module  
Ehlers, Eklund, & Munk, 2007 Open Access Proceedings Tracks on Ocean Informatics Environment, ICSOBI  
Ehlers and Eklund, Informatics for the Long-Term: Ecological Information Management: Special International Conference on System Science Proceedings, ICSOBI, IEEE Computer Society, 2004  
Ehlers, Eklund, Munk and Eklund, Informatics for the Long-Term: Ecological Information Management: Ways of Knowing in a Networked Environment. In International Handbook of Science Research, Knowledge, Methods and IT Strategy (eds. Springer)




### Gizmo

Gizmo is an application that enables teleconferencing for individuals and groups using desktop computers and the internet.

Participant: Ryan Rykaczewski

URL: <http://cce.lternet.edu/aboutus/communication>

Distance Communicating




### Picture-of-the-Day

During a remote field study such as an oceanographic cruise, daily contact provides a uniquely local real-time point of engagement for students, classrooms, teachers, staff, family and friends.

Participant: Andrew King

Real-Time Field Experience




### Pier Walks

When CCE LTER research program participants walk on the SIO pier with students and public, a unique marine science experience is created.

Participants: Andrew Taylor, Melissa Garren


Local Field Experience



### Event Logging

The Event Logger is one of the initial elements of the data stewardship process for shipboard data. Event numbers are provided and logged from a planned list of events but may also be assigned after the cruise when an informal activity benefits from more formal identification (i.e. a targeted net tow).

Data Stewardship & Sampling Design



### CCE Home Page

The CCE LTER web site is a central location from which other information including data and activities may be located.

URL: <http://cce.lternet.edu>

Community Website

2008-08-10

92

Poster date: 2008-11-17

Title: CalCOFI Biological Data Management

Description: An information system designed for working with multiple oceanographic biological data collections is presented. DataZoo is an extensible system that supports data discovery, access, query, and exchange for data such as the CalCOFI integrated biological data.

Authors: Karen Baker, Mason Kortz, James Connors, Lynn Yarmey



## CalCOFI Biological Data Management



Karen S. Baker, Mason Kortz, James Connors, Lynn Yarmey  
 Ocean informatics, Scripps Institution of Oceanography, UCSD, 92093-0218, USA

**Abstract:** DataZoo, an information system designed for working with multiple oceanographic biological data collections, is an extensible system that supports data discovery, access, query, and exchange for data such as the CalCOFI integrated biological data and bottle measurements from hydrographic casts. The system is a data and metadata repository designed to meet the needs of researchers, policy makers and the public. It is a publishing forum that includes a dataset catalog, personnel directory, and metadata system. Dictionaries and controlled vocabularies play a key role and facilitate data integration. The metadata schema takes into account local and community standards including the Ecological Metadata Language, augmenting it with local unit, attribute, and qualifier dictionaries. DataZoo is organized into three web-based functional units: data, resources, and management. A suite of resources extend the information system interface to the desktop so local participants can manage their own data - and in turn consider their individual data practices in relation to a project repository made readily apparent via web interfaces and web services. A community information system creates a data curation commons that highlights shared technical components, organizational arrangements, and collective practices, all central elements to growth of a local information infrastructure able to bridge projects, communities and networks.





### DataZoo

a local information repository

<http://oceaninformatics.ucsd.edu/datazoo>

**What is it?**

- An extensible information system and data repository
- Multi-project publishing system for heterogeneous scientific data
- Collection of multiple datasets of one table identifiable by project & study
- Dataset catalogue with data templates and metadata
- Application supporting data and metadata
- Mechanism for gathering, managing and preserving datasets
- Local approach to data stewardship and design

**What does it do?**

- Provides long-term access to data and metadata
- Enables web data query and data integration
- Supports data and metadata management over the web
- Provides metadata forms and elicitation
- Provides resources such as documentation, dictionaries, and tools
- Represents collections within multi-application Dataspaces

**Who uses it?**

- Local data providers/co-designers: LTER CCE & PAL; CalCOFI SIO & NOAA
- Diverse co-designers/stakeholders: Ocean Informatics community & public

**What's in it?**

- Elements: datasets (>90); studies (>425); personnel (>2500); storage (<1Tb)
- Metadata: collection, temporal, spatial, personnel, methods/analysis, qc
- State: Continuing to grow and redesign

**How is it built?**

- Technology: Apache, MySQL, PHP, PERL, XML/XSLT
- Standards: Templated long-term datasets, EML metadata, local standards
- Data practices: Dictionaries, term sets, augmented attribute qualifier system
- Shared libraries: YUI, JGraph, GoogleMaps
- Design practices: API-based, agile methods, articulation work

**Ocean Informatics Initiative**  
 designing a sociotechnical conceptual framework  
 for a local information environment

**Question: Interested in a tour?**  
 If you are interested in a guided tour of the system, please write below your name, email, and any comments you may have about particular interests.

25

Poster date: 2009-05-13

Title: CCE LTER: An Oceanographic Eventlogger as One Part of an Information Environment  
Description: The CCE LTER initiated at SIO in 2004 enabled launch of "Ocean Informatics", a new approach to design of information infrastructure in support of interdisciplinary science. CCE works synergistically with Palmer Station LTER and with California Cooperative Fisheries Investigations.  
Authors: Karen Baker, Mason Kortz, James Connors

The poster is divided into two main sections. The left section, titled "An Oceanographic Event Logger", is further divided into "A FIELD PERSPECTIVE" and "A DESIGN PERSPECTIVE".

**A FIELD PERSPECTIVE:**

- Field Practices:** An event logger promotes conventions, conversations, & connections at sea between diverse data efforts.
- Purpose:** A logger coordinates activities across groups and with the ship's bridge.
- Methods:** An event log program is installed on standalone or networked computers.

**A DESIGN PERSPECTIVE:**

- Information Infrastructure:** An SIO Ocean Informatics team supports the design, development, deployment, and enactment of effective data practices as part of the infrastructure for a site-based information environment (Baker et al., 2006).
- Design and Development:** Collaborative and continuing design together with classification and information theory contribute to development of event logger features:
  - Configuration files for local flexibility
  - Authoritative lists of shared vocabulary
  - Event numbers as field practices that travel from sea to land
- Acknowledgements:** [List of names]

**Central Diagram:** A central hexagon contains "Event Number" and "GPS Timestamp". It is connected to six surrounding boxes: "Bridge Activities", "Cruise Activities", "Lab Activities", "Macro-Material Activities", "CUES Activities", and "Cruise Event Log".

**Right Section: "An Information Environment"**

- A Design Studio:** The design studio provides a center - a table or community commons - for participatory design. The studio is a place available; time is prioritized for dialogue, investigation, and formalization; it is a place to commentate, wonder and know-it, where there are always chairs, chalkboard, and laser.
- An Information Architecture:** Site-based information management facilitates a reconceptualization of system development as architecting infrastructure, ensuring continuity, and facilitating learning.
- An Informatics Team:** An informatics team has multiple expertise - technologists, database designers, data managers, semantics workers, interface designers, programmers, mediators, analysts, information architects & scientists - creating an integrated information environment.

A large orange arrow points from the "An Oceanographic Event Logger" section towards the "An Information Environment" section.

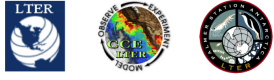


Poster date: 2009-09-14

Title: LTER: A Web of Repositories


Description: The movement and exchange of data are frequently described using a 'flow' or a 'pipeline' model. We differentiate a uni-directional data 'flow' from an alternative model, a web-of-repositories. A web-of-repositories is a federation of diverse nodes.

Authors: Lynn Yarmey, Karen Baker



## A Web of Data Repositories

Lynn Yarmey and Karen Baker  
Scripps Institution of Oceanography, UCSD, La Jolla, CA



---

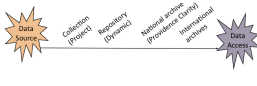
**Abstract**

The movement and exchange of data are frequently described using a 'flow' or a 'pipeline' model. We differentiate a uni-directional data 'flow' from an alternative model, a *web-of-repositories*. A *web-of-repositories* is a federation of diverse nodes where communication, connections, and data exchange are multi-directional. Each node has a unique *sphere-of-context* with technical, organizational and social dimensions. In this poster we explore a multi-repository data landscape.


**Data stewardship and data curation**

Data stewardship and data curation are neither problems to be solved nor solutions in and of themselves. Rather these concepts represent dynamic learning arenas. Through context-aware data curation, the possibility arises of designing, federating and sustaining an interoperable "*web-of-repositories*", fulfilling the ultimate goal of data stewardship (Baker and Yarmey, in press).

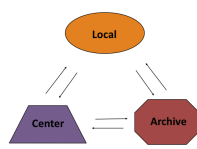
---



A linear view of various stages in the data flow from the data source (collection by field personnel) to data availability to the scientific community and access to the public. The simplistic pipeline model does not reflect the ongoing access and interdependent work at each point along the assembly line.



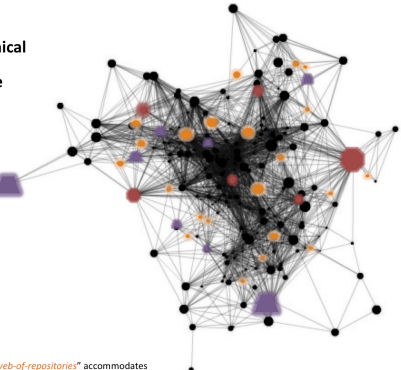
A hierarchical view of data flow where access occurs all along the line or during the life cycle of the data. While the relationships between repositories are displayed and the distributed access responsibility is shown, the figure implies a hierarchy of repositories and nests data activities.



An exemplar web of three representative repository types. The local, center and archive shown here can be thought of in parallel with the NSF data collection types of research, resource and reference (National Science Board, 2005). Repository differences are reflected in what may be described as repository goals that serve different audiences with differing stakeholder interests that define local task outputs: local management is attuned to data use for planned research, centers to current data reuse within the discipline, and archives to future data reuse. Note that the arrows indicate bidirectional exchange.

### Web-of-Repositories

- Federated
- Non-hierarchical
- Collaborative
- Diverse
- Inclusive
- Flexible
- Distributed
- Coordinated
- Partnered
- Sustainable



Conceptualizing a "*web-of-repositories*" accommodates a variety of repository types and represents an ecologically inclusive approach to data curation. One strength of a web model is the inclusion of a wide range of activities and feedback loops. The web portrays in particular a non-hierarchical set of pathways that represents the complex set of data flows that occur in practice. The web as an element of the infrastructure requires new understandings of scientific practices, data practices, and curation practices as it enables distributed collective practices (King, 2006), scientific data collections (SDC), and the "conceptualizing of SDCs as distributed in nature and practice" (Cragin and Shankar, 2006).

**References**

Baker, K. & Yarmey, L. Data Stewardship: Environmental Data Curation and a Web-of-Repositories. IDC, in press.

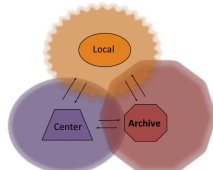
Cragin, M.H. & K. Shankar (2006). Sovereign Data Collections and Distributed Collective Practice. *CSCW* 13:185-204.

King, J.L. (2006). Modern Information Infrastructure in the Support of Distributed Collective Practice in Transport. *CSCW* 13:111-121.

NSB (2005). National Science Board: Long-lived Data Collections: Enabling Research and Education in the 21<sup>st</sup> Century. NSF NSB-05-40.

### Spheres-of-Context

The measurement 'context' refers in part to the properties of the broader physical environment in space and time and is recorded in the accompanying metadata. The context (and thus the metadata) includes the technical and social environments composed of instruments, people, traditions and organizational entities associated with obtaining the measurement as well as the later processing, storage, use and reuse of the resulting data. In practice, data curation can be imagined as a shifting contextual window. From a multiple repository view, these contextual windows may be described as interrelated *spheres-of-context* (Baker and Yarmey, in press).



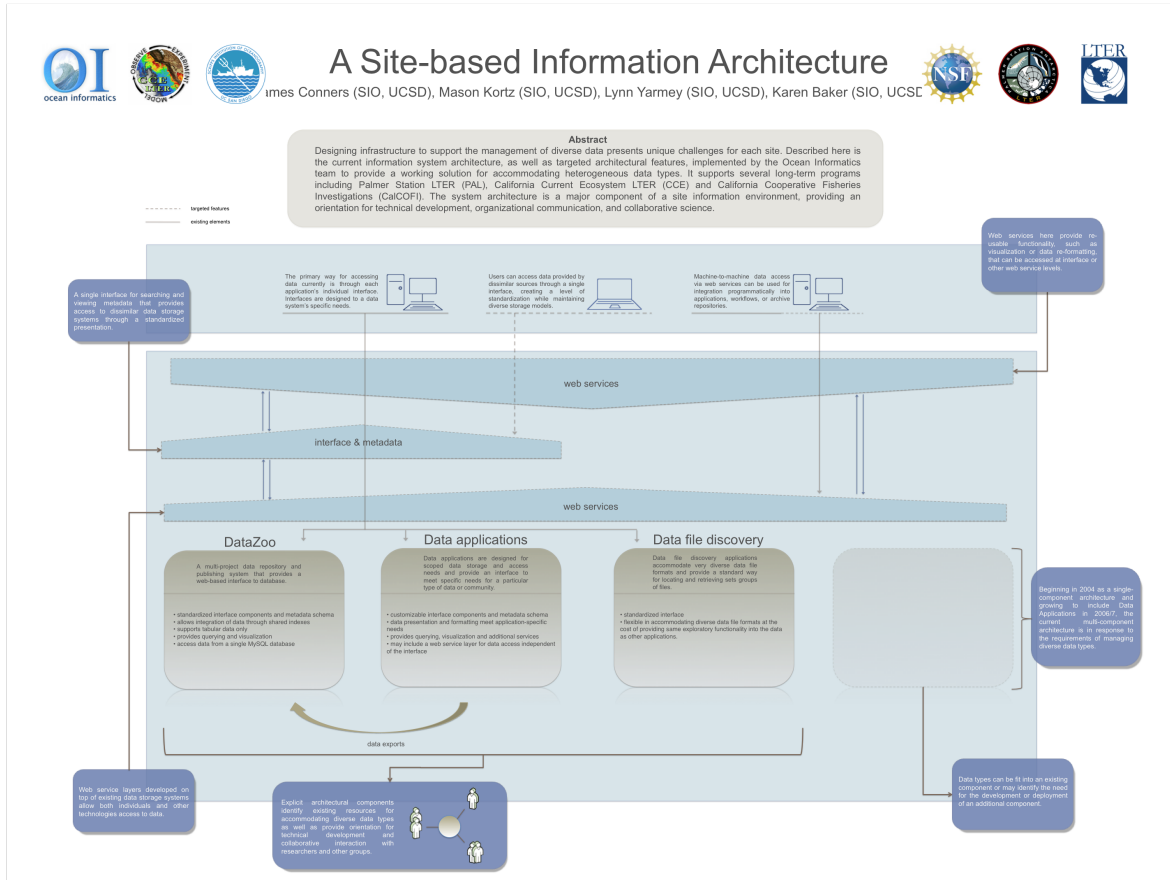


Poster date: 2009-09-14

Title: PAL & CCE LTER: A Site-Based Information Architecture

Description: Designing infrastructure to support the management of diverse data presents unique challenges for each site. Described here is the current information system architecture, as well as targeted architectural features, implemented by the Ocean Informatics team.

Authors: James Conners, Mason Kortz, Lynn Yarmey, Karen Baker

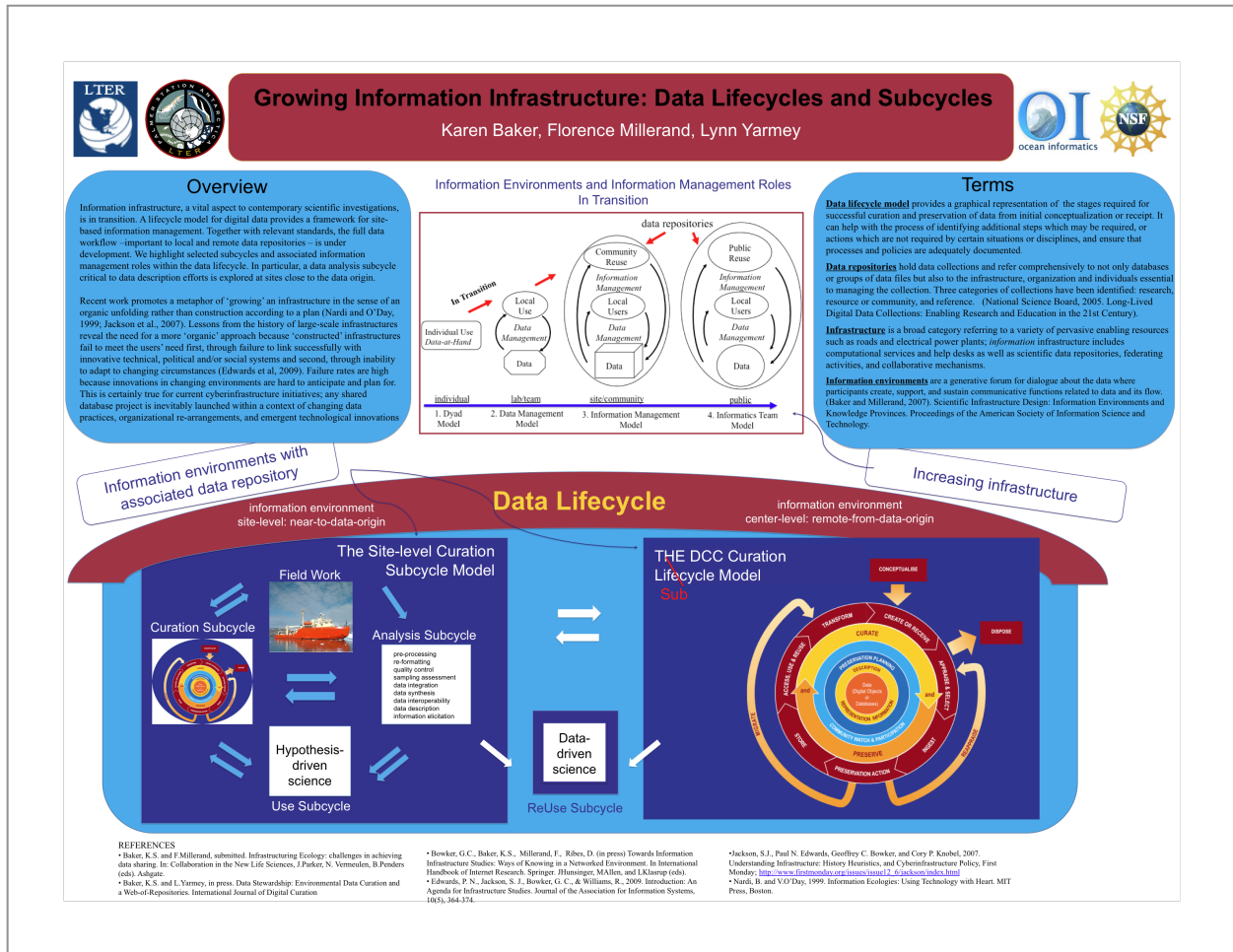


Poster date: 2009-09-14

Title: LTER Growing Information Infrastructure: Data Lifecycles and Subcycles

Description: Information infrastructure, a vital aspect to many contemporary scientific investigations, is in transition. A lifecycle model for digital data together with plans for standards provide a framework for site-based information management

Authors: Karen Baker, Florence Millerand, Lynn Yarmey



Poster date: 2009-09-14

Title: LTER Unit Dictionary & Unit Registry

Description: Units of measurement are a fundamental element of scientific discourse and data integration. The LTER Unit Working Group has developed two initiatives to promote consistent use of units throughout the network including the Unit Dictionary.

Authors: Mason Kortz, Lynn Yarmey, James Conners, Todd Ackerman, Karen Baker



LTER Information Management Committee  
Unit Working Group Projects



## Unit Dictionary & Unit Registry

Mason Kortz<sup>1</sup>  
Lynn Yarmey<sup>1</sup>  
James Conners<sup>1</sup>  
Todd Ackerman<sup>2</sup>  
Karen Baker<sup>1</sup>

(1) PAL & CCE, SIO, UCSD, La Jolla, CA  
(2) NWT, CU-Boulder, Boulder, CO

### Unit Dictionary

The Unit Dictionary comprises the list of scientific units in use by the LTER network, the practices involved in managing these units, and the emergent standards governing their use.

The intent of LTER Unit Dictionary project is both to preserve site autonomy by permitting use of site-specific units and to facilitate standardizing activities at the same time. Thus the barrier to entry for content is low, but there are guidelines and processes for review specified in the practices as well. A community dictionary entails the need for an accepted interpretation of use of the dictionary in particular situations.

The Unit Dictionary addresses an important step in integrating diverse data. When units are described and recorded following common practices, it is possible to convert and evaluate interrelations of measurements. Moreover, such conversions are not single-use or site specific but are generally applicable to the wider network.

#### Quantities, Units and Attributes

Quantities, units, and attributes are all used to describe scientific measurements. The Unit Dictionary builds on the SI concept of quantities, and builds towards the idea of an Attribute Dictionary. The following definitions show the relationship between these concepts:

**Quantities:** A quantity is a way of describing a type of phenomenon. The quantity speed, for example, describes movement over a distance in a period of time.

**Units:** A unit is a particular instance of a quantity. All units of a quantity measure the same type of phenomenon – for example, meter per second and foot per year both measure speed, and grams per liter measures mass density.

**Attributes:** An attribute is the representation of a specific phenomenon in a specific unit. Grams of carbon per liter and air speed of a seagull in meters per second are examples of attributes.

### A Living Dictionary

The Unit Dictionary is a living dictionary, designed to grow and change with the community that uses and supports it. Three principles that contribute to the dynamic aspect of a living dictionary are:

**Visibility:** A living dictionary must be visible to the community, along with the process for updating both the practices and the content. The Unit Dictionary practices are available via digitally distributed report documents, and the contents are available through the Unit Registry.

**Openness:** The users of a living dictionary are also participants, able to inform its continuing design. The content and practices associated with the Dictionary are continually being developed by the LTER community in working groups, video teleconferences, and through tools like the Unit Registry.

**Provenance:** A living dictionary is built iteratively, with each version building on the previous. Because of this, the history of the practices and content both must be recorded. The history of the Unit Dictionary, and the decisions made during its lifecycle, are preserved along with the current state of the Dictionary.

### Further Reading

For more information and discussion on the UWG's projects, check the web site at: <http://intranet.lter.edu/im>. Current articles include:

**Unit Dictionary: Best Practices:** Guidelines on creating complete and consistent units for use in the LTER network.

**Unit Registry: Technical Specification:** An overview of the Unit Registry software, including use cases, design considerations, data models, and implementation details.

**Units: Background:** A historical review of the development of the Unit Dictionary and Unit Registry efforts and the changes they have seen.

More articles on the Unit Working Group and related topics can be found in the LTER IM newsletter, DataBits, and the website.

### Practices & Technology

The Unit Dictionary practices and the Unit Registry technology influence each other in a setting of continual design. The Registry software has been designed to meet the needs of the Dictionary project, without imposing technical barriers that limit or contradict the practices. In some cases, however, the technology informs the practice – for example, the use of a centralized Registry model brought about the idea of vetting as a cross-site committee process, rather than something to be handled at individual sites.

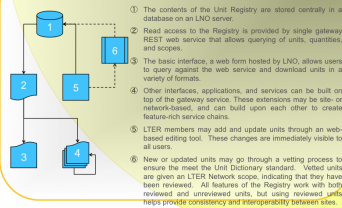
### Unit Registry

The LTER Unit Registry is a centralized web service that allows querying of units in the Unit Dictionary. This service can be accessed through a web site or incorporated into site applications. LTER sites can add and edit units, which are then available for use by all sites.

Units can be compared and verified across sites, enabling better data interoperability and eliminating the need for each site to track units separately. Network metadata tools can leverage the Registry to ensure that units are used consistently across sites.

The Registry also tracks all changes made to units and can alert a site when a unit is added, modified, or deprecated. These warnings can prevent conflicting unit definitions from propagating across sites.

### Deployment Architecture



### Site/Network Configuration

The Unit Registry is supported at both the site and network levels, allowing the development process to leverage the strengths of both environments.

LTER Sites contribute to the continuing development of the Unit Registry by holding VTC design sessions, writing and maintaining code, and creating useful extensions to the service. All content in the Registry is also provided and reviewed by site members.

The LTER Network Office provides a central deployment environment for the Unit Registry software. Sites do not need to maintain individual installations. The LNO allows provision of a Subversion repository for version control of the source code, making it available for cross-site development.

### Unit Working Group

The Unit Working Group is a group of site and network information managers concerned with challenges and solutions surrounding the use of units in a widely distributed network such as the LTER. Our work ranges from social issues (e.g. standards-making, shared resources, the value of local participation) to technical issues (e.g. database and interface design, code management, online tools).

If you're interested in the Unit Working Group, please take a look at our community forum on the information management website at:

<http://intranet.lter.net.edu/im/>


To join the Unit Working Group mailing list, please contact one of the co-chairs:

Poster date: 2009-12-07

Title: Toward Integrated Data: Web Access to CalCOFI Ichthyoplankton Data

Description: IchthyoDB (<http://oceaninformatics.ucsd.edu/ichthyoplankton>) is a queryable web application that provides data about abundance of fish eggs and larvae sampled as part of the CalCOFI program.


Authors: Karen Baker, Mason Kortz, Ed Weber, Rich Charter, Susie Jacobson, Sam McClatchie, Bill Watson, Tony Koslow



Toward integrated CalCOFI data ...

## Web Access to CalCOFI Ichthyoplankton Data

Karen S. Baker<sup>1</sup>, Mason Kortz<sup>1</sup>, Ed Weber<sup>2</sup>, Rich Charter<sup>2</sup>, Susie Jacobson<sup>2</sup>, Sam McClatchie<sup>2</sup>, Bill Watson<sup>2</sup>, and Tony Koslow<sup>1</sup>  
<sup>1</sup> Scripps Institution of Oceanography, UCSD; <sup>2</sup> Southwest Fisheries Science Center, NOAA



Making ichthyoplankton data publicly available is an important first step toward integrating CalCOFI physical and biological data. IchthyoDB (<http://oceaninformatics.ucsd.edu/ichthyoplankton>) provides a queryable web interface to the abundance of fish eggs and larvae. The application serves data from all cruises, 1950 to present, including mesozooplankton displacement volume and individual ichthyoplankton species captured in oblique, surface, vertical, or depth-stratified net tows. IchthyoDB was made available to the public in June 2009. It is part of a larger project led by the Ocean Informatics team at Scripps Institution of Oceanography working collaboratively with the NOAA Southwest Fisheries Science Center to develop a new generation of information infrastructure for the CalCOFI program. The project is already providing diverse CalCOFI datasets in a variety of publicly accessible formats via an architecture for highly structured data (<http://oceaninformatics.ucsd.edu/datazoo>) that supports data filtering, plotting, integration, and exchange. IchthyoDB data are published into Datazoo, thus providing an alternative web interface that co-locates the data with other CalCOFI datasets. We are currently developing approaches and applications that better integrate datasets in response to the needs of researchers, policy makers, and the public.

### IchthyoDB, a specialized interface

<http://oceaninformatics.ucsd.edu/ichthyoplankton>

- A specialized application that allows users to view, query, and output data and metadata from a complex collection of studies
- Data downloaded to the desktop, exported for exchange and/or published into Datazoo for integration with other CalCOFI data
- A database of 308 studies, 539 cruises, 53,797 tows for fish eggs and larvae.

#### IchthyoDB - Egg and Larvae Counts

Note: Several important changes have occurred in sampling methods for collecting ichthyoplankton. In 1969, tow depths extended from 140 m to 210 m, and nets were changed from 0.55-mm-mesh silk to 0.505-mm-mesh nylon. In 1977, oblique tows were changed from using 1-m bridled ring nets (denoted C1 in the data) to 0.71-m bridless borgo nets (denoted C3). See Hewitt 1980, Brinton and Townsend 1981, and Chinn and Smith 1985 (References, below) for details.

Summary		Taxonomies	
900911 rows by 26 columns <small>The number of rows in this result set exceeds the limit on some versions of Excel</small>		<b>Code</b>	<b>Scientific Name</b>
Unit:	NumberPer10m2	239	<i>Digenichthys atlantica</i>
ShortTowType:	C1,CB	31	<i>Engraulis mordax</i>
CruiseTypeCode:	C	31	<i>Engraulis mordax</i>
Year:	1950-2008	9942	Large oceanic fish
Month:	1-12	9998	1950-1984 (month, station numbers less than or equal to station 70 by 1985 cruise only); 2007-present (month and borgo, all stations, all cruises)
TowBegin:	0:00-24:00	991	<i>Merluccius productus</i>

Codes	
Type	Code
Cruise	C
Tow	C1
Tow	CB

#### Data published into Datazoo

<http://oceaninformatics.ucsd.edu/datazoo>

- An information system with data and metadata models creating an integrative architecture for viewing, querying, outputting and exchanging datasets.
- Datasets accessible as cruise collections and as time series across cruises.
- A publication system for specialized applications.
- A work in progress....

#### IchthyoDB - Egg and Larvae Counts

**Tow Type Filters**

Oblique  Vertical  Manta  MODNESS

**Time Period Filters**

Year:  to   
 Month:  to   
 Time of Day:  to

**Species Filters**

Common Name:

Code	Scientific Name	Common Name	Type	Occurrences
31	<i>Engraulis mordax</i>	Northern anchovy	Larvae	18091
1028	<i>Microstomus pacificus</i>	Rockfish	Larvae	38
31	<i>Engraulis mordax</i>	Northern anchovy	Eggs	1028

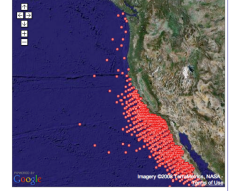
**Output Options**

include negative tows

**Geographic Filters**

By Latitude/Longitude: Lines  to   
 By Line/Station: Stations  to


**Species Distribution**



**Data URL**

#### Datazoo - Integration System

- Ichthyoplankton Datasets
- Hydrographic Bottle Dataset
- HPLC Dataset
- Additional Datasets



100

Poster date: 2010-10

Title: Metadata Database Models and EML Creation at LTER Sites

Description: Overview of LTER site IM Systems using entity-relationship diagrams.

Authors: M.Gastil-Buhl et al (KBaker, MKortz, JConners)

**Metadata database models and EML creation at LTER sites**

— compiled by M.Gastil-Buhl (MCR) from contributions by D.Henshaw & S.Remillard (AND),  
 J.Laundre (ARC), J.Walsh (BES), P.Tarrant (CAP), K.Baker, M.Kortz & J.Conners (CCE/PAL), D.Bahauddin (CDR), J.Chamblee (CWT), L.Powell (FCE), W.Sheldon (GCE), J.Campbell (HBR), E.Boose (HFR), K.Ramsey (JRN), S.Bohm (KBS), A.Skibbe (KNZ), E.Melendez-Colom (LUQ), S.Welch (MCM), C.Gries (NTL), H.Humphries (NWT), H.Garritt (PIE), M.O'Brien (SBC), K.Vanderbilt (SEV), N.Kaplan (SGS), J.Porter (VCR), I.San Gil (LNO/NBII)

<series>
<coverage>
<attribute>
<dataTable>
<contact>

<dataset>
<project>
<access>
<citation>
<purpose>
<unit>

<maintenance>
<eml>
<methods>

### Abstract

The purpose here is to spark discussion. Preparing for data integration, we will each examine our IM System to ask if it will meet potential new metrics. Some LTER sites already generate PASTA-ready EML. Will their design work at my site?

### Survey

Please correct or update your site information. Write directly on this participatory poster.

Site	How is metadata stored?	How is EML created?	How is EML stored?	How is EML distributed?	How is EML accessed?
AND	MySQL	Scripts	MySQL	Web	Web
ARC	MySQL	Scripts	MySQL	Web	Web
BES	MySQL	Scripts	MySQL	Web	Web
BNZ	MySQL	Scripts	MySQL	Web	Web
CAP	MySQL	Scripts	MySQL	Web	Web
CCE	MySQL	Scripts	MySQL	Web	Web
CDR	MySQL	Scripts	MySQL	Web	Web
CWT	MySQL	Scripts	MySQL	Web	Web
FCE	MySQL	Scripts	MySQL	Web	Web
HBR	MySQL	Scripts	MySQL	Web	Web
HFR	MySQL	Scripts	MySQL	Web	Web
JRN	MySQL	Scripts	MySQL	Web	Web
KBS	MySQL	Scripts	MySQL	Web	Web
KNZ	MySQL	Scripts	MySQL	Web	Web
LUQ	MySQL	Scripts	MySQL	Web	Web
MCM	MySQL	Scripts	MySQL	Web	Web
NTL	MySQL	Scripts	MySQL	Web	Web
NWT	MySQL	Scripts	MySQL	Web	Web
PAL	MySQL	Scripts	MySQL	Web	Web
PIE	MySQL	Scripts	MySQL	Web	Web
SBC	MySQL	Scripts	MySQL	Web	Web
SEV	MySQL	Scripts	MySQL	Web	Web
SGS	MySQL	Scripts	MySQL	Web	Web
VCR	MySQL	Scripts	MySQL	Web	Web

### Summary

19 sites use a relational database system for metadata (6 MySQL, 7 SQL Server, 4 Oracle). Of these, 14 sites generate EML from their metadata RDB. 6 sites plan to use Drupal.

5 sites serve a local data catalog from EML. 13 sites have multiple data tables within single EML documents.

### Scope

Here we focus only on metadata, not the data per se. Metadata-data congruency can be enhanced when the data are coordinated within the metadata system. So this is an incomplete picture. GIS is not covered here.

### Commonalities

All LTER sites share common things. Entity-Relationship diagrams show how these things are related. Each thing corresponds to one or more EML elements.

All sites need to present metadata on websites, in EML documents and other uses, such as other metadata standards.

### Longevity and Continuing Design

Some LTER sites' models designed in the 1990s are still in use today, such as VCR and AND, having migrated to new servers and new applications as technology changed. They remain useful because their schema inherently model the characteristics of metadata and through continuing design to keep pace with evolving standards.

### Mature Models

DataZoo at CCE/PAL, GCE Metabase and AND Metadata Database are three examples of mature models, in production, and part of a larger IM System at these LTER sites. These models continue to undergo improvements. Web page display is just one of their uses. EML is currently generated by scripts from all three of these metadata databases. The AND and GCE metadata model designs pre-dated EML; the extraction of EML was developed after the initial design. EML is just one of several metadata standards these three are designed to serve. All three undergo continuing development.

EML generated from the constrained model of a database is more likely to meet future metrics, especially if the data itself is filtered through a connected system.

Metabase collects data descriptions as part of a data ingest application.

DataZoo uses a data access layer to synchronize data with its metadata.

### Generic

### Future

Web services add options for development and use of data and metadata. The Unit Registry web service will soon be followed by the Controlled Vocabulary of Keywords and then subsequently by the NIS Administrative modules (bibliography and personnel). With this approach, sites may connect to services, replacing or synchronizing those parts of their local database. How will this affect our metadata database architecture?

Several sites are looking to participate in future development of metadata data models.

The GCE Metabase has been adopted by CWT and is planned to be ported to PostgreSQL at MCR and SBC.

Six LTER sites (LUQ, SEV, PIE, ARC, NTL, VCR) are pooling resources to develop a Drupal-based metadata storage, display and EML creation system.

### Drupal Environmental Information Management System (DEIMS)

### EML Specification

### Acknowledgement

IMs at AND, CCE/PAL, GCE, JRN, LUQ, SBC, VCR and NBII/LNO contributed additional background material and advice.

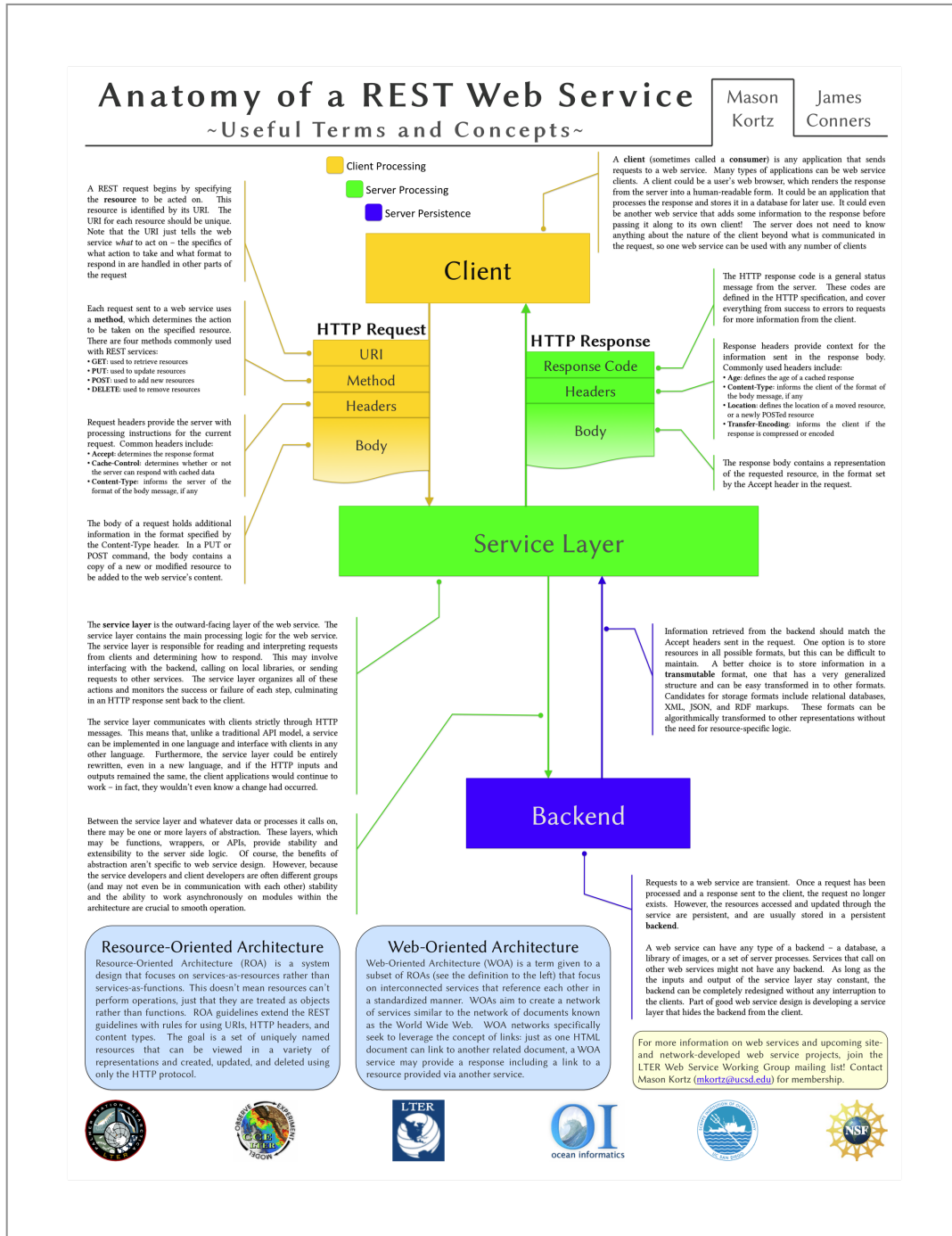
This research was supported by the US National Science Foundation's Long Term Ecological Research program

Poster date: 2010-10

Title: Anatomy of a REST Web Service

Description: Presenting a resource-oriented architecture as an augmentation of the web-oriented architecture at the LTER IMC Annual Meeting.

Authors: Mason Kortz, James Connors






Poster date: 2010-10

Title: Toward Data Sharing and a Web-of-Repositories: CalCOFI Information Management and Data Delivery

Description: Data flow from specialized interfaces to data published into DataZoo.

CalCOFI program PICES Symposium.

Authors: Karen Baker, Ed Weber, Tony Koslow



Toward data sharing and a web-of-repositories ...

## CalCOFI Information Management & Data Delivery

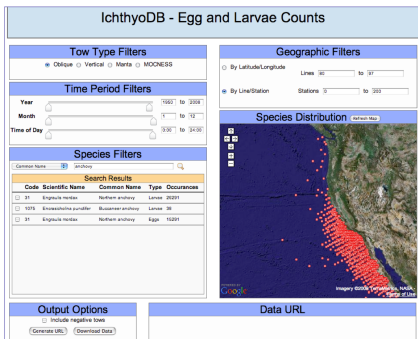
Karen S. Baker<sup>1</sup>, Ed Weber<sup>2</sup>, and Tony Koslow<sup>1</sup>  
<sup>1</sup>Scripps Institution of Oceanography, UCSD; <sup>2</sup>Southwest Fisheries Science Center, NOAA

**IchthyoDB, a specialized interface**  
<http://oceaninformatics.ucsd.edu/ichthyoplankton>

- A specialized application that allows users to view, query, and output data and metadata from a complex collection of studies
- Data downloaded to the desktop, exported for exchange and/or published into DataZoo for integration with other CalCOFI data
- A database of 308 studies, 539 cruises, 53,797 tows for fish eggs and larvae.

**Diverse, local data published into DataZoo**  
<http://oceaninformatics.ucsd.edu/datazoo>

- An information system with data and metadata models creating an integrative architecture for viewing, querying, outputting and exchanging datasets.
- Datasets accessible as cruise collections and as time series across cruises.
- A publication system for specialized applications.
- A data repository delivering data to the web-of-repositories



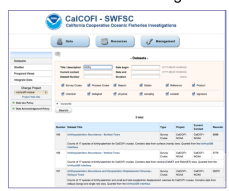
The DataZoo information environment aggregates heterogeneous data (e.g. two and three-dimensional physical and biological data sampled on a variety of scales), enhancing data access and contributing to the coherence and quality of the long-term CalCOFI data. Currently, data and associated metadata can be browsed, queried and visualized before download by individual users. DataZoo includes datasets ranging from species level counts and hydrographic profiles to biogeochemical measurements and ancillary datasets such as marine birds. It includes the core CalCOFI data sets as well as data from partner programs such as the California Current Ecosystem Long-Term Ecological Research Program. A recent redesign enables delivery of larger files including profile data not previously included in DataZoo. Data delivery and exchange services are under development to meet the future goals of improved and automated access to CalCOFI datasets in coordination with other ocean observing programs.

The CalCOFI program has been developing a web-based information-management system known as DataZoo since 2007. DataZoo expands upon existing CalCOFI data management practices to allow the worldwide community of scientists and the general public to use CalCOFI data effectively. DataZoo is a substantial advance over publishing raw databases because it includes additional elements that, in combination, make it the central feature of an "information environment"

DataZoo - Integration System


- Ichthyoplankton Datasets
- Hydrographic Bottle Dataset
- HPLC Dataset
- Additional Datasets

DataZoo Data Catalog



**Local data to "Web-of-Repositories"**<sup>13</sup>

- A national information infrastructure for discovery
- Datasets and collections of datasets
- A federated, non-hierarchical set of data repositories
- A diverse, inclusive, distributed, collaborative system



<sup>13</sup>KSBaker and LYarney, 2009. Data Stewardship: Environmental Data Curation and a Web-of-Repositories. International Journal of Digital Curation 4(2):12-27  
[http://interoperability.ucsd.edu/docs/09BakerYarney\\_DCC.pdf](http://interoperability.ucsd.edu/docs/09BakerYarney_DCC.pdf)

PICES, October 2010

Poster date: 2010-12

Title: CalCOFI Data Management: Unique Identifiers for Integrating Data

Description: Unique identifiers for co-ordinating and integrating diverse datasets.

Authors: Mason Kortz, Ed Weber, James Conners, Jim Wilkinson, Karen S. Baker, Tony Koslow

### CalCOFI Data Management: Unique Identifiers for Integrating Data

Mason Kortz<sup>1</sup>, Ed Weber<sup>2</sup>, James Conners<sup>1</sup>, Jim Wilkinson<sup>1</sup>, Karen S. Baker<sup>1</sup>, and Tony Koslow<sup>1</sup>  
<sup>1</sup>Scripps Institution of Oceanography, UCSD; <sup>2</sup>Southwest Fisheries Science Center, NOAA

**Abstract:** The CalCOFI cruise program has been providing a wide array of physical and biological oceanographic data for more than 60 years. Many CalCOFI data analysis projects require that these data be integrated for comparative studies. However, the evolution of sampling and data management practices over six decades often makes accomplishing this integration difficult due to differences in nomenclature such as cruise names, station designations, and methods of grouping related measurements.

In 2009, information managers from SIO and SWFSC began a collaboration to produce a set of unique identifiers to allow physical and biological data to be quickly matched by cruise, station, and sample for both past and future CalCOFI data. The process has been an iterative one. Each iteration has improved the quality and reliability of these matches; from roughly 70% of samples matching in the first attempt in 2009 to 95% with the latest set of identifiers. During the development of common data indices, the components necessary to resolve data relationships have distilled into three distinct keys. As other data relationships are established, the need for additional components may become apparent. This poster illustrates a set of unique identifiers that support data integration, establishing a standard-fit baseline of merged data for general use and an approach that permits alternative match choices depending on the needs of the individual researcher.

**NOAA Data**

Net tows  
(Bongo, Paravet,  
Manta)

**Scripps Data**

Physical Data  
(Bottle samples,  
CTD casts, etc.)

#### Data Integration Issues:

**Typical Station**

**Typical Station Variations**

#### Data Integration Solution:

The **StationID** identifies a station occupation within a cruise. A station occupation includes all activity between station arrival and departure. The StationID, which is composed of a line number, station number, and order occupied, provides a way to group related activities. The StationID can be used to reconstruct the timeline of a cruise. It also allows datasets to be quickly and accurately matched at the station level.

Within a station occupation, each measurement has a **SampleID**. The SampleID designates data as part of a set of samples taken at a similar time and location. This allows researchers to quickly find comparable data across multiple CalCOFI datasets. The SampleID field consists of a timestamp for the set of samples and a numeric identifier (in case there are multiple, unrelated sampling activities occurring at the same time).

Each set of **data fields** will be associated with a CruiseID, StationID, and SampleID. Currently, CalCOFI data managers are working to assign IDs to sixty years of historical physical and biological data in order to aid in cross-dataset analyses.

In addition to the CruiseID, StationID, and SampleID fields, each dataset will continue to have a full set of **metadata** fields. The new ID fields augment, not replace, the existing metadata. The ID fields provide an easy and reliable way to match data across CalCOFI datasets, but other matching criteria can still be used instead of or in conjunction with the IDs.

CruiseID	StationID	SampleID	Latitude	Longitude	TowDate	...	Engraulis mordax	Sebastes spp.	...
2001-10-25-C-32NM	080.0-060.0-065	045400-1	34.148	-121.153	2001-11-08	...	66.84	0.00	...
2001-10-25-C-32NM	080.0-055.0-066	083400-1	34.295	-120.8	2001-11-08	...	2500.00	140.00	...
2002-01-24-C-31JD	093.3-026.7-001	122600-1	32.958	-117.303	2002-01-24	...	3.79	18.95	...
2002-01-24-C-31JD	093.3-028.0-002	151200-1	32.915	-117.393	2002-01-24	...	0.00	4.91	...

## Full Abstracts of Posters by Category

### A. POSTERS about Ocean Informatics

#### **10. Title: LTER Growing Information Infrastructure: Data Lifecycles and Subcycles**

Author(s): Karen Baker, Florence Millerand, Lynn Yarmey

Date: 2009-09-14

Description: Information infrastructure, a vital aspect to many contemporary scientific investigations, is in transition. A lifecycle model for digital data provides a framework for site-based information management. Together with relevant standards, the full data context, important to local and remote data repositories, is under development. We highlight selected subcycles and associated information management roles within the data lifecycle. In particular, a data analysis subcycle critical to data description efforts is explored at sites close to the data origin.

#### **20. Title: INTEROP Scientific Infrastructure Design: Information Environments and Knowledge Provinces**

Author(s): Karen Baker, Florence Millerand

Date: 2007-10-19

Description: Conceptual models and design processes shape the practice of information infrastructure building in the sciences. We consider two distinct perspectives: (i) a cyber view of disintermediation where information technology enables data flow from the field and on to the digital doorstep of the general end-user, and (ii) an intermediated view with bidirectional communications where local participants act as mediators within an information environment. Drawing from the literatures of information systems and science studies, we argue that differences in conceptual models have critical implications for users and their working environments. While the cyber view is receiving a lot of attention in current scientific efforts, highlighting the multiplicity of knowledge provinces with their respective worldviews opens up understandings of sociotechnical design processes and of knowledge work. The concept of a range of knowledge provinces enables description of dynamic configurations with shifting boundaries and supports planning for a diversity of arrangements across the digital landscape.

#### **24. Title: LTER Environmental Data Management: Infrastructure Studies Insights**

Author(s): Florence Millerand and Karen Baker

Date: 2007-08-02

Description In the mist of major changes in ecological data collecting, managing and sharing, an interdisciplinary team of information, ecological, and social scientists has been brought together at LTER PAL and CCE sites to facilitate the growth of site-based information infrastructure. While research endeavors traditionally focus either on the technical or on the social aspects of information systems design, this project addresses simultaneously the technical, social, and organizational dimensions of the development, usage, and maintenance of community information infrastructures in ecological science. Issues include design methodology, change mechanisms and interdisciplinary collaboration as well as participant engagement, articulation processes, and data stewardship. The poster presents the project, the research area (Infrastructure Studies), and findings from a case study on the design, development and implementation

processes of the Ecological Metadata Language standard in the LTER community. A conceptual framework based on the notion of enactment from organization theory is presented to broaden the understanding of large-scale information infrastructure deployment. Initiated a decade ago, Infrastructure Studies appears to be a new and promising research area for the digital needs emerging in the natural sciences.

**26. Title: LTER: Long Term Informatics**

Author(s): Karen Baker, Cyndy Chandler, Anna Gold, Florence Millerand, Jerry Wanetick  
Date: 2007-08-02

Description: With the information age as one of the many ramifications of the Internet, our understandings, cultures, and communities are undergoing change. LongTermInformatics.org is a loose network forming in response to contemporary information environment needs and expectations. Participants include local informatics and information infrastructure teams, each adapting to its own environmental data niche. These capacity-building efforts include earth science informatics, library informatics, and social informatics.

**28. Title: LTER: Research in Infrastructure Studies: Social & Organizational Perspectives on Ecological Data Management**

Author(s): Florence Millerand and Karen Baker  
Date: 2006-09-20

Description: In the mist of major changes in ecological data collecting, managing and sharing, an interdisciplinary team of information, ecological, and social scientists has been brought together at LTER PAL and CCE sites to facilitate the growth from site-based to larger-scale federating infrastructures. While research endeavors traditionally focus either on the technical or the social aspects of information systems design, this project addresses simultaneously the technical, social, and organizational dimensions of the development, usage, and maintenance of large-scale information infrastructure in ecological science. Such dimensions include design methodology, change mechanisms and interdisciplinary collaboration as well as participant engagement, articulation processes, and data stewardship. The poster presents the project, the research area (Infrastructure Studies), and findings from a case study on the design, development and implementation processes of the Ecological Metadata Language standard in the LTER community. A conceptual framework based on the notion of enactment from organization theory is presented to broaden the understanding of large scale information infrastructure deployment. Initiated a decade ago, Infrastructure Studies appears to be a new and promising research area for the digital needs emerging in the natural sciences.

**33. Title: Initiating the Data Dialogue: 2005 CalCOFI Conference Interactive Poster**

Author(s): Karen Baker  
Date: 2005-12-06

Description The interactions surrounding the 2005 CalCOFI Data Management poster are captured through photographs of updates and additions made to the poster during the poster session of the annual conference. In addition, the data management workshop held during the conference is shown.

**34. Title: CalCOFI Data Management: Overview and Reflection**

Author(s): Karen Baker, Karen Stocks

Date: 2005-12-05

Description A CalCOFI White Paper (2005) provides an overview of the current state of data and its management within the California Cooperative Ocean Fisheries Investigations (CalCOFI) program. The report presents steps and recommendations for building towards an integrated, online information system for CalCOFI. In addition to discussing how this effort could scale, the white paper considers present efforts within the context of the emerging Pacific Coast Ocean Observing System (PaCOOS) as well as other community efforts. As one of the longest-running, multidisciplinary ocean monitoring and observing programs in existence, the emphasis of data management within CalCOFI has focused on the twin goals of (a) quality control and curation of individual datasets collected on CalCOFI cruises and (b) data availability for researchers and fisheries managers through printed reports and requests to the data curators. Today, a new goal is emerging of having CalCOFI datasets available online and, eventually, interoperable with other CalCOFI-related datasets and within the larger, developing federation of the Ocean Observing System data. In this poster we provide a summary of concrete recommendations for moving forward in addition to inviting participants to consider their datasets in the context of a collection of CalCOFI datasets.

## **B. POSTERS by Ocean Informatics: Conceptual**

### **6. Title: LTER: A Web of Repositories**

Author(s): Lynn Yarmey, Karen Baker

Date: 2009-09-14

Description: The movement and exchange of data are frequently described using a 'flow' or a 'pipeline' model. We differentiate a uni-directional data 'flow' from an alternative model, a web-of-repositories. A web-of-repositories is a federation of diverse nodes where communication, connections, and data exchange are multi-directional. Each node has a unique sphere-of-context with technical, organizational and social dimensions. In this poster we explore a multi-repository data landscape.

### **7. Title: LTER Information Management History Database (HistoryDB)**

Author(s): Robert Petersen, Sean Wiley, Nicole Kaplan, Eda Melendez, Karen Baker

Date: 2009-09-14

Description: Organizational history requires a facility to manage, archive and present event details as well as narratives that provide perspective to the events. While events form a historical thread, storied narratives weave these threads together into a retrospective. The LTER Information Management Committee has recognized that working collaboratively to understand their history is a tool for exploring how they function within the LTER organizational structure. Such a tool provides valuable input to the development of governance procedures for community-level efforts. The Information Management Committee Governance Working Group is designing and developing HistoryDB as a platform to record and publish significant events related to the development of Information Management within the LTER network. This work is prompted by the recognition of how our future may well be informed if we are able to remember and discuss our past.

### **14. Title: LTER Information Managers: A Community of Practice**

Author(s): Karen Baker, Nicole Kaplan, Inigo San Gil, Margaret O'Brien, Florence Millerand  
Date: 2008-08-10

Description: Communities of Practice are groups of people who share a concern or a passion for something they do and who want to learn more about how they do it. Such a community is more than a group of people having the same job or a network of connections between people. Three elements characterize a Community of Practice: 1) the domain, 2) the community, and 3) the practice. Regular interaction such as with an annual meeting is a key integrative mechanism that brings into play elements of practice including agenda setting, knowledge management, professional development, advocacy, and resource mobilization. The history and multi-dimensional aspects of Communities of Practice provide a framework for considering information management organizationally through structures that facilitate communication and learning. We explore the Long Term Ecological Research Information Management Committee in particular as a Community of Practice. Examples of how the information management role has emerged and is defined within the Long Term Ecological Research community will be presented. How the committee as a collective fits within this framework will be considered by taking into account interests, activities, and relations. Active membership, professional engagement, and collective learning are needed to ensure relevance as well as long-term sustainability.

### **15. Title: LTER Information Infrastructure: Emergent Roles, Responsibilities and Practices**

Author(s): Lynn Yarmey, Karen Baker  
Date: 2008-08-10

Description: Human activities together with technical elements and collective practices are core elements for growing local infrastructure as well as for bridging with other communities and networks. Site information management activities create a shared data curation experience where data curation refers to managing the capture, use and preservation of the data. Identifying and elaborating upon local data activities opens up the complex set of arrangements that comprise site information management, including the variety of roles emerging to address mediation and collaboration. Any one activity may be carried out in practice by different participants at each site. That is, what one site considers an information management role may be carried out by a researcher, technician, analyst, or education coordinator at another site. The diverse distributions of responsibilities at each site are a result of meeting local scientific needs with a mix of local participants and practices. Comparing and contrasting different site infrastructure arrangements prompts discussion that deepens our understanding of data and data curation. Insight into data activities and their associated roles and responsibilities may be seen as a preparatory step for conscientiously designing an effective data network.

### **31. Title: LTER IM Articulation Work: Developing Community Web Recommendations**

Author(s): Nicole Kaplan, Karen Baker, Barbara Benson, John Campbell, Corinna Gries, James Laudre, Jeanine McGann, Eda Melendez-Colom, Marshall White  
Date: 2006-09-20

Description: Over the past two years, the Web Site Design Recommendations Working Group developed recommendations for web sites in response to challenges of first generation LTER web sites. They worked to align a set of social, technical and organizational elements. Articulation work is described as work that enables other work such as within a task, within a project, or across organizational entities. Articulation work refers to the interrelating of parts or

the alignment of work elements, often involving a range of planning, coordinating, and negotiating efforts. The Web Site Design Recommendations working group's efforts are an example of articulation work involving both explicit elaboration and attention to alignment of multiple elements. Social and organizational elements were considered while addressing the needs of web site users, organizational and technical elements influenced recognizing successful navigational and organizational components, and community and technical elements were used to create designs and links to communicate each site as being part of the LTER Network. The recommendations are currently being presented to the LTER Executive Board. This working group's work will need to continue - key is the need for review and update in order to accommodate changes in technology and delivery mechanisms as well as in conceptual understandings, organizational categories, social perspectives, community elements, and synthesis strategies. Future plans thus include planning both for updated web design and for the attendant articulation work.

### **25. Title: LTER: Data Integration in the Decade of Synthesis**

Author(s): Mason Kortz, Lynn Yarmey, James Connors, Karen Baker

Date: 2007-08-02

Description As data availability, findability, and even queriability become more ubiquitous, the need to make sense of data from multiple, disparate sources increases. Data integration and data synthesis allow extension of the scope of data beyond local use, creating a whole that is greater than the sum of its parts. This poster/demo examines the similarities and differences between integration and synthesis, taking PAL and CCE site-level data integration projects and their role in the LTER network data synthesis efforts as case examples. The poster also describes the possibility of recursive integration and synthesis and discusses the role of metadata in data integration.

## **C. POSTERS by Ocean Informatics: Technical**

### **1. Title: CalCOFI Data Management: Unique Identifiers for Integrating Data**

Author(s): Mason Kortz, Ed Weber, James Connors, Jim Wilkinson, Karen S. Baker, and Tony Koslow

Date: 2010-12-03

Description: The CalCOFI cruise program has been providing a wide array of physical and biological oceanographic data for more than 60 years. Many CalCOFI data analysis projects require that these data be integrated for comparative studies. However, the evolution of sampling and data management practices over six decades often makes accomplishing this integration difficult due to differences in nomenclature such as cruise names, station designations, and methods of grouping related measurements.

### **2. Title: CalCOFI Information Management and Data Delivery**

Author(s): Karen Baker, Ed Weber and Tony Koslow

Date: 2010-10-22

Description: The CalCOFI program has been co-developing a web-based information-management system known as DataZoo since 2007. DataZoo expands upon existing CalCOFI data management practices to allow the worldwide community of scientists and the general public to use CalCOFI data effectively. DataZoo is a substantial advance over publishing raw databases because it includes additional

elements that, in combination, make it the central feature of an ?information environment?. The DataZoo information environment aggregates heterogeneous data (e.g. two and three-dimensional physical and biological data sampled on a variety of scales), enhancing data access and contributing to the coherence and quality of the long-term CalCOFI data. Currently, data and associated metadata can be browsed, queried and visualized before download by individual users. DataZoo includes datasets ranging from species level counts and hydrographic profiles to biogeochemical measurements and ancillary datasets such as marine birds. It includes the core CalCOFI data sets as well as data from partner programs such as the California Current Ecosystem Long-Term Ecological Research Program. A recent redesign enables delivery of larger files including profile data not previously included in DataZoo. Data delivery and exchange services are under development to meet the future goals of improved access to CalCOFI datasets in coordination with other ocean observing programs.

### **3. Title: Metadata database models and EML creation at LTER sites**

Author(s): M.Gastil-Buhl (MCR) from contributions by D.Henshaw & S.Remillard (AND), J.Laundre (ARC), J.Walsh (BES), P.Tarrant (CAP), K.Baker, M.Kortz & J.Conners (CCE/PAL), D.Bahauddin (CDR), J.Chamblee (CWT), L.Powell (FCE), W.Sheldon (GCE)

Date: 2010-09-23

Description: The purpose here is to spark discussion. Preparing for data integration, we will each examine our IM System to ask if it will meet potential new metrics. Some LTER sites already PASTA-ready EML. Will their design work at my site?

### **4. Title: Anatomy of a REST Service: Useful Terms and Concepts**

Author(s): Mason Kortz, James Conners

Date: 2010-09-23

Description: An overview of the basic concepts and technology of a REST web service.

### **5. Title: CalCOFI Toward Integrated Data: Web Access to CalCOFI Ichthyoplankton Data**

Author(s): Karen Baker, Mason Kortz, Ed Weber, Rich Charter, Susie Jacobson, Sam McClatchie, Bill Watson, Tony Koslow

Date: 2009-12-07

Description: IchthyoDB (<http://oceaninformatics.ucsd.edu/ichthyoplankton>) is a queryable web application that provides data about abundance of fish eggs and larvae sampled as part of the CalCOFI program. The application serves data from all cruises, 1950 to present, including mesozooplankton displacement volume and individual ichthyoplankton species captured in oblique, surface, vertical, or depth-stratified net tows. IchthyoDB was made available to the public in June 2009. It is part of a larger project led by the Ocean Informatics team at Scripps Institution of Oceanography working collaboratively with the NOAA Southwest Fisheries Science Center to develop a new generation of information infrastructure in support of the CalCOFI program. The project is already providing diverse CalCOFI datasets in a variety of publicly accessible formats through Datazoo, an information system for highly structured data that supports data filtering, plotting, integration, and exchange (<http://oceaninformatics.ucsd.edu/datazoo>). IchthyoDB data are published into Datazoo, thus providing an alternative web interface that co-locates the data with other CalCOFI datasets. We are currently developing approaches and applications that better integrate datasets in response to the needs of researchers, policy makers, and the public.

### **8. Title: PAL & CCE LTER: A Site-Based Information Architecture**

Author(s): James Conners, Mason Kortz, Lynn Yarmey, Karen Baker

Date: 2009-09-14



Description: Designing infrastructure to support the management of diverse data presents unique challenges for each site. Described here is the current information system architecture, as well as targeted architectural features, implemented by the Ocean Informatics team to provide a working solution for accommodating heterogeneous data types. The system architecture is a major component of a site information environment, providing an orientation for technical development, organizational communication, and collaborative science.

### **9. Title: LTER Unit Working Group Projects: Dictionary and Registry**

Author(s): Mason Kortz, Lynn Yarmey, James Connors, Todd Ackerman, Karen Baker

Date: 2009-09-14

Description: Units of measurement are a fundamental element of scientific discourse and data integration. The LTER Unit Working Group has developed two initiatives to promote consistent use of units throughout the network. One is the LTER Unit Dictionary, comprising the set of units in use by the LTER sites and the best practices that support them. The other is the Unit Registry, a software solution for online access to the Unit Dictionary. This poster provides an overview of both efforts, including motivations, progress made, and future plans.

### **11. Title: CCE LTER: An Oceanographic Eventlogger as One Part of an Information Environment**

Author(s): Karen Baker, Mason Kortz, James Connors

Date: 2009-05-13

Description: The CCE LTER initiated at SIO in 2004 enabled launch of "Ocean Informatics", a new approach to design of information infrastructure in support of interdisciplinary science. CCE works synergistically with Palmer Station LTER and with California Cooperative Oceanic Fisheries Investigations (CalCOFI) at Scripps and at NOAA Southwest Fisheries Science Center. Major activities of the CCE LTER Information Management to date have been to develop an information environment that includes: a) a cross-project, open source framework that provides collaborative tools and activities; b) a project web site (<http://cce.lternet.edu>) with dynamic elements such as personnel and bibliography modules; c) an information system (<http://oceaninformatics.ucsd.edu/datazoo>) serving as a local data repository providing both data access and integration; d) a multi-component architecture anchored by data dictionaries and metadata; and e) a suite of resources supporting local data handling, analysis, and visualization. Local informatics research focuses on discursive practices, sociotechnical systems design, and the semantic work required at the human-information interface while network activities include participation in a dictionary working group, governance working group, and the Databits Newsletter. The event logger used at sea as part of the data flow process is being demo'd during the LTER Science Council Pier Walk at Scripps Institution of Oceanography.

### **12. Title: CalCOFI Biological Data Management**

Author(s): Karen Baker, Mason Kortz, James Connors, Lynn Yarmey

Date: 2008-11-17

Description: An information system designed for working with multiple oceanographic biological data collections is presented. DataZoo is an extensible system that supports data discovery, access, query, and exchange for data such as the CalCOFI integrated biological data and bottle measurements from hydrographic casts. The poster will provide answers to: What is DataZoo?; What does it do?; Who uses it?; What's in it?; How is it built? The system is a data and metadata repository designed to meet the needs of researchers, policy makers and the public. It is a publishing forum that includes a dataset catalog, personnel directory, and metadata system. Dictionaries and controlled vocabularies play a key role and facilitate data integration. The metadata schema takes into account local and community standards including the Ecological Metadata Language, augmenting it with local unit, attribute, and qualifier

dictionaries. DataZoo is organized into three web-based functional units: data, resources, and management. A suite of resources extend the information system interface to the desktop so local participants can manage their own data - and in turn consider their individual data practices in relation to a project repository made readily apparent via web interfaces and web services. A community information system creates a data curation commons that highlights shared technical components, organizational arrangements, and collective practices, all central elements to growth of a local information infrastructure able to bridge projects, communities and networks.

### **13. Title: LTER Abstracting Functionality and Access: Facilitating Data System Manageability and Site Coordination**

Author(s): Mason Kortz, James Connors, Karen Baker

Date: 2008-08-10

Description: As the functionality of site data systems increases, frequently so does the complexity. Organizing system functionality through distinct layers of abstraction, from low-level system access to high-level user access, is key to maintaining a manageable system. Toward this end, a data system that is an interdependent set of databases, files, and other resources can often be abstracted into a relatively compact set of data access methods. Abstraction layers allow developers to leverage not only the content of a data system but the organizational logic as well. Leveraging may take the form of facilitating local site reuse or sharing across projects and sites. Abstraction enables the development of multiple applications, accessing the same data system - and its data - via a single interface layer. This poster explores three models by which data access methods may be abstracted and shared: application programming interfaces, remote procedure calls, and resource state transfers. Each model is defined in general as well as illustrated by examples designed, developed, and deployed at two Long-Term Ecological Research sites (Palmer Station and California Current Ecosystem).

### **16. Title: Scientific Communication and Information Infrastructure**

Author(s): Karen Baker, Beth Simmons, Ryan Rykaczewski, Alison Cawood, Peter Davison, Moira Decima, Melissa Garren, Andrew King, Andrew Taylor, Jesse Powell, Melissa Soldevilla, Mike Stukel

Date: 2008-08-10

Description: Scientific communication is central to collaborative scientific endeavors. A shared information infrastructure facilitates communication and collaboration. Digital information infrastructure occurs in multiple forms. The poster presents examples of CCE LTER communication: data publishing with information system DataZoo, story publishing with a children's book, community designing with a design studio, multi-media publishing with a picture gallery, referencing with an online bibliography, real-time field experiences with a picture-of-the day, local field experience with pier walks, data stewardship & sampling design with an event logger, and a community website with the CCE Home Page.

### **17. Title: CalCOFI Local Metadata: Augmenting the Ecological Metadata Language**

Author(s): Lynn Yarmey, Karen Baker, James Connors

Date: 2007-11-17

Description: Metadata is an integral and necessary part of data sharing; the enactment of a metadata standard not only guides the creation of local metadata documents but is also a link between local and broader communities. A full metadata record, including but not limited to descriptions of the field environment, detailed accounts of analytical methods, and summaries of quality control procedures, is essential to the understanding and use of any dataset. Without the context of the data, measurement values are subject to misinterpretation and misuse. A rich local metadata standard prompts consideration of the range of information necessary to form a complete metadata record. Such a standard creates a structure

and format that provide those knowledgeable about a dataset a place to record unique as well as common elements. Standardized metadata functionally makes possible automated comparisons and visual presentation of datasets. In addition to establishing a local foundation for data sharing, a standard becomes an integrative bridge when developed in parallel with community and national standards. The Ecological Metadata Language (EML) provides a metadata specification with growing acceptance in environmental science communities. In this poster, we discuss adaptations and augmentations made to EML for the Ocean Informatics community information system (DataZoo) in order to ensure the local metadata structure, while still linked to the broader community, is optimized to capture any complexity associated with local oceanographic datasets.

### **18. Title: CalCOFI Data Management: Developing Community Standards**

Author(s): James Wilinson, Karen Baker, Rich Charter

Date: 2007-11-17

Description: CalCOFI represents a partnership of multiple agencies conducting quarterly joint oceanographic cruises, CalCOFI field team members work as a cohesive cross-agency unit to accomplish the cruise goals. Associated participants frequently integrate their field measurements and sampling with the long-term core CalCOFI measurements and samples. Once a cruise concludes, however, this cohesive unit disperses; individuals return to their respective agencies and labs to process samples and analyze data. Each group uses lab or agency specific methods and software to generate data products in local formats. These diverse data processing methods, products, and storage formats create challenges for merging datasets. Development and incorporation of shared data management practices or joint standards enable data integration. Shared practices include a) Standard, persistent vocabulary and formats e.g. use of the same labels for the same data columns with translation tables for different units; b) Standard, persistent date & position formats; c) Standard line & station designations for gridded data e.g. 93.3 120.0; d) Sequential station numbering e.g. order-occupied; e) Event numbers e.g. when needed for resolving station activities; f) Distribution of data in non-proprietary format e.g. tab delimited text or csv Metadata i.e. details of context, measurements & equipment; g) Designating common columns, such as order occupied or event number, and adding them to existing data products allows heterogeneous datasets to be related and ingested into relational databases or into data analysis and visualization applications.

### **19. Title: CalCOFI & Ocean Informatics DataZoo: A Multi-Project Data Publishing System**

Author(s): Mason Kortz, James Connors, Karen Baker

Date: 2007-11-17

Description: The DataZoo information system is a hub in the Ocean Informatics learning environment that creates a central forum for data exchange, collaborative design, and community building. It is a central repository for data and metadata of member projects, providing data aggregation, ingestion, description, visualization, download, integration, and standardized exchange. It serves as a publishing arena for datasets from individual project members and from project groups. A number of design features facilitate scientific work. For example, local work benefits from data availability and queriability while community work benefits from alignment with metadata standards. The flow of data from the field to a local repository is supported through cross-project extensibility, dataset ingestion templates, and time-series storage of study collections. Data integration and exchange are enabled by the use of study-specific internal indexing, cross-project dictionaries, and augmented metadata describing data to a column level. Ancillary related tools are being developed such as project-specific sampling grid converters, dataset joining tools, and a date-time calculator. Working together with LTER and CalCOFI participants to develop a local information system creates the opportunity to improve capture of data and metadata as well as to understand community needs.

### **21. Title: CCE LTER Information Infrastructure**

Author(s): Jerry Wanetick, Karen Baker, Nate Huffnagle, Lynn Yarmey, Mason Kortz, James Conners

Date: 2007-09-17

Description: Information Infrastructure is an arrangement of computational systems, an iTeam, information systems and partnerships associated with a core interest in informatics. Ocean Informatics is defined as the work at the intersection of oceanography, social science and information science.

## **22. Title: Ocean Informatics Information System: One Element of an Information Infrastructure**

Author(s): Karen Baker, Mason Kortz, James Conners, Jerry Wanetick

Date: 2007-09-17

Description: Focus is on an Information system for managing data - DataZoo 2.0 -at the heart of a configuration of computational systems, an iTeam, informatics work, and a complex set of partnerships.

## **23. Title: A working Standard: Augmenting the Ecological Metadata Language**

Author(s): Lynn Yarmey, Karen Baker

Date: 2007-09-17

Description: Metadata standards are an integral and necessary part of data sharing as they provide a structure and format to allow comparisons of data context. A full and complete metadata record is essential to understanding and using any dataset, as without the context of the data, values are meaningless. A metadata standard not only prepares for future dataset comparisons and integrations, but also prompts the user to consider of all parts of a complete metadata record, from descriptions of the field environment to detailed accounts of any and all analytical methods and quality control procedures performed. A standardized metadata format also allows for quick automated or visual comparisons of datasets and begins to lessen the impact from any workflow articulation differences. The Ecological Metadata Language (EML) is a standard with growing acceptance in the scientific realm, it's strengths include attribute-level descriptions and a flexible architecture. In this poster, we discuss the adaptations and augmentations made to EML to better encapsulate the complexity inherent to our local datasets.

## **25. Title: LTER: Data Integration in the Decade of Synthesis**

Author(s): Mason Kortz, Lynn Yarmey, James Conners, Karen Baker

Date: 2007-08-02

Description: As data availability, findability, and even queriability become more ubiquitous, the need to make sense of data from multiple, disparate sources increases. Data integration and data synthesis allow extension of the scope of data beyond local use, creating a whole that is greater than the sum of its parts. This poster/demo examines the similarities and differences between integration and synthesis, taking PAL and CCE site-level data integration projects and their role in the LTER network data synthesis efforts as case examples. The poster also describes the possibility of recursive integration and synthesis and discusses the role of metadata in data integration.

## **27. Title: CalCOFI: An Oceanographic Event Logger**

Author(s): James Wilkinson, Karen Baker

Date: 2006-12-04

Description: Local data management, informed by field sampling and data use, supports community coordination at the interface of data collection and data curation. An oceanographic event logger recently deployed on a series of research cruises extends data management into the data collection arena. The event logger system consisting of a digital tablet, a community eventlog, and a unique index - is designed to promote conventions such as standard vocabulary and to establish relations between diverse data

efforts at the time of collection. The event logger addresses issues of time, space and categorization that assist subsequent data integration.

**29. Title: CCE LTER: Information Management (2004-2006)**

Author(s): Karen Baker, Lynn Yarmey, Mason Kortz, Jerome Wanetick

Date: 2006-09-20

Description: The California Current Ecosystem information management efforts were launched with inquiries into existing data practices. This was followed by design, development and deployment of elements of an information infrastructure including secure web and file services as well as a platform for exploration of collaborative software applications from content management systems to shared plotting tools. A set of core technical services have been developed including extensive file storage capacity, disk sharing technologies, and planning toward single sign-on directory services. Sociotechnical services have included development of an Ocean Informatics conceptual framework supporting infrastructure process-building, design teams, and forums within the Integrative Oceanography Division at Scripps Institution of Oceanography. Two initial information system elements include database development organized in coordination with field use of an electronic event logger and a web site designed to include dynamic elements such as a bibliography module, media gallery, regional mapping application, and station location converter. Work on both metadata and quality assurance proceeds synergistically with local organizational partners Palmer LTER guided by the LTER community standards, the California Cooperative Oceanic Fisheries Investigations (CalCOFI) with an emerging regional program Pacific Coast Ocean Observing System (PACOOS), the Southern California Ocean Observing System (SCCOOS) and support communities such as Quality Assurance of Real-Time Oceanographic Data (QARTOD) and the Marine Metadata Interoperability Project (MMI).

**30. Title: Palmer LTER: Design of a Queriable Ocean Information System**

Author(s): Karen Baker and Shaun Haber

Date: 2006-09-20

Description: Field data, originating with domain understandings and practices that shape sampling and collection, has informed development of the PAL LTER information system. In becoming digitally preserved, data capture may in turn be influenced by an information system's organizing principles and structure. Focusing on the goal of an automated web service able to browse datasets in hierarchical arrangements, to generate automated queries and plots, and to meet community metadata and exchange standards, design has involved both exploring potential system assumptions and constraints as well as on articulating their ramifications in terms of requirements for data to adapt to such a system. In moving from a data system that makes data accessible to an information system that makes data queriable, the PAL LTER data structure makes use of templates for dataset type definitions, of attribute dictionaries referenced to unit dictionaries, and of quality assurance procedures as central to the capacity for automating traversals through the system. In terms of developing understandings of data and its availability in digital repositories, information system design (and redesign) may be considered an important part of data stewardship.

**32. Title: Ocean Informatics: Conceptual Framework for Marine Science Information Management**

Author(s): Karen Baker, Jerry Wanetick, Shaun Haber, Lynn Yarmey, Mason Kortz, Florence Millerand, Jesse Powell, Jim Wilkinson, Robert Thombly, Julie Thomas, Beth Simmons

Date: 2006-04-01

Description: The work of Ocean Informatics is represented at the union of oceanography, information science and social science domains. Participants range from data and information managers to technical

specialists, archivists, scientific researchers, educators, as well as those working in science and infrastructure studies.

**35. Title: Palmer LTER: Information Flow and Management**

Author(s): Karen Baker, Anna Gold, Frank Sudholt, Helena Karasti, Geoffrey Bowker

Date: 2003-09-18

Description: Organizational repositories are being constructed today to address the needs of scientific information management in a digital environment. Given the social aspects of information, building useful information systems requires infrastructures that reflect the unified and expressive relationships of data, documents, people, institutions and partnerships. The Palmer Long-Term Ecological Research (LTER) program information management is working in partnership to explore articulation of the LTER community information management practices and to prototype a co-construction of a low barrier bibliographic referatory/repository.

## 11 Appendix: Ocean Informatics Event Gallery

A table summarizing events is given below followed by event flyers created as a one-page reminder of visitors and events occurring during the visit.

Date	Title	Description
2003-12-10	US Joint Global Ocean Flux Study and Data Systems	JGOFS Visit: Cyndy Chandler
2004-10-18	PACOOS-CalCOFI Data Management Meeting	PACOOS-CalCOFI
2004-11-05	SIO, WHOI, and Informatics	JGOFS Visit: Cyndy Chandler
2004-11-17	CalCOFI Annual Symposium Data Management Workshop	CalCOFI
2005-12-05	CalCOFI Annual Symposium Data Management	CalCOFI
2005-12-07	CalCOFI Annual Symposium DM Workshop Survey	CalCOFI
2005-12-05	CalCOFI Annual Symposium DM Workshop Handout	CalCOFI
2006-03-16	Controlled Vocabularies to Ontologies and Concept Maps Too	LTERR & Science Studies: Deana Pennington
2006-08-18	Cyberinfrastructure, Ocean Informatics, and Data Management	JOGS visit: Cyndy Chandler
2006-12-08	Ocean Informatics, Design Sessions and a Video	JGOFS visit: Cyndy Chandler
2007-01-23	CCE LTER Information Infrastructure and the Data	CCE LTER
2007-03-01	Ocean Informatics, Cyberinfrastructure and CalCOFI Handout	CalCOFI: Handout
2007-03-01	Ocean Informatics, Cyberinfrastructure and CalCOFI	CalCOFI: Tony Koslow
2007-07-23	Ocean Informatics, Data Integration and EML	LTERRNBII: Inigo San Gil
2007-08-18	Data issues, Roles, and Uptake	Library Visit: Anne Grahame
2007-09-05	Data Issues, Roles, and Library Support for E-Science	Library Visit: Anna Gold
2007-11-02	CICSE and SIO: CalCOFI IMECOCAL	CalCOFI IMECOCAL
2007-11-11	DataZoo, Drupal, and APIs	Ocean Informatics: Shaun Haber
2007-11-26	CalCOFI Conference: Information Management	CalCOFI Conference
2008-01-11	Dataturbine, open source, and site specifics	LTERR MCR Visit: Sabine Grabner
2008-03-20	Information environments and communication	Library Visit: Kristin Yarmey
2008-04-02	DataZoo and Classroom Use	SIO Education
2008-04-03	Ocean Informatics and Information Systems	WHOI Teleconference: Cyndy Chandler
2008-05-15	Conversations on Metadata	LTERR NBII Visit: Inigo San Gil
2008-05-29	UC-LTERR Graduate Student & Post-doc Symposium	LTERR CCE, SBC, MCR
2008-05-31	Information Management cross-site visit	LTERR NTL visit: Barbara Benson
2008-06-16	Ocean Informatics Monograph Write Session	Science Studies: Florence Millerand
2008-07-17	Source code and Sociotechnical Programming Practices	Science Studies: Stephane Couture PaCOOS: Johnathan Phinney, Karen Baker, Sharon Mesick
2009-06-09	Regional Zooplankton Workshop	Science Studies: Sonja Palfner
2009-06-15	Ocean Informatics Exchange	Science Studies: Sonja Palfner
2009-07-13	Site Exchange, Cross-synthesis Comparison, and Governance	LTERR SGS Visit: Nicole Kaplan
2010-03-06	Information Exchange and Information System Elements	LTERR MCR Visit: Mary Gastil
2010-05-24	LTERR Unit Registry	LTERR KBS Visit: Sven Bohm
2010-08-01	LTERR Unit Registry	LTERR SEV Visit: Ken Ramsey
2010-11-12	Units and Governance	LTERR LUQ Visit: Eda Melendez
2010-12-04	Information Systems	LTERR MCR Visit: Mary Gastil
2010-12-10	Site-Site Discussion	LTERR CAP Visit: Philip Tarrant
2011-03-11	Music, Business and Scientific Digital Delivery	Ocean Informatics: Shaun Haber

1 Event Flyer

Date: 2003-12-10

Title: US Joint Global Ocean Flux Study and Data Systems

Description: JGOFS Visit: Cyndy Chandler

# US Joint Global Ocean Flux Study and Data Systems

SIO/IOD-LTER/OIC & WHOI/JGOFS

Integrative Oceanographic Division, Scripps Institution of Oceanography, UCSD  
Data Management Office, US Joint Global Ocean Flux Study, Woods Hole  
Oceanographic Institute

*10 December 2003*



Karen S. Baker, Jerry Wanetick, Steve Jackson, Bren Mills  
Dawn Rawls, Charleen Johnson, Cyndy Chandler





3

Event Date: 2004-10-18

Title: PACOOS-CalCOFI Data Management Meeting

Description: PACOOS-CalCOFI New

Author: Karen Baker

## CalCOFI Data Management Meeting Summary – 18 Oct 04

Karen Stocks, Karen Baker, David Allison, Richard Charter, Sherri McCann, Stephen Diggs, Ralf Goericke, Mati Kahru, Gregory Mitchell, Mark Ohman, Fernando Ramirez, Christian Reiss, Jennifer Sheldon, Jerome Wanetick, Jim Wilkinson, Dave Wolgast



This meeting brought together participants at SIO and SWFSC interacting with CalCOFI-related datasets in diverse ways. Karen Stocks and Karen Baker, tasked with forward planning in support of ongoing CalCOFI data management teams, coordinated the meeting. They opened with a presentation that covered 1) a consideration of local efforts within the context of emerging institutional, regional and national partnerships (Figure A); 2) an outline of generic data system with elements from the ingestion of multiple data types through user query and integration (Figure B); and 3) the activities Baker and Stocks are undertaking and the products they are developing.

A round-table discussion followed during which several themes emerged. First, that the integration of data is important for supporting analysis and visualization. Second, that metadata, standards, exchange protocols, and core categories (that is, joint nomenclature and language) play a critical role in data systems. Although the range of tasks is large and the data/system/organizational relationships complex, participants initiated two key processes: shared infrastructure and information flow. A series of products and processes were discussed including a recent PACOOS proposal to fisheries and a data management presentation at the upcoming CalCOFI Conference (Figure D). In addition, a personnel directory and a dataset inventory were handed out with a request for edits and updates. A follow-up technical mini-meeting will be planned for December after the November cruise, at which standards, metadata, and data integration will be considered in detail. Meanwhile, a critical issues list was begun and will be circulated prior to the next meeting.

### -Critical Issues

- Inventory local datasets, data types, and data sizes as well as expectations and resources
- Prioritize data tasks and expectations with respect to resources
- Identify and prioritize user community participant groups and products
- Develop mechanisms for participants to engage in identifying requirements and designing the system
- Articulate and bridge local individual metadata standards and formats
- Identify national metadata standards and exchange protocols
- Identify a data system model and mechanisms to interface with local systems
- Create a shared information infrastructure for diverse groups and diverse data types
- Consider how to build out and support partnerships
- Identify critical field collection-data system factors such as station name conventions and reporting formats
- Establish assessment mechanisms to ensure the system meets and continues to meet user needs

3

Event Date: 2004-11-05

Title: SIO, WHOI, and Informatics

Description: JGOFS Visit: Cyndy Chandler

Author: Karen Baker



# SIO, WHOI, and Informatics

November 05, 2004

CCS conference room (at the foot of the SIO pier)



The Ocean Informatics Exchange Workshop continues a dialogue initiated last year between folks managing oceanographic field data at SIO and WHOI. We are doing some forward planning with respect to the multiple dimensions of infrastructure and the design process for a contemporary information environment appropriate for communities in general as well as for organizations such as the SIO Integrative Oceanography Division (IOD).

-Karen Baker, Jerry Wanetick, and Cyndy Chandler

## Agenda

9:00 Agenda & Logistics (Baker)  
 Welcome (Guza)  
 9:15 Introductions (round-table)  
 9:30 Reviewing Past and Present  
 - Conceptual Model  
 - Tensions/Balances  
 IOD Context (Wanetick)  
 10:30 Break  
 11:00 WHOI Context (Chandler)  
 11:15 Semantics and Terms  
 Informatics: Historical Perspective  
 Informatics: Domain Perspective  
 11:45 Wrap-up  
 12:00 Meeting review (round-table)  
 12:15 Lunch

## Participants

Karen Baker, SIO, [kbaker@ucsd.edu](mailto:kbaker@ucsd.edu)  
 Cyndy Chandler, WHOI, [cchandler@whoi.edu](mailto:cchandler@whoi.edu)  
 Art Gaylord, WHOI, [agaylord@whoi.edu](mailto:agaylord@whoi.edu)  
 Shaun Haber, SIO, [srhaber@ucsd.edu](mailto:srhaber@ucsd.edu)  
 Florence Millerand, LCHC, [fmiller@ucsd.edu](mailto:fmiller@ucsd.edu)  
 Dawn Rawls, SIO, [dawls@ucsd.edu](mailto:dawls@ucsd.edu)  
 Uwe Send, SIO, [usend@ucsd.edu](mailto:usend@ucsd.edu)  
 Karen Stocks, SDSC, [kstocks@sdsc.edu](mailto:kstocks@sdsc.edu)  
 Wayne Suiter, SIO, [wsuiter@ucsd.edu](mailto:wsuiter@ucsd.edu)  
 Julie Thomas, SIO, [jthomas@ucsd.edu](mailto:jthomas@ucsd.edu)  
 Jerry Wanetick, SIO, [jwanetick@ucsd.edu](mailto:jwanetick@ucsd.edu)  
 Lynn Yarmey, SIO, [lyarmey@ucsd.edu](mailto:lyarmey@ucsd.edu)



4

Event Date: 2004-11-17

Title: CalCOFI Annual Symposium Data Management Workshop

Description: CalCOFI New

Author: Karen Baker

*Southwest Fisheries Science Center NOAA Fisheries  
Scripps Institution of Oceanography  
California Department of Fish and Game*

## **CalCOFI Annual Conference 2004**

### **California Cooperative Oceanic Fisheries Investigations**



### **CalCOFI Data Management**

17 Nov 2004 Wednesday

CalCOFI Data Management: Today and Beyond  
Karen I. Stocks and Karen S. Baker  
1440-1500 Presentation, Sumner Auditorium

Integrating CalCOFI Datasets in a Web-Based Browser  
Steve Diggs and Christian Reiss  
1530-1730 Poster, IGPP Munk Conference Room, in IGPP

Data Management Workshop  
K.Stocks, K.Baker, S.Diggs, C.Reiss, R.Charter  
1630-1800 Helen Raitt Room, 3rd floor SIO library



5

Event Date: 2005-12-05

Title: CalCOFI Annual Symposium Data Management

Description: CalCOFI

Authors: Karen Baker

## CalCOFI Conference and Data Management

CalCOFI Conference (<http://calcofi.org>)  
Scripps Institution of Oceanography, Sumner Auditorium  
5-7 December 2005  
Southwest Fisheries Science Center, NOAA Fisheries  
Scripps Institution of Oceanography  
California Department of Fish & Game



### Agenda with Data Management Components Highlighted

Monday, 5 December 2005

Mon: 11:00am-1:00pm CalCOFI Registration Opens: Sumner Auditorium Portico

Mon: 1:00pm-1:30pm Welcome & Opening Remarks: Elizabeth Venrick & Charleen Johnson

Scripps Institution of Oceanography, La Jolla, CA

SESSION I: STATUS OF THE CALIFORNIA CURRENT

SESSION II: STATUS OF CALIFORNIA'S FISHERIES

RECEPTION / DINNER

Mon: 5:30pm-9:00pm Reception & Catered Buffet Dinner  
(Upper Campus - Eucalyptus Point Conference Center)



Tuesday, 6 December 2005

Tues: 08:00am-8:30am Registration – Sumner Auditorium Portico

SESSION III: Symposium of the Conference: "CalCOFI: The sum of the parts"

Moderator: Elizabeth Venrick - Scripps Institution of Oceanography, La Jolla, CA

Tues: 1:30pm-2:10pm California Current Ecosystem, the newest Long Term Ecological Research

(LTER) site – Mark D. Ohman, Scripps Institution of Oceanography, La Jolla, CA

Tues: 4:30pm-7:30pm POSTER PRESENTATIONS, Wine & Cheese Reception  
(Vaughan Hall 100; north side of Sumner Auditorium)

Distribution of CalCOFI Report, Volume 46

DM Poster:

DM Poster:

SESSION V: WORKSHOPS

Wed: 1:30pm-2:30pm DATA MANAGEMENT WORKSHOP - Karen S. Baker<sup>1</sup>,

Richard Charter<sup>2</sup>, and

James Wilkinson<sup>1</sup>

<sup>1</sup>Scripps Institution of Oceanography, La Jolla, CA

<sup>2</sup>Southwest Fisheries Science Center, La Jolla, CA

Wed: 2:30pm-3:30pm PACOOS SAMPLING STRATEGIES ESTABLISHING REGIONAL COHERENCE WITHOUT SACRIFICING LOCAL TIME SERIES - Bill Peterson<sup>1</sup> and Dave Checkley<sup>2</sup>

<sup>1</sup> NOAA/NMFS/NWFSC, Hatfield Marine Science Center, Newport, Oregon

<sup>2</sup> Scripps Institution of Oceanography, La Jolla, CA



6

Event Date: 2005-12-07  
CalCOFI Annual Symposium  
DM Workshop Survey  
CalCOFI  
Karen Baker



**10 Question DM Survey**  
CalCOFI Conference - DM Workshop  
Scripps Institution of Oceanography  
December 7, 2005



**Community Building**

1. Are you familiar with the topic of information management in the context of a scientific research program?  
(circle one) Yes   Somewhat   No
2. When would you be interested in discussing your data management practices?  
(circle one) Now   Soon   Later   Never
3. How would you be interested in discussing IM, ie listserves, meetings, community forums  
\_\_\_\_\_
4. Are you using or interested in using web-based collaborative applications, ie Blog, Wiki, etc?  
(circle one) Yes   Maybe   No
5. What element(s) of an annual CalCOFI DM meeting interest you the most?  
\_\_\_\_\_  
\_\_\_\_\_

**Data**

6. What types of data do you work with?  
\_\_\_\_\_
7. What data formats do you work with?  
\_\_\_\_\_
8. What approach to data interoperability and exchange do you use now or aim to use?  
\_\_\_\_\_  
\_\_\_\_\_

**Wrap-Up**

9. List key words/phrases that capture what CalCOFI DM means to you/your work.  
\_\_\_\_\_  
\_\_\_\_\_
10. Is there anything else you would like to share?  
\_\_\_\_\_  
\_\_\_\_\_

Name \_\_\_\_\_  
Email \_\_\_\_\_  
Organization/Dept/Div \_\_\_\_\_  
Disciplinary Specialty \_\_\_\_\_  
Other DM communities you are familiar with \_\_\_\_\_  
Other DM communities you coordinate with \_\_\_\_\_

*Thank you for your input!*

*Please return survey to Florence  
Millerand or Karen Baker  
2252 Sverdrup Hall; 858-534-2350*

Event Date: 2005-12-05

CalCOFI Annual Symposium DM Workshop Handout

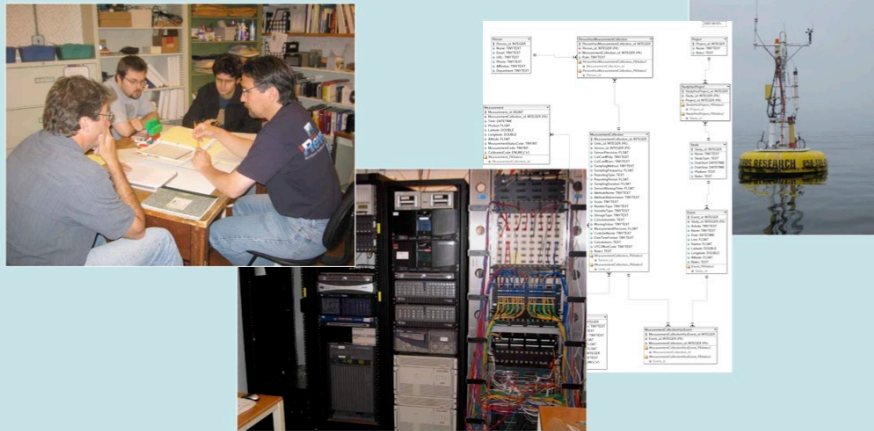
CalCOFI

Karen Baker

## CalCOFI Conference: DM Workshop Handout

*A Conceptual Framework for Marine Science Information Management*

Toward shared **Infrastructure** Promoting Resource and Data Interoperability  
organizational • technical • community • local



### Joint Database Design Teams

*Heterogeneous Data Types:* shipboard underway, event logging, survey data, in situ automated & manual shore stations, moorings, mets, and gliders

### Cyberinfrastructure Endeavors

*Federated Data Flows:* centers, applications, services, transport, and information systems support

### Data/Information Managers as Designers/Mediators

*Community Building Mechanisms:* workshops, standards, best practices, reading groups, MMI, SIO Webheads



<sup>1</sup>Baker, Jackson, Wanetick, 2005 System Science Proceedings  
Towards an Ocean Informatics Environment, HICSS05



8

Event Date: 2006-03-16

Controlled Vocabularies to Ontologies and Concept Maps Too

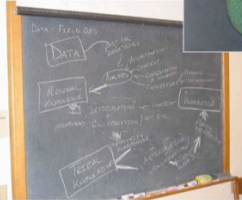
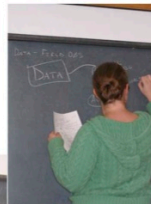
LTER & Science Studies: Deana Pennington

Karen Baker

## Controlled Vocabularies to Ontologies and Concept Maps Too

UCSD/SIO/IOD 16 March 2006

Working Group and Workshop  
with Deana Pennington



### Agenda:

0700-0800 Breakfast  
0830-0930 HSD Poster & Dictionary (FM,KB)  
9:30-1000 Knowledge Representation (Deana)  
1000-1100 John Porter Conference Call  
1115-12:30 Data to Knowledge (OI Workshop)  
1230-1330 IOD Luncheon  
1330-1400 break  
1400-1500 Cyberinfrastructure (LTER CCE)  
1530-1630 coffee& debrief

### Artifacts:

Informatics Concept Map (OI Report)  
Information Infrastructure Design Matrix (Florence, Karen)  
controlled vocabulary, dictionary, metadata standard, ontology  
Metaphor Drawings  
Workshop Concept Maps  
Paper Outline Sketch  
Recordings & Photos



### Creating a Knowledge Framework

1. The pieces (dp)-what structural elements add what functions  
uncontrolled vocab, controlled vocab, taxonomy, thesaurus, dictionary, ont
2. controlled vocabulary (jp) query interface group
3. unit registry & attrib dictionary (kb)
4. ontology
5. shared vision - how we link together we get a plan and let folks know

Event Date: 2006-08-18

Cyberinfrastructure, Ocean Informatics, and Data Management

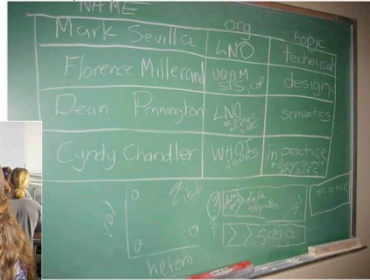
JOGS visit: Cyndy Chandler

Karen Baker



# Cyberinfrastructure, Ocean Informatics, and Data Management

UCSD/SIO-/IOD 18 August 2006  
Cyberinfrastructure, Ocean Informatics, And Data Management with Cyndy Chandler



### Agenda:

- 3:30-4:00 Introductions
- 4:00-5:30 Cyberinfrastructure Overview
- 5:30-6:00 Wrap Up and Round-Table
- 6:00-7:00 Surfside TGIF

### Sustaining the Dialogue

- 1) Infrastructure
- 2) Cyberinfrastructure
- 3) Workshop vs local CI
- 4) Meeting the needs of data collectors, managers, users

### Artifacts:

- OI Visiting Scholar Table
- CSIG06 Topic Prompts
- Data Collection-Processing-Delivery Distribution Arrow
- Photos

**Summer Institute CSIG06 - GEON**  
cyberinfrastructure for geoscientists  
<http://www.geongrid.org/CSIG06/>



- \* Focus: introducing geoscientists to commonly used as well as emergent technology
- \* Topics: Web Services, Workflows, Knowledge Representations, and Geographic Information Systems
- \* Key Words : ontology and ontology-based search



Event Date: 2006-12-08

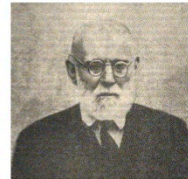
Ocean Informatics, Design Sessions and a Video

JGOFS visit: Cyndy Chandler

Karen Baker



## Ocean Informatics, Design Sessions and a Video



UCSD/SIO 8 December 2006  
Ocean Informatics, Design Sessions and a Video  
with Cyndy Chandler (WHOI)

### Sustaining the Dialogue

- 1) Dialogue
- 2) Dictionaries
- 3) Information Systems
- 4) Joint Writing

### Artifacts:

- Paper OI: A Process in Practice
- List: Lessons learned
- Design Studio
- Video
- Photos



### Cross-case analysis: JGOFS-PAL Twelve Lessons Learned

#### Guiding Principles

- 1) Informatics support
- 2) Data teams
- 3) Joint products
- 4) Data reuse

#### Local practice

- 5) Community communication
- 6) Data description
- 7) Data quality
- 8) Data policy

#### Local approaches

- 9) Information environment
- 10) Information infrastructures
- 11) Adaptive strategies
- 12) Local repositories

### Agenda

- 7:30-9:00 Breakfast
- 9:00-9:30 Prepare Lessons Learned
- 10:00-11:00 Design Session: Participant Lists
- 12:00-1:30 Lunch and Lessons Learned
- 1:30-2:15 Poster Discussion: Event Logger
- 2:15-2:45 Design Session: Parameter Dictionaries
- 2:45-3:15 Design Session: Gazetteers
- 4:00-5:30 Paul Otlet video
- 5:30-6:30 Surfside TGIF



Event Date: 2007-01-23

CCE LTER Information Infrastructure and the Data

CCE LTER Annual Meeting

Karen Baker



TUESDAY, JAN. 23rd, 2007		
MUNK Lab Conference Room (Rm 303 - western most room)		
Time	Participant	CCE element/topic
8:45 a.m.	ALL	Coffee and bagels
9:00	Mark Ohman	Overview and developments in the National LTER Network
9:15		
9:30	Ralf Goericke	<b>Augmented CalCOFI measurements</b>
9:45		Progress, results and problems
10:00	Lihini Aluwihare	Dissolved organics - DOC/DON data
10:15	Andrew King	Phytoplankton & biogeochemistry of Fe
10:30	Melissa Soldevilla	Marine mammal acoustics
10:45	Dave Checkley	Bongo-LOPC data (also process cruises)
11:00	***COFFEE BREAK***	
11:15	Alison Cawood	ZooScan analyses
11:30	Mati Kahru	<b>Other time-series meas.</b> - Remote sensing
11:45	Hey-jin Kim	Nearshore long-term observations of surface CHl
12:00	Russ Davis/Mark Ohman	gliders
12:15		DISCUSSION: INTEGRATION OF MEASUREMENT PROG
12:30	***LUNCH BREAK***	Catered deli box lunches/drinks
		<b>GRAD. STUDENT meeting/lunch</b> (see A. King and B. Hopkinson)
1:30	Mike Landry	<b>Process Cruises</b>
1:45		Overview and results
2:00		DISCUSSION: PROCESS CRUISE DESIGN AND RESULT
2:15	Peter Franks	<b>Modeling</b>
2:30	Steven Bograd	Control volume concept
2:45	Pascal Riviere	A high resolution nested modeling framework
3:00	Hey-jin Kim	Stratification/upwelling cells + nutrient supply
3:15		DISCUSSION: MODELING APPROACHES AND INTEGRA
3:30	***COFFEE BREAK***	
3:45	Karen Baker	<b>Information Management</b>
4:00		Data & the CCE information system
4:15		DISCUSSION: AN INTEGRATED INFORMATION SYSTEM
4:30	Beth Simmons	<b>Education and Outreach</b>
4:45		Outreach venues, projections & synopsis
5:00	Kathy Barbeau	Recruitment of REUs
5:15	Ralf Goericke	CCE - Japan exchange
5:30	Mark Ohman	Priorities for 2007 Supplement requests
5:45 p.m.	refreshments	
Please note venue change and earlier start time on Wed.!!		
WEDNESDAY, JAN. 24th, 2007		
REVELLE Lab Viz Center (4th floor, NE side on footbridge level) - IGPP		
Time	Participant	CCE element/topic
8:15 a.m.	ALL	Coffee and bagels
8:30	Small groups	Break-out session regarding manuscript writing and other collaborations



**Ocean Informatics**

March 2007  
Lunch Meeting Handout

**in 2002**

Information science  
metadata  
short-term  
structured  
unstructured  
long-term  
data  
domain science

**in 2003**

Science  
service  
IM  
manage  
design  
Data  
Technology

Information Management Tensions  
Baker and Karasti, 2002

Information Management Role  
Karasti and Baker, HICSS 2004

**in 2007**

complex  
im  
cyber  
dm  
small  
large  
simple

Information Infrastructure: a Federating Perspective  
Baker and Millerand, ASIST 2007

Event Date: 2007-03-01

Ocean Informatics, Cyberinfrastructure and CalCOFI

CalCOFI: Tony Koslow

Karen Baker

## Ocean Informatics Event UCSD/SIO - 1 March 2007

Ocean Informatics, Cyberinfrastructure and CalCOFI  
with Tony Koslow and Florence Millerand

### Sustaining the Dialogue



*Ocean Informatics events are held routinely though aperiodically as a long-term forum for communication, cooperation, and camaraderie regarding all things informatics, design, and/or digital related.*

### Sustaining the Collaboration

#### Attendees

Tony Koslow - CalCOFI Director  
 Florence Millerand - OI Human Social Dynamics  
 TJ Moore - SWFSC/NOAA GIS  
 Shaun Haber - OI Programmer/Associate  
 Beth Simmons - Edu/Outreach PAL&CCE LTER  
 Jerry Wanetick - OI & IOD Comp. Infrastructure  
 Jesse Powell - Grad Student CalCOFI/CCE  
 Andrew Taylor - Bio Technician, CCE/Landry  
 James Connors - OI Programmer/Student  
 Mason Kortz - OI Programmer/Analyst  
 Robert Thombley - CalCOFI Programmer  
 Lynn Yarmey - OI Programmer/Analyst  
 Laurie C - Bio Technician, CCE/Aluwihare  
 Jim Wilkinson - CalCOFI Programmer/Analyst  
 Charleen Johnson - OI Logistics & Transcription

### Agenda

12:00 Lunch  
 12:15 Roundtable Introductions  
 12:20 Tony Koslow: welcome and background  
 12:30 Florence Millerand: welcome back & updates  
 12:45 IOD web logo: The Swash (poster presentation)  
 12:45 Digital Landscape Figure  
 Roundtable Discussion  
 All Day Interviews

### Artifacts:

Handout Digital Landscape  
 -Tensions  
 --DM Role  
 --Cyberinfrastructure  
 The Swash Logo Poster  
 Pirate, Knights, and small figures  
 Photos

Event Date: 2007-07-23

Ocean Informatics, Data Integration and EML

LTERNBII: Inigo San Gil

Karen Baker

## Ocean Informatics, Data Integration and EML with Inigo San Gil

UCSD/SIO - 23-24 July 2007

### NBII, Network & Site Coordination



*Ocean Informatics events are held as a long-term forum for communication, cooperation, and camaraderie regarding all things data and design, informatics and information infrastructure.*

### Sustaining the Dialogue

#### Attendees

Inigo San Gil - LTER Network Office & NBII  
 Karen Baker - PAL& CCE IM  
 Tony Koslow - CalCOFI Director  
 Jerry Wanetick - OI & IOD Comp. Infrastructure  
 James Connors - OI ProgrammerAnalyst  
 Mason Kortz - OI ProgrammerAnalyst  
 Robert Thombly - CalCOFI Programmer  
 Lynn Yarmey - OI ProgrammerAnalyst  
 Jim Wilkinson - CalCOFI ProgrammerAnalyst  
 Rich Charters - SWFSC  
 Susie Campbell - SWFSC  
 David Jorgensen - OI ProgrammerAnalyst  
 Andrew Taylor - Bio Technician, CCE/Landry  
 Charleen Johnson - OI Logistics & Transcription

### Agenda

#### Monday

09:00 DataZoo & EML data publishing  
 12:00 Lunch Meeting (Vaughn Hall)  
 01:30 Metadata, unit/attribute dictionary & data integration  
 02:30 IOD visit and coffee  
 03:30 PASTA and data provenance  
 08:00 Dinner

#### Tuesday

09:30 Web Modules  
 10:00 Keywords, thesaurus, and FGDC-BPD  
 12:00 Lunch & Pier Walk  
 01:30 Wrap-up

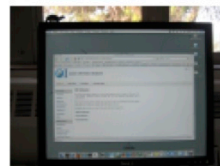
#### Artifacts:

Handout Data Integration  
 -ASM03: Decade of Synthesis!?  
 -LTER Decades  
 -Key issues & Key views

Turtles, Tortises, and Frogs

Photos

DataZoo 2.0



Event Date: 2007-08-18

Data issues, Roles, and Uptake

Library Visit: Anne Graham

Karen Baker

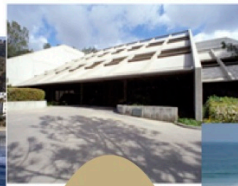
## Library Informatics: Data Issues, Roles, and Uptake with Anne Graham

UCSD/SIO - 18 August 2007

### Sustaining the Dialogue

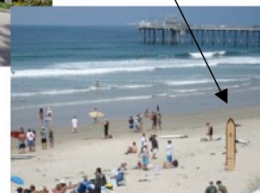


*Ocean Informatics events are held routinely though aperiodically as a long-term forum for communication, cooperation, and camaraderie regarding all things informatics, design, and/or digital related.*



Snippets: roles, semantics, data structure, library as repository, cultural side, uptake, datasets for publication, Digital curation, archive vs preservation, DSpace, Dataverse, rdf

long board



### Agenda

- 12:00 GeoSciences Summer Institute end
- 2:30 SIO Shuttle
- 2:45 Ocean Informatics Design Studio Introductions
- 3:00 SIO Library Tour; Peter Breugerman
- 3:30 OI-UCSD-MIT Library Discussion
- 5:15 Adjourn
- 5:30 La Jolla Shores Dinner
- 7:30 Airport

### Attendees

- Anne Graham - MIT Librarian
- Karen Baker - SIO Information Manager
- Lynn Yarmey - SIO Program/Analyst
- Mason Kortz - SIO Programmer/Analyst
- James Connors - SIO Programmer/Analyst
- Jerry Wanetick - SIO/IOD Computation Facility
- Declan Fleming - UCSD IT
- Ardys Kozbial - UCSD Library
- Robert McDonald - SDSC Archive

### Listening In:

"Librarians are all about access. Is there going to be a culture clash with the data world? I'm not from the hiding school but from the sharing information school."

"Archive is a co-opted term."

"There's raw data and then there's the article level"

"NSF has begun to ask for data management plans."

"Messy data is one of the nuances of science."

"roles...a liaison to laboratory practices"

"There's the Digital Curation Conference 3 and the Preservation and Added Value Conference PV2007"



Event Date: 2007-09-05

Data Issues, Roles, and Library Support for E-Science

Library Visit: Anna Gold

Karen Baker

# Library Informatics: Data Issues, Roles, and Library Support for E-Science

with Anna Gold

UCSD/SIO - 05 September 2007

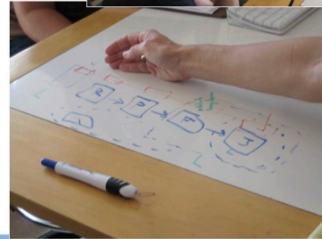
## Sustaining the Dialogue



Reading: Association of Research Libraries Draft Report of Joint Task Force on Library Support for E-Science



*Ocean Informatics events are held throughout the year as a long-term forum for communication, cooperation, and camaraderie regarding all things informatics, design, and/or digital related.*



## Agenda

- 10:00 Design Studio Introductions
- 10:15 Meet Mark Ohman, Lead PI CCE
- 10:30 Coffee
- 10:45 Presentation & Discussion
- 12:00 Round Table Close

## Attendees

- Anna Gold - MIT Head Librarian
- Karen Baker - SIO/OI Information Manager
- Lynn Yarmey - SIO/OI Program/Analyst
- Mason Kortz - SIO/OI Programmer/Analyst
- James Connors - SIO/OI Programmer/Analyst
- David Jorgensen - SIO/OI Volunteer

## Listening In:

- "there seems to be a lumping together of technical and conceptual issues"
- "it depends on the scale of the science"
- "the idea of a data package"
- "my understanding of e-Science, or at least what I perceived it as, changed"
- "there's so much data work to be done and more than enough room"



surf's up

cam's up

Event Date: 2007-11-02  
 CICESE and SIO: CalCOFI IMECOCAL  
 CalCOFI IMECOCAL  
 Karen Baker

## CalCOFI Data Management: CICESE and SIO

UCSD/SIO - 02 November 2007



### Starting the Dialogue



*Ocean Informatics events are held periodically as a forum for communication, cooperation, and camaraderie. This event brought together participants working with data associated with California Cooperative Fisheries Investigations (CalCOFI) from SIO in La Jolla and from CICESE/IMECOCAL near Ensenada.*

- Facilities:**  
 Design Studio  
 Computational Facility
- Demos:**  
 Eventlogger  
 DataZoo

### Agenda

- 11:00 Introductions
- 11:15 Welcome (Mark Ohman)
- 11:30 Computer Systems (Jerry Wanetick)
- 12:30 Eventlogger Demo
- 13:00 CICESE & SIO data
- 13:30 DataZoo Demo
- 16:00 Departure



### Attendees

- Tim Baumgartner - CICESE/IMECOCAL
- Norma Ramirez - CICESE/IMECOCAL Data Manager
- Pablo Torres - USABC Computer Science
- Karen Baker - SIO/Ocean Informatics
- Mason Kortz - SIO/OI Programmer/Analyst
- James Connors - SIO/OI Programmer/Analyst



UCSD/SIO/Integrative Oceanography Division:

- <http://calcofi.org>
- <http://cce.lternet.edu>
- <http://pal.lternet.edu>
- DataZoo**
- <http://oceaninformatics.ucsd.edu/datazoo>

CICESE: Centro de Investigacion Cientifica y de  
 Educacion Superior de Ensenada  
 IMOCOCAL: Invetigaciones Mexicanas de la Corriente  
 de California

- <http://www.cicese.mx>
- <http://imecocal.cicese.mx/>
- <http://www.uabc.mx>



Event Date: 2007-11-11

DataZoo, Drupal, and APIs

Ocean Informatics: Shaun Haber

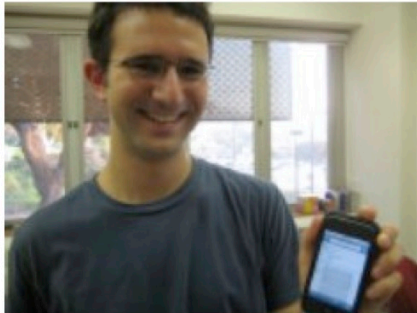
Karen Baker

## DataZoo, Drupal, and APIs



UCSD/SIO - 11 November 2007

### Continuing the Dialogue



*Ocean Informatics events are held periodically as a forum for communication, cooperation, and camaraderie. This event brought together participants working to implement technical solutions and to design new solutions in production environments, whether they be scientific research or the music industry.*

**Demos:**  
Drupal  
DataZoo  
Facebook

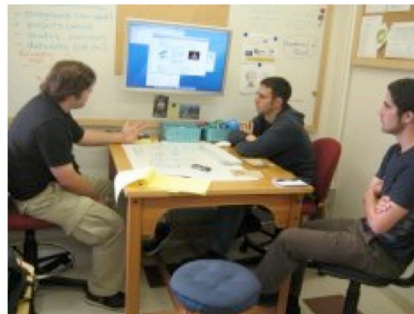
**Outcome:**  
Clarity in identifying a goal to create API interfaces for DataZoo modules, ie PeopleZoo and units.

### Agenda

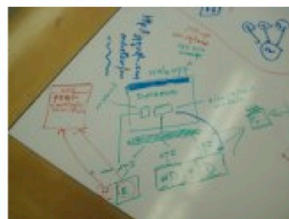
- 12:00 Lunch
- 12:30 Drupal & DataZoo
- 17:00 Music Industry and Tech Development
- 17:30 Departure

### Attendees

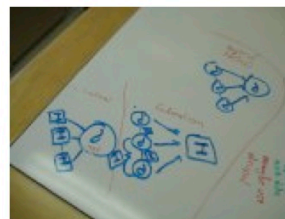
- Shaun Haber - Warner Brothers
- Karen Baker - SIO/Ocean Informatics
- Mason Kortz - SIO/OI Programmer/Analyst
- James Connors - SIO/OI Programmer/Analyst



Protocol Stack



API, Web Services and Test Console



Centralized, federated and existing hybrid case

19

Event Date: 2007-11-26

CalCOFI Conference: Information Management

CalCOFI Conference

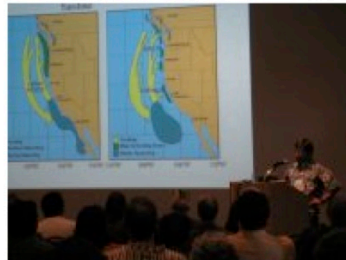
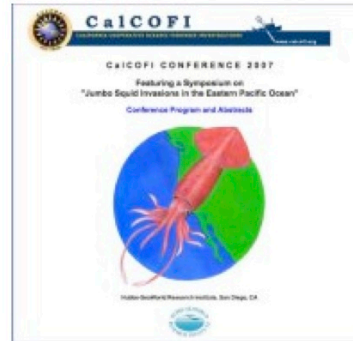
Karen Baker

## CalCOFI Conference: Information Management

26-28 November 2007

-----

with participation  
by CCE LTER  
and Ocean Informatics



State of the Current presentation by Ralf Goericke.



### Ocean Informatics posters:

- CalCOFI Data Management: Developing Working Standards - Jim Wilkinson, Karen Baker and Rich Charters
- Ocean Informatics Datazoo: A Multi-Project Data Publishing System - Mason Kortz, James Connors, and Karen Baker
- Metadata Standard: Augmenting the Ecological Metadata Language - Lynn Yarmey, Karen Baker, and James Connors
- Oceanographic Event Logger - Jim Wilkinson and Karen Baker

Event Date: 2008-01-11

Dataturbine, open source, and site specifics

LTER MCR Visit: Sabine Grabner

Karen Baker

## Open source software and site specifics

UCSD/SIO - 11 Jan 2008

CCE, PAL, and MCR LTER  
with Sabine Grabner



### Open source software and site specifics



### Agenda - Friday

- 11:00 Introductions
- 12:00 Lunch
- 01:30 IOD Computational Infrastructure
- 02:30 Debrief SDSC & Data Turbine
- 03:30 MCR Information Management Overview
- 04:30 PAL, CCE DataZoo Overview
- 05:30 TGIF
- 07:00 Dinner



*Ocean Informatics events are held periodically as a forum for communication, cooperation, and camaraderie regarding all things data and design, informatics and information infrastructure.*

### Sustaining the Dialogue

#### Participants

- Sabine Grabner - MCR LTER IM
- Karen Baker - PAL & CCE LTER IM
- James Conners - OI ProgrammerAnalyst
- Mason Kortz - OI ProgrammerAnalyst
- Lynn Yarmey - OI ProgrammerAnalyst
- Jerry Wanetick - OI & IOD Comp.Infrastructure
- Mark Ohman - CCE LTER Lead PI



### Example Outcomes

Cross-site translation table:

- MCR - OceanInformatics
- observable - column
- parameter - attribute

Two potential Databits topics

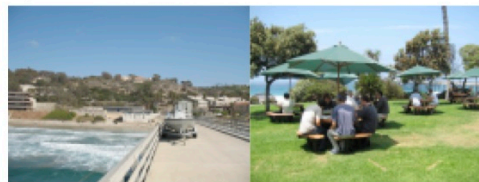
- QARTOD qa/qc working group
- LNO-site data request coordination

### Memorable Quotes:

Re Ocean Informatics SIO: *"It seems like a data or information management sanctuary. It's more like a data management sanctuary than a data management mafia"*

MCR Site Review on Information Environments:

*"An information environment isn't something you finish; it's something you start."* (from Baker and Bowker, 2007. Information Ecology: Open System Environment for Data, Memories and Knowing. JIIS)



Event Date: 2008-03-20

Information Environments and Communication

Library Visit: Kristin Yarmey

Karen Baker



## Information environments and communication UCSD/SIO - 20 Mar 2008



Archive Matters  
with Kristen Yarmey

### Information environments and communication



*Ocean Informatics events are held periodically as a forum for communication, cooperation, and camaraderie regarding all things data and design, informatics and information studies.*

### Agenda - Thursday

- 2:30-3:30 Round Table discussion
- 3:30-4:00 Ecosystems
  - EBM: ecosystem based management
  - IBM: infosystem based management
- 4:00-5:00 Ocean Informatics, Data and DataZoo
- 5:00-5:30 Wrap Up

### Sustaining the Dialogue

#### Participants

Kristen Yarmey - UMaryland;  
College of Lib & Info Studies  
Karen Baker - OI Information Manager  
James Conners - OI PA/IM/Design  
Mason Kortz - OI PA/IM/Design  
Lynn Yarmey - OI PA/IM/DM



Archivists.org

#### Memorable Moments:

Information, time, and boundaries: "It seems there are two world systems: active use and archive ... and preservation is becoming more inclusive.  
A view of repositories: "A repository is a general name for archives and special collections."

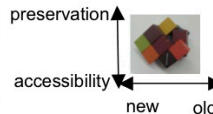
#### Issues

iSchools and archival programs  
Paper: Dear Mary Jane by Fleckner  
Focusing on the container or the content  
Authority headings and metadata tags

#### Vocabulary

Appraisal-cost to keep it  
Benign neglect-an approach

Consider a letter of George Washington and then a dataset....

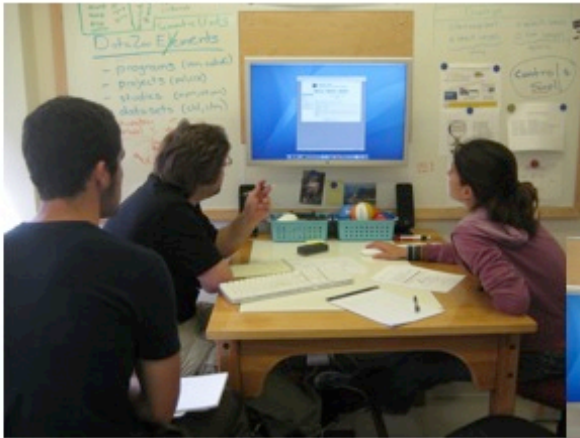


22

Event Date: 2008-04-02  
DataZoo and Classroom Use  
SIO Education  
Karen Baker



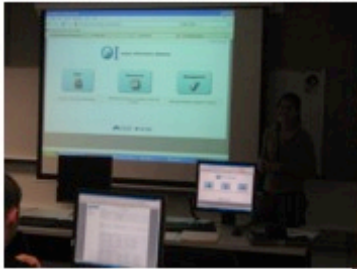
### Becoming familiar with DataZoo



**DataZoo is a multi-project, multi-agency site-level information system providing a variety of data services.**



### Taking DataZoo into the Classroom



**Data Access,  
Discovery,  
and Reuse**



Event Date: 2008-04-03

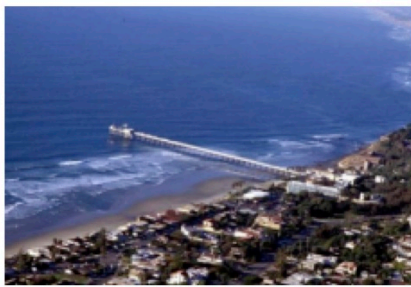
Ocean Informatics and Information Systems

WHOI Teleconference: Cyndy Chandler

Karen Baker

## Ocean Informatics and Information Systems

SIO-WHOI - 03 Apr 2008



Scripps Institution of Oceanography



Woods Hole Oceanographic Institute

*Ocean Informatics events are held periodically as a forum for communication, collaboration, and camaraderie regarding all things data and design, informatics and information studies.*

### Agenda - Thursday

- 8:00-8:10 Introductions (11ET)
- 8:10-8:15 What is Ocean Informatics? (11:10 ET)
- 8:15-8:20 What is an Information System? (11:15 ET)
- 8:20-9:20 What is DataZoo? (11:20 ET)
- 9:20-9:30 Wrap Up (12:20 ET)



### Sustaining the Dialogue

#### Participants

Karen Baker - SIO Information Manager  
 Mason Kortz - SIO PA/IM/Design  
 Lynn Yarmey - SIO PA/IM/DM

Chyndy Chandler - WHOI Data Manager  
 Bob Groman - WHOI Data Manager  
 Dicky Allison - WHOI Data Manager  
 Julie Allen - WHOI ApplWeb Programmer  
 Dave Glover - WHOI Data PI



Information System Architecture and Design



### Memorable Overviews:

**Which metadata standard is best to use?: "All of them"**  
**BCO-DMO Development:** JGOFS and map server to a new system with RDF triples, Drupal, and DataTracker.  
**OI DataZoo:** colocated with community; publishing system; dictionaries, term sets and IOD infrastructure





## Ocean Informatics Event UCSD/SIO - 15 May 2008

Conversations on Metadata  
with Inigo San Gil

### Considering information environments and systems

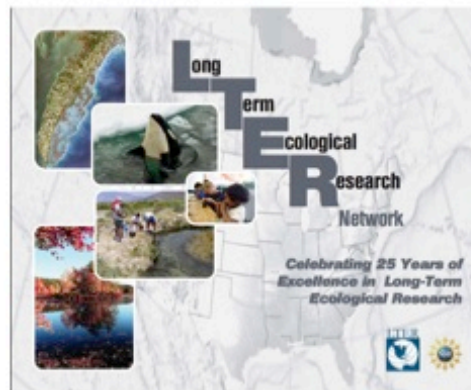


*Ocean Informatics events are held periodically as a forum for communication, collaboration, and camaraderie regarding all things data and design, informatics and information studies.*



Mobility and Tools

## Bringing Together NBII and LTER Efforts



Event Date: 2008-05-29

UC-LTER Graduate Student and Post-doc Symposium

LTER CCE, SBC, MCR

Karen Baker

## UC-LTER Graduate Student & Post-doc Symposium 29 May 2008 at SIO, UCSD



LTER sites represented at the meeting:

CCE – California Current Ecosystem

SBC – Santa Barbara Coastal

MCR – Moorea Coral Reef

**The purpose of the symposium was to allow graduate students and post-docs to share their LTER-related research with each other and the SIO/UC community.**

*There are several topics unique to LTER research in the marine environment (e.g. population connectivity by advective and biological processes, flexibility in the size structure of primary producers and grazers, carbon flow through the marine ecosystem, close relationships between physical forcing and ecosystem structure). Knowledge of the research conducted by other students at marine sites will facilitate cooperation and discussion. Students from each of the UC-LTER sites plan to continue meeting to share research on an annual basis. There were approximately 30 participants throughout the day's meeting in 4500 Hubbs Hall.*



CCE LTER graduate student representative, Ryan Rykaczewski, planned and opened the symposium.

### POSTER SESSION



A poster session was held in addition to presentations. Grad. students Darcy Taniguchi (right) and Andrew Taylor (left) discuss CCE plankton communities.

### AGENDA

- 9:30-12:15 Introductions to each LTER site and related research talks
- 12:15-12:45 Lunch
- 12:45-2:00 Poster session
- 2:00-3:30 CCE presentations, cont'd
- 3:30-3:45 Discussion
- 4:00 Symposium ends

26

Event Date: 2008-05-31

Information Management Cross-site Visit

LTER NTL visit: Barbara Benson

Karen Baker



Ocean Informatics Event  
UCSD/SIO 31 May 2008

Barbara Benson  
North Temperate Lakes LTER



Site-to-Site: NTL CCE PAL



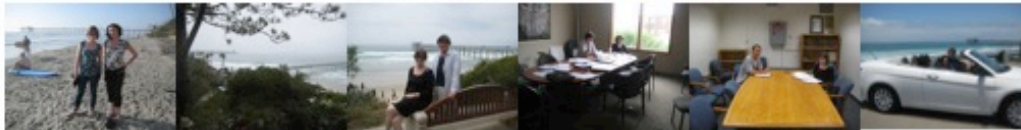
**Information Management Topics**

- Dynamic database access
- GLEON for instrumented sensors
- Considering controlled vocabularies
- Best practices
- Sharing code across sites
- QA/QC approaches & surprise theory
- Funding opportunities
- Processes for software development



Event Date: 2008-06-16  
 Ocean Informatics Monograph Write Session  
 Science Studies: Florence Millerand  
 Karen Baker

## Ocean Informatics Monograph Meeting 16-19 June 2008 - SIO, UCSD



### AGENDA

- |           |  |
|-----------|--|
| 13 Jun:   | arrival FM   |
| 14-15 Jun | CIP: OI at meetings: 4S & EIMC   |
| 15 Jun    | arrival CA   |
| 16 Jun    | OI Design Studio tour<br>DataZoo demo by JC<br>Lunch at SIO<br>DM/IM targeted analysis |
| 3-5       | Science Studies unColloquium   |
| 5-6       | UCSD Tour  |
| 17 Jun    | Coop ZooSpace demo by MK<br>Report Outlining & Diagram<br>Lunch at SIO                 |
| 3:30      | Graduate student re IM   |
| 18 Jun    | Report writing   |
| 19 Jun    | CA departure<br>J.Wanetick re OI<br>FM departure                                       |



### Meeting Activities:

- Gather existing Ocean Informatics ethnographic materials
- Review the current status of Ocean Informatics endeavors through demos and discussions
- Carry out joint analysis and synthesis of materials
- Complete report outline and begin monograph writing
- Meet with Science Studies Program participants and tour UCSD



Demoing:  
DataZoo and  
the Cooperative  
ZooSpace



Analysis & Synthesis:  
At the design table  
and in an analysis  
session



Event Date: 2008-07-17

Source code and Sociotechnical Programming Practices

Science Studies: Stephane Couture

Karen Baker



Design Studio Discussion



Design Table Demos & Dialogue

**Agenda**

Thursday  
10:30-11:00 Introductions  
11:00-12:00 Discussion: Thesis  
12:00-12:45 Demo DataZoo  
12:30- 1:30 Lunch  
1:30- 2:00 Tour IOD & Computational Infrastructure  
2:00- 2:30 Coffee  
2:30- 3:30 Demo DataSpace  
6:00 Dinner

Friday  
10:30-11:00 Working  
12:30- 1:30 Lunch  
01:30- 3:30 Aquarium  
6:30- TGIF Cliffside



**Some Discussion Topics**  
Beautiful code; functional code  
Code, structure, and architecture  
Design Studio activities  
Agile vs extreme programming  
Participatory Design  
Ethnographic open-ended interviews  
Tolerance  
[Ocean Informatics](#)

**Memorable Moments**  
"Style of programming expresses yourself rather than you ideas."  
"I would use 'elegant' - readability, brevity, and efficiency - rather than 'beautiful'."

- Participants**  
Stephane Couture - Graduate Student  
Karen Baker - OI Information Manager  
James Connors - OI PA/IM/Design  
Mason Kortz - OI PA/IM/Design  
Lynn Yarmey - OI PA/IM/DM  
Jerry Wanetick - OI PA/Systems/DM  
Nate Huffnagle - OI PA/Systems  
Tony Koslow - Director CalCOFI



Event Date: 2009-06-09

Regional Zooplankton Workshop

PaCOOS: Johnathan Phinney, Karen Baker, Sharon Mesick

Karen Baker



### PACOOS Zooplankton Workshop

#### Data and Information Infrastructure

for the CA Current Large Marine Ecosystem (CCLME)  
including Mexico, USA and Canada  
June 9-10, 2009, SIO/UCSD, La Jolla







Co-sponsored by  
Jonathan Phinney, SWFSC/NOAA  
Karen Baker, SIO/UCSD  
Sharon Mesick, NCCDC/NOAA

**Pacific Coast Ocean Observing System - <http://pacoos.org>**

**PaCOOS Zooplankton Workshop: Data and Information Infrastructure**  
**AGENDA - June 9-10 2009**

**Tuesday June 9, 2009**

**Session 1 (3.5 hrs) 8:30 – 12noon**

- Introductions and overview of meeting objectives (Jonathan Phinney)
- Presentations
  - Regional Ecosystems Data Management: using standards (Sharon Mesick, NCCDC)
  - Case Study: Zooplankton Atlas (Todd O'Brien)
  - The IOOS Data Integration Framework (Jeff de La Beaujardière, IOOS Program Office)
- Break
- Presentations
  - Review of existing community practices and standards (Karen Baker)
    - Zooplankton Data Organization and Online Delivery (Mark Ohman)
    - Zooplankton data Cross Comparative Work (Baldo Marinovic)
  - Review Survey Results (Jonathan Phinney; Karen Baker)

**12noon-1PM Lunch Break**

**Session 2 (3.5 hrs) 1:00-4:30 PM**

- Working groups form and rotate
- Renssemble and review group findings

**Social Hour (5-7PM)**

**Wednesday June 10**

**Session 3: (3.5 hrs) 8:30-12noon**

- Recap prior session, synthesize results
- Summary and next steps for this group

**12noon-1PM Lunch Break** (lunch brought in on site)

Working Group Questions:  
 (1) Community Issues: Infrastructure requirements for PACOOS Network, e.g. What are zooplankton standards?  
 (2) Data Policy Issues: Zooplankton Information Sharing Goals, e.g. What are data sharing issues and responsibilities?  
 (3) Data Documentation Issues: Infrastructure Methods, Models and the Development Process, e.g. What are representative data types, characteristics, and metadata?

**Data in a Bottle**

Zooplankton research involves creating and maintaining collections of both physical samples as well as digital counts.




**Data from a Bottle**

**Highlights...**

- Large Marine Ecosystems: PACOOS, the California Current LME-scale community enabling science and facilitating collaborative development of products through data management activities and knowledge sharing.
- Biological data are complex and difficult to convert into well-described data objects given limitations in today's data standards developed for physical data.
- Focusing on partnership and horizontal data integration at the local, national and international levels.
- Envisioning a zooplankton community of scientists and data managers.
- Inclusion of data management into proposals explicitly for data work.





30

Event Date: 2009-06-15

Event Date: Ocean Informatics Exchange

Science Studies: Sonja Palfner

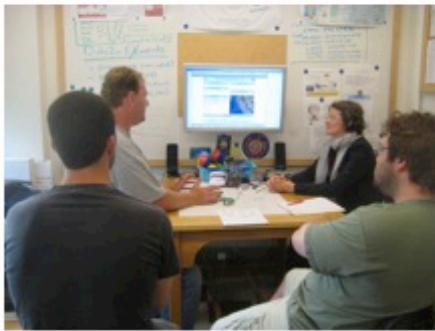
Karen Baker



## Ocean Informatics Event UCSD/SIO - 15 - 18 Jun 2009

Sonja Palfner

Technische Universität Darmstadt, Germany  
Cyberinfrastructure Grid



At the Design Table



Informatics Reading Group

### Agenda

#### Monday

Work overviews

#### Tuesday

Welcome Mark Ohman, CCE LeadPI  
OI Reading Group  
Lunch beach walk  
Demo DataZoo

#### Wednesday

Databits News article  
Aquarium Visit  
Boogie Boarding

#### Thursday

OI Reading Group  
OI Physical Infrastructure  
Supercomputer Center Visit

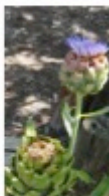
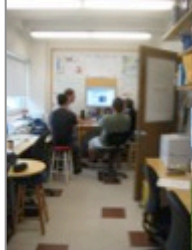


### Visit Highlights

Databits Publication: Cyberinfrastructure Travels: Sharing & Shaping Time, Space and Data. "In the case I am studying, the role of an information manager is not defined as part of the whole project. But the work must be done. So, that means the work has to be done by someone, this work that seems to be mostly invisible and thereby unacknowledged. These situations are often related to institutional framings; and indeed the power of institutions in distributing work and defining work roles should not be underestimated."

### Participants

Sonja Palfner - Visiting Researcher  
Karen Baker - PAL/CCE Information Manager  
James Conners - OI Information Architect  
Mason Kortz - OI Database Design  
Lynn Yarmey - OI Information Manager  
Tony Koslow - Director CalCOFI  
Mark Ohman - LeadPI CCE LTER  
Mathew Bietz - Science Studies Researcher  
Robert Peterson - OI REU, geophysics  
Sean Wiley - OI REU, computer science  
Anna Lara-Lopez - Zooplankton Post Doc  
Jerry Wanetick - OI Computer Infrastructure



Event Date:  
Visit: Nicole Kaplan  
Karen Baker



## Ocean Informatics Event UCSD/SIO - 13-16 Jul 2009

Nicole Kaplan  
LTER Cross-Site Visit  
SGS - CCE & PAL



At the Design Table



Informatics Reading Group

### Agenda

#### Monday

Proposals discussion

#### Tuesday

OI Design studio tour  
Welcome Mark Ohman - CCE LeadPI  
Informatics reading group  
Unit registry/dictionary discussion  
Design session LTER IM HistoryDB

#### Wednesday

Databits & Network News articles  
Lunch Tony Koslow, CalCOFI  
Governance work  
Aquarium visit & Shakespeare play

#### Thursday

Governance materials convergence  
Demo DataZoo & OI Architecture



### Reading Group Discussion

Grudin, 1988. Why CSCW Applications Fail

- Disparity between who does the work and who get the benefit
- Breakdown of intuitive decision-making
- Difficulties of evaluation

Ocean Informatics, 2009. How Design Might Succeed.

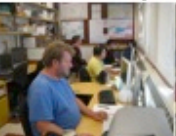
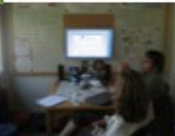
- Re-consider scoping & problem formulation
- Scaling via participatory/collaborative design
- Evaluation through continuing design

### Visit Note

A history of the LTER IMC at best is like a governance play book that provides insight into options and past actions, into outcomes and justifications.

### Participants

Nicole Kaplan - SGS Information Manager  
Karen Baker - PAL/CCE Information Manager  
James Conners - OI Information Architect  
Mason Kortz - OI Database Design  
Lynn Yarmey - OI Information Manager  
Tony Koslow - Director CalCOFI  
Mark Ohman - LeadPI CCE LTER  
Robert Peterson - OI REU, geophysics  
Sean Wiley - OI REU, computer science  
James Hayes - OI Undergraduate, history





32

Event Date: 2010-03-06

Information Exchange and Information System Elements

LTER MCR Visit: Mary Gastil

Karen Baker



Ocean Informatics Event  
UCSD/SIO 6 Mar 2010

M. Gastil-Buhl  
Moorea Coral Reef LTER



MCR



CCE



PAL



**Information Management Topics**

Perspectives:  
Cruise and time series data  
Enterprise and prototype systems  
IM in- a-box and Integrative Information Environments

Digital Moorea: biocode, genetic materials

Data packages and validation tools




Event Date: 2010-05-24

LTERR Unit Registry

LTERR KBS Visit: Sven Bohm



Karen Baker



## Ocean Informatics Event

### UCSD/SIO 24 - 27 May 2010

Sven Bohm  
Kellogg Biological Station KBS  
LTERR Unit Registry



At the design table

### Site-to-Site



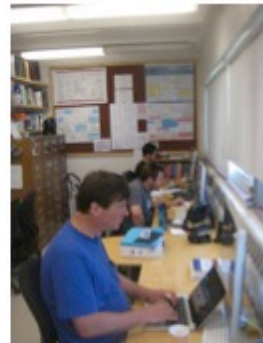
KBS



CCE



PAL



At the keyboard

### Agenda

Monday

- Review agenda
- Review unit query and management interfaces
- Scope KBS unit registry client
- Discuss REST implementation decisions

Tuesday

- Review feedback for updated web service features
- Ingest KBS units to dictionary
- Browse and discuss site web sites

Wednesday

- Review new features for registry web service
- Visit Ocean Informatics Computational Services
- Discuss LTERR web service standards

Thursday

- Compare Information system elements

### Visit Highlights

- 19 KBS custom units added to Unit Registry
- KBS unit table redesigned to pull information from Unit Registry
- KBS online metadata form using Unit Registry for unit selection (see image below)
- New and updated features added to Unit Registry (changelog, media type precedence)
- Best Practices for adding, updating, and deprecating of units revised



Name	weight	description	
DATE	1	harvest date	angstrom
Trt	2	treatment	are
Rep	3	replicate	atmosphere
Crop	4	species or group sampl	bar
Yield-bu_A	5	crop kernel/seed harv	becquerel
Yield-kg_ha	6	crop kernel/seed harv	britishThermalUnit
			bushel
			bushelPerAcre
			bushelsPerAcre deprecated: bushelPerAcre)
			calorie
			candela

KBS use of web service in metadata form dropdown




Event Date: 2010-08-01

LTER Unit Registry

LTER SEV Visit: Ken Ramsey

Karen Baker





## Ocean Informatics Event

### UCSD/SIO 3-5 Aug 2010

Ken Ramsey, Justin Jensen

Jornada Basin LTER  
LTER Unit Registry







At the design table

### Site-to-Site



JRN



CCE



PAL

### Agenda

Tuesday

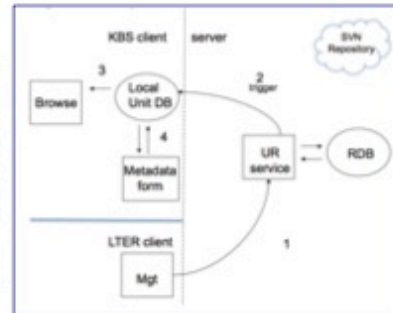
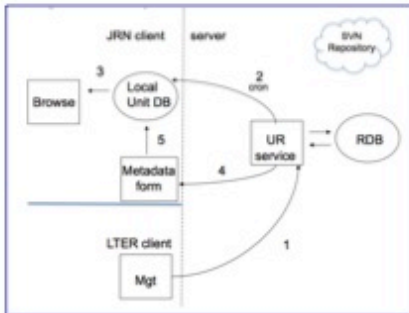
Review existing docs and management interfaces  
Design focus: scope of coding & products  
Development focus: prototype code

Wednesday

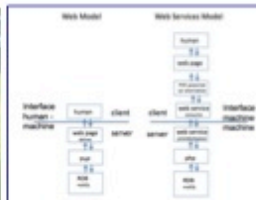
Continue design and development  
Site-Site exchange and web visits  
Browse and discuss site web sites

Thursday

Review registry web service client application  
Visit Computational Infrastructure  
Pier walk



The web service resides on a server at LNO with the code in a SVN repository. The diagrams illustrate the two different Unit Registry web service clients implemented by JRN and KBS.



Event Date: 2010-11-12

Units and Governance

LTER LUQ Visit: Eda Melendez

New Karen Baker



**Ocean Informatics Event**  
**UCSD/SIO 12-13 Nov 2010**

Eda Melendez-Colom  
 Luquillo LTER






**At the design table**

**Site-to-Site**



**LUQ**



**CCE**



**PAL**

**Agenda**

Friday  
 SIO Information Management meeting  
 Ingestion of LUQ Units  
 Site-Site exchange and web services  
 Drupal Framework & Unit Registry

Saturday  
 Governance Working Group Terms of Reference  
 Governance Working Group Databits article  
 Signature Datasets Databits article  
 Good Read Databits





From the rain forests of Puerto Rico to the Eifel Tower in Paris to the field station at AND LTER for a GIS meeting prior to stopping at the seaside at Scripps Institution of Oceanography (University of California San Diego), this visit was filled with site-to-site exchange on information management issues relating to units, web services, and governance.






36

Event Date: 2010-12-04  
Information Systems  
LTER MCR Visit: Mary Gastil  
Karen Baker

## Ocean Informatics Event UCSD/SIO 3 Dec 2010

M. Gastil-Buhl  
*Moorea Coral Reef LTER*

**Discussing Information Systems, Practices, and Standards in the Design Studio**

### Articulation Work & Developing Stories



- IM vision, research and data practices
- Language and sociotechnical issues
- Lessons learned & assessment strategies
- Concepts of standards & standard making
- Attribute characterization
- Databases and dataspace
- Infrastructure growth & local-level information environments

### Elements of Infrastructure

### Site-to-Site

-  MCR
-  CCE
-  PAL

### Agenda for a MicroRetreat

Friday morning  
DataZoo and information systems  
-Data access layer  
-Table templates design  
-Management interfaces  
Information system models  
Lunch walk to Cheese Shop



Friday afternoon  
Site-site exchange  
Historical perspective  
Computational infrastructure tour  
DataBits history & guide



Friday Evening/Saturday Morning  
Dichotomies: universal - situated  
Generic ERD metadata model for LTER systems

37

Event Date: 2010-12-10

Site-Site Discussion

LTER CAP Visit: Philip Tarrant

Karen Baker



Ocean Informatics Event  
UCSD/SIO 02 May 2008

Philip Tarrant  
*CAP LTER*

Information management approaches  
and process improvement work



Site-to-Site

- CAP
- CCE
- PAL

World class solutions  
... and mythbusters



Ocean-side  
lunch



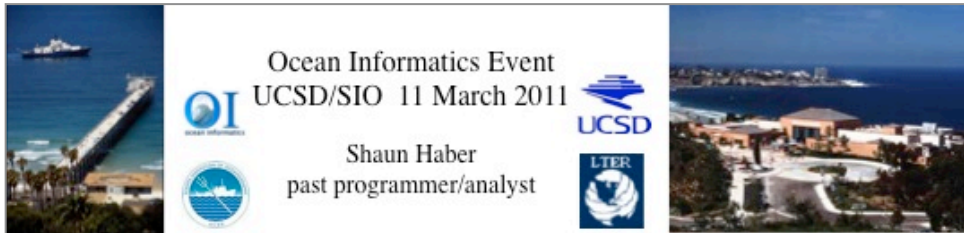
38

Event Date: 2011-03-11

Music, Business and Scientific Digital Delivery

Ocean Informatics: Shaun Haber

Karen Baker



Back in the design studio after walking to the Cheese Shop



Web programmer on the move:

