Ocean Informatics: Information Management in Practice

Karen S. Baker[1] and Cynthia L. Chandler[2]

1 Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093-0218, USA

2 Woods Hole Oceanographic Institution, MS 43, Woods Hole, MA 02543 USA

Abstract

Technological changes during the latter part of the twentieth century in combination with an increase in interdisciplinary, global, ocean science research initiatives are driving a new understanding of and appreciation for the role of information management as an integral part of scientific research. Technological capabilities in both computer science and communications have enabled collection and processing of data sets that are increasing in both volume and complexity. Changing expectations of data availability require data and information managers to collaborate with a diverse team of professionals to develop the infrastructures required to support current research programs with requirements for multi-scale data integration and interdisciplinary synthesis efforts. Examples of information management in practice are drawn from two long-term, oceanographic programs, the Palmer Station Long-Term Ecological Research program and the U.S. Joint Global Ocean Flux Study. Reflection on the changing role of information management and the importance of developing the requisite infrastructure, leads to discussion of Ocean Informatics as the application of informatics to the domain of Ocean Science, work that occurs at the intersection of oceanography, social science and information science. While focusing on data set collection and information system building, we are exploring interdisciplinary mechanisms and strategies that create the learning environment required for creating and sustaining a global data collection of measurements representing the natural world.

1. Information ecology: *from data availability to interoperability*

Studying ecosystems that change over decadal time frames brings forward data management issues such as data storage and access as well as data use and integration. Traditionally data has been available formally through publications and informally through personal exchanges. With the advent of affordable digital storage and the Internet, some data are available over the long-term beyond a single project or an individual investigator's career through submission to national archives. The relatively recent practice of institutionalizing data submission and reuse via local digital repositories brings with it a number of changes in organizational arrangements and resource uses, in perspectives and policies, as well as in roles and collaborative strategies (Star and Ruhleder, 1996; Birholtz and Bietz, 2003; Zimmerman, 2003). The notion of data integration is only now beginning to be unpacked from a monolithic technical concept into a multifaceted set of interrelated issues involving data types, structures, and standards relating to information infrastructures, work practices, and community processes (e.g. Kling and Jewett, 1994; Sheth, 1999; NSB, 2005).

Today's work with multi scale data and broader synthetic efforts is changing the focus of local data management (data capture and use) to community information management (data capture, archive, and access) in order to meet the need for data integration and federation. Figure 1 illustrates this transition from individual or single project data management to a multi-scientist, multi-project information management paradigm. Data comes in many forms - streamed or point samples, gridded or irregularly sampled - in space and time. Typical examples of integrating mechanisms include domain catalogues for data discovery, coordinated dictionaries for data inter-comparability and metadata standards providing data contextualization. All of these support automated use.

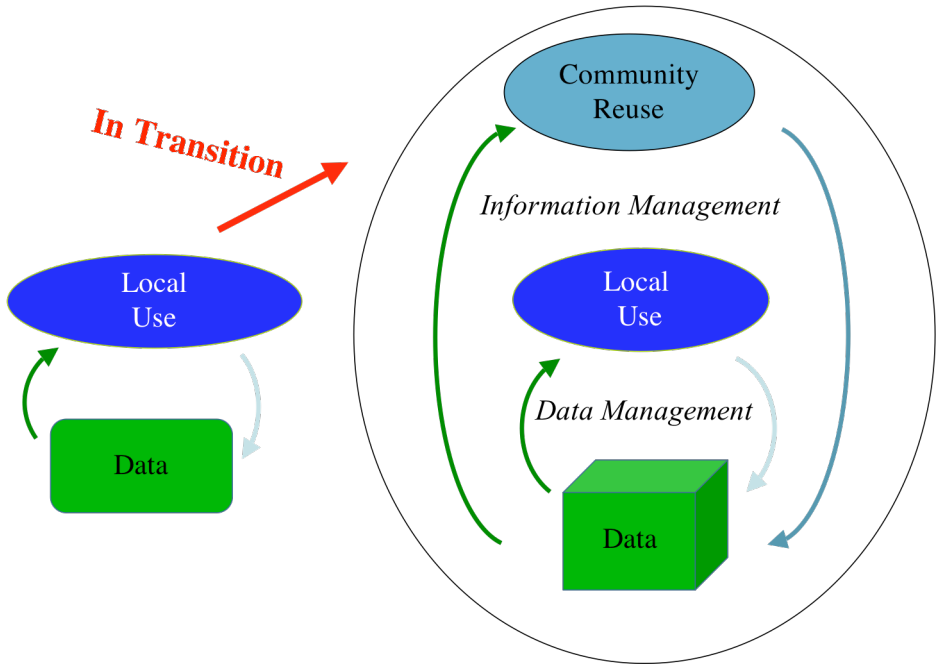

Figure 1. Scientific data practices are in transition. The under-appreciated, ongoing transition from data use to data reuse. In this example, local data management is concerned with local data use but community participation brings broader information management concerns such as providing context through common language, semantic relations, and community standards.

In addition to facilitating data integration, there is discussion and increased effort toward development of interoperable data systems. The IEEE Standard Computer Dictionary (1990) defines interoperability as the ability of two or more systems or components to exchange information and to use the information that has been exchanged. We use the term to describe both technical (syntactic and structural) as well as semantic interoperability, while noting that often one of the system 'components' is a person. Syntactic and structural interoperability are concerned with the technical aspects of representation and exchange of information (e.g. organization and format of data and metadata) (Veltman, 2001; Visser et al., 2000), while semantic interoperability has to do with the meanings embedded in the information exchanged, and the ability to accurately interpret those meanings (Friesen, 2002).

The past decade of expanding expectations coupled with advances in technology has prompted an ongoing period of problem formulation in the oceanographic arena in terms of elements and requirements for next generation community information systems (Thorley and Trathan, 1994; Baker et al., 2000; Brunt et al., 2002; Chandler, 2004; Glover et al., 2006). The design of controlled vocabulary in coordination with the ecological, oceanographic, digital library, and information system communities is preparing communities for the significant epistemological and ontological issues that arise with data synthesis efforts, both human and automated (Ribes and Bowker, 2005; Gruber, 1993; Guarino and Welty, 2000; Smith, 2003; Smith and Welty, 2001).

Issues of data provenance and data governance emerge when long-term data can be made readily available. Data provenance is concerned with documenting what has been done to the data and by

whom. Within the realms of information science and informatics, such topics are the subject of investigation and inquiry (Greenwood et al., 2005; Simmhan et al., 2005). Our understanding of data provenance is deepening as forums develop to compare and contrast what is done to the data in a variety of locations and circumstances. Consideration must be given to how data transformation from raw to calibrated to product-specific is documented along with who is responsible and where modifications occur. Data governance, a phrase used to describe who is responsible for data at its various stages, involves both community strategies and domain agreements. The language in general and definitions in particular for differing dataset types and products are being worked out now although existing arrangements influence current work practices. In making data accessible and interoperable, reliability and quality depend upon attention to both data provenance and data governance concerns.

Experience with the first generation Palmer Station LTER data system (Baker, 1998) and the U.S. JGOFS data management system (Glover et al., 2006) provides an opportunity to reflect upon information management and community information infrastructure. Considering collaboratively the lessons learned over a decade of data management experience, system development and use within the Palmer Station LTER and U.S. JGOFS communities draws out cross-program insights that can inform future design efforts.

2. Background

A master directory is a catalog providing a description of data and its location in a standardized format. It provides an example of a contemporary web-based organizational structure to support scientific data discovery and thus data reuse. Tagged elements describing the data constitute the metadata. Of particular note, the Global Change Master Directory (GCMD) includes a regional organizationing structure, e.g. the Antarctic Master Directory portal (AMD; http://gcmd.nasa.gov/Data/portal_index.html) for the Southern Ocean. These directories are organized by topic though searchable by locations, instruments, platforms and projects as well as by free text.

Data Description for Multiple Contexts

Metadata is key to data discovery, integration and query (Michener and Brunt, 2000; Cook et al., 2001). Describing data in anticipation of usage within multiple contexts remains an elusive enterprise and the subject of ongoing work in many communities. This work is often conducted in parallel with the ontological work of identifying relations between data elements. One approach to metadata is to consider it a process of documentation. Developing extensive metadata in order to enable automated access and query, the Federal Geographic Data Committee (FGDC) approved a metadata content standard in 1998 for geospatial data and a biological data profile in 1999. Many disciplines recognize and are addressing the need for community specific metadata standards. Such standards shape and provide guidance about data description content to networks of local repositories that complement existing national archives such as the discipline specific US National Oceanographic Data Center (NODC; http://www.nodc.noaa.gov). In addition to content standards, there are protocols for transporting data across a digital network such as the OPeN Data Access Protocol (OPeNDAP, Cornillon et al., 2003) and data delivery environments such as the UniData

Thematic Realtime Environmental Distributed Data Services (Domenico et al., 2002; http://www.unidata.ucar.edu). Further, data coordination efforts are informed and supported by standards developed by committees such as the International Standards Organization (ISO; http://www.iso.org) addressing a range of information topics including geospatial and temporal information. In response to the array of new information needs, community specific organizations have emerged. The National Center for Ecosystem Analysis and Synthesis (NCEAS, http://nceas.ucsb.edu) supports ecological data synthesis and research while the nascent Marine Metadata Interoperability Project (MMI, http://marinemetadata.org/) provides a community forum for sharing robust marine metadata practices. In addition, QARTOD (http://www.qartod.org) is a multi-agency effort to establish guidelines for quality assurance of real-time oceanographic data. Each of these initiatives involves long-term collaborative efforts focusing on informatics issues such as classification agreements and standards development within and between scientific fields. Such efforts not only influence existing practices, but they require a new type of conceptual and semantic readiness in addition to an appreciation of process building. For example, there may be new metadata forms to fill out, but there are also established names to map and new metadata elements to define.

Oceanography: an Earth System Science

Oceanography began as a global science with its first oceanographic expedition, the three and a half year voyage of the H.M.S. Challenger (1872-1876) (Oreskes, 2002). As an earth system science today, it has grown from laboratory and field studies, empirical and theoretical research, and species and population investigations to include disciplinary, interdisciplinary and ecosystem approaches. From the days of the International Biological Program (IBP, 1964-1974) and subsequently with the Long-Term Ecological Research Program (LTER, 1980-ongoing), ecological science has managed synergistic component studies (from bacteria to primary producers to predators) with whole system views of material and energy flows through ecosystems. More than four decades after the International Geophysical Year (IGY 1957-1958) prompted a suite of global activities, a variety of multi-year and multi-sited global ocean science research projects have been initiated. In the 1980s and 1990s this included the Joint Global Ocean Flux Study (JGOFS) and the GLOBal ocean ECosystem dynamics (GLOBEC, http://www.globec.org/) project, both sponsored by the Scientific Committee on Oceanic Research (SCOR, 1987) and identified as core programs of the International Geosphere-Biosphere Programme (IGBP). Efforts at coordinating regional, computer-mediated partnerships focusing on the oceans include the recent NSF sponsored Ocean Research Interactive Observatory Networks (ORION, see http://coreocean.org) and the NOAA program for Coastal Ocean Observing Systems (COOS, http://www.csc.noaa.gov/coos) that coordinates globally with the Global Ocean Observatory System (GOOS; http://www.ioc-goos.org/). Large-scale, interdisciplinary efforts in the earth sciences are developing with the United States Long-Term Ecological Network (http://lternet.edu), the US continental scale National Environment Observatory Network (NEON, http://www.neoninc.org/) and within the Geosciences Network (GEON; http://www.geongrid.org). International coordination includes programs such as the International Long-Term Ecological Research Program (ILTER; http://ilternet.edu).

Collectively, these developments augment notions of the solitary scientist and the independent project, recognizing the practice of ocean science as a socially complex, globally distributed, and highly mediated form of distributed collaborative practice. In modeling programs from the U.S.

JGOFS Synthesis and Modeling Program (Doney et al., 2002) to the ongoing PARADIGM effort (Rothstein et al., 2006), there is a developing understanding of the need to consider multiple scales (including local and global), multiple time frames (from immediate to past and future), and multiple perspectives (from technical to social). The vision of ocean science collaboratories includes information technology used effectively to foster the technical, scientific, and social requirements of scientific collaboration (NRC, 1993).

Case Examples

LTER Information Management: The concept of the LTER Network emerged from the IBP Program (Golley, 1993; Smith, 1968) in terms of a community organization that could address ecological events occurring over multi-decadal timeframes in coordination across a variety of ecosystems (Hobbie et al., 2003). A network of individual LTER sites was established in 1981 along with a Network Office to serve as a communication hub. The network started with six sites and has grown to 26 sites with each site consisting of 10 to 40 investigators from a variety of disciplines collaborating to study a particular biome. As the program was starting, the NSF called for establishment of electronic networking between the sites, and data management was the topic of a workshop held in 1982 (Gorentz, 1992).

Following focus on long-term research in the 1980's and on larger-scale research in the 1990's, the LTER scientific community designated 2000-2010 the decade of synthetic research. The LTER information management committee recognized the need for data access and integration by endorsing and adopting the Ecological Metadata Language (EML). First released in 2002, EML was developed with NSF support as a specification using XML schema and supporting a domain catalog (Jones et al., 2001; McCartney and Jones, 2002). LTER sites are in the midst of designing methods to enact and use EML locally (Millerand and Bowker, forthcoming). The process of EML design, development, deployment, and enactment has played an important role in providing the LTER community (scientists, information managers, technologists, students) with an understanding of the under-appreciated concept of data interoperability. Interoperability is often perceived as binary or bimodal: data either is or is not interoperable. In practice, a continuum results from the differences inherent in scientific field data collection, including issues associated with heterogeneity in general and packaging, granularity, and semantics in particular (Cornillon et al., 2003). It is the work of science to consider and take into account these differences. In studies of complex biotic-abiotic earth systems, data differences are introduced by the many field sampling techniques representing differing scientific emphasis – from surveys and behavior studies to point samples to streamed data and so forth. Thus even data within a single domain are typically geographically, temporally, and/or methodologically disparate.

Information management and data sharing is planned as an integral part of LTER research site programs with the goal of developing sustainable infrastructure processes to facilitate the flow of data from field measurements to a local repository. In the case of the Palmer Station LTER (PAL; Smith et al., 1995; Ducklow et al., 2006), data are available online from a centralized repository designed in 1991 to provide local data access (http://pal.lternet.edu; Baker, 1998). To meet new requirements for data query and to prepare for federation efforts, the site has initiated design and deployment of an updated information system requisite for providing additional services such as

data query and exchange, taking a holistic approach captured by the concept of an information ecology (Davenport, 1997). Transition involves moving from local forms and static text file delivery to an open architecture web services model. As the role of data management expands into one of information management (Baker et al, 2000), a relational database approach incorporates multiple data dictionaries designed in collaboration with community partners. An informatics conceptual model, designed in partnership with social scientists with interests in science and technology studies and action research (Whyte, 1991), takes into account technical, organizational, and social factors and provides a framework for addressing the organizational, semantic and collaborative issues integral to today's larger scale scientific endeavors (Atkins, 2003; Finholt, 2004; NSF AC-ERE, 2003).

U.S. JGOFS Data Management: Two global-scale, oceanographic research programs of the 1980s (Tropical Ocean-Global Atmosphere (TOGA) study and World Ocean Circulation Experiment (WOCE)), designed to understand the role of the oceans in affecting climate change, were primarily focused on physical processes. Marine geochemists and biologists of that time felt that physical transport models alone would be insufficient to describe the oceans' role in climate change, specifically with respect to greenhouse gases and especially atmospheric carbon dioxide. Discussions at a 1982 NATO meeting on the chemistry of the upper ocean and a 1984 workshop (NAS, 1984; Brewer et al., 1986) provided the initial guidelines and scientific framework for a program initially called the Global Ocean Flux Study (GOFS). GOFS soon expanded into U.S. JGOFS - a larger, coordinated ocean science research program. In addition to being an element of the U.S. Global Change Research Program, U.S. JGOFS was also a component of the multi-national Joint Global Ocean Flux Study (JGOFS), sponsored by the Scientific Committee on Oceanic Research (SCOR) and the International Geosphere-Biosphere Programme (IGBP) (U.S. JGOFS Steering Committee, 1990; Fasham et al., 2001; NRC, 1999). The program infrastructure supported and encouraged biological, chemical, geological and physical oceanographers and modelers in an interdisciplinary investigation of the pools and fluxes of carbon and associated biogenic elements in the ocean (Buesseler, 2001). The U.S. JGOFS research program comprised six cooperating, complementary activity elements: a dedicated project and data management office, four basin-scale process studies, two long-term time-series programs, a global carbon dioxide (CO_2) survey, satellite observations of surface ocean pigment distribution and finally a synthesis and modeling project.

U.S. JGOFS Scientific Steering Committee members and U.S. National Science Foundation Ocean Sciences Division program managers recognized early on (NAS, 1984; U.S. GOFS Steering Committee, 1986) that a coordinated, multi-disciplinary, long-term research program would also require a data management strategy that addressed the needs of participating investigators as well as those of the overall program. A U.S. JGOFS data manager was identified in 1988, and the Data Management Office (DMO) was created in 1994. From the beginning, DMO staff members worked closely with investigators funded to conduct U.S. JGOFS related projects. Much of the collaboration was in the form of assistance with data parameter definitions as well as metadata creation and quality control of data contributed to the collection. The Data Management Office staff worked closely with investigators on the collection and subsequent publication of complete metadata for every data entity contributed by U.S. JGOFS investigators. In addition to geospatial and temporal metadata, data acquisition and processing methodology were made available. The

complex interdisciplinary investigations that were the hallmark of JGOFS were facilitated by the availability of metadata records.

All process study data were ingested into an object-oriented, relational database (Flierl et al., 1992; Glover, 2001) and made available via what was at the time the nascent Internet and eventually via the World Wide Web. The DMO also took responsibility for final contribution of all data to the National Oceanographic Data Center (NODC) as well as for publication of the final data report (http://usjgofs.whoi.edu/publications/FinalDataRpt.html). As the U.S. JGOFS research program transitioned from the process-oriented field studies to modeling (Sarmiento and Armstrong, 1997), the data system was extended to include a customized Live Access Server (Hankin et al., 1998). Synthesis and model results, larger in volume and often global in scope as opposed to basin-specific (Doney et al., 2002), required a more graphically oriented user interface, and extended visualization capabilities (Glover and Chandler, 2001). DMO staff worked closely with investigators to provide timely availability of data during the active research phase and to ensure preservation of the completed data collection as an important part of the JGOFS legacy.

Cross case Observations

Common sampling grid

For both PAL LTER and U.S. JGOFS projects, research conducted aboard sea-going vessels has involved pre-cruise development of a shared sampling scheme. The logistics of planning for use of a shared platform initiates a geographic and temporal coordination that carries through to the data and to relations in a database. Project sampling arrangements play an important role in oceanographic fieldwork. PAL LTER defined a geographic grid of standard stations that has been sampled annually since 1991 (Waters and Smith, 1992); the U.S. JGOFS community of investigators defined standard station locations for one of the four basin studies in order to facilitate integration of data from cruises. Cooperative planning of cruise sampling strategies initiates cross-study discussions, creates a shared understanding of measurements, and informs subsequent data handling.

Event log

There is a tradition of keeping a ship's log with periodic notations of ship location and sea conditions. Scientists have found it is useful to extend this tradition to include recording events in what has been designated an 'event log'. This task is frequently carried out with the help of bridge personnel who are responsible for overseeing all over-the-side instrument deployments. In an event log, each sampling event during a research cruise is assigned a unique identifier that serves as a coordinating index between events and ultimately as a relational index in a database. In the absence of an event log, seemingly small differences in how data are gathered in the field – from unsynchronized clocks to differing station-naming conventions - become progressively difficult to reconcile as the ship returns to port and data is carried by investigators to different laboratories. Further, 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. A unique time and location stamp recorded with an event number or perhaps encoded within it becomes part of an ordered cruise history.

Dictionaries and Methods

The process of fully describing data necessitates creation of a parameter or attribute dictionary. The concept of a controlled vocabulary with common parameter names was eventually initiated in both programs although no general community standard exists. Dictionaries consist of names associated with parameter type, sampling specifics (e.g. such as methods) as part of the definition and units of measurement. For example, Palmer defines nitrateDissolvedInorganic_uM/l as an attribute name with two associated short names, nitrate and NO3. Associated with this attribute, is a definition, storage type, measurement scale, number type and precision. For U.S. JGOFS, NO3 is the parameter name representing Nitrate as determined by the methodology described in the core measurement sampling protocols (UNESCO, 1994) and reported in micromoles per liter. In both programs, controlled vocabularies evolved in order to define each parameter name uniquely.

Lessons Learned

Experience garnered during Palmer LTER and U.S. JGOFS has yielded many 'lessons learned'. Twelve of these lessons have been selected for discussion below and will be related to lessons summarized by earlier communities (Glover et al., 2006; Stonebraker, 1994; Spencer, 2006)

In Practice: Data Management Work

1. Collaborative Research: Research is ongoing with respect to what constitutes effective collaborative science. New fields have emerged as scientific collaboration has grown, e.g. information systems design, collaboratories, and computer supported cooperative work (Khazanchi and Munkvold, 2000; Gasson, 1998; Finholt, 2004; Schmidt and Simone, 1996). Community awareness, communication, and articulation with respect to data practices and process building are all important to defining and establishing robust data management practices in a community arena. Within the LTER, each site's data manager is a member of the LTER Information Management Committee. This forum represents a community-of-practice, a group that meets regularly and shares issues and activities in addition to building common understandings. In a recently formulated review criteria for LTER site information management, an annual review of the system is recommended. A data management advisory committee is one type of organizational structure. Thorley and Trathan (1993) spoke early on of the need for a clearly defined management structure with allocated roles and responsibilities. JGOFS established a Data Management Task Team (DMTT) while PAL LTER created a DM Advisory Committee comprised of a data manager and scientific investigators.

2. Data Policy: Although agency data policies may exist such as the NSF requirement for access to data within two years of the time it is collected, collaborative development of a site data policy addressing agency, project, institution, and international issues is critical for community data management efforts. Policy-making represents an opportunity for data managers and scientists to consider the implications of data sharing and the mechanisms for meeting data management goals. A published data policy with respect to data use and with respect to acknowledgement of use serves to align expectations of all members of collaborative groups within the project or program. U.S. JGOFS lesson 4.1 (Glover et al, 2006) noted that such policies represent a 'shift in culture'.

3. Data Description: Data description in the form of controlled vocabularies, shared dictionaries, registered ontologies and robust metadata records is critical to long-term data reuse and is enhanced if mechanisms exist for compliance with standards - although community, national, and international standards are in development. Metadata-enabled data discovery is a key to secondary use of project data, as well as the public accessibility of data that is understandable by those not affiliated with the project or investigator responsible for the original data acquisition.

4. Data Quality: Data quality assurance and quality control exist in a variety of forms including explicit, tacit, and contextually implicit. The work of quality assurance prior to data acquisition and quality control of acquired data is to review and document data procedures and to free data of errors. This process is frequently iterative, engaging instrumentation developers and researchers who utilize the data in conducting laboratory, field, modeling and theoretical investigations. The creation of integrated data products when started early in a project life cycle, serves as an important quality assurance agent. The products are an important diagnostic tool that is part of the ongoing process of ensuring final data quality.

In Practice: Information Management Work

5. Data Reuse: While data reuse is not a new concept, the scale of reuse has changed. Moving from the realm of an individual investigator using their own data or sharing data from a publication with colleagues, the complexity of data description is non-linear. As data travels further from its origin, the amount and types of data description required to define the data context increases. In addition, complete information about data sampling, handling and analysis practices is not traditionally articulated and does not travel with the data so it must be meticulously recorded as part of the metadata. Finally, data access involves a multitude of distributed users and user types, of managers and curators as well as developers and technologists. The roles associated with data are expanding in parallel with work emerging in the digital realm and identified from a variety of overlapping perspectives, e.g. from the arenas of informatics, information systems, information sciences, computer science, library science, information technology, infrastructure studies and science and technology studies (Abbott, 1988; Karasti and Baker, 2004, Baker and Bowker, 2005). Communication across and between all the data roles is critical to enhance data capture, preservation, use, and reuse.

6. Informatics Support: In today's changing digital landscape, it is important to build understanding of data and information management as science-driven processes involving new types of problem formulations associated with data system design and infrastructure building as well as with data analysis and use. Scientists are currently trained and supported to generate data in a format useful for hypothesis driven research. A widely held (if rarely voiced) consensus exists regarding the role of data and information management within the practice of ocean science as a whole: data work is seen as essentially supportive, a necessary but taken-for-granted prop to the central work of field observation, experimentation, and theory-building. Data management, providing a service in terms of data handling for local collection and analysis, emerges as a process with sensitivity to and respect for local diversity and particulars of collections. There is an

understanding that new insights arise from grappling with heterogeneity, with creating work-a-rounds for exceptions to fit within data structures.

7. Data Mediators: A team is needed to address the multiple facets of information management. Data stewardship for the full data lifecycle is concerned with project design through data collection, description, analysis, integration, exchange, usage, and preservation. Friedman (1989) provides taxonomies for some of the multiple roles that are part of a data team. One data management scenario is a centralized data office (U.S. JGOFS lesson 4.5: people). The LTER Network consists of a geographically distributed group of twenty-six data managers coordinated as a committee. Organizational arrangements provide support for data management through team communication and scientific oversight through communication with active investigators or a steering committee. Data management can be seen as more than 'science driven', operating in support of science and in practice with science. Ideally, data and information management includes involvement of investigators in problem formation and project scoping with respect to data repository development, interface construction and communication of practices through education and outreach efforts. Mechanisms must exist for timely communication through newsletters and more formal publications and personal contact at meetings and project-specific workshops.

8. Data Integration: Increasing expectations for integration of heterogeneous data types in support of today's scientific research are changing information management practices. There is increased effort aimed at automated data integration to replace labor-intensive manual data integration techniques. Identifying and building scientific and project-based data products is often an important component of data integration and collaborative work. The combination of unique event numbers and complete metadata from an event log enables generation of 'merged data products'. Cruise-level merged data products from like sampling devices can then be combined. In U.S. JGOFS, a single integrated product was formed for each basin (e.g. all data reported from the Arabian Sea Niskin bottle samples).

In Time: Sustainability

9. Local Repositories: The aim of local repositories is availability of data collections over time. Establishment of local repositories facilitates the data quality control process and further enables local knowledge-building. Local repositories and associated personnel represent an interface between a diversity of individual investigators or projects and the national archives that address longer-term requirements of permanent data archive. Distributed data management organizations are in a position to build relationships of trust between community members through shared practices, experiences and joint planning, e.g. a vibrant and collaborative community of practice.

10. Adaptive Strategies: The aim of adaptive strategies is sustainability of a community information system. Ever-changing technology, organizational structures, and personnel frequently introduce under-appreciated ramifications. A U.S. JGOFS data management timeline (U.S. JGOFS lesson 4.6: This is hard) provides an illustration of the backdrop of changes in technology and personnel during their data efforts. A comparative, modular and extensible design approach with an eye to data migration enables incorporation of new technologies and techniques. In terms of social

adaptation, an environment supporting long-term information infrastructure-building, community-building, and continuing learning is one that fosters local participant engagement and innovation.

11. Information Infrastructures: A primary aim of information management infrastructures is sustainability of data flows. Careful construction of the local infrastructure as part of an ongoing process is required to support both short-term communications and data work as well as long-term data access and cultural memory instantiated as a digital record. Stonebraker (1994) described infrastructure as necessary, time-consuming and very expensive in reflecting upon the first three years of the multi-year Project Sequoia. Star and Bowker (2002) capture the activeness of the concept through use of the verb 'to infrastructure'. More recently Spencer et al. (2006) have described cyberinfrastructure as a process, a living entity. Multiple intertwined facets of infrastructures include social, cultural, human, technical, procedural (e.g. protocols and tools), mechanical, historical, and data system support. As infrastructures evolve, there is a continuous negotiation and alignment required in response to and in support of the environment.

12. Information Environment: One important aim of an information environment is sustainability through continuing mutual learning. The challenge and pleasure of striving to represent the world (natural and/or cultural) in digital form requires design, development and support of new types of information environments. Co-evolution of collaborative cultures is an important part of the development process. An informatics environment includes multiple infrastructures, simultaneously created and utilized in a co-dependent evolutionary process. Technological advances have driven these changes to some degree, but it is the interdisciplinary requirements of modern scientific endeavor that are the important driving factor.

3. Ocean Informatics

Today, in the United States, "informatics" is used in a variety of senses often associated with data management, computer science, information science, information technology, information systems design, communication science, human computer interface, societal interactions with all of the proceeding elements and the research science striving to observe the processes inherent in all these components. We take informatics as the application of information science and the use of information technology in ways that promote communication and incorporate organizational and social interfaces. When conjoined with a specific domain, we suggest it holds the potential to enable new approaches to information flow. "Ocean Informatics" is used by us both in the broader meaning discussed above and also to highlight some of the aspects of oceanographic data management that make it distinct from data management in other sciences (Baker et al., 2005).

Ocean Informatics is the application of informatics to the domain of Ocean Science; the work that occurs at the intersection of oceanography, social science and information science. It is a conceptual framework bringing together theory and practice for those working with ocean data. Ocean informatics itself is embedded in an environment of research-participation-education-training inhabited and shaped by the data and its structure, information users and their partnerships, information systems designers and managers, side by side with students and their teachers. The

specific goal of Ocean Informatics is to create an environment wherein we design, develop, and enact data and information management products and processes in order to 1) make oceanographic data available to the scientific community according to mutually agreed upon requirements and in support of interdisciplinary research and 2) to enable better understanding of the data and information at hand. Understandings of the notions of both design and of articulation are central to the ocean informatics approach (Jackson and Baker, 2004; Baker and Millerand, in press).

Data support within oceanography has traditionally been performed on a project-, collection-, and in some cases cruise-specific basis, with few efforts to establish common cross-collection platforms and community-wide protocols. Collections are held and managed primarily as local entities with bridges to datasets housed at other ocean research centers built and maintained on a more or less ad hoc basis. Resource constraints contribute to the way things develop. Funding for information management in the largely soft-money world of ocean science has been secured on a project-by-project basis producing an organizational barrier to cross-project data integration. Ideally, data access systems could be designed to dissolve socio-political barriers to collaborative research.

A number of characteristics of data and of data handling arise and are the subject of ongoing research, e.g. data classifications and data integration, data heterogeneity and system complexity as well as infrastructure. Interdisciplinary work involves accounting for the semantics of multiple domains as well as developing collaborative work mechanisms. Heterogeneity describes multiple sampling scales and data types but incorporates a variety of approaches, organizations, and technologies involved. There is an interdependence of data and its supporting environment in terms of technologies and of institutions. Complexity arises in terms of digital representational schemes as well as in time frames of community conceptual readiness (Kaplan and Seebeck, 2001).

There is a tendency to take a technologically optimistic view of data handling in spite of the complex issues of collaborative science, data reuse, data findability, and data queriability that are the subject of active research. Perhaps because of this tendency, we have constructed for ourselves a constant reminder about the work of information management and information systems as a role that blends the logic of earth sciences and the interpretive skills involved in data contextualization. Our reminder takes the form of a gentle manifesto inspired by Venturi (1966):

A Gentle Manifesto for Data Management

> We celebrate complexity and contradiction in the organization of data. We do not like the incoherence or arbitrariness of incompetent information architecture nor the precious intricacies of uncontextualized information systems. Instead, we speak of a complex and contradictory array of flexible local arrangements resulting from the articulation of relations between local practices and community standards. We facilitate emergent design processes as part of experiential learning that elicits and preserves the richness and ambiguity of data collecting experience, including that experience which is inherent in managing data within communities.

4. Conceptual Frameworks

Infrastructure

Infrastructure in the digital realm involves issues of computational resources, networks and standards, while in the cultural realm involves organizational arrangements, community activities and educational supports. Discussions in Science and Technologies Studies focus on infrastructure as an emergent field of study (Star and Bowker, 2002; Edwards et al., 2006). There are inherent difficulties in planning for small data tables and large data streams, for data ingestion and data delivery, as well as for the expected and the emergent. In addition, Figure 2 captures the multiple time frames to be taken into account in planning for data collection. The figure depicts the coexistence of multiple scales: the short-term concerns of data collection and project coordination that require concurrent balancing with long-term concerns for preservation and reuse.

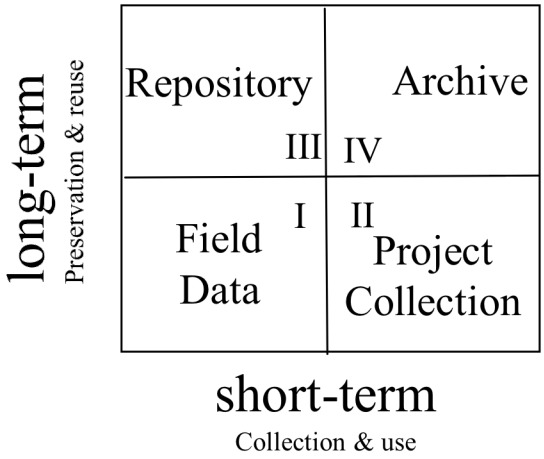

Figure 2. The simultaneous interplay of short-term and long-term in data timeframes that take into account both data use and data reuse.

Common data space

Recent database work opens up the data management landscape conceptually - from databases to dataspaces (Franklin et al., 2005). The 'dataspaces' approach is presented as both a new agenda and an architecture that allows for multiple ways of addressing data issues and framing questions of information management. In addition, development is recognized as taking place over multiple timescales: "One of the key properties of dataspaces is that semantic integration evolves over time and only where needed. The most scarce resource available for semantic integration is human attention." In addition to making the role of mediation visible, the dataspace concept umbrellas explicitly a continuum of organizational and semantic arrangements that handle diverse data types, states and approaches. Franklin et al. (2005) describe the value of bringing data collections together, even if just in proximity and loose association: "Dataspaces are not a data integration approach; rather, they are more of a data co-existence approach". As we seek tighter collaborative configurations in both ecological science and informatics endeavors, the dataspaces concept highlights the notion of a development process over time for data and information management, for theory and practice.

The proliferation of data repositories and the need for their interface, brings into focus requirements for compatible data transport protocols and for data standards. Mechanisms for data transport and database federation are the subject of discussions that depend upon development of new, shared vocabularies. Research is only beginning to delve into classification analysis and semantics in an effort to learn by identifying differences. Sensitivity to difference in turn may foster recognition of the value of inclusiveness with respect to elements outside the norm.

Common information space

A scientific community is a community by virtue of member agreement about overarching goals and through member participation in planning and shared activities. Technical aspects of a community may include shared digital spaces in the form of web pages, project portals and shared disks. Organizational elements frequently include project bibliographies, controlled vocabularies and ontologies. Social elements include annual meetings, shared events and ongoing dialogue.

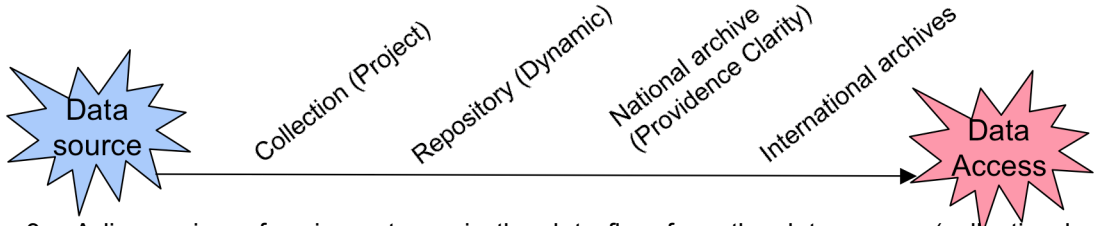

Figure 3a. A linear view of various stages in the data flow from the data source (collection by scientists) to data availability to the scientific community and access to the public.

Data flow may initially be thought of as linear – from data collection to data storage to data access as illustrated in Figure 3a. We suggest that data access occurs at points all along 'the line'. A model of the data source nested within the realms of projects, repositories, national and international archives provides an alternative representation of the multiple points of access and of the complex relations between data at any point in the life cycle (Figure 3b).

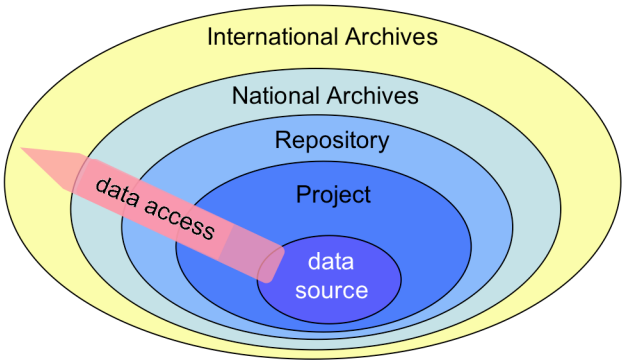

Figure 3b. A nonlinear view of data flow where access occurs all along the line or during the life cycle of the data.

An example of the cross-over between efforts is given by the interdisciplinary, multi-year experiment conducted at the Konza Prairie Research Station. Local repository efforts were augmented by collaboration with the Oakridge National Laboratory DAAC, a member of the multi-node NASA data consortium (Strebel et al., 1989). Such an example, suggests the larger information space shown in Figure 4 where data flows to local repositories but moves to domain centers, national archives and/or directly to the public. Between these various 'users' of the data, there are frequent bidirectional feedbacks.

Driven by the needs of science for combining heterogeneous data and data types integrated over multiple time frames, Ocean Informatics creates a conceptual framework with an explicit goal: an information environment that supports data management, information systems design and a multi-faceted infrastructure. From the data perspective, local management, data quality and data integration efforts are taken into account. From the information systems perspective, data relations, shared dataspaces and design processes are fundamental concepts. Finally, an understanding of

infrastructure as intertwined technical, social and organizational elements, spurs new approaches to collaboration, community and shared learning.

Interdisciplinary research programs are by their nature more heterogeneous with respect to content, geography, personnel and vocabulary. All of these factors conspire to complicate the necessary collaboration. The trend toward increasingly diverse and complex data collections will likely continue (Doney, 2004; IMBER, 2005; NRC, 1992; SOLAS, 2004). Information systems designed in support of these programs benefit from modularity, flexibility and extensibility but most importantly will remain useful when framed by science-driven initiatives with actively involved investigators and grounded by local data management in practice. Information system design is a continuous process, responsive to the changing needs of domain science, information science and technology.

While we view data curation as dealing with the steps, procedures and products of data, it is with the activity of information management in concert with informatics and data stewardship that larger, long-term issues of data flow and sustainability, data reuse and interpretation begin to be recognized and addressed (Karasti et al, 2006). Our perspective is one of products and processes that enhance data quality and support oceanographic science. Short-term products include the handling of data and work with information systems. A longer-term vision involves creating an environment that explains and maintains existing information processes while researching and designing new ones. Taking a step back from the local laboratory and field programs, we glimpse the larger context of a global digital arena. The goals of such a digital environment also entail products and processes. The product involves an information model capable of federating while allowing for difference. Establishing a local informatics approach anchored 'in practice' to oceanographic research would be a process to ensure sensitivity to today's anomalies that may be tomorrow's insights. It is a combination of science and informatics efforts that can ensure the data that are preserved and made accessible represent well the data that are collected as representative of the natural world.

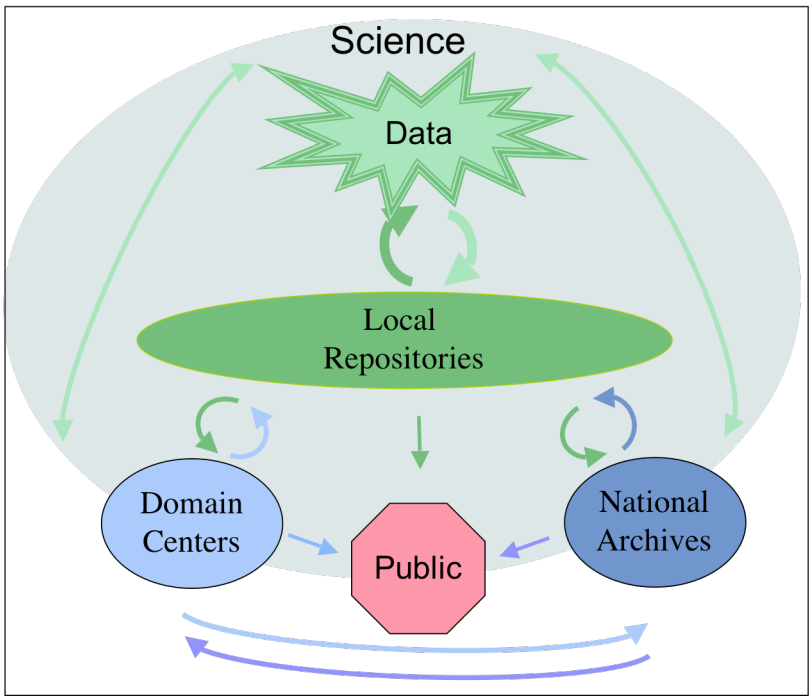

Figure 4. An organizational view of the scientific data collections that are created for scientific purposes. Local repositories, project or domain centers, national archives and public access are depicted. The arrows represent information flow direction and places where requirements must be communicated and expectations set.

Whole Earth Ecosystem

The object of study of oceanography, the global ocean, in and of itself, is a vast medium; its surface covers more than 70% of Earth's surface while its greatest depths exceed the dimensions of Mount Everest. However, the ocean is never "in and of itself" because the waters of the global ocean interact with the planetary atmosphere, land forms and deep-earth tectonic activity. To study the ocean is to study the abundant plant and animal life found there, the physical and chemical properties of the water and the myriad interactions and mutual influences of all these aspects of the global ocean system. To study the ocean is to work on vast spatial scales and temporal scales that reach into the Earth's earliest stages of evolution. To study the ocean is to recognize the interface of the human species and our constructed environment with the natural environment and take into consideration anthropogenic effects and their ramifications. New approaches to studying the global oceans are under consideration (NRC, 2003).

Whole Earth Trajectories. An increasing number of interdisciplinary and cross-domain approaches today are beginning to recognize and incorporate explicitly the interconnections of human and environmental systems within a context of co-evolving trajectories as we move from what might be called the 'Machine Age' into the 'Systems Age' (Ackoff, 1997). Drawing on visioning by the NSF Advisory Committee on Complex Environmental Systems (NSF AC-ERE, 2003) and the Human Dimensions Committee of the LTER (LTER Newsletter, Spring 2005; Waltner-Toews et al., 2003), the LTER inclusive ecosystem model can be augmented by explicitly representing information systems as interfacing with the human dimension of a whole earth ecosystem model (Figure 5). This fits within the context of contemporary work on socio-technical aspects of information systems, with the increasing attention in science and technology studies to the role of the seemingly everyday mundane work of keeping routines 'routine' as they adapt to changes and with digital library work developing access and preservation of information in dramatically new ways. Modeling an environmental ecosystem as a closed system with defined inputs and outputs represents a complex scientific enterprise; modeling a whole earth ecosystem as an open, natural system with emergent characteristics, promises to be even more challenging.

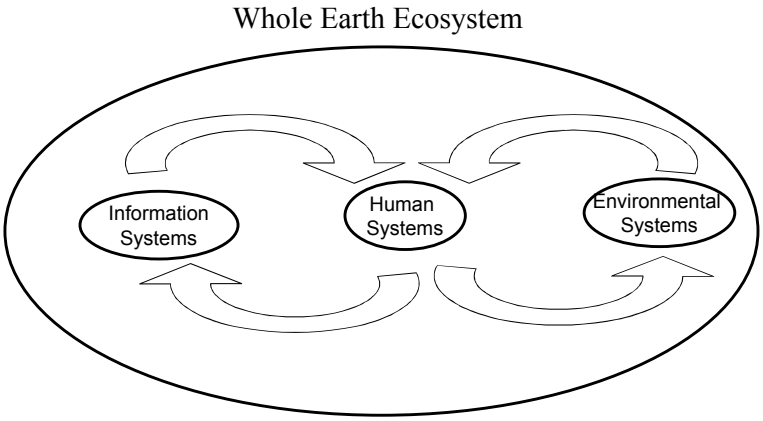

Figure 5. Representation of the whole earth as an ecosystem, a system of complex systems, taking into account environmental, human, and information dimensions.

Acknowledgements

The LTER work (NSF/OPP #02-17282 and #04-05069) synergizes with work supported by NSF/SBE/SES Human Social Dynamics grant #04-33369 as well as the Ocean Informatics Environments supported at Scripps Institution of Oceanography and Woods Hole Oceanographic

Institution. Work conducted by the U.S. JGOFS Data Management Office (NSF OCE-8814310 and OCE-0097291) was carried forward by the Ocean Carbon and Biogeochemistry Coordination and Data Management Office (NSF OCE-0510046) and will continue supporting ocean science research through the Biological and Chemical Oceanography Data Management Office (NSF OCE-0646353).

References

Abbott, A., 1988. The System of Professions: An Essay on the division of expert labor.

Ackoff, R. 1997. Systems, messes and interactive planning. In: Trist, E., F.Emery, H.Murray (eds), The Social Engagement of Social Science: A Tavistock Anthology. The Socio-ecological Perspective, Vol. III: 417-438. University of Pennsylvania Press, Philadephia, PA.

Atkins, D. and NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure, 2003. NSF-AP Report: Revolutionizing Science and Engineering Through Cyberinfrastructure.

Baker, K. S., 1998. Palmer LTER Information Management, in Data and Information Management in the Ecological Sciences: a resource guide. Michener, Porter and Stafford (eds), University of New Mexico, p 105-110.

Baker, K.S., S.J. Jackson and J.R. Wanetick, 2005. Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment. Proceedings of the 38th Hawaii International Conference on System Sciences. HICSS38, IEEE Computer Society, 2-6 January 2005, Big Island, Hawaii

Baker, K.S., B. Benson, D.L. Henshaw, D. Blodgett, J. Porter and S.G. Stafford, 2000. Evolution of a Multi-Site Network Information System: the LTER Information Management Paradigm, BioScience 50(11):963-983.

Baker, K.S. and G. Bowker, (in press). Information Ecology: Open System Environment for Data, Memories and Knowing. Journal of Intelligent Information Systems. BDEI Special Series.

Baker, K.S. and F. Millerand, (in press). Articulation Work Supporting Information Infrastructure Design: Coordination, Categorization, and Assessment in Practice. Proceedings of the 40th Hawaii International Conference on System Sciences (HICSS), 2007, 3-6 January, Big Island, Hawaii, pp. 1-10, IEEE Computer Society, New Brunswick, NJ.

Birnholtz, J.P. and M.J. Bietz, 2003. Data at Work: Supporting Sharing in Science and Engineering. M. Tremaine (ed) Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work (GROUP'03; 2003 November 9-12). ACM Press. V34: 339-348.

Brewer, P.G., K.W. Bruland, R.W. Eppley and J.J. McCarthy, 1986. The Global Ocean flux Study (GOFS): Status of the U.S. GOFS Program. EOS **67**(44): 827-832.

Brunt, J.W., P. McCartney, K.S. Baker and S. Stafford, 2002. The future of Ecoinformatics in Long Term Ecological Research, in Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL. edited by NCallaos, JPorter and NRishe. IIS 7:367-372.

Chandler, C.L., 2004. U.S.JGOFS Data Management: Then and Now. In U.S.JGOFS Newsletter 12(4), November 2004

Cook, R.B., R.J. Olson, P. Kancruk and L.A. Hook, 2001. Best practices for preparing ecological data sets to share and archive. Bulletin of the Ecological Society of America 138-141.

Cornillon, P., J. Gallagher, T. Sgouros, 2003. OPENDAP: Accessing Data in a Distributed, Heterogeneous Environment. Data Science Journal 2: 164-174.

Davenport, T.H., 1997. Information Ecology. New York, Oxford University Press.

Domenico, B., J. Caron, E. David, R. Kambic and S. Nativi, 2002. Thematic Real-time Environmental Distributed Data Services (THREDDS): Incorporating Interactive Analysis Tools into NSDL. Journal of Digital Information 2(4)): Article 114, 2002-05-29.

Doney, S. C., 2004. Ocean carbon and climate change (OCCC) : an implementation strategy for U.S. ocean carbon research. Boulder, CO, U.S. Carbon Cycle Science Scientific Steering Group, U.S. Carbon Cycle Interagency Working Group, University Corporation for Atmospheric Research**:** 108.

Doney, S. C., J.A. Kleypas, J.L. Sarmiento, P.G. Falkowski, 2002. "The US JGOFS Synthesis and Modeling Project - An introduction." Deep Sea Research Part II: Topical Studies in Oceanography **49**(1-3): 1-20.

Ducklow, H. W., K. S. Baker, D. G. Martinson, L. B. Quetin, R. M. Ross, R. C. Smith, S. E. Stammerjohn, M. Vernet, and W. Fraser, 2006. Marine ecosystems: The West Antarctic Peninsula. *Philosophical Transactions of the Royal Society of London, Special Theme Issue, "Antarctic Ecology: From Genes to Ecosystems."*

Edwards, P.N., S.J. Jackson, G.C. Bowker, and C. Knobel, 2006. Understanding Infrastructure: Dynamics, Tensions, and Design. NSF Human Social Dynamics Report.

Fasham, M. J. R., B. M. Baliño, M.C. Bowles (eds.), 2001. "A new vision of ocean biogeochemistry after a decade of the Joint Global Ocean Flux Study (JGOFS)." Ambio special report, No 10: 30.

Finholt, T.A., 2004. Collaboratories. Annual Review of Information Science and Technology 36:73-107. Medford NJ, E B. Cronin

Flierl G., Bishop J.K.B., D.M. Glover, S. Paranjpe, 1992. Data management for JGOFS: Theory and design. 229–24921 in Churgin J, ed Proceedings of Ocean Climate Data Workshop. 18–24921 Feb 1992. Greenbelt, MD; Springfield (VA): National Technical Information Service

Franklin, M., A. Halevy and D. Maier, 2005. From Databases to Dataspaces: A New Abstraction for Information Management. SIGMOD Record 34(4): 27-33.

Friedman, A. L., 1989. Computer Systems Development: History, Organization and Implementation. Chichester, UK: Wiley.

Friesen, N, 2002. Semantic Interoperability, Communities of Practice and the CanCore Learning Object Metadata Profile. In Proceedings International WWW Conference (11), Honolulu, Hawaii, USA.

Gasson, S., 1998. Framing design: a social process view of information system development. Proceedings of the Hawaii International Conference on Information Systems, 224-236.

Glover, D.M., C.L. Chandler, S.C. Doney, K.O. Buesseler, G. Heimerdinger, J.K.B. Bishop and G.R. Flierl, 2006. The US JGOFS Data Management Experience. Deep-Sea Research II 53:

793-802.

Glover, D.M., 2001. Taking Care of the Legacy: Data Management in U.S. JGOFS. Oceanography 14(4): 106-107.

Glover, D. M. and C. L. Chandler, 2001. An Update On Data Management In U.S. JGOFS. U.S. JGOFS News. U.S. JGOFS Planning Office, Woods Hole, MA. **11**(3): 16.

Golley, F. 1993. A history of the ecosystem concept in ecology: More than the sum of the parts. Yale University Press. New Haven.

Gorentz, J., 1992. Data management at biological field stations and coast marine laboratories, Report of an invitational workshop, W.K. Kellogg Biological Station; 22-26 Apr 1990; Lansing (MI): Michigan State University. Contains 1982 report as Appendix. http://www.lternet.edu/ecoinformatics.

Greenwood, M., C.A. Goble, R. Stevens, J. Zhao, M. Addis, D. Marvin, L. Moreau and T. Oinn, 2003. Provenance of e-Science Experiments - experience from Bioinformatics, In: *Proceedings of UK e-Science All Hands Meeting 2003*. S.J. Cox, ed., East Midlands Conference Centre, Nottingham, pp. 223-226.

Gruber, T.R., 1993. A Translation Approach to Portale Ontology Specification. Knowledge Acquisition 5: 199-220.

Guarino, N and C. Welty, 2000. A Formal Ontology of Properties. Knowledge Engineering and Knowledge Management: Method, Models and Tools. 12th International Conference (EKAW 2000). R. Dieng and O. Corby. Berlin/New York, Springer: 97-113.

Hankin, S., J. Davison, J. Callahan, D. E. Harrison and K. O'Brien, 1998. A configurable web server for gridded data: a framework for collaboration. In 14th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology. AMS Providence, RI, pp. 417–418.

Hobbie, J.E., S.R.Carpenter, N.B.Grimm, J.R.Gosz and T.R.Seastedt (2003). The US Long Term Ecological Research Program. BioScience 53(2), 21-32.

IMBER, 2005. Science Plan and Implementation Strategy. IGBP Report No. 52, IGBP Secretariat, Stockholm, Sweden. 76pp.

Institute of Electrical and Electronics Engineers, 1990. *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries.* New York, NY.

Jackson, S.J. and K.S.Baker, 2004. Ecological Design, Collaborative Care and Ocean Informatics. Proceedings of the Participatory Design Conference 2: 64-67. July 27-31, 2004, Toronto, Canada.

Jones, M.B., C.Berkley, J.Bojilova and M.Schildhauer, 2001. Managing Scientific Metadata. IEEEE Internet Computing 59-58.

Kaplan, S. and L. Seebeck, 2001. Harnessing Complexity in CSCW. Proceedings of the Seventh European Conference on Computer Supported Cooperative Work, 16-20 September 2001, p.359-397

Karasti, H. and K. Baker, 2004. Infrastructuring for the Long-Term: Ecological Information Management. Hawaii International Conference for System Science Proceedings, 2004.

Khazanchi, D. and B.E. Munkvold, 2000. Is Information Systems a Science? An Inquiry into the Nature of the Information Systems Discipline. The Database for Advances in Information Systems 31(3): 24-42.

Kling , R. and T. Jewett, 1994. The Social Design of Worklife with Computers and Networks: A Natural Systems Perspective.Advances in Computers 39: 239-293.

McCartney, P.H. and M.B. Jones, 2002. Using XML-encoded Metadata as a Basis for Advanced Information Systems for Ecological Research. Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL 2002. IIIS 7:379-384. Orlando, FL N.Callaos and J.Porter (eds).

Michener, W.K. and J.W. Brunt, 2000. Ecological Data: Design, Management and Processing. Oxford, Blackwell Science.

Millerand, F. and Bowker, G.C. (forthcoming). Metadata Standards. Trajectories and Enactment in the Life of an Ontology. In S.L.Star and M.Lampland (Eds), Formalizing Practices: Reckoning with Standards, Numbers and Models in Science and Everyday Life.

NAS, 1984. Global Ocean Flux Study: Proceedings of a Workshop. National Academy of Sciences. Washington, D.C., National Academy Press**:** 360.

NRC, Ocean Studies Board Commission on Geosciences, Environment and Resources, 1992. Oceanography in the Next Decade: Building New Partnerships, National Academy Press, Washington, D.C., 216 pp.

NRC, Committee Toward a National Collaboratory: Establishing the User-Developer Partnership, 1993. National collaboratories: applying information technology for scientific research, Computer Science and Telecommunications Board, Commission on Physical Sciences, Mathematics and Applications, National Research Council, Washington, D.C., National Academy Press.

NRC, Committee on Major U.S. Oceanographic Research Programs, 1999. Global Ocean Science: Toward an Integrated Approach, National Academy Press, Washington, D.C., 184 pp.

NRC, Committee on Exploration of the Seas, 2003. Exploration of the Seas: Voyage into the Unknown, National Academy Press, Washington, D. C., 213 pp.

NSB, 2005. Long Lived Digital Data Collections: Enabling Research and Education in the 21st Century, National Science Board (NSB-05-40, Revised May 23, 2005).

NSF AC-ERE, 2003. Complex Environmental Systems: Synthesis for Earth, Life, and Society in the 21st Century. Pfirman, S. (ed). NSF Advisory Committee for Environmental Research and Education.

Oreskes, N., 2002. History of Earth Sciences. Encyclopedia of Physical Science and Technology, Third Edition, Volume 4. Academic Press

Ribes, D. and G. Bowker (in press). Ontologies and the Learning Trajectory. Journal for the Association of Information Systems .

Rothstein, L.M., J.J. Cullen, M.A. Abbott, E.P. Chassignet, K. Denman, S.C. Doney, H. Ducklow, K. Fennel, M. Follows, D. Haidvogel, E. Hoffman, D.M. Karl, J. Kindle, I.Lima, M. Maltrud, C. McClain, D. McGillicuddy, M.J. Olascoaga, Y. Spitz, J. Wiggert and J. Yoder,

2006. Modeling Ocean Ecosystems: The PARADIGM Program. Oceanography 19(1): 22-51.

SCOR, 1987. The Joint Global Ocean Flux Study: Background, Goals, Organization and Next Steps. Report of the International Scientific Planning and Coordination Meeting for Global Ocean Flux Studies, Paris, 2/17—19/87, Available from SCOR Secretariat, Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1, 42pp.

Sarmiento, J. L. and R. A. Armstrong, 1997. U.S. JGOFS Synthesis and Modeling Project Implementation Plan: The Role of Oceanic Processes in the Global Carbon Cycle. Woods Hole, MA, U.S. JGOFS Planning and Coordination Office, Woods Hole Oceanographic Institution**:** 73.

Schmidt, K. and C. Simone, 1996. Coordination Mechanisms: Towards a Conceptual Foundation of CSCW Systems Design. Computer Supported Cooperative Work: The Journal of Collaborative Computing 5, 155-200p.

Sheth, A.P., 1999. Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. In Interoperating Geographic Information Systems. MGoodchild, MEgenhofer, RFergeas, CKottman (eds). Kluwer Academic Publishers, Boston.

Simmhan, Y.L., B. Plale and D. Gannon, 2005. A Survey of Data Provenance in e-Science, SIGMOD Record, 34(3), 32-36.

Smith, B., 2003. Ontology. In Blackwell Guide to the Philosophy of Computing and Information. LFloridi. Oxford, Blackwell: 155-166.

Smith, B. and Welty, C., 2001. Ontology: Towards a New Synthesis. In Proceedings of FOIS'01.

Smith, F.E., 1968. Proceedings of the National Academy of Sciences of the United States of America 60(1): pp5-11.

Smith, R. C., K. S. Baker, W. R. Fraser, E. E. Hofmann, D. M. Karl, J. M. Klinck, L. B. Quetin, B. B. Prezelin, R. M. Ross, W. Z. Trivelpiece and M. Vernet, 1995. The Palmer LTER: A long-term ecological research program at Palmer Station, Antarctica. *Oceanography*, *8* (3), 77-86.

SOLAS, 2004. Science Plan and Implementation Strategy, *IGBP Report No. 50*, IGBP Secretariat, Stockholm, Sweden, p. 88.

Spencer, B.F, R. Butler, K. Ricker, D. Marcusiu, T. Finholt, I. Foster and C. Kesselman, 2006. Cyberenvironment Project Management: Lessons Learned http://neesgrid.ncsa.uiuc.edu/documents/CPMLL.pdf

Star, S.L. and Ruhleder, K., 1996. Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. Information Systems Research 7(1).

Star, S.L. and G.C. Bowker, 2002. How to Infrastructure. L.A. Lievrouw and S.Livingston (eds The Handbook of New Media. London, SAGE Publications.

Stonebraker, M., 1994. Sequoia 2000: A reflection on the first three years. Pages 108-116 in French J, Hinterberger H, eds. Seventh International Working Conference on Scientific and

Statistical Database management, 28-30 Sep 1992, Charlottesville, VA. Los Alamitos, CA, IEEE Computer Society Press.

Strebel, D. E., Newcomer, J. A., Ormsby, J. P., Hall, F. G. and Sellers, P. J., 1989. Data management in the FIFE information system. Proceedings of IGARSS'89 Symposium. Vancouver B.C. 42-45.

Thorley, M.R. and P.N. Trathan, 1994. Biomass data set documentation. Cambridge, CB3 OET, United Kingdom British Antarctic Survey, Natural Environment Research Council.

UNESCO, 1994. Protocols for the Joint Global Ocean Flux Study (JGOFS) Core Measurements. IOC Manuals and Guides 29: 170 pp.

U.S. GOFS Steering Committee, 1986. U.S. Global Ocean Flux Study Steering Committee Minutes. 26-27 June 1986, URI GSO, Narragansett, RI. U.S. JGOFS Planning Office, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 8 pp.

U.S. JGOFS Steering Committee, 1990. U.S. Joint Global Ocean Flux Study Long Range Plan, The Role of Ocean Biogeochemical Cycles in Climate Change. U.S. JGOFS Planning Report Number 11. U.S. JGOFS Planning Office, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 71 pp.

Veltman, K. H. (2001). "Syntactic and Semantic Interoperability: New Approaches to Knowledge and the Semantic Web." The New Review of Information Networking **7**: 159-183.

Venturi, R. 1966. Complexity and Contradiction in Architecture. The Museum of Modern Art (MoMA) Papers on Architecture, NY, c.1966, 1977, 1988.

Visser, U., Stuckenschmidt, H., Wache, H., Vögele, T., 2000. Enabling technologies for interoperability. In: Workshop: Information Sharing: Methods and Applications at the 14th International Symposium of Computer Science for Environmental Protection, Visser, U. and Pundt, H., Eds., TZI, Univ. of Bremen, Bonn, Germany, pp. 35-46.

Waltner-Toews, D., J.J. Kay, C. Neudoerffer and T. Gitau, 2003. Perspective Changes Everything: Managing Ecosystems from the Inside Out. Frontiers in Ecology and the Environment 1:23-30.

Waters, K. and R.C. Smith, 1992. Palmer LTER: A sampling grid for the Palmer LTER Program. Antarctic Journal of the United States, 27%95): 236-239.

Whyte, W.F. (1991) Participatory Action Research. SAGE publications, London.

Zimmerman, A.S., 2003. Data Sharing and Secondary Use of Scientific Data: Experiences of Ecologists. PhD Thesis, The University of Michigan, Ann Arbor, 272p.