Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment

Karen S. Baker
Scripps Institution of Oceanography
University of California at San Diego
kbaker@ucsd.edu

Steven J. Jackson
Department of Communication
University of California at San Diego
sjjackso@ucsd.edu

Jerome R. Wanetick
Scripps Institution of Oceanography
University of California at San Diego
jwanetick@ucsd.edu

Abstract

This paper considers the elements and challenges of heterogeneous data management and interdisciplinary collaboration, drawing from the literatures on participatory design, computer-supported cooperative work, and science studies to support information design efforts within the rapidly evolving world of large-scale science projects.

Certain tensions are embedded in such collaborative projects, being rooted in distinctive disciplinary knowledge interests brought to the table and expressed in occasionally divergent understandings of project rationale, identity and success. A continuum of strategies exist for dealing with such tensions. In this paper, we discuss two of these: a strategy of 'mindful variety' built around an appreciation of disciplinary, organizational and biographical heterogeneity of collaborative ventures; and attention to the proliferation of 'boundary objects' and shared languages between and within adjacent communities of practice. These strategies are considered specifically through the lens of an Ocean Informatics Environment (OIE), a concept that joins ocean, information, and social scientists working to construct locally responsive, adaptive and scalable information infrastructures for the practice of ocean science. Our team seeks to design an environment supporting reflexive and heterogeneous data practices responsive to the multiple work worlds of ocean science. We consider the development of an ethic of collaborative care, offered as a working principle for the identification, preservation and bridging of disciplinary difference in the cooperative design of scientific work settings as one of several strategies emerging from ongoing work.

1. Introduction

Like other domains of the earth sciences, the practice of ocean science has come to rest in recent years on an expanding and rapidly changing web of institutional relations, data networks, and advanced information systems. Changes in the type, scale and complexity of questions posed by ocean scientists have driven (and in some cases, been driven by) broader shifts in the information technology and computational landscapes. Past years have seen efforts at re-scaling the object(s) of ocean science with research, transitioning from single cruise efforts and wide deployment of a standardized platform to multi-platform, multi-cruise basin-wide studies. Researchers have sought to make the leap to multi-project integration over time for the long-term interdisciplinary study of global ocean systems whose 'parts', loose and heterogeneous assemblies of local processes and nested sub-systems, inevitably sum to larger, more complex, and still poorly understood 'wholes'.

Shifts to 'whole ocean' thinking have been accompanied by a move to increased disciplinary plurality. In addition to the traditional core areas of oceanographic research, scattered across fields drawn from the physical, chemical, biological, geological, and atmospheric sciences that share certain common approaches to data handling, modeling, and visualization, research partnerships now encompass expanding technology, education (training, formal, informal, and outreach) and community engagements with local stakeholders and policymakers. Because oceanography depends upon interdisciplinary collaboration and management of diverse data, it is an appropriate realm for utilizing strategies that support and improve these activities of collaboration and data management.

2. Oceanography: an Earth System Science

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 the juxtaposition of component studies (from bacteria to primary producers to predators) with whole system views of material and energy flows through ecosystems. For oceanography, more than four decades after the International Geophysical Year (IGY 1957-1958) prompted a flurry of global activities, a variety of multi-year and multi-sited global projects have been initiated, including in the past decade the Joint Global Ocean Flux Study (JGOFS;

http://usjgofs.whoi.edu) and the International Geosphere-Biosphere Programme (IGBP) Global Ocean Ecosystem Dynamics (GLOBEC, http://www.pml.ac.uk/globec/main.htm). 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 Systems (GOOS; http://ioc.unesco.org/goos/). 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.nsf.gov/bio/neon) 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 changes have further challenged already suspect notions of the solitary scientist and the independent project, revealing the practice of ocean science as a socially complex, globally distributed, and highly mediated form of distributed collaborative practice

2.1. 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, 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 then 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. 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 it's greatest depths exceed the dimensions of Mt Everest.

But 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. 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 protocols. Collections are held and managed as primarily local entities with bridges to datasets housed at other ocean research centers built and maintained on a more or less ad hoc basis. More generally, 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. Under this conception, funding for information management in the soft-money world of ocean science is built and organized on a project-by-project basis producing a further institutional barrier to integration (Mukerji, 1989).

3. Challenges Within Ocean Informatics

3.1. Heterogeneous Data

Data-centered, seagoing research projects involve assembling, deploying, and tending field equipment that may be launched by teams from shore, small boats or large ships. Data may be retrieved from field storage devices or streamed real-time via line-of-sight or satellite telemetry. Logistics dominate field activities, demanding time and resource driven coordination of sampling designs, ship schedules, equipment preparations, and unexpected contingencies in remote field locations. Post fieldwork involves tending equipment, physical samples and data analysis. Such activities commonly produce a diversity of data sets and archives (Bowker, 2000). A group of research programs, even when housed within a single institution, frequently maintains a diverse array of data holdings.

In recent years, questions of data integration, documentation, storage and access have been framed in debates over the utility of metadata – literally, data about data – which ideally packages enough information about the context of data to extend its usefulness beyond the immediate time, place and circumstances of the initial research project. But metadata solutions to the problem of information management and access may underestimate the 'layered' qualities of data, its location within nested hierarchies of databases, schemas, ontologies, languages and institutions. More recently, scholars have pursued data integration through the mechanism of schema integration, i.e. by finding semantic correspondences and integration points across multiple schemas as a basis for resolving nontrivial differences in semantics, units, precision, resolution, protocols, and aggregation (Wang et al, 2004); from this perspective, data collections are staged for interoperability if local protocols for dealing with semantic functions and conditions for one-element to multiple-element matches are developed.

3.1.1. Heterogeneous data in Practice

The Scripps Institution of Oceanography (SIO) at the University of California, San Diego (UCSD) research programs take as their object of study the internal and interactive dynamics among complex, large-scale and multidimensional earth/ocean/atmospheric systems centered upon (but not exclusively restricted to) ocean processes. In addition to well-established research traditions in biological and physical oceanography, the Integrative Oceanography Division (IOD) at SIO has ongoing interests in marine chemistry and archeology, geology, and information systems. Unlike some more theoretically oriented strains of ocean science, the research programs maintain a firm grounding in the practice of field observation, with data collection from near shore waters to deep ocean sampling to model generated data.

A research scientist, juggling an abundance of hypotheses, finds time a major constraint. Data gathered into project specific databases eases data access and availability. Today's aggregation of diverse collections creates new scenarios, bringing together data so that new questions may emerge and be addressed. Considering changes in data practices and information flows as well as managing and maintaining diverse data repositories are elements of today's digital transitions. An abundance of data means more data may be used with a single hypothesis or more hypotheses may be examined more quickly. In either case, we have yet to see how more data adds to or relaxes time as a constraint.

In focusing on three data collections with distinctly separate origins collocated within the IOD data center, questions of data interoperability similar to those being asked at a national level by federated partnerships arise. In addition to sharing computational infrastructure, each collection holds in common: the centrality of a project sampling scheme over time, the role of data in service to a single project, and a growth over the last decade to larger partnership science. The IOD Data Zoo houses current and historical California coastal oceanographic data sets utilized by scientists at SIO and by the coastal oceanographic community at large (Wanetick and Browne, 1999). The California Cooperative Ocean Fisheries Investigations CalCOFI database archives more than 50 years of periodically sampled fisheries-related data (Ohman and Venrick, 2003). Finally, the Palmer LTER research program since 1991 has gathered data at Palmer Station focusing on questions of ice influences on the Antarctic marine ecosystem (Smith et al, 1995).

3.2. Complex Systems

Post-war systems approaches, broadly defined in recent decades, have come to play an important and distinctive role in casting inquiries into the context of "closed" and "open" systems. In closed systems, processes are considered internally consistent and predictive. They range from the simple to the complicated, involving many steps and rules characterized by both linear and hierarchical flows as well as whole system and component feedback loops. The modern meta-science of cybernetics attempts to use the bounded characteristics of closed systems to predict future events based upon known present and past states. Earth and biological sciences use the closed systems approach in applying the principles of conservation of energy, mass, and nutrients, to study processes such as ocean currents or the carbon cycle. Social sciences also use closed systems theory, assuming dynamics can be explained without reference to external influences, when the discrete or closed 'society' is adopted as a unit of sociological analysis.

Open systems, defined by complex, non-deterministic processes are radically *in*complete, characterized by their sensitivity to external inputs and responses to external stimuli, e.g., the input of solar energy, in the form of light into biological systems driving photosynthesis and in the form of heat flux driving ocean currents. In open systems, notions of turbulence, noise, and chaos have been introduced as counterpoints to the more ordered world of relationships defined for closed systems.

An example of the tension between these two approaches, closed verses open, complicated verses complex, can be found in an oceanographic field campaign. We plan for the complicated and are often surprised by the complex. Deploying a profiling instrument over the side of a ship is a 'complicated' operation in normal circumstances; it involves tending to equipment calibration and readiness, personnel preparation and availability, bridge coordination and navigation, often in a remote location or aboard a ship in the middle of the ocean. There are a variety of explicit rules and procedures that define these activities. When weather or seas, equipment or personnel vary outside the 'norm', however, deployment becomes 'complex', contingent on multiple social and/or environmental factors that lead to a diverse set of adjustments and evaluations as priorities are (often tacitly) reordered to take into consideration the safety of both personnel and equipment in addition to the finely tuned optimization of the data gathering process. Sociologists of science have long noted the distinction between explicit and implicit (or 'tacit') forms of knowledge, and the long-standing neglect of the latter within mainstream accounts of the work of science – an insight with important implications for the design of scientific information systems. These views contrast with the notions of linear progression from data to information to knowledge, highlighting the centrality of local knowledge of the data gatherer/user (Polanyi 1956; Snowden, 1999; Kaplan and Seebeck, 2001; Gasson 2004; Bowker and Baker, 2004).

The challenge of designing an ocean informatics environment is informed by the heterogeneous character of such systems where the important variations, between knowledge (or data) held locally in an idiosyncratic environment and made globally accessible using universally accepted schema and methodologies for dissemination, are recognized as differing perspectives. Such perspectives can be understood and improved by moving back and forth between the complex and the complicated, as part of an ongoing process of describing observed phenomena within digital structures.

3.3. Interdisciplinarity

Oceanography has traditionally called for interdisciplinary collaboration because the systems studied present not only a heterogeneous cast of components for scientific study but also a complexity of interaction among those components. The interdisciplinarity within a collaborative research team is expressed through the different discipline interests brought to the table by each participant. These are structured in part through the organizational positions and career incentives determined by each participant's placement with specific institutional and disciplinary matrices. Research methods, apparatus and data management are likely to vary across disciplines. Critically, unspoken assumptions can become verbalized after a research program is designed and underway. One example is the always slippery language of 'success'. By what criteria are we to assess the processes and outcomes brought about by an interdisciplinary project? Research facilitated and papers published? Hardware and software developed or implemented? Hits to the website and bytes served? But how does one get at the 'softer', less tangible benefits that might emerge? New conversations and collaborations between previously distant colleagues, geographically or disciplinarily? Can and should an ocean informatics system be designed to facilitate all these measures of success?

Interdisciplinary projects tend to produce heterogeneous data sets as scientists investigate open, complex systems that foster emergent behavior. Unanticipated discoveries as well as deviations from well laid out research programs can occur. Ocean informatics is challenged to facilitate success, surprises and even unsuccess. When project segments or initiatives don't 'work', research participants can learn important understandings about the nature of collaboration and the practice of ocean informatics. From this perspective, an entirely plausible one within the social study of science, the telling failure, the spectacular unsuccess, may be a research finding of the first order; yet this will provide small comfort to a project scientist or data manager left to pick up the pieces when a field campaign fails or an established system proves inflexible or is simply unused.

3.3.1. Interdisciplinarity in Practice

Both interdisciplinarity and data heterogeneity, with all their attendant challenges, are part of the work of ocean informatics. Such challenges when viewed as informative and productive rather than primarily disruptive, provide important opportunities for learning from diversity and for building flexibility, adaptability, and ultimately sustainability into the long-term practices of informatics. Participatory design techniques have figured centrally in our methodological mix, with ethnographic analysis, participant observation and iterative design approaches deployed to draw out, identify, and support the real data practices of IOD information managers, researchers, field technicians, graduate assistants, administrators, educators and learners. A working principle of the group has been the understanding that there exists no 'perfect perspective' on ocean informatics, no single institutional, epistemological or technical position from which the full complexities of community data practices are automatically visible. Instead, there exists only a collection of partial perspectives, situated 'takes' on the practice of ocean informatics that can (and should) be elucidated through a careful blend of social, institutional and technical analysis and action.

As this brief overview suggests, data issues with respect to combining datasets and collections are both 'complicated' and 'complex', in the sense of those terms advocated by Kurtz and Snowden (2003). To address this complexity, the Ocean Informatics concept has brought data and information managers (with long experience working with the SIO community and datasets) together with social science perspectives drawn from the fields of communication, science and technology studies (STS), and information science. In this regard, the heterogeneity of the data itself is matched by heterogeneity in the methods, orientations and analytic tools employed.

From the beginning, cross-disciplinary collaboration on the Ocean Informatics concept has faced infrastructural challenges of the most mundane sort, from the challenge of fitting into established and still generally discipline-bound funding structures, to the organizational challenge of coordinating work across separate administrative units within the university, to the simple geographical separation between SIO located on the ocean and the "upper campus" at UCSD located on the mesa overlooking SIO. Could social scientists be convinced to go 'down the hill' to Scripps? Could ocean scientists and information managers be convinced to make the trip to main campus?

4. Strategies to Facilitate Collaboration

Certain tensions are embedded in collaborative projects, being rooted in distinctive disciplinary knowledge interests and practices and expressed in occasionally divergent understandings of project rationale, identity and success. A continuum of strategies exists for dealing with such tensions. In this section we discuss two of these: a strategy of 'mindful variety' that implies an attention to the divergence and heterogeneity inherent in collaborative ventures, and the strategy of highlighting 'boundary objects' and languages between and within adjacent communities of practice.

4.1. Mindful Variety

Recent work by Weick and Sutcliffe (2001) has described the distinctive set of processes adopted by High Reliability Organizations (HROs) for dealing with risk and uncertainty. For organizations in this category (nuclear power plant operators, aircraft carriers, space shuttles, etc.), the consequences of system breakdown and communication failure may be, quite literally, catastrophic, making opportunities for organizational learning through traditional trial and error methods infeasible. How then do HROs learn? A strategy commonly adopted by the most successful of these organizations has been to shift institutional focus from norms to anomalies, exchanging a fixation on strict and ordered procedure for a finely-tuned sensitivity to moments of deviance, divergence, and departure. Thus, apparently trivial anomalies in non-mission critical systems are investigated as potential harbingers of larger and more serious vulnerabilities; minor and apparently harmless instances of technical or organizational breakdown are treated as potential indicators of underlying system pathologies.

Beyond their value as potential early warning indicators, anomalies in highly ordered systems can be seized upon as opportunities for assessment and organizational learning. Viewed in this light, the (non-catastrophic) departure from routine may stand less as organizational nuisance than fortunate accident, a valuable (and in the tightly-controlled world of HROs, rare) opportunity for reexamination of organizational systems and processes. When seen as 'failures', some of our little discussed indicators of need for attention to system robustness are lost as learning opportunities. Weick and Sutcliffe name this general sensitivity to instances of divergence and 'noise' within high-risk systems the principle of 'mindful variety'.

4.1.1. Mindful Variety with Data

Although early thermal and ozone mapping satellites effectively broadened awareness (successful mindfulness) of 'norms' for a generation of remote sensing participants, satellite data analysis failed to identify both an El Nino event in the Pacific Ocean and the first appearance of the ozone hole in the Antarctic. In each case data were 'masked out' as 'no data' since they were outside the limits defined as normal in the analysis process. Given the overwhelming influx of new data, these crucial data were ignored. Fortunately other measurements brought attention to these critical data. A ship returned to provide the first report of warmer Pacific temperatures; a British ground-based ozone instrument was first to 'discover' the Antarctic ozone hole. Subsequent re-examination and analysis of the full suite of unmasked satellite data verified these ground-based findings.

An early Antarctic cruise demonstrates the positive potential in the principle of mindful variety. Scientists proposed to investigate impacts due to the changing ultraviolet radiation (UV) on diatom blooms reported as an Antarctic Spring ice-edge phenomena. When open-ocean sea-ice edge sampling did not reveal major diatom blooms, scientists with limited ship time faced the dilemma of deciding whether to remain at the established sampling site to investigate UV effects on an unanticipated phytoplankton community or to steam to a region where the proposed diatom population could be found and would allow for experiments referenced to past literature. The final decision was to remain at the original site and study this departure from the anticipated. The team reported a 17% decrease in water column productivity due to the ozone hole in Science the next year (Smith et al, 1990). The decision to attend to, i.e. be mindful of, the divergence outside the original understandings of plankton population history, i.e., variety, produced scientific insight and discovery.

4.1.2. Mindful Variety with Partnerships

With interdisciplinary partnerships, we extend the mindful variety strategy to include mindfulness not only to deviation from a norm but to differences between norms of different disciplines that are interfacing. This strategy then requires an organization be both difference-alert and time-prepared. Cooperation among Ocean Informatics team members has been tested by such divergences in the working methods and cultures of the participating disciplines. For instance, early grant writing efforts were hampered by confusion stemming from different understandings of the nature and role of hypothesis-making and testing: could the project proceed (and get funded) with a loosely-defined set of research questions, trusting to the principle of ethnographic emergence, or should the project start from a more strictly defined and ideally falsifiable set of hypotheses that could be 'tested' rather than 'explored'? This speaks in turn to larger questions of empirical design, evidence and preparation. What would count as legitimate evidence for the variety of claims and projects advanced under the Ocean Informatics label? Given the emergent nature of the project, is it possible or advisable to define project benchmarks and assessment strategies in advance, and if so, how strictly should these be adhered to? Within the field-oriented culture of IOD, what constitutes the 'field' of Ocean Informatics, and who is its audience?

4.2. Creoles and Boundary Objects

Coming together from multiple projects, disciplines, communities and/or domains to work collaboratively requires building a shared understanding within a Community of Practice (Lave and Wenger, 1991). Artifacts and objects that are used by multiple members of the community can be described as 'boundary objects'. Developing and sharing joint objects, such as a field sampling grid or a data diagram, is a strategy for establishing common reference points in order to draw out explanations (to make visible) and bring together an articulation of heterogeneous information or views. An object affords the opportunity for repeated reflection upon a jointly held entity that may prompt an evolution of individual views or an understanding of alternate views.

The concept of boundary objects, an analytic tool for considering that which is shared, brings to the foreground frequently unarticulated individual assumptions and community understandings (Star and Griesemer 1989; Star 1990). Four categories of boundary objects have been described: repositories, forms, ideals types and terrain with coincident boundaries.

Once conceived, the storehouse of boundary objects appears inexhaustible as they arise in a wide variety of forms and circumstance. Within oceanography, sharing an event such as a cruise or field campaign provides a common experience from which to draw. Such a shared experience may provide enough common contexts to permit participants to understand and work with differing points of view embedded in interdisciplinarity.

An established ship or study sampling grid is a boundary object that establishes a common geography. Goodwin (1995) remarks 'the sampling grid establishes the basic rhythms that structure the life of scientists working on the ship' and provides for 'convergent diversity'. This coordinated understanding persists through post cruise work representing a flexible shared framework of language and experience that these interdisciplinary researchers use as a reference and basis for comparisons. Such a local, shared geographic form defines a common sampling arena and facilitates integration and synthesis of data, concepts, and stories.

Each boundary object carries with it a story of its own. For instance, extensive discussions and negotiations led to establishment of the Palmer LTER sampling grid (Waters et al, 1992). Because lines of longitude converge to the pole, a map projection was adopted that created a rectilinear pattern off the west coast of the Antarctic Peninsula, providing a cognitively comprehensible station-naming schema. Thus the sampling grid station nearest Palmer Station with coordinates (64.9333 degrees South, 64.4000 degrees West) converts to the more memorable (600.040) representing 600 km North and 40 km West of an agreed upon origin. The grid nomenclature seemed an internal project artifact until it began to appear in presentations of other researchers. Creating a web accessible program able to calculate Palmer station nomenclature given any combination of geographic latitude and longitude made the grid available to all, effectively enlarging the community.

The concept of boundary objects can expand outside the physical realm to include the social and cultural, extending the concept of 'physical terrain with coincident boundaries', to 'conceptual terrain with coincident boundaries'. Just as Goodwin (1995) identified the sea as an oceanographic shared boundary object, the ecosystem active within a specified geographical terrain serves to define a shared conceptual 'terrain' for ecologists.

Boundary objects are ubiquitous. In addition, the boundaries to be crossed may be those of projects, communities, disciplines, or domains. As *'boundary' objects* they perform the valuable function of focusing our attention on multiple relationships to a common reference point rather than solely on the object itself.

4.2.1. Shared Overarching Questions

Shifting from discipline focus to community considerations, the development of consensus around shared overarching questions and principles creates an important high level program prompt, providing guidance and cohesion, creating a shared conceptual terrain.

Contemporary education frameworks that build on the concept of inquiry learning recognize critical questions as integrative factors and as contextual reference points important in providing a unifying umbrella for diverse inquiries (Wiggins and McTighe, 1998). Overarching questions for long-term oceanographic research might include:

- What is the **value** of oceanographic research?
- How does a **long-term research** approach change our perspective?
- How does an **ecosystem view** impact research?
- What are the history and the nature of **collaborative science**?
- What are **the roles of information and technology** for work in the field, laboratory, community, and in cultures?

Such shared concepts provide focus over time and foster discussions among participants and across communities.

4.2.2. Shared Languages

In his study of the experimental worlds of high-energy physics, Peter Galison (1997) notes the crucial role of linguistic and material intermediaries in managing and bridging cross-disciplinary encounters. Drawing on a colonial metaphor, Galison demonstrates the importance of 'pidgins' and 'creoles' grounded in the material culture of experimental practice in establishing a shared realm of reference and meaning standing between, yet distinct from, the disciplinary and situational knowledge of each participating group. From this perspective, the production of ‘trade languages’ stands as a crucial and frequently overlooked part of the work of scientific collaboration. Building on this insight, the Ocean Informatics group has actively sought to support the linguistic work of creolization, seeking to build hybrid terms and concepts shared among the teams of ocean, information, education, and social science participants.

A significant word in this creole is ‘infrastructuring’. “Infrastructure” usually conjures images of physical constructions such as bridges, roads, transmission cables, telephone lines, sewer systems and other public facilities. So too in the domain of computer science infrastructure usually refers to computer hardware and software systems, and, at times, the facilities housing the computer equipment. “Infrastructuring”, a term coined to emphasize the active nature of infrastructure work rather than the more traditional image of a physical entity (Star and Bowker, 2002) refers to an interwoven technical and social system of machines and people, of products and processes. The term ‘infrastructuring’ is adopted in our work with ocean informatics to highlight the combined ongoing involvement in interdependent digital (cyber) and intellectual construction needed to create and sustain a functional ocean informatics environment.

4.2.3. Ocean Informatics Mixing Layers

Examples of the multiple types of these boundary crossing enablers can be found within the Ocean Informatics community playing an important integrative role: shared databases and publications are repositories of data and knowledge, respectively; a data dictionary and glossary are negotiated forms; a metadata language in the form of local or national standards is an ideal type; and a cruise sampling grid that divides the ocean ‘terrain’ into a pattern of well-defined geographic (hence replicable) stations and transects. Each of these objects represents a zone of active mixing across boundaries that separate different concepts, understandings, and traditions.

Additional boundary objects for Ocean Informatics now include ethnographic fieldwork in the form of interviews and recording sessions carried out during year 2003-2004, as they become incorporated in group projects. We have focused on process while carrying out situated activities including co-construction of specific products, i.e., facilities statements and web pages, proposals and papers which create ongoing dialogue and articulated analyses. Preliminary prototyping with collaboration software is adding new dimensions to this interdisciplinary mix (Baker et al, 2004).

While an important and encouraging approach, the time and patience required for this bridging / translation work of constructing boundary objects has constituted a project challenge in its own right. From this perspective, the work of collaboration depends upon the production of local information ecologies – exploring and designing jointly. Taking a holistic approach places information within a broader ‘ecological’ context (Davenport, 1997), shaped by social and technical factors, supporting domain or company goals, managing data, and designing for technology through attention to how information is created, gathered, aggregated, used, and distributed. As an open system for learning, an information ecology is adaptive and changes over time. The information ecology metaphor serves as an important reminder about:

- the perspective of an information culture
- multiple system components & their interfaces
- the goal of integrating across domains
- nonlinear data/knowledge/information strategies

5. Ocean Informatics Environment (OIE): Emergent Concepts

The ocean informatics team within the Integrative Oceanography Division at Scripps Institution of Oceanography envisions a transition from separate data management of independent data sets to an ocean information environment (OIE), a locally adaptive information system offering flexible access to shared, archived and documented data that will remove a significant barriers to work with existing data and will work on design of methods for incorporating new and legacy datasets. Rather than expand an existing data facility or push the local repositories out to national centers distant from the data originators, we consider design strategies for transforming local data facilities into a much needed OIE characterized by information infrastructuring focused on contemporary data handling techniques, responsive to human/technical dynamics and conducive to training and education for a variety of participants through participatory design and iterative assessment. Our research is motivated by a practical interest in building and sustaining effective information ecologies, as learning communities with a shared philosophy and accepted practices for content, collaboration, and communication (Davenport, 1997; Karasti and Baker, 2004). To this end, all participants play an active role in the multiple aspects of the project. Participants from the ocean and information sciences observe (the traditional participant observer role of the social scientist) while social scientists are involved in community activities as engaged participants.

A major tenet of the ocean informatics environment design is a recognition of the benefits to creating an intellectual commons where there can be an articulated interplay of differing modes of knowledge and facets of information management. Our informatics efforts, which combine ethnographic and action research approaches

(Greenbaum and Kyng, 1991; Karasti, 2001; Schuler and Namioka, 1993) are grounded within the local context of field measurements and data analysis, recognizing and supporting the multiple forms of knowledge structuring the practice of ocean science. Our goal is a learning enabled environment which may be viewed at the individual level as a form of 'in-reach' or training and at the community level as a mechanism for adapting to change (Bransford et al, 1998).

A unifying element of the OIE is designing with the philosophy of Open Source in mind (Schweik and Semenov, 2003; Schweik and Grove, 2000). For a collaborative design-centered environment, the seventeen lessons on open source endeavors provide some guidance (Raymond, 1999). The OIE affords a test bed for use of open-source software with important oceanographic databases that will increase availability of tools across economic barriers and provide a common scientific training for information managers, scientists, students, and general public. Using open source software in a sense places OIE in the realm of open systems, as new software will become available and appropriate interfaces developed. This emphasis has a broader impact in allowing others to adopt similar software with little additional expense.

The effort to establish and sustain a collaborative environment infrastructure brings to the fore the challenges of integrating heterogeneous data, of working with open, complex systems, and of drawing together an interdisciplinary working team. In designing an OIE these challenges and strategies must be engaged from the earliest steps of planning.

5.1. Collaborative Care

A wide variety of strategies might be adopted for dealing with the situation of collaborative heterogeneity described above. One apparently simple solution would be to *erase* it: to construct, as far as possible, an overarching ocean informatics identity capable of sublimating and transcending the more specific knowledge interests of the individual participants. Under this scenario, the collaboration becomes a sum of its parts, but does so thoroughly that its specific composition, the particularity of the parts, fades to insignificance. One can *interface it*: constructing a well-defined central system to which individual stakeholders can connect. A key distinction for this case is whether the central system is imposed from above or designed collaboratively by community participants. Or one can *prioritize*, arrange the plurality of participant knowledge interests into mutually recognized hierarchies: certainly A, maybe B and C, and if we're really lucky, D, E and F.

There are real efficiencies to be found down any of these roads, which perhaps explain their common (and no doubt frequently appropriate) use. But there are also real costs, measured in participants who see their interests downgraded or overwritten and who will therefore drift away, or withhold full commitment and participation. At the project level, this can lead to a general dissipation in creative tension, the jarring yet provocative dislocations, that make collaborations under the right set of circumstances such frustratingly productive experiences.

The model of collaborative care proposed here (Jackson and Baker, 2004) trades hierarchical solutions for an ethics of care founded on the histories of collaborative interaction – an approach paralleling Weick and Sutcliffe's (2001) call for 'mindful variety'. It recognizes heterogeneity and divergence as natural properties of collaborative endeavors, and treats these as assets, rather than obstacles to be overcome. At the same time, it acknowledges the frequently significant *costs* of collaboration, and seeks to accommodate these under a regime of mutual concern shared among the various project participants.

Securing financial support and creating timelines for such an approach requires a broadening of traditional perspectives in planning research projects. One important aspect of this is a shared commitment to interstitial work, the slow and ongoing practice of translation that respects the integrity of disciplinary originals – the interests of the physical oceanographer are *not* perfectly coincident with those of the biological oceanographer; the interests of the social scientist are *not* perfectly coincident with those of the data manager or ocean scientist, and vice versa – even while developing languages and practices that smooth the sharpest edges of disciplinary disjuncture. The interests of one project are *not* perfectly coincident with those of another project – even when collocated with respect to discipline and geography.

Care implies as well a mutual respect for the diversity of needs that participants bring to the collaboration, along with an openness to compromise, including the occasional willingness to relax or amend one's own interests in the collaboration to accommodate the pressing needs of another participant. As suggested above, in the absence of authoritative solutions to the challenge of heterogeneity (as would exist, say, in a traditional department contracting relationships), the grounding for this ethical model is to be found ultimately in the relations of trust and care that grow from the experience of collaboration itself. This constitutes an important and under-recognized 'moment' in the building of research collaborations more generally, and adds yet another layer to the 'thickness' of infrastructure described below.

5.2. Thick Infrastructure

Beyond pragmatic considerations, the embedded quality of data points to what we have come to call the *thick infrastructure* of ocean science. In contrast to thin understandings, in which the problems of information management are cast as purely technical phenomena, matters of hardware and software, etc., thick perspectives recognize the mutual constitution, and sometimes interchangeability, of the human and the technical (Star and Ruhleder, 1996; Star, 1999; Star and Bowker, 2002). The historical depth of this relationship mitigates against any easy (and certainly any global) answers to the problem of data integration. If data were a purely technical phenomenon (thin infrastructure), it would perhaps be amenable to the quick technical fix. To the extent that it has grown into and out of the social worlds it frames, the problematic of data is a good deal more complicated.

5.3. Collaborative Design

As scientific themes broaden, integrative efforts arise that prompt collaborative teams of increasing diversity, information systems with more complex requirements, and infrastructures of thicker sociotechnical composition. Such work brings an increasing need for new approaches to both collaborative science and interactive systems design. With internet technologies providing the method to establish a communication network for distributed computers, 'the grid' is a term that refers to networking technologies that provide distributed computational services at a global scale on the Internet. Looking to future scenarios with grids (Futrell, 2003), we consider whether a growing understanding of the multiple facets and layers of infrastructure can open up the questions asked from consideration of distributed access points to discussion of modes and mechanisms of participation with federated processes. We suggest that in addition to asking 'how to build a useful grid', we must ask community questions such as 'how to grid' – and 'how to collaborate' over time. In so doing, when Foster (1999) states that "grid technologies need to enable flexible, controlled resource sharing on a large scale", a full suite of technical, organizational, and community considerations become evident and may be recognized as choices with respect to limited resources, limiting standards, and non-federated services.

Developing understandings of everyday practices presents a rich tradition from which to build understanding of scientific work both in theory and in practice (Star and Strauss, 1999; Star and Greismer, 1989; Boland and Tenkasi, 1995; Bowker, 2000). As experience with collaborative science matures, the shift in questions asked within information and knowledge management (Kling, 1999; Davenport, 1997; Orlikowski, 1996) is reflected in the concept of design manifested in the emergence of the computer supported cooperative work community itself as well as of user-centered, participatory design efforts. A fundamental element of the ocean informatics environment is the recognition of benefits in establishing an active process of 'collaborative design', a research approach to design that we have elsewhere termed ecological design (Baker, 2004; Jackson and Baker, 2004). This approach engages participants in a community organized to consider both the framing of questions and processes as well as to provide answers to questions and delivery of products.

6. Environments by Design

We have sought in this paper to explore a set of characteristics, strategies, and processes capable of sustaining and building upon the diversity that accompanies deeply interdisciplinary collaborations in the world of ocean science. With this in mind, we have proposed the notion of 'thick infrastructure' to name the difference that endures *and should*, together with the concept of *collaborative care,* as an ideal for the simultaneous preservation and bridging of that diversity. Collaborative Design is an approach to work that is emerging to support both thick infrastructure and collaborative care. We have identified the OIE as a collaborative process rather than a product and have presented pertinent strategies of mindful variety and boundary object development. While the paper reports and reflects on our early experience with the Ocean Informatics Environment, we believe such experiences speak to a much wider and growing dynamic of collaboration, both inside and outside the world of science – and indeed, go to the heart of participatory design philosophies and practices in general.

7. Acknowledgement

Our special thanks to G.Bowker and H.Karasti for their contributions to developing data care concepts, to D.Rawls for critical input and editting, to B.Simmons for education bridging, and to the Ocean Informatics, Science Studies and IOD community participants. NSF grants OPP-02-17282 and SBE/SES & OPP-04-05069 in addition to NSF/DEB/LTER and Scripps Institution of Oceanography support this work.

8. References

[1] K. S. Baker, G. Bowker, and H. Karasti, "Designing an infrastructure for heterogeneity in ecosystem data, collaborators, and organizations," *Proceedings of the Second National Conference on Digital Government Research, Los Angeles, CA*, 2002, pp. 141-144.

[2] K. S. Baker, "Ecological Design: An Interdisciplinary, Interactive Participation Process in an Information Environment," *Proceedings of the Workshop on Requirements Capture for Collaboration in e-Science, 14-15 January 2004, Edinburgh*, 2004, pp. 5-7.

[3] K. S. Baker, S. R. Haber, and M. White, "Postnuke Portal Software: Community, Content, and Collaborative Management System," *LTER Databits Newsletter*, http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/04spring/#3fa, 2004.

[4] H. S. Becker, *Tricks of the trade: how to think about your research while you're doing it*. Chicago: University of Chicago Press, 1998.
[5] R. J. Boland and R. V. Tenkasi, "Information Management for the Intelligent Organization: Roles and Implications for the Information Professions," Digital Libraries Conference (March), 1995.
[6] G. C. Bowker, "Biodiversity Datadiversity," *Social Studies of Science*, vol. 30, 2000, pp. 643-684.
[7] G. Bowker and K. Baker, "Information Ecology: Open System Environment for Data, Memories and Knowing (accepted)," *Journal of Intelligent Information Systems. BDEI Special Series*, 2004.
[8] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn: Brain, Mind, Experience, and School, Expanded Edition*. Washington, D.C.: National Academy Press, 2000.
[9] T. H. Davenport, *Information Ecology: Mastering the information and knowledge environment*. New York: Oxford University Press, 1997.
[10] J. Futrell and AC-ERE, "Environmental CyberInfrastructure: Tools for the Study of Complex Environmental Systems," http://www.nsf.gov/ere 2003.
[11] P. L. Galison, *Image and logic : a material culture of microphysics*. Chicago: University of Chicago Press, 1997.
[12] S. Gasson, "The management of distributed organizational knowledge," Proceedings of the Hawaii International Conference on Information Systems, 2004.
[13] C. Goodwin, "Seeing in depth," *Social Studies of Science*, vol. 25, 1995, pp. 237-274.
[14] J. Greenbaum and M. Kyng, *Design at work*. Hillsdale: Lawrence Erlbaum, 1991.
[15] S. J. Jackson and K. S. Baker, "Ecological Design, Collaborative Care, and Ocean Informatics," Proceedings of the Participatory Design Conference, Toronto, 2004.
[16] S. Kaplan and L. Seebeck, "Harnessing Complexity in CSCW," Proceedings of the Seventh European Conference on CSCW, 2001.
[17] H. Karasti, Increasing Sensitivity Towards Everyday Work Practice in System Design, Ph.D. thesis, University of Oulu, 2001.
[18] H. Karasti, K. S. Baker, and G. C. Bowker, "ECSCW 2003: Proceedings of the Computer Supported Scientific Collaboration Workshop, Eighth European Conference on Computer Supported Cooperative Work, Helsinki, Finland, 14 September 2003," *Research Papers Series: A34*. Finland: University of Oulu, 2003, pp. 55.
[19] H. Karasti and K. S. Baker, "Infrastructuring for the long-term: ecological information management," in *Proceedings of the Hawai'i International Conference on System Sciences (HICSS) 2004, 5-8 January, Big Island, Hawaii*. New Brunswick, NJ: IEEE, 2004.
[20] R. Kling, "What is Social Informatics and Why Does it Matter?," *D-Lib Magazine*, 1999, pp. 1-23.
[21] C. F. Kurtz and D. Snowden, "The new dynamics of strategy sense-making in a complex and complicated world," *IBM Systems*, vol. 42, 2003, pp. 452-483.
[22] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*: Cambridge Press, 1991.
[23] C. Mukerji, *A Fragile Power: Scientists and the State*. Princeton: Princeton Unvierstiy, 1989.
[24] M. D. Ohman and E. L. Venrick, "CalCOFI in a changing ocean," *Oceanography*, vol. 16, 2003, pp. 76-85.
[25] W. Orlikowski, "Improvising organizational transformation over time: a situated change perspective," *Information Systems Resarch*, vol. 7, 1996, pp. 63-92.
[26] E. S. Raymond, *The Cathedral and the Bazaar*. Cambridge: O'Reilly, 1999.
[27] D. Schuler and A.Namioka, *Participatory Design: Principles and Practices*. New Jersey: Lawrence Erlbaum Associates, 1993.
[28] C. M. Schweik and J. M. Grove, "Fostering Open-Source Research via a World Wide Web System," *Public Admin. and Management: An Interactive Journal*, vol. 5, 2000.
[29] C. M. Schweik and A. Semenov, "The Institutional Design of Open Source Programming: Implications for Addressing Complex Public Policy and Management Problems," *First Monday*, vol. 8, 2003.
[30] R. C. Smith et al., "Ozone depletion: Ultraviolet radiation and phytoplankton biology in Antarctic waters," *Science*, vol. 255, 1992, pp. 952-959.
[31] R. C. Smith, 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, "The Palmer LTER: A long-term ecological research program at Palmer Station, Antarctica," *Oceanography*, vol. 8, 1995, pp. 77-86.
[32] D. Snowden, "Three metaphors, two stories and a picture," *Knowledge Management Rev,* vol. 7, 1999, pp. 30-33.
[33] S. L. Star and J. R. Griesemer, "Institutional Ecology, "Translations," and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39," *Social Studies of Science*, vol. 19, 1989, pp. 387-420.
[34] S. L. Star, "The structure of ill-structured solutions: boundary objects and heterogeneous distributed problem solving," in *Distributed Artificial Intelligence, Vol. 2*, L. Gasser and E. M. N. Huhns, Eds. London: Morgan Kaufmann Publishers, Inc., 1990, pp. 35-54.
[35] S. L. Star and K. Ruhleder, "Steps toward an ecology of infrastructure: design, access for large information systems," *Information System Research*, vol. 7, 1996, pp. 111-134.
[36] S. L. Star, "The Ethnography of Infrastructure," *American Behavioral Scientist*, vol. 43, 1999, pp. 277-391.
[37] S. L. Star and A. Strauss, "Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work," *CSCW*, vol. 8, 1999, pp. 9-30.
[38] S. Star and G. Bowker, "How to Infrastructure," in *The Handbook of New Media*, L. A. Lievrouw and S. L. Livingstone, Eds. London: SAGE, 2002, pp. 151-162.
[39] J. R. Wanetick and D. R. Browne, "Access to California Coastal Historical and Real-Time Data via the World Wide Web and its Applications," Proceedings of the Fifth California Islands Symposium, 1999.
[40] G. Wang, Y.-K. Nam, and K. Lin, "Critical points for interactive schema matching," in *Technical Report CS2004-0779, 31 January 2004*: UCSD Department of Computer Science, 2004.
[41] K. J. Waters and R. C. Smith, "Palmer LTER: A sampling grid for the Palmer LTER program," *Antarctic Journal of the US*, vol. 27, 1992, pp. 236-239.
[42] K. E. Weick and K. M. Sutcliffe, *Managing the Unexpected, Assuring High Performance in An Age of Complexity*. San Francisco: Jossey-Bass, 2001.
[43] G. Wiggins and J. McTighe, *Understanding by Design*. Virginia: ASCD, 1998.