Title

Ocean Informatics Monograph. Ocean Informatics Initiative: an Ethnographic Study (2002-2006). Part 1: Report

Authors

Millerand, Florence
Baker, Karen S

Publication Date

2011-06-01

OCEAN INFORMATICS MONOGRAPH

Ocean Informatics Initiative: an Ethnographic Study (2002-2006) Part 1: Report

Florence Millerand[1] and Karen S. Baker[2]

[1]Department of Social and Public Communication, University of Quebec, Montreal (UQAM)
[2] Scripps Institution of Oceanography, University of California, San Diego (UCSD)

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

Scripps Institution of Oceanography Technical Report
June 2011

Table of Contents

Abstract

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

1 Introduction

Significant technological advances coupled with increasing interdisciplinary large-scale scientific programs are driving profound transformations in oceanographic research practices (e.g. Alverson, 2007; NRC, 2003) as evidenced by programs and partnerships such as the Ocean Observing Initiative, the Integrated Ocean Observing System, the National Oceanographic Partnership Program, and the Global Ocean Observing System (Ocean Observations, 2003). New technological capacities (e.g. new instrumentation, sensor networks, and grid computing) allow for increasingly large and complex data collecting and processing not encountered in the past. These projects require development of new types of collaboration and exchange mechanisms as well as expertise in data handling and coordination. Much is expected from information technology to help with augmenting local and global digital arrangements as well as with disciplinary and multidisciplinary science (Atkins, 2003; NSB, 2005).

Cyberinfrastructure projects have flourished during the last decade especially in the natural sciences along with workshops, conferences and report publications on the subject. With the aim of fostering global interdisciplinary research through shared resources (e.g. instruments, data, and so forth), large-scale technological projects prevail given the cyberinfrastructure vision. However, the Ocean Information Technology Infrastructure Working Group Report (Ocean ITI, 2004) states that "Adequate information technology infrastructure capabilities at each local research institution are essential for designing and implementing new scientific programs and new observing systems that are needed to address complex, interdisciplinary problems in ocean research" (p.2). Thus in addition to the highly visible large-scale cyberinfrastructure deployments, 'local-scale' initiatives are emerging within research communities. The research presented in this report is about one such initiative within an oceanographic institution.

The report presents Ocean Informatics (OI), a local information infrastructure initiative enabling long-term interdisciplinary oceanographic research at Scripps Institution of Oceanography in San Diego. The OI Initiative was launched in 2002 with the idea of creating an **information environment** for managing data and information while at the same time defining a conceptual framework for the theory and practice of informatics as applied to the domain of Ocean Science. OI spans multiple organizational and disciplinary units, yet it developed framed by a keystone relationship with the scientific community within which it is embedded. OI works within the context of an array of interconnected scientific projects and practices in a seagoing, data-centric institution where research scientists address an increasingly interdisciplinary set of research questions. As a summary of progress rather than a comprehensive monograph, this report focuses largely on the emergence and evolution of the Ocean Informatics information environment (2002-2006). Four key characteristics distinguish the OI Initiative: focus on infrastructure growth (i.e. involving technical services, design strategies, and information management), a set of data practices, an informal organizational structure, and a conceptual approach. These characteristics contribute to defining a framework for local-scale information infrastructure development.

Using ethnographic methods such as participant observation, document analysis, and interviews, we observed and analyzed the development of the OI Initiative over a period of 4 years from 2002 to 2006. Ethnography is a qualitative research approach to understanding, articulating, and reflecting upon everyday practices and change through field observations, in our case focusing on ocean information management and infrastructure. A total of 18 participants involved in various ways with Ocean Informatics were interviewed, including ocean scientists, data managers, and programmer analysts.

In the next section, we provide some background by describing the institutional context and early history of Ocean Informatics, and we present the main theoretical concepts for information infrastructure drawn from the Science and Technology Studies and Social Informatics literatures. In section 3, we present the participants' description of the Ocean Informatics information environment highlighting key characteristics. In the fourth section, we discuss the specificity of the initiative, notably in comparison with other types of initiatives. The report then concludes and considers some future lines of research and development. The projects and institutions mentioned along with associated web links are summarized in Table 1.

Table 1. Projects and Programs

Abbrev	Project Name	Web Page	Organizational/Funding Context
BCO-DMO	Biological and Chemical Oceanography Data Management Office	http://www.bco-dmo.org/	NSF Division of Ocean Sciences
CalCOFI	California Cooperative Fisheries Investigations	http://calcofi.org	NOAA, California Ocean Protection Council
CCE	California Current Ecosystem LTER at SIO	http://cce.lternet.edu	NSF Division of Ocean Sciences
CCS	Center for Coastal Studies	http://ccs.ucsd.edu	UCSD Research Unit, past (see IOD)
CIP	Comparative Interoperability Project	http://interoperability.ucsd.edu	NSF Social and Economic Division
CLIVAR	Climate Variability and Predictability of the World Climate Research Programme (WCRP); Climate and Carbon Hydrographic Office at SIO	http://www.clivar.org/carbon_hydro/	World Climate Research Programme (WCRP)
CRSEO	Center for Remote Sensing and Environmental optics, UCSB	http://www.geog.ucsb.edu/history/	UCSB Research Unit, past (see ICESS)
DataZoo	OI DataZoo	http://interoperability.ucsd.edu/datazoo	Ocean Informatics Information System
ICESS	Institute for Computational Earth System Science, UCSB	http://icess.ucsb.edu	UCSB Research Unit
IOD	Integrative Oceanography Division, SIO, UCSD	http://iod.ucsd.edu	SIO, UCSD Research Division
LTER	Long Term Ecological Research	http://lternet.edu	NSF Biological Directorate
OOI	Ocean Observatories Initiative of Consortium for Ocean Leadership	http://www.oceanleadership.org/programs-and-partnerships/ocean-observing/ooi/	NSF Division of Ocean Sciences
OI	Ocean Informatics	http://oceaninformatics.ucsd.edu	Multi-Project SIO Initiative
PAL	Palmer Station LTER led out of Marine Biological Lab, WHOI; information management SIO	http://pal.lternet.edu	NSF Office of Polar Programs
R2R	Rolling Deck to Repository	http://www.rvdata.us/	NSF funded project
SCCOOS	Southern California Ocean Observing System at SIO	http://www.sccoos.org	NOAA Integrated Ocean Observing Program
SIO	Scripps Institution of Oceanography	http://sio.ucsd.edu	UCSD Directorate
UCMBO	University of California, Marine Bio-Optics	see Appendix 2	UCSB Research Unit, past
UCSD	University of California, San Diego	http://ucsd.edu	University of California system
Zoo	CCS DataZoo	http://zoo.ucsd.edu	Center for Coastal Studies data system

2 Background

2.1 Institutional Context

Scripps Institution of Oceanography (SIO), a leading oceanographic institution founded in 1903 by William Ritter, became part of the University of California, San Diego (UCSD) in 1912 (Raitt and Moulton, 1966). Directed by the UCSD vice chancellor for marine sciences, SIO is comprised of approximately 100 faculty, 300 researchers, and 220 doctoral students as well as 800 staff.

SIO oceanographers conduct research in a highly dynamic research environment supported by grants and contracts from funding agencies and foundations. Faculty members are supported for teaching and university participation while scientists carry out research supported by research grants that typically have a one to several year duration framing the project-by-project nature of oceanographic investigations as Shor (1978) described in a general history of events and Mukerji (1989) in a history of technological drivers. This difference in assuredness or continuity of funding for teaching and research support is profound and captured by the terms 'hard-money' and 'soft-money' positions. Research faculty may be fully or partially 'hard-money' while research staff (non-faculty) are typically soft-money positions. A basic research unit is the single investigator laboratory led either by a professor who teaches or by a scientific researcher supported by research grants. A variety of scientists – biologists, chemists, physicists, geoscientists, and climatologists – are involved in field research that contributes to making SIO a data-rich, analysis-intensive work arena. Field and laboratory studies involve a wide range of instrumentation and field sampling techniques as well as a variety of computational and data management arrangements.

Researchers are joined into a number of divisions (Center for Marine Biotechnology and Biomedicine, Integrative Oceanography Division, Marine Biology Research Division, Geosciences Research Division, Institute of Geophysics, Climate-Atmosphere Science & Physics Oceanography, and Marine Physical laboratory) that in turn form three sections represented on the Director's Executive Council at SIO today: Earth, Ocean & Atmosphere, and Biology. Faculty and students are joined into curricular groups (Geosciences, Climate-Ocean-Atmosphere, and Ocean Biosciences). Centers exist as an alternative organizational structure created to focus on designated research themes. Data are managed independently, largely on an ad hoc basis and may be posted publicly or sent to other individuals, centers, and/or archives. Funding for information management in the soft-money world of ocean science is shaped by the project-by-project nature of research funding and the institution's organizational structure. Usually, funding for data work is included indirectly in individual project proposal (typically funded 1-3 years) as a specification for equipment, at-sea data collection salary and/or a programmer's salary for data analysis.

The Integrative Oceanography Division (IOD) was created in 1998 as a result of mergers of a center, a group, and a division (the Center for Coastal Studies, CCS, the Marine Life Research

Group, MLRG, and the Marine Research Division, MRD). With approximately 40 researchers (excluding students and postdocs) designated Principle Investigators (PIs) today, IOD is atypical as it is one of the only divisions having both biological and physical oceanography members. The Center for Coastal Studies (CCS), formed in 1977 within a single building at the foot of the SIO pier, provided a shared infrastructure comprised of administrative and computational services for its members at the time. With the IOD merger, the CCS administration became the administrative office for all IOD. Computational services remained in place informally for previous CCS scientific members and administrative staff. IOD researchers who were not CCS members continued to take care of their own computational needs. Throughout this time, both UCSD and SIO experienced changes in their computational facilities and services[1].

Oceanography is a field with a tradition of interdisciplinary work recognizing the necessity of sampling across a variety of spatial and temporal scales (Powell and Steele, 1995; Steele, 1989; Stommel, 1963). Oceanographic research like other field observation oriented research is anchored by data collection and use of scientific data where the work of data organization and data analysis are central. Today, there is an additional key factor – data sharing. Data sharing has become part of the data work process not only for 'monitoring' studies but also for research studies in response to contemporary expectations associated with reuse of data that goes beyond the original, hypothesis-driven collection intent of scientific research. This new factor involves formation of data collections staged for accessibility and exchange (NSB, 2005). A number of significant difficulties exist at this time for sharing and preserving data formally over time due to the complexity of study of living systems, the heterogeneous nature of biological data and sampling, and the differences in semantics or language use across institutions, fields, and domains.

Interdisciplinary research approaches together with extreme heterogeneity of field data represent a real challenge for data comparison and integration. These data challenges inspired the idea of investigating a new approach to data and information management that would be conducive to support theory and practice of both oceanographic research and data management – including data organization, systems design, information managers' professional development, and so forth. The Ocean Informatics Initiative was born in response to these needs. As an Initiative that spans multiple projects, it represents an alternative to the traditional programmer venue at SIO

[1] Within the institutional context, the university today has an academic computing and media services (ACMS) focused on supporting students and faculty. In 1985, a large reorganization of the UCSD Computer Center resulted in formation of Administrative Computing and Telecommunications (ACT) and Academic Computing Services (ACS) to deal with the three major stakeholders: research, instruction, and administration. A Network Operations services group (NetOps) also emerged at this time in response to the need for focus and flexibility in developing internet-related infrastructure. NetOps was absorbed by ACT in 2007. ACS and Media Services joined in 2009 to become ACMS. Through the years at SIO an administrative group developed that became responsible for campus networking as well, a media group, and a web group, with each group reporting to the SIO Director's Office. These changed in terms of size, responsibilities, and organizational arrangements over the years, coalescing in 2010 into a single information technology group. Research groups having local needs differing from or beyond those provided by the SIO-wide services groups or the centralized UCSD ACS have traditionally developed local solutions. These solutions, of course, change over time as scientific requirements, technology and organizational arrangements co-evolve.

defined either by a single investigator's laboratory or a single project. While creating an information management aware effort, OI developed with a strong relationship with the scientific environments within which it is embedded and supports, initially the Long-Term Ecological Research program (LTER) and subsequently also the California Cooperative Fisheries Investigations (CALCOFI).

2.2 Ocean Informatics History in Brief

The concept and reality of a shared information infrastructure occurs in the early history of both Ocean Informatics co-founders Karen Baker and Jerry Wanetick. Baker and Wanetick are long-time employees at SIO, Baker since 1975 for 35 years and Wanetick since 1977 for 33 years. Baker's early work focused on bio-optical oceanography and remote sensing carried out from 1976-1983 first at the off-campus SIO Visibility Laboratory and then subsequently from 1987 on-campus at SIO in conjunction with the Institute for Computational Earth System Science (ICESS) originally known as Center for Remote Sensing and Environmental Optics (CRSEO) at UCSB[2]. A multi campus group (UCMBO, University of Marine Bio-Optics) linked Baker's work at SIO (UCSD) with that of bio-optical oceanographer Raymond Smith at ICESS (UCSB) via an infrastructure physically located at UCSB (Appendix 2). Each of these research environments featured a *shared* computational and administrative infrastructure. They could be described as boundary spanning in that they cross project boundaries and incorporate interests in data, its organization and analysis. Wanetick's work was at the Center for Coastal Studies (CCS) since its origins in 1977, the center that in merging with a research group and a division created IOD. IOD is an SIO division with participants who incorporate their view of the dynamic, ongoing nature of bringing together multiple scientific research perspectives by placing an active participial adjective 'integrative' into the division name rather than the passive form 'integrated' form referring to a fait accompli. The division web page (Appendix 1) states:

> *"The ongoing principle of the Integrative Oceanography Division (IOD) resides in a shared commitment to collaborative, interdisciplinary science. "Integrative" denotes our philosophy that multiple approaches are important in creating a better understanding of the ocean system."*

Wanetick served as system administrator for the CCS shared computer systems, participated as a dive-certified technician contributing to field research, and was programmer-data manager for a number of research projects including the decade-long project involving moorings and their data (Wanetick and Larsen, 2010)

Today, Karen Baker leads information management for two Long-Term Ecological Research sites (Palmer Station, PAL, and California Current Ecosystem, CCE) and is affiliated with the

---

[2] With a background in physics and computer programming, Baker developed first as a data analyst and then contributed to bio-optical oceanographic research. She worked at the Visibility Laboratory (VisLab) on Point Loma from 1976 to 1983 where the focus was on atmospheric and oceanic visibility studies using optical tools and image processing. In 1983 her collaborator Raymond C. Smith left SIO to join the Geography Department at UCSB, and she left VisLab to become part of the UC Institute of Marine Research at SIO that subsequently became Marine Research Division prior to being disbanded in 2003.

California Cooperative Fisheries Investigation (CalCOFI). The US Long-Term Ecological Research (LTER) program is a large collaborative effort funded by the US National Science Foundation involving more than 2000 scientists and students distributed across 26 sites organized around study of a biome, investigating ecological processes over long temporal and broad spatial scales. LTER brings a non-traditional continuity in funding and participation in networks. It is unique in having embedded data management at local sites and in establishing distributed sites as a network from the program's start in 1980. LTER has been summarized elsewhere both in terms of its science (Hobbie et al., 2003) and in terms of information management (Callahan, 1984; Baker et al, 2000). Baker and Karasti (2004, 6) report "This approach creates a forum within which the information manager role is co-constructed by participants (e.g. scientists and technicians, research assistants and students) simultaneously within the local site, the network, and the information manager community as data, technology, and science community needs are addressed." The LTER program creates a unique context with long-term funding and a safe harbor for research site development. Further, as the first oceanographic site within the LTER ecological network, the ecological perspective brought new, critical attitudes and insights relating to dynamics and sustainability of living systems that were also likely to be relevant to digital configurations. With its configuring of multiple sites as a network, LTER has played a pivotal role – practical and theoretical - in the development of the Ocean Informatics Initiative. Collaboration with CalCOFI in 2005 broadened long-term partnering with a scientific program that has a history of more than a half-century of joint oceanographic fieldwork as well as with practical experience in multi-agency cooperation (Ohman and Venrick, 2003).

Since 2008, Jerry Wanetick oversees the Computational Infrastructure Services (CIS), a recharge facility within IOD. In 2010 he also became director of information technology for the Scripps Institution of Oceanography. CIS is an institutionally recognized recharge (cost-recovery) facility available to members of IOD providing services such as desktop support, remote disk storage, backup services, and so on. In January 2011, approval was given to initiate an a recharge facility for information management. In retrospect, the legacy of shared infrastructure of the UCSD/SIO Center for Coastal Studies and the UCSB Institute for Computational Center for Earth System Science combined with the LTER-informed approach to information management led to the Baker and Wanetick joint vision for the Ocean Informatics Initiative (Baker, Jackson, and Wanetick, 2005) and the growth of a local information environment (Baker and Wanetick, 2010).

As the information manager for the Palmer Station LTER site, Karen Baker is in charge of information management and information systems. PAL, the seventeenth LTER site and the network's first oceanographic site, began in 1990 to study the marine ecosystem off the West Coast of the Antarctic Peninsula (Smith, 1995; Ducklow, 2007). The site researchers are distributed geographically across the United States with the information management component based at SIO. Information management efforts are summarized in the literature (Baker, 1998; Jackson and Baker, 2004; Baker and Chandler, 2008). The first PAL data system was launched online in 1998. The call by participants for a new system as well as a new approach came some years later at the 2002 annual Palmer Principal Investigators meeting. At that time, the existing system consisted of a hierarchically structured aggregation of data files providing for availability and sharing of data. However, an architecture that allowed for data query was lacking. For instance, it was not possible to get subsets of data by time, location, and data type, such as 'the temperature data for 1994-1997 at a designated station location off the coast of Palmer station'. Instead, one would download the 1994, 1995, 1996, and 1997 datasets, each containing dozens of station locations and many variables in addition to temperature. Palmer's need for a new information system architecture spurred plans for development of a contemporary infrastructure as well. Thus, the OI initiative first emerged as a joint computational and information management undertaking, developing concurrently as a conceptual framework. The Ocean Informatics Initiative was named and a team identified – composed of Baker, Wanetick and two programmers/analysts (Nate Huffinger and Shaun Haber). A timeline of OI development is shown in Table 2 and can be traced from OI related publications (Appendix 3). Occasional requests for action at an institutional level with respect to information management and infrastructure were made (Appendix 4) but engagement was at a project and network level rather than at the institutional level.

Table 2. Timeline of Ocean Informatics (OI) Context and Activities

1990: PAL site begin
1991: PAL site IM Karen Baker
1998: PAL data system launched online
2000: PAL participates in Network IS design & publication (Baker et al, 2000)
2002: BDEI NSF Proposal funded; STS PostDoc H.Karasti (2002); focus LTER IM
2002: PAL Information System redesign requested; infrastructure reconsidered
2003: IOD incorporated division MRD
2003: STS Grad Student S.Jackson; focus Ocean Informatics
2003: Ocean Informatics Initiative begin
2004: Ocean Informatics shared server installed (iOcean)
2004: OI Design Studio created in 2252 SVH (Baker's Lab)
2004: Ocean Informatics NSF Proposal 1 (declined)
2004: Interoperability NSF Proposal funded 2004-2006
STS PostDoc F.Millerand; Grad Student D.Ribes; focus: infrastructure, standards
2004: CCE LTER begin; OI multi-project effort begins
2005: DataCat: information system Generation 1
2005: Ocean Informatics Reading Group
2005: CCE site begin with KBaker as IM; CalCOFI collaboration begin
2006: PAL LTER computational infrastructure move UCSB to UCSD completed
2006: OI Design Studio established at 2266 SVH (student C.Arseault; focus Ocean Informatics)
2006: UCSD/Science Studies Program formal affiliation (K.Baker)
2006: DataZoo design begin; Grad Student B.Lindseth re eventlogger
2006: EML standard endorsed and adopted by IMC
2006: Unit dictionary process initiated; unit registry prototyped
2006: Ocean Informatics NSF Proposal 2 (declined)
2007: Ocean Informatics NSF Proposal 3 (declined)
2007: Ocean Informatics second server for collaborative work installed (iSurf)
2007: DataZoo launch: information system Generation 2; Grad Student E.Aronova re Big Data
2008: IOD Ocean Informatics Computational Recharge Facility Launched
2010: Ocean Informatics third server supporting virtual machines (vSurf)
2010: DataZoo: information system Generation3; Grad Student J.Donovan re Design Studio
2011: IOD Information Management Recharge Facility formation approval received

For Baker, both the LTER environment and a partnership with scholars in Science Studies specialized in study of information infrastructures for the sciences, contributed to framing the Ocean Informatics Initiative at practical and conceptual levels. Further, from an overview paper co-written with members of the information management committee emerged a realization of the value of collaborating with researchers from social science disciplines who could provide expertise addressing a noticeable lack of vocabulary and discursive capacity, of conceptual frameworks and work practices analysis (Baker et al, 2000). From 2002, Baker began a partnership with scholars in the UCSD Science Studies Program, a postgraduate program bringing together the four disciplinary departments of communication, sociology, history, and philosophy. An NSF-funded research project with Baker, an STS scholar (Geoffrey Bowker) and a postdoc (Helena Karasti)[3] influenced the framing of the Ocean Informatics Initiative.

In 2003, a few SIO researchers joined to formulate and propose a new marine LTER site. Prompted by the LTER requirement for including an information manager, Karen Baker was invited to join as the information manager for the proposed site. The proposal was successful; the California Current Ecosystem site (CCE LTER) was launched in 2004. CCE LTER has a synergistic relationship with the California Cooperative Fisheries Investigation (CalCOFI), and OI members began collaborating with CALCOFI in 2005 (Stocks and Baker, 2005). The original Palmer LTER data system designed for one project was reformulated as a system able to serve PAL and CCE LTER as well as CALCOFI.

At the same time, Baker developed two new research projects with Science Studies researchers, one supporting a three community comparative study (GEON, LTER, and Ocean Informatics)[4] and the other investigating organizational structures of informatics initiatives[5]. Efforts, involving an STS scholar (Geoffrey Bowker), a post doc (Florence Millerand) and graduate students (David Ribes, Steve Jackson, Brian Lindseth), helped shape Ocean Informatics as a practice-based initiative promoting support for sustainable scientific data management and infrastructure growth. Meanwhile, three subsequent proposals submitted to NSF in calls for computational tools and infrastructure were received with interest but not funded, often considered mismatches with the funding program calls focused on large-scale or cutting-edge computer science products[6].

Ocean Informatics development of shared infrastructure unfolded at the local level within the broader IOD community through two seminar presentations titled "IOD Data Center: Building CyberInfrastructure in an Ocean Informatics Environment (OIE)" in 2005 by Wanetick et al. and "Data, Cyberinfrastructure, and Interoperability: Highlights from Infrastructure Studies" in 2006

---

[3] A one year grant: Designing an Infrastructure for Heterogeneity of Ecosystem Data, Collaborators and Organizations (National Science Foundation NSF/EIA/BDEI 2001-2003).

[4] A three year grant: Interoperable Strategies for Scientific Cyberinfrastructure: A Comparative Study (NSF SES; 2003-2006).

[5] A one year grant: Organizational Informatics: Interdisciplinary Infrastructuring and Agents of Change Project (OPP/SBE; 2004-2005).

[6] Three multi-year proposals: 'Ocean Informatics: Integrating Heterogeneous Data with User-Centered Information Infrastructure' (NSF CISE SEI+II(IIS); 2004); 'Ocean Informatics: A Contemporary Environment for Comparative Ecosystem Studies' (NSF DBI Database Activities; 2006), and 'Ocean Informatics Learning Environment: Data Stewardship and Data Integration in Practice' (NSF OCI Data Interoperability Networks; 2007).

by Baker, Millerand, and Ribes as well as at the network level through a number of LTER task-oriented working groups (e.g. the LTER IMC Site Description Directory Working Group in 2002, the SIO Personnel Module Working Group in 2003, the LTER IMC Dictionary Working group in 2005, the LTER IMC Unit Working Group in 2007, the LTER IMC Web Services Working Group in 2009). Steve Jackson, a Science Studies student in the summer of 2003, took the UCSD shuttle from upper campus down the hill to the foot of the SIO pier to work with Wanetick and Baker to draft an IOD resource statement that is now used to describe the IOD computational context on the division web page, in job descriptions at UCSD and in proposals for NSF (see footnote 6). In addition, periodic reading groups were held (starting in 2004 monthly and evolving into summer weekly meetings) (Appendix 5). And Science Studies Program students engaged (Brian Lindseth 2006, Elena Aronova 2007-2008, Joan Donovan 2010-2011) on topics relating to a shipboard eventlogger, the history and concept of big data, and the OI Design Studio. These efforts fostered community building around the Initiative by engaging participants as well as providing an education venue. Further, the reading group helped to address the need for discussion of theoretical as well as practical concerns in planning for the development of the Ocean Informatics Initiative.

Physical growth of Ocean Informatics was evident with installation of computer servers (iOcean in 2004, iSurf in 2006, vSurf in 2010) at IOD, a dedicated space (a 'design table') in Baker's laboratory in 2004 followed by a move into a new Design Studio in 2006 (Donovan and Baker, 2011), the development of an information system given the name DataZoo 2.0 in 2004, and a more robust and flexible redevelopment of the information system in 2007 and then 2009. At the moment, the Ocean Informatics DataZoo information system houses current and historical oceanographic datasets from the Southern California Bight and the Western Antarctic Peninsula contributed by LTER and CALCOFI related scientists. DataZoo has grown to a multi-component information system with not only highly structured datasets but also highly complex data and very large homogeneous data file collections (Baker, Kortz, Conners 2011). The Ocean Informatics team has grown incrementally working closely from its start with biological oceanographers and ecologists amidst an array of institutionally-based data management efforts often having initially focused on physical data but eventually expanding to include biological data whose collection originates at SIO.[7] The focus of the Initiative is shaped initially by support from the two LTER sites funded by NSF that flows into well-delineated budgets for information management with the site information manager given administrative responsibility. Subsequently support has been provided periodically from two CalCOFI sites funded by NOAA, creating effectively an integrated, multi-agency effort.

2.3 Research Work on Infrastructure and Innovation

[7] For instance, a few examples of the national efforts with developing interests in handling biological data are SCCOOS (Southern California Coastal Ocean Observing System) established as a regional program at SIO funded by NOAA as part of an Ocean Observing System (OOS), CLIVAR & Carbon Hydrographic Data Office at SIO funded by NSF and growing out of the World Ocean Circulation Experiment (WOCE), and Biological and Chemical Oceanography Data Management (BCO-DMO) at WHOI funded by NSF growing out of the Joint Global Oceans Flux Study (JGOFS).

Emergence of so called ‘e-science’ or ‘cyberinfrastructure’ projects are major contemporary technological changes in the sciences (Atkins, 2003; NSFCC, 2007). Large-scale information infrastructures are being built aiming at fostering interdisciplinary research through shared resources (e.g. instruments, databases, and so forth) – thus crossing institutional and national boundaries. There are a number of issues associated with such endeavors relating to, for instance traditions of scientific independence, difficulties of sharing implicit knowledge, discipline-specific language, and differing worldviews. In terms of more formal organizational barriers, how Ocean Informatics plays out amidst issues related to technology and infrastructure development (e.g. Edwards et al, 2007) along with scaling and collaboratories (Bos et al., 2007) remains to be seen. Bos et al. (2007, 2) noted in their inventory of the ‘collaboratories’ that emerged in the last 25 years, “only a few of these efforts have succeeded in sustaining long-distance participation, solving larger-scale problems, and initiating breakthrough science”. Under what conditions would such initiatives be more likely to succeed? What would be the most suitable organizational arrangements enabling infrastructure-related projects? How about technological development processes, or community and user involvement? It is of common parlance for those who face the enormous difficulties of developing any useful new collaboration platform or communication tools, to stress the importance of ‘organizational issues’. Yet, these issues remain little understood. They are the objects of studies of several research fields (Appendix 6.1). Two of the literatures that provide valuable insight on these issues will be explored further in this report: Science and Technologies Studies (STS) and Innovation Studies.

2.3.1 Infrastructure Studies

A recent line of research within Science and Technologies Studies has focused on the study of information infrastructures (e.g. e-grids, help desks, computational resources, data repositories, and so forth). Pooled under the label ‘Infrastructure Studies’, these works are beginning to provide new understandings of *technological dynamics*, *organizational concerns,* and *social dimensions* associated with contemporary developments of information technologies for the support of science (e.g. Bowker et al., 2010; Ribes and Lee, 2010; Edwards et al, 2009). Within the scope of this report, we focus on three main contributions from this line of research:

- an organic metaphor for envisioning infrastructural development
- an understanding of technological change in terms of hybridity and continuity
- advocacy for a sociotechnical approach to organizational change

a) An organic metaphor for envisioning infrastructural development

Lessons from the history of large-scale infrastructures reveal the contributions of a metaphor of **‘growing’ infrastructure** in the sense of an organic evolutionary process rather than construction according to a well-defined plan (Edwards et al, 2007). First, ‘constructed’ infrastructures frequently fail to meet the users’ needs, through failure to link successfully with innovative technical, political and/or social systems and second, through inability to adapt to changing circumstances (Edwards et al, 2009). Failure rates are high because innovations in changing environments are hard to anticipate and plan for. This is certainly true for current cyberinfrastructure initiatives; any shared database project is inevitably launched within a context of changing data practices, organizational re-arrangements, and emergent technological innovations. Closely related to the metaphor or growing an infrastructure is the concept of ‘**information ecology’** (Nardi & O’Day, 1999) that allows for an understanding of the environment dynamics as a system of “people, practices, values and technologies” (Nardi and O’Day, 1999, 1). A key property of an information ecology is its locality or situatedness. Technologies get defined in the way they are used at a certain place and at a certain moment; they then could be defined in other ways depending upon the temporal and spatial contexts in which they evolve. Technologies exist within a “habitation” formed of networks of relationships between people and other technologies.

b) Understanding technological change in terms of hybridism and continuity

Despite persistence of the view that there is a single technological ‘revolution’ of such strength that it subsumes all others, ‘new’ technologies are never totally new but rather built on preexisting ones. Star and Ruhleder discuss this phenomenon and refer to it as a property of infrastructure that always happens, that is, to ‘build on an installed base’:

> “Infrastructure does not grow *de novo:* it wrestles with the "inertia of the installed base" and inherits strengths and limitations from that base. Optical fibers run along old railroad lines; new systems are designed for backward-compatibility; and failing to account for these constraints may be fatal or distorting to new development processes”» (Star and Ruhleder, 1996, 113).

Following those lines, Suchman suggests the concept of ‘**artful integration’** to account for an understanding of technological change in terms of hybridism and continuity rather than in terms of unification and rupture:

> “in place of the vision of a single technology that subsumes all others (*the* workstation, *the* ultimate multifunction machine), we assume the continued existence of hybrid systems composed of heterogeneous devices. Powerful technical systems on this view comprise not hegemonies but artful integrations.” (Suchman, 2002, 99).

Technologies are not discreet and decontextualized artifacts but results of the “collective achievement of new, more productive interactions between devices, more powerful integration across devices and between devices and the settings of their use” (Suchman, 2002, 99). Thus, development of an information infrastructure is envisioned as an artful integration where constraints of any kind (technological, organizational or social) got aligned in stable relationships allowing for take up and use by user communities.

Understanding of the organizational complexity associated with producing information infrastructure represents one of the major challenges for researchers in infrastructure studies. One of the most useful concepts for comprehending the interactions between technologies and organizations has been the **‘web of computing’** concept (Kling and Scacchi, 1982) where developing a technology (or implementing it, or having it adopted and work, etc.) implies to develop at the same time the organizational venue. This understanding implies abandoning the idea of linear cycles of technology development (from its design to implementation to usage), to instead, look at **configurations** of user communities, technologies and organizations at each of their development stages (Ribes et al, 2005; Baker et al, 2005). For instance, a comparison of strategies deployed for insuring data interoperability in three scientific communities required a comparison of the simultaneous mobilization of community, organizational and technical resources (including roles and responsibilities involved) as the strategies were developed and deployed (see: Comparative Interoperability Project http://interoperability.ucsd.edu). In one case, a few members of the community were more involved than expected, prompting some redistribution of roles and responsibilities between technologists and information manager members of the community (Millerand and Bowker, 2008, 2009; Millerand and Baker, 2010).

As we have noted previously

> "The web construct contrasts with the analytic simplicity of a discrete entity model that takes computer resources as independent and socially neutral tools. Introducing such a dynamic highlights the translation, negotiation and mediation processes occurring throughout information systems development processes." (Millerand and Baker, 2010, 138).

Envisioning the 'organizational issues' related to technological change with the 'web of computing' lens then implies a focus on the processes by which both technological and organizational 'issues' are defined interactively, as well as by the specificity of the contexts (including the institutional arrangements) in which the development and use of the technology is embedded (Kling, 1996).

c) Advocacy for a sociotechnical approach to organizational change

Along this line of thought, studies of technological change called for adoption of a **sociotechnical approach** (Bijker and Law, 1992). In contrast to a simplistic view of a one-way relationship where people (or users) are dependent on technology, a sociotechnical approach focuses on the reciprocal relationships by which technologies (e.g. a database system), social contexts (e.g. a center where the database is developed or a research community where it is used) as well as organizational and institutional arrangements (e.g. a formal requirement to publish the data in the database) are constituted as a single seamless web. Work practices are understood as composed of multiple elements, as "networks of people, tools, organizational routines, documents and so forth" (Berg, 1999, 89) where the technology is one element among others in the network configuration.

In order to understand technological change, one needs to observe processes of stabilization and destabilization of the networks: "a technology is stabilized if and only if the heterogeneous relations in which it is implicated, and of which it forms a part, are themselves stabilized. In general, then if technologies are stabilized, this is because the network of relations in which they are involved -- together with the various strategies that drive and give shape to the network -- reach some kind of accommodation." (Bijker and Law, 1992, 3).

But developing an information infrastructure would never be as simple as 'installing and using the new infrastructure'. As Berg notes: "The sociotechnical approach argues that development projects should be seen and managed as being the politically textured, organizational change processes that they inevitably are. " (Berg, 1999, 94). Technological and organizational changes are inevitably political (Winner, 1985); they likely contribute to redistribution of tasks and responsibilities, are used in disputes and conflicts, and/or embed values that may differ for various groups of actors.

2.3.2 Innovation Studies

Innovation Studies have emerged in the fields of management and economics as researchers sought out understandings of the processes of innovation. Since innovation is considered a major driver of the economy, the factors that may lead to innovation are considered to be critical. A pioneer in these fields is Von Hippel whose early studies focused on determination of the roles of the concerned actors, namely of users and manufacturers in innovation processes. One of Von Hippel's key findings is that a large number of innovations are made by users. A 1976 study on scientific instrument development showed that "approximately 80% of the innovations judged by users to offer them a significant increment in functional utility were in fact invented, prototyped and first field-tested by users of the instrument (i.e. the scientists themselves) rather than by an instrument manufacturer." (Von Hippel, 1976, 212). Innovation is often local, situated, and closely tied to users needs. This seems to be even more true for technological innovation in scientific contexts where scientists have a long tradition of developing technological tools (instruments, software, and so forth) either on their own or in collaboration with engineers to further their particular needs - that Segal (2009) framed as a 'culture of professional end-user developers'.

Two phrases coined by Von Hippel – **'lead users'** and **'innovation community'** - are pertinent to the development of a local information environment. The definition of lead users is as follows:

> "Lead users are users whose present strong needs will become general in a marketplace months or years in the future. Since lead users are familiar with conditions which lie in the future for most others, they can serve as a need-forecasting laboratory for marketing research." (Von Hippel, 1986, 791).

The definition of innovation communities highlights having "user-led innovations as a collective process" (Van Oost et al. 2009, 184). The overarching concept of innovation communities is defined as

> "… an organized cooperation in the development, testing, and diffusion of user-initiated innovations. Users as well as manufacturers can be members; the innovation community can be purely functional but may also fulfill the role of a social (virtual) community providing sociability, support, a sense of belonging, and social identity" (Von Hippel, 2005, 96).

Drawing from the work of Von Hippel, Van Oost et al. (2009) suggested the concept of **'community innovation'** to describe the type of emergent, user-initiated project in which the community itself is an essential element of the innovation. Inversion of the word order – from innovation community to community innovation – reflects the idea that the community doesn't necessarily precede the innovation but rather may be formed simultaneously with the innovation. In other words, at the same time that the technological innovation is developed, the social community develops as part of a process of co evolution between the two.

2.4 On Methods

Qualitative methods (Strauss, 1987) are well suited to provide grounded understandings of emergent phenomena. They are able to provide 'thick' descriptions (Geertz, 1973) of the object under study, which in our case is an information infrastructure initiative in the course of its development. In addition, they provide the views and meanings of the object for the actors involved, in this case including the founders of the Initiative, the team of developers, the expected users (i.e. the scientists) and so on. Data presented in this report have been collected with ethnographic techniques such as participation observation, interviews, and document analysis. In a nutshell, ethnography refers to observations carried out 'in the field' where each study may represent an extended case study or a briefer case example. An ethnographer creates a *qualitative description* of human social phenomena, based on fieldwork, and emphasizes the study of the phenomena *in action* (as they occur and develop), ultimately producing an account of the phenomena in the form of a text (such as this report). In the natural sciences, an equivalent of this type of social science field study might be called a field study or a project report.

The involvement of the authors of this report with the Ocean Informatics Initiative is ongoing, and data are still being collected. The data presented in this report, however, have been limited largely to those collected over a period of four years (from 2002 to 2006). We conducted 26 interviews with 18 participants including founders of the Initiative, developers of the infrastructure (programmers/analysts, systems administrators, technicians) as well as data contributors and other users (research scientists, graduate students, post-doc, educators, institution representatives or administrative participants). Techniques used involved individual and small group interviews mainly using a semi-structured format. All the interviews were recorded and transcribed. The quotes for this paper are marked as follows: information manager (IM), and scientist whose primary focus was not information management (S). For ethical concerns, all names and places are anonymized. Table 3 provides roles, working titles, categories, and final products for some of these interviews.

Table 3. Final products and interviewee roles, working titles, categories.

FINAL PRODUCT	ROLE	TITLE	CATEGORY
processed data	data analyst	programmer analyst	DM/IM
information system	system developer	programmer analyst	DM/IM
dataflow	data processor	staff research associate	DM/IM + technician
database system	system developer	programmer analyst	DM/IM
dataset package	field technician, data manager	programmer analyst (SRA before)	DM/IM + technician
structured and processed data	data analyst and data user	post doc (scientist)	DM/IM + scientist
database system	system developer	programmer analyst	DM/IM
data	data producer and data user	scientist	scientist
database system	database designer	programmer analyst	DM/IM
processed and analyzed datafiles	data producer, processor, analyst and user	staff research associate	DM/IM + technician
*DM=data manager			
*IM=information manager			
*SRA=staff research associate			

Participant observation was carried out in participants' offices, on docked ships during cruise preparations, at workshops, and at local and community meetings as well as at 'design sessions' in the Ocean Informatics Design Studio where system development could be observed 'in action' along with work processes, collaborative interactions, and communication mechanisms. Observation of design sessions and demos provided useful insights into the systems developed and the design strategies. An earlier organizational monograph for the LTER information management community (Baker and Karasti, 2004) provides an overview of information management within the LTER community within which Ocean Informatics participates.

The ethics of many of the data and information issues is complex (Winner, 1986; Goguen, 1997; Feng, 2000; Carusi, 2006; Carusi and Jirotka, 2007). As part of the ethnographic research process, Institutional Review Boards (IRB) review ethics approval requests at universities. An IRB was applied for and obtained at UCSD for this study. Interviews and events are frequently audio recorded so consent forms are required that inform and prompt participants to consider their rights and the potential ramifications of recordings. Developed in response to ethical concerns relating to observational studies, the process is now applied to participant observation venues as well where relations may range from partnered collaboration to unforeseen adversarial situations. Of course, central to participant observation are the notions of unanticipated results and of critical inquiry, pursued collaborative, joint investigations aiming at understanding and thereby informing ongoing work practices – the elements and their characteristics, the strengths and the weaknesses of work arrangements.

Document analysis included digital artifacts (such as information system documentation, prototypes, interfaces, tools, etc.) and textual documents (reports, publications, proposals, laboratory archives, event flyers, etc.). Grounded theory methodology including theoretical sampling (Strauss and Corbin, 1998; Glaser, 1978) guided the data collection and analysis process. Grounded theory, a qualitative analysis approach used with ethnographic materials, focuses on defining social phenomena through systematic methods and generation of theory from empirical investigations. Theoretical sampling is a non-random sampling technique frequently used in small-scale qualitative research studies that is explicitly emergent and targeted (Glasser and Strauss, 1967). Analysis included data coding for theme identification and analytical category elaboration. Ethnographic materials including field hand-outs, interview questions, and memos are collected in Appendix 6. An overview of qualitative analysis software and the unsuccessful attempt to use the software package Semato are given in Appendix 7.

Authors of this study look at the Ocean Informatics Initiative with both insider and outsider perspectives since Baker is one of the Initiative founders as well as a primary source who developed over time into a field-based ethnographic co-investigator of sorts. As a participant and an observer, she was able to provide grounded meanings of the data, data management and infrastructure practices. Acting more as a 'classical' ethnographer, Millerand provided input on reframing analysis and interpretations in a broader context. In contrast to more traditional participant observation through regular interventions and active participation in some events, she added a more active, collaborative element, an approach captured somewhat by the notion of action research (Rogers, 1997; Lau, 1999). For instance, she wrote an 'imaginary' dialogue between a social scientist and an ecological scientist as a pedagogical tool to explain the meaning of the term 'social', providing concrete examples of the social scientist's scientific contribution when studying technology use and development in a research community such as LTER (Appendix 6.1). It is precisely this active, multi-dimensional collaboration that stimulates reflection and supports mediation across distinct disciplinary fields in addition to intermediation between levels of granularity that aids in the identification and unfolding of work understandings (Agre, 1997).

3 The Ocean Informatics Initiative: Participant Perspectives

The OI Initiative is developing during a time prior to existence of a shared vision with respect to key elements of a national information infrastructure. It is a time when national and international organizations are focusing on standards and national archives reinventing themselves (e.g. Higgins, 2009; NOAA/NESDIS 2008). It is unknown whether a data repository 'link up' will be hierarchical, web-like, or some combination of these and other arrangements. It is not known how data sharing and data production will alter the focus and distribution of scientific work or what roles will develop to support new types of data and data practices as well as the increase in scientific collaboration and community participation. Views on change exist at all levels, bottom-up, top-down and middle-out though the data sharing impetus from funding agencies is well-recognized as summarized by one scientist:

> *"(…) some of it is dictated top down, um, strings attached to the NSF monies these days are 'thou shalt share'." (S)*

We present four characteristics of the Ocean Informatics Initiative identified by participants in interviews:

- An information infrastructure (from a sociotechnical standpoint)
- A set of data practices (from a data-centric perspective)
- An informal organization (from a human and institutional standpoint)
- A conceptual approach to infrastructure growth (from a design perspective)

These categories represent different lenses through which Ocean Informatics participants define and make sense of the Initiative as it is developing from 2002 to 2006. As mentioned earlier, participants include founders of the Initiative, developers (programmers/analysts, systems administrators, technicians) as well as data contributors and other users (scientists, graduate students, post-doc, educators, institutional representatives). We will refer to all of them as participants while we distinguish quotes from two groups: IM (for information managers, referring to founders and developers) and S (for scientists and students, referring to data contributors and other users).

3.1 Ocean Informatics as an Information Infrastructure

Ocean Informatics is seen by some of the participants from a predominantly technical perspective, particularly with respect to its work with computer systems and physical infrastructure, and by others from a broader perspective including organizational and social dimensions. This aspect of OI elaborates on the central importance of design of the information systems (of how they are created as part of the Initiative) as well as of the standards used.

3.1.1 Computer systems and physical infrastructure

To the environmental researchers who work with the Ocean Informatics team, the information system is the most visible part of the OI Initiative, and most needed in order to make their data publicly available (a requirement for some projects and from some funding agencies). Therefore, from the researchers' perspective, the definition of Ocean Informatics as an entity is built on the notion of computers and technology, and consists mainly of a database that they would use to access oceanographic data:

> *"Um, but what Karen and Jerry, Jerry Wanetick, are proposing to do … will create an infrastructure where different kinds of data can go in, and the data should be able to be integrated. And people want to look at the variability, a biological entity in relation to a physical process. They should have all the resources to extract that information back out." (S)*

A related perspective common to both researchers and data managers is Ocean informatics defined as a "set of tools". Those tools, mostly oriented towards data aggregation and integration include dataset templates, shared metadata information and common data format rules, in other words, standards. The local characteristic of these tools is highlighted, as well as the critical need for standards:

> *"Ocean informatics is [a] group of people here at Scripps who want to work together to build some tools from the ground up, meaning kind of build their own tools and not rely only on applications and tools developed from higher funding sources" (IM)*

> *"It's very important to reach agreements of how to process the data, how to insert the data into the database, which fields we should use in order standardize everything. I mean standardization is basic." (S)*

The Ocean Informatics Initiative is viewed as a key actor in developing standards to facilitate data interoperability and data integration at Scripps for some biological programs, particularly given the extreme heterogeneity of ocean data:

> *"Ocean Informatics is a kind of a goal I guess. It's a goal to take dissimilar heterogeneous data of all sorts of formats and somehow connect them or establish some standards that will allow the connection of data. Things like the unit dictionary, the standard vocabulary, some data format you know rules. (…) So I kind of see Ocean Informatics as kind of a, like I said, a goal to establish some rules that facilitate the data connections." (IM)*

3.1.2 Organizational infrastructure and infrastructure-building

Rather than a purely technical effort, Ocean Informatics Initiative is also envisioned as a combination of social (human) elements and technical (systems) ones. As the following quotes illustrate, the importance of the 'human' side of infrastructure is key for members of the OI team:

> *"Well I would say Ocean Informatics is [laughing] a combination of a team of people, a set of design strategies, and a set of goals." (IM)*

> *"I don't know, I mean, it does make sense to me that the infrastructure doesn't really exist without people. Or it would exist but it would be completely unfunctional. And so if there is an infrastructure that doesn't function is it really an infrastructure? So I can see, I mean it's clear to me that the people are part of the infrastructure." (IM)*

.

Ocean Informatics' technical particularities also reside in the design strategies adopted by its team, focusing on usability and long-term sustainability. Those strategies imply incorporating members of the community into the design process rather than handing them final products. To take into account the researchers' needs all through the development of their tools, the team has chosen to include them in the process at very early stages of the design process in order to get their input and comments on the progression of the projects. Participatory design[8] is a key strategy influencing the Ocean Informatics design approach. The people implicated in database design, for instance, work side by side and constantly share information about their progression. They also modify their individual tasks from project to project, in order to see different aspects of their work and become more polyvalent:

> *"As a design strategy I think it's about really trying to incorporate members of the community into our design process, not trying to just hand them finished products and telling them but to work with them. Really inviting them in, you know at really early stages in the design you know. Even when certain projects are still in what you would call a beta state, getting the researchers in there and looking at and letting them tell us what they think about it. So I mean I think it's this community-based design." (IM)*

Ocean Informatics design strategies deal with the long-term in the sense that such strategies are embedded in both the coding methods and the communication practices. For instance, the programmers would use modular programming, planning for code reuse rather than narrow task-specific or one-time use code:

> *"And also there are other design strategies we've really emphasized in the last year. Like designing with API's instead of just writing procedural code. You know our new focus on documentation. And then I'd say its also a set of goals we have in terms of educating users in addition to just providing tools." (IM)*

Rather than simply presenting finished products, designers present options and development phases in order to open up discussion. This focus on the long-term brings its own challenges, mainly in terms of resource allocation and funding. As this information manager explains, development of applications or systems is usually envisioned as a short-term goal:

---

[8] Participatory design is a design approach that emerged in the 1960s in Scandinavian countries with the aim of actively involving all stakeholders (notably end users) in the design process to help ensure that the system designed meets their needs and is usable. (e.g. Schuler and Namiok, 1993; Blomberg et al, 1996).

"And you know I think (…) the easy way is to go short term, to go with one-use applications. One use programs. Hire a programmer for three hours to fix whatever specific problem you're looking at, as opposed to creating an environment that will address all of those things in a community fashion or in a group situation where its the longer term. It's a bigger picture kind of situation. I think a lot of scientists are so in tune with the funding and with the short-term answer, with the need to get things done yesterday and to put out papers, grants and proposals … and all of those things that the long-term is really hard to look at for practical and for emotional reasons. So I think there are a lot of challenges." (IM)

Finally, Ocean Informatics also refers to infrastructure building strategies where infrastructure is envisioned as a sociotechnical approach i.e. thinking of both technological, organizational, and social development at the same time (see theoretical section 2.3). Multiple dimensions of some Ocean Informatics infrastructural elements have been formalized recently by its two founders, highlighting the interdependencies among the facets of information infrastructure: social (individual roles, team, communities), organizational (strategies, resources, policies), and technical (design approaches, solutions, applications, systems) (Baker and Wanetick, 2010).

In setting the conditions so that information management expertise could develop and better serve contemporary oceanographic science, OI aims at providing information management services to scientists through an integrative team effort. As one participant explains, information management effort and expertise is usually very fragmented, and Ocean informatics could be a 'place' where common expertise and resources could be developed:

"I mean there all these different projects that all have different funding sources and nobody is really working together with each other. They're just kind of working in their own little world. So actually it's all very disparate. And the idea behind Ocean Informatics is to create a place for people to come together to share and to communicate with each other." (IM)

3.2 Ocean Informatics as a Set of Data Practices

The Ocean Informatics Initiative is also defined as a set of data practices where the phrase 'data practices' refers to specific ways of working with scientific data. Data practices are very diverse and understudied in general. They include various aspects of data management, e.g. data processing, analyzing, formalizing, formatting, organizing, standardizing, and so on. Achieving interoperability, i.e. making data comparable across studies and disciplines, is an ultimate goal. We found three main ways of seeing OI as a set of data practices with different outcomes:

- A service to support and enable scientific work; this view is an expectation among institutional and scientific participants;
- A shared practice tying together members of the OI team, understood as a community of practice; this view is found among programmers, information managers, and students;

- An applied practice with scientific underpinnings aimed at identifying data issues, managing the data and the technical constraints associated with digital records as well as formulating and implementing informatics approaches.

3.2.1 A service

For most participants the Ocean Informatics Initiative is seen as a service to support and enable scientific research. Ocean Informatics is recognized as a means by which scientific work is accomplished because it provides for data access, use, and sharing. From this perspective, OI remains a black box with 'raw data' as an input and 'manageable data' as an output. What these data practices *are* is not the main focus, but rather why they are useful.

> *"You know they [the scientists] see, they still have like a very modular view. OK they have their lab, and they produce their data, and then it kind of goes into a black box. And then it's on the web. You know, and (...) the black box being, yea the Ocean Informatics. So I think there are still a fair amount of people who have the view that, you know, ocean informatics is a service they use but not something they are really part of." (IM)*

The Ocean informatics initiative is envisioned as providing services both technical (e.g. data storage, data processing) and informatics related (e.g. strategies for organization of data, for analysis of data, and so on). In a discussion with a researcher about where to put the description about information management on the project web site (work carried out by Ocean Informatics members), the question was whether to put it under the 'data' page or under the 'scientific research' page. The researcher pointed out that the information management content should go on both pages because information management work is closely tied to scientific research. Information management is by definition a service to a research domain and it implies close interactions between scientist and information managers on a daily basis. For instance, helping the scientists with the analysis of the data requires developing analytic tools as well as assessing associated results. This type of work necessarily involves a large set of discussions and exchanges. In contrast, providing technical support involves mainly the intervention of a technician without requiring close work with the scientists, as this participant explains:

> *"Should it be with the other components or should it be with the data component? And I'm saying it should be with the science component, with the scientific research (...) [because information management] goes both ways. The technical services goes only one way." (S)*

3.2.2 A shared practice

For most data and information managers, the Ocean Informatics Initiative is defined as shared practices, and not primarily a service to scientific research. Shared data practices are what ties them together. As such, the OI Initiative works as a community of practice, i.e. as a "group of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly" (Wenger, 2006) (Appendix 6.9). Information managers view themselves primarily as practitioners engaged in a shared practice which is essentially about managing scientific data in a standardized way through common ways of doing and thinking about data:

”I’m unsatisfied with the way that data is managed and recorded and used at my own personal level. I mean even with the way I do it, especially with the way I do it. (laughter) And it’s my realization that using the data is complex and hard to do, even getting access to it. And so I see Ocean Informatics as a way of standardizing or formalizing data management and data processing. Data use. And so there’s lots of different tools that are under that broad heading. You know there’s standardized metadata, there’s you know how there’s data formats, and there’s how data is transported, how data is checked.” (IM)

Data practices have still to be developed and agreed upon. There are many approaches to any particular aspect of working with data. The OI Initiative provides a space for information managers to discuss these approaches and to choose an appropriate one to establish for common use:

“It’s a goal to take dissimilar heterogeneous data of all sorts of formats and somehow connect them or establish some standards that will allow the connection of data. Things like the unit dictionary, the standard vocabulary, some data format you know… rules. (…) Everything from the metadata, the data collection to the data processing and analysis and interpretation to data archives and data collections. Like I said, there’s a bunch of them [data collections] scattered around that are related but aren’t necessarily connected.(…) So I kind of see Ocean Informatics as kind of a goal to establish some rules that facilitate the data connections.” (IM)

From this perspective, although sharing common data practices implies working with common technical infrastructure and tools e.g. dictionaries or information systems, the main focus is on common ways of doing data management:

“[Ocean Informatics is] something that makes it easier for me to bring in lots of different data. And the technical things that make that happen are interesting but are of secondary importance to me right now so.” (IM)

”Ocean Informatics, by placing the spotlight on how you actually deal with the data, forces you to be more professional about it. And I think so when Karen brings up these broader issues of metadata and what does it mean when you say O_2 concentration, it forces me to think in those terms of how can it be used and how it can be accessible to other people. And then makes my work product cleaner. Easier for me to go back and use myself (…) because I’ve taken the steps to formalize it or to standardize it.” (IM)

3.2.3 An applied practice with scientific underpinnings

OI provides new opportunities for members to work with data and data curation, technology and systems design, and other communities and their participants, while adding some form of rigor and validity to work with data. Informatics is an applied field, in the OI case applied to oceanography, as has been summarized in a report by one of its founders: “Ocean Informatics is the application of informatics to the domain of Ocean Science. It is the work that occurs at the intersection of oceanography, science studies and information management." (Baker, 2005, see Figure 3.1).

It is common in many research communities[9], that an information manager doesn't have a formal expertise in informatics. Coming from very different backgrounds (from computer science to biology to oceanography), they have learned data management on the job, thus leading to various levels of knowledge and know-how, typically well-connected to local circumstances. They then share a common interest and engagement in improving data management, information systems, and the larger information environment based upon a combination of available and emergent techniques, standards, and scientific principles. In recounting how he got involved in the community, a data manager explains that he was attracted by the OI "standard based" and "right way to do it":

> *"And two things that appealed to me about working with Karen is one that she understands sort of the importance of looking at how things are done, not just whether they are done. And also that her team is looking at things from a computer science perspective and not from a (…) everyone over there is a biologist first and foremost. And they got their training in biology and have only come into computers through their work. So they have learned on their own (…) and it creates a really informal, unstructured approach to system design, whereas people over here (…) have all had computer science backgrounds. So its very structured and very you know very standard based and the right way to do it." (IM)*

3.3 Ocean Informatics as an Informal Organization

From the interviews with data and information managers emerged the vision of OI as an informal organization, i.e. a loosely structured group of people sharing common goals, expertise, and resources, typically crossing institutional boundaries (disciplinary divisions, universities, government laboratories and so on).

> *"Ocean Informatics is more, well [another group] is a set group of people whereas Ocean Informatics is more of an idea that has sort of an ebb and flow of individuals." (IM)*

From this perspective, Ocean Informatics is in the first place a group of people – the individuals behind the tools – which constitutes the "OI team".

> *"When I think of Ocean Informatics I think of it as a group of people." (S)*

The team represents an array of expertises, and Ocean Informatics represents a coordinated access to this expertise:

> *"We spend, you know, we work together every afternoon. So we're pretty comfortable with each other's levels of expertise so we can make the assumption that OK Shaun's*

[9] On information management communities, see: Karasti and Baker (2004); Baker and Karasti (2004).

> *going to know what I'm talking about if I just start saying whatever, you know, random database terms or something. Also I mean a lot of the other technical people I've met have at least roughly the same level of expertise. So I think in general it's a little safer then to make assumptions about kind of a shared background. And so you can kind of dive right in to talking about whatever the meat of what you want to discuss is". (IM)*

The OI Initiative remains unrecognized organizationally. They don't think of it as a facility, a project, a laboratory or even an organized group. It tends to represent a place to find an individual with whom they are familiar or as a source of skills and independent services (e.g. programming, databases, data management) that are available from a set of individuals rather than from an organized group.

Ocean Informatics is defined more as a network than as a formal group of people, in the sense that it's an ad hoc community that forms and reforms, spread out across a department that exists over time. Its identity and recognition within IOD (its institutional division) is beginning but it remains largely invisible at an institutional level, which puts it in a fuzzy category:

> *"Well Ocean Informatics is kind of this ground up ad hoc community. So a challenge is convincing other people, you know, about exactly what we are doing. And I think a lot of people tend to get confused: what is Ocean Informatics? Because it's not a project, you know it's not a source of funding or anything. It's very much this invisible community that we're trying to make visible. We're trying to make visible, you know how Karen is documenting meetings, and you have the blog... But there is, there is really no centralized Ocean Informatics location or place. Its very spread out and kind of just forms and then it might form again. You know just when people meet." (IM)*

Even if the initiative remains invisible institutionally in the sense that it has no official institutional status, there is a certain level of organization within the group, for instance in terms of management (e.g. hiring, budgets, proposals, committees), communication and coordination mechanisms (e.g. common calendar, shared wiki for documentation), visual identity (e.g. a logo, an Ocean Informatics Web site), activities and events archiving, periodic review of tasks and work distribution, and common venues (e.g. participation in the annual LTER information management meeting). Yet, non hierarchical decision making, work autonomy, collaboration, and informal communication are core values within Ocean Informatics.

The Ocean Informatics team represents a learning environment for most data and information managers. Being part of the team allows for learning from each others experience, for acquiring new knowledge and expertise first by learning and then by implementation, second by developing sensitivity to and teaching participants about ramifications of choices in data packaging, information management strategies, and so on.

> *"Since I joined the Ocean Informatics community I think I've been much more aware also of the outside, of things outside of Ocean Informatics. I can kind of use the Ocean Informatics model or environment, as a lens, by which to look at other programs and to have a greater insight into what problems and solutions they're trying to deal with. It's been exciting; it's really fun, and I really enjoy it." (IM)*

Finally, one of the most important aspects of the Initiative is that it provides a data commons. It contributes to building ties between researchers and institutions in putting together data that were rarely aggregated before. OI is not a typical research project for Scripps (i.e. a project led by an individual researcher over a defined period of time – typically one to three years), it's an initiative that brings together disparate and heterogeneous research environments (two LTER sites, two CalCOFI sites, and individual investigator projects) for the long term.

> *And the idea behind Ocean Informatics was (...) not so much a top down method where UCSD could say OK you ought to meet here and figure something out [about data]. It's a ground up effort all volunteer. I mean if you want to come and talk, scientist A and scientist B want to come meet together and talk and then they figured hey we should try to get scientists C and D to join us because they work with those other projects. And we could all figure what we are all doing. And maybe we can help each other out somehow. We can you know, we don't have to duplicate any efforts here, we could work together in accomplishing something even though we kind of all come from different places. You know we're all located here in the same place so we should be sharing some efforts. So that's the idea behind it." (IM)*

3.4 Ocean Informatics as a Conceptual Approach to Infrastructure Growth

The Ocean informatics Initiative is defined as a conceptual approach to infrastructural growth from a design perspective. We found this view depicted via three different themes described by the OI participants:

- An integrative undertaking
- A long-term, networked, and collaborative effort
- A scientific, investigative activity

3.4.1 An integrative undertaking

Ocean Informatics represents a conceptual approach that frames the work of the group. As the group works with data from oceanographic projects that focus on field research and the ecology of the marine ecosystem, it is used to dealing with disparate data types and projects arrangements. On the website, the idea of Ocean Informatics is described more concretely as a multi-project endeavor:

> *"At UCSD Scripps Institution of Oceanography, Ocean Informatics is a multi-project endeavor that invites designers and researchers into collaborative arrangements such that scientific data practices, information management, and application design are co-constructive. The aim is to provide data access and local integration in the short-term while researching and improving understandings of data organization, description, and federation. These then inform data stewardship efforts including local repositories, community standards, and sustainable information infrastructure. The Ocean Informatics initiative*

includes design and population of the DataZoo information system together with a variety of other applications and infrastructure building activities."[10]

OI is a kind of short-hand or buzzword that packages together a way of carrying out data work together with designing an information infrastructure that reaches across projects. As a service to the scientists whose data is organized typically within individual laboratories, OI brings new approaches to formulating projects that enlarge the scope of studies and thereby for field science enlarge the scope of data handling. Information management in conjunction with information infrastructure may be viewed as providing overarching concepts inclusive of data management, data analysis, information systems, informatics, and networking. As this participant expresses it, there is a well-recognized need 'to bring it together':

"The way that you are able to easily access large datasets and how they are annotated, and basically bring datasets into your own desktop without much trouble. So, that's how things should work. Things should be easy and it should be easy to grab information, know what you are grabbing, know that details of the information, the metadata. What are the potential pitfalls of the data? And so that's where the informatics comes in I think (…) to make you aware of the nature of your data basically." (IM)

"There's not a commercial driving force that is really pushing people to bring it [all these different datasets] together. I think that there would be two main forces. One you can get better science papers by having different datasets to work with (…) The other way would be that people just can't afford to manage their own data anymore (…) because their datasets are getting so large. So they have to adopt some sort of common practice, data management practice, that makes it much easier for the data to come together." (IM)

The need for moving from an 'ad hoc basis' to an 'integrated design' for data arrangements is recognized by this participant, with the understanding that data integration also brings new challenges for scientists:

" [data integration] will require a culture change. And require much more integrated design of data recording, of data processing, and conversion of those data into useful information.(…) In the past we have done this kind of thing usually on an ad hoc basis as individual investigators. But at this juncture we have to think about the different pieces of the data, how they fit together, and how they can be communicated in a way that will be useful to us internally, and to the outside world. So that it is the integration of data, a new, very new, element for us." (S)

As an example, while one of the first steps recognized was the need to create a 'happy format', that is, a format useful and useable by several participants, such cross-purposing is recognized as increasing the effort involved:

Well, data integration means various types of information sources, in just one format and this is the first thing (…) normally integration implies create a format, but also that this format is

[10] Ocean Informatics, http://oceaninformatics.ucsd.edu/, visited 24 July 2010

useful for all the people, not only for you, for your mind, but also for other minds. So, you have to cross purpose. Yeah. So that's why the effort is double. To integrate the data where your format is a happy format for other scientists. (S)

3.4.2 A long-term, networked, and collaborative effort

Ocean Informatics creates a collaborative space, providing a shared mechanism to move beyond individual project or program identities to build collaborative work:

> *"Ocean informatics is a term we use to describe any time that there is some sort of collaboration between two projects. (...) He and I met the other day to talk about some database issues. Because you know he's trying to accomplish something similar to what I am, and similar to what [another person] had been helping out with. And you could say that when the three of us get together and come up with an idea, that was an ocean informatics accomplishment. So it's not just any one person. It almost feels like any time (chuckle), so any time two people meet, at least two people meet from different affiliation sources, then that's Ocean Informatics." (IM)*

The OI Initiative is shaped by the long-term nature of its four major projects (two LTER sites and two CALCOFI sites) as well as by its organization as a network of nodes:

> ***"****The Ocean Informatics model that I had in mind very much mirrors the science model within the LTER which is a network of nodes doing site-based science and yet collaborating as a network. So within Scripps then Palmer could be seen as a node, CCE could be seen as a node (...) CalCOFI could be seen as a node. And could we create a network of nodes so that the network, the conceptual umbrella for it, would be Ocean Informatics." (IM)*

Networking is seen as a way to address the question of scaling the Initiative from a local environment to 'something' broader:

> *How will that scale? I don't know. (...) That's why networking has an appeal. We don't have to grow really big. We can grow by growing our relations and our partnerships. (IM)*

The long-term is what constitutes the identify raison-d'etre for Ocean Informatics and provides continuity for envisioning the sharing of datasets across space and time, as this participant explains:

> *We're tasked with making datasets available for long-term scientific use. So it's a growth from what it used to be within the university, from one dataset for one particular question. Some of the scientific questions themselves are changing, they're broadening and as a result of that broadening the scientists need not just the data set they collected (...) the data needs to be shared with colleagues (...). One can still exchange files person-to-person, point-to-point, but now more often there's this notion of a central repository of*

some kind, enough to foster consideration of computational arrangements and requirements, and of successes and failures over long periods of time. (IM)

Long-term data becomes a player in and of itself. An information manager describes how it takes on a life of it's own, and how it creates a different set of aims both for scientists and information managers:

When you are talking about long term data, that focusing on the data actually creates a different set of aims than focusing on a hypothesis and a research question. Sometimes you have to prioritize, whether you are going to finish that project or whether you are going to do something that leaves the data in a state for re-use or for preservation …because of the closeness with the data … Before maybe you were thinking about the trajectory of the research question that was part of the trajectory of the career of a scientist you are working with, which is what allows research to continue, whereas now the data comes in as a player in and of itself. Now suddenly there are technical issues and social issues to doing things for the betterment of the data. And the data living longer than the scientists has a big impact. So suddenly the work I had to do bifurcated. (IM)

This participant articulates what is a recurring theme arising throughout this investigation, that is an awareness of the importance of close association of information management at the data collection field level with the scientific efforts.

3.4.3 A scientific, investigative activity

When considered as a conceptual approach for infrastructural design, Ocean Informatics is regarded as a field of research and development relating to information management and informatics. This view of OI as a 'whole science by itself', which is mainly found among data and information managers, is expressed by this participant:

We are using a lot of data. And so the data management is a very important part because we have a LOT of data, so I think that it has to be included with science and not be excluded. Because it is a whole science by itself to manage the data because you have a lot of data. It's not an isolated project, a small isolated project, in which the management even can be done by the scientists. But in this case you have to be an expert in data management. Because they are working very tightly with scientists to be part of the science. It's also science, the science of managing the data. (S)

The OI vision involves a research-based informatics approach working hand-in-hand with its service-oriented data management responsibilities. As described by its founders in a 2007 proposal:

"Ocean Informatics Environment provides an arena for defining informatics and technology products that support the biological goal of cross biome comparative scientific studies by taking into account multiple data types and scales and by making heterogeneous data accessible and queriable. Many times technical goals dominate database and infrastructure proposals. Table 4 is provided in order to describe the technical but also to capture explicitly the organizational and social products stemming

from this proposed work. (…)

"This project's long-term approach emphasizes an interactive user-centered information infrastructure as part of an ocean informatics environment where participatory design intermingles with training and education to create new approaches to information systems design. The OIE established within this project represents a test bed for design practices and data flow applicable in many data rich environments faced with multiple divides and perspectives."[11]

A distinction is made between a technical service that can be described, standardized, and made routine through standard operating procedures and that of the investigative aspects of work that require investigation in order to define the questions, the design activities, etc.:

[Information management] is a technical service when we tell her [our information manager] what to do. And she does what we want. But we are saying, if it's a collaboration or a science interaction then we influence how she does things and she influences how we do things. So it's a two-way thing. (S)

Also, information managers need to know about science practices, for instance to know when and how to accommodate anomalies, to understand how every dataset is different. As this participant explains, information managers have to understand and follow rigorous methods and protocols while also being aware of, and attentive to, variations and anomalies:

And scientific experience is seen as a key in data work. Experience is the key thing, not only in science, in medicine. For example you cannot make an automatic decision about your patients like machines. Why not? Because every patient is different.. (…) this is very important, two things: to follow some kind of methodology, a standard methodology, and also to know about the different variations (…) because you can have anomalies, and if you don't incorporate those anomalies into your knowledge (…), you never know about this new kind of variation (…) so, it's very important to be open minded, be flexible in these kind of things. (S)

The size of data collections and the amount of data work associated with new field instrumentation introduces new requirements for processing, analysis, organization, description, integration, delivery, and visualization. This creates new types of work in the scientific arena that may be delegated to students, technicians, and/or information managers:

Its a brand new system and so we got the data back from that cruise and I was, and I had to process it, and what I wound up doing was I 'binder-ized' it in a way that. Let me, I'll just show you. Before I got here, all the data files were just in some random directory and there was no documentation. And what did I do with that (…) I had a very extensive data set description and then I did plots of all the profiles. So that a PI can go back and look at printouts and kind of get a general feeling for what was going on at that time. (…) As

[11] 'Ocean Informatics Learning Environment: Data Stewardship and Data Integration in Practice', 2007 NSF OCI Data Interoperability Networks unfunded proposal (see footnote 6).

> *I've gotten better at my job, I've taken the view of final end product for the PIs. That's what they're really looking for. They hate it; they don't really like processing data. 'I don't like processing the data!' They want to just get their hands on the data and publish it. (IM)*

The delegation of data processes that are complex to information managers is described by a graduate student aware of local modeling efforts:

> *One of our grad students is using a model (…) and he found some big error in it. It's a relatively new model so it hasn't stabilized. But so many of these science projects are based on more and more complicated code and modeling (...). What that means is that because they are so complex, they're almost certain to have problems (…). And that burden is put on to the data processors or the information managers. (GS)*

Rather than seeing the OI Initiative simply as a service to scientific research, that is as a means allowing scientific work to be done, information managers tend to see data practices as an 'end-product' per se. Thus it's not merely about working with data to enable scientific work, it's mostly about working to design and develop better data practices, and so doing, contributing to the advancement of the theory and practice in informatics.

> *I think that's a big part of informatics is, you know, trying to enable, from both a technical and social point, the interoperability of data. And then I think there is also a research element to the ocean informatics project which looks at, for both the technical and social aspects of interoperability, why does it work. Why doesn't it work? You know when interoperability breaks down, why is it that it breaks down. And what, you know, what do you need to keep it from breaking down in the future (…). And then also kind of a research element of saying, well what is interoperability and when do we know we have achieved it? And what can we do to make it easier in the future? [The research aspect is about] answering I guess the kind of higher level questions, the less immediate questions. (IM)*

In many ways, OI is a 'field experiment', i.e. an attempt to design while being responsive to scientific support needs, using professional management and programming methods to generate solutions and products in a nimble manner, while integrating – in an experimental manner – additional knowledge with emergent, practice-based knowledge about information management. Though business and computer science models exist, Ocean Informatics doesn't develop according to a pre-existing model that has been already validated:

> *Another thing is, you know, a lot of times we sit down and we make design decisions and we kind of say, we come to a point where we say we're not entirely, we're not a hundred percent sure this is the way we want to implement it, but we're just going to go forward with it. We're going to try it out; we're going to see what happens. And if we have to fix it later, we do. Whereas my impression, talking to friends of mine who do similar work in corporate fields, it's like there's much more lengthy planning periods ahead of time. And there are many more white papers written up and plans compared and I think there's, I think things tend to move a little faster through the planning stage here. (IM)*

4 Ocean Informatics: an Information Environment in Formation

The Ocean Informatics Initiative emerged in 2002, yet it is still evolving and transforming itself. From our ethnographic account of participants' views on the Initiative, we summarized the variety of perspectives as four key characteristics of Ocean Informatics: an information infrastructure (from a sociotechnical standpoint), a set of data practices (from a data-centric perspective), an informal organization (from a human and institutional standpoint), and a conceptual approach to infrastructural design (from a design perspective). These characteristics are closely intertwined in practice; for instance, the conceptual approach to design directly influences the information infrastructure that is developed, and the informal structure allows for emergence of collaborative practices in data work.

> The Ocean Informatics Initiative forms an '**information environment**' in the making. The concept of an 'environment' was first incorporated in a presentation of the Ocean Informatics concept (Baker, Jackson, and Wanetick, 2005). The phrase 'information environment' was highlighted in a 2005 report by one of the Initiative's founders (Baker, 2005):
>
> *"The Ocean Informatics team (…) 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."*

In this current report, the notion of an 'information environment' is used to refer to Ocean Informatics' work to develop a setting that provides conditions and services (e.g. information systems, tools, approaches, strategies, etc.) which enable science participants to find, manage, and exchange data and data products efficiently. In addition, it creates a venue for the continuing, interdependent processes of learning and investigation into information management. The term 'environment' is used to refer to the shared physical and conceptual space of the Initiative – where datasets are managed, information systems are developed and applications are used and evaluated. The term 'information' is used to refer to the raw material at the center of the Initiative (inclusive of data, information, knowledge and the relations among them), while alluding to information systems, information sciences, and informatics, all objects and domains of Ocean Informatics.

For the four characteristics articulated by initiative participants in Section 3 as a starting point, we consider related aspects evident from our observational study. In defining and setting up an information environment, the OI Initiative appears to be a **community innovation**, i.e. a users-led innovation that leads to new **roles for information managers** and that builds upon a **networked approach to collaborating and scaling** while being deeply **rooted in local contexts**. Drawing on the literatures of Infrastructure Studies and Innovation Studies (see Section 2), we elaborate on each of these points in the following subsections.

4.1 Enabling Community Innovation and Development

The information environment that the OI Initiative is gradually setting in place is the product of an "innovation community" – where innovations are user-led and result from a collective process (Von Hippel, 2005). The founders of the Initiative can be seen as 'lead-users' (Von Hippel, 1986), in the sense that they anticipated some needs occurring within their work arena that would become needed more widely and over the long-term, i.e. information management needs for contemporary oceanographic science efforts. Examples of such needs include arrangements and systems to facilitate aggregation and integration of highly heterogeneous datasets. Having said that, the community that the Initiative has shaped – in the form of an information environment – allows for the development, testing and sharing of tools and applications, design approaches and strategies as well as collaborative understandings that bridge distributed efforts and are relevant to local-network configurations in particular.

The OI information environment is innovative in many ways. It distinguishes itself by developing local approaches to infrastructure design and development as well as its own way of organizing itself. For instance, a 'continuing design' approach to infrastructure is experimented with as an adaptation of a participatory design approach with a focus on the long term (see: Karasti et al, 2010). Work processes are highly collaborative and take place in a 'design studio' that provides a physical and conceptual space for much needed discussions about the elements involved: the types of data submitted and their associated data contexts including both the characteristics and categories of data, sampling, methods, and analysis. Vocabulary is developed and semantics clarified around the design table while category-building and perspective-broadening occur. The Design Studio makes room for data-centric discussions that contribute to decisions about how the data will be represented online at the local-level, whether authoritative sources need to be referenced and how standards are interpreted. Local insight may contribute to either or both particularizing and generalizing observations. Strongly user-oriented, the OI approach has many similarities to the 'culture of professional end-user developers' observed in the sciences (Segal, 2009), where scientists have a long tradition of developing technological tools (instruments, software, and so forth) on their own to further their particular needs. Products and processes developed within the OI information environment are local, situated and closely tied to users' needs, understandings, interests, and anxieties. OI participants may build from existent models and tools while modifying and developing to a scale appropriate for use at a local-level and/or as members of a network for use at the community-level.

Though an 'innovation community', the OI information environment can also be seen as an 'community innovation' (Van Oost et al, 2009) in the sense that it is not only about a community of people enabling innovation, but also a community that emerged and takes shape within a context where specific technical and organizational innovations are developed. As such, the information environment doesn't preclude the continuing design approach or the design studio (to name only a few), but rather emerges simultaneously as design approaches and work practices are developed and experimentally tested. Both innovation and the community emerge simultaneously within a process of co-construction.

4.2 Expanding Information Management Roles

The OI information environment is a vivid example of current transformations affecting the way in which information management is done in scientific settings. As we mentioned early on, information management strongly depends upon scientific project funding as it is often included indirectly in individual project proposals (for instance as a specification for equipment, at-sea data collection salary, and/or a programmer's salary for data analysis). Further, data and information managers have extremely diverse levels of expertise, and they are often charged with a vast array of tasks and responsibilities (from data processing to data visualization to data delivery). The specificity of their work remains mostly taken for granted – even by the scientists with whom they work.[12] Their understanding of the role and tasks at hand develops in practice as does their ability to identify issues, articulate approaches and negotiate plans.

Often in the absence of formal professional training, information managers have learned data management on the job, with skills shaped in practice by the routine crossing of many disciplinary fields in their daily work (dealing with technology, management, information sciences, computer science, oceanographic science, biology, statistics, data analysis, and so on). As a result, it is hard for them to identify with one research field or another much less with one professional association or another. Thus the work of information management is often ill defined and little understood within the realm of scientific research. As both data-taking capabilities and data-integration expectations explode, responsibilities are spilling into roles sometimes designated as information manager, computer scientist, and technologist or programmer into an array of titles old and new such as information scientist (Garfield, 1987), project scientist, data steward, information specialist, and data scientist (Swan and Brown, 2008).

The information environment provided by the OI Initiative aims precisely at filling in the gaps of some of information management professionalization issues. Though hired with the training to fulfill present needs, the concepts, tools, and techniques required over time by an information manager change rapidly and must be acquired. There is an increasing number of information schools within universities and data management threads within library schools; they tend to focus on larger-scale information infrastructures or institution-wide data centers, respectively. More recently, there are roles emerging relating to data/information management and curation (e.g. Swan and Brown, 2008; Walters and Skinner, 2011; Williams and Jaguscewski,

[12] The theme of 'invisible work' has been recognized and elaborated upon in fields requiring close collaboration across multiple professional fields with differing roles (e.g. Star and Strauss, 1999; Suchman, 1995; Star, 1991; Blomberg et al., 1996). Status is influenced by an institution's organizational focus. Medicine deals daily with the interface of doctors, nurses, and physical therapists treating patients and researchers developing new techniques. In education, one finds specialists trained to handle special needs children interfacing with classroom teachers and curriculum developers. Minority studies have made visible the changeable perceptions of the roles of women and of foreigners (Suchman and Jordan, 1989).

forthcoming). Information management when performed locally, appears to incorporate informatics. Baker (2005) defines informatics as an applied approach to information management using a suite of information science and social science knowledge in conjunction with a specific scientific field (e.g. bioinformatics, ecoinformatics or ocean informatics). Providing training for information management recalls earlier times and alternative professions where apprenticeship and mentoring occur. Beyond this local scale training, information manager training is not readily available on the topics and scales required.

From its inception, the OI Initiative put a strong emphasis on the need to set in place the conditions for two functional capabilities: providing information management services as well as establishing a learning environment. That is, learning by information managers would be encouraged and enabled professionally as part of (and perhaps as required by) the scientific process in contrast to acquired in an ad hoc, nonprofessional or amateur manner. Note, this notion emerged within a university setting. Many universities can be seen today as organizations with dual purposes, that of educating students within well-bounded intervals ranging from 4 to 10 years and also of carrying out scholarly research through a well-developed system involving students and professors. Such a system is frequently known but not explicitly recognized for the ongoing education of professors as well as students through their duties with teaching, dialogue with students, and supervision of graduate students. The university, however, does not provide in kind for the continuing education of information specialists who today require more than a three-session university training class in Microsoft Access or a night class in web page design taught by a business consultant. The LTER Network with its arrangement of 26 sites with 26 information managers who meet and communicate regularly, provided an early example of information managers learning from each other as part of a Community of Practice (CoP; see section 3.2.2). This recognition stimulated the proactive development of Ocean Informatics community traditions such as expression through writing of notes, memos, posters, and papers (Figure 3; Appendices 4, 6 - 10), and event flyers (Appendix 11) as well as technical reports, newsletter articles and papers (Appendix 3). Communication efforts also took the form of dialogue (design sessions, reading groups (Appendix 5) and leadership (local, institutional, and network-wide initiatives, events, working groups, and committees).

Nesting within the LTER network where information management has always been a recognized component of this research program, the OI Initiative is strongly influenced by the LTER information management trajectory while also being a direct contributor to its evolution. Figure 1 (Baker and Millerand, 2010; Figure 1) portrays an evolution of the information managers' roles and organizational arrangements in four stages. The progression begins with an individual technician or analyst having data responsibilities and working in close coordination with a researcher. When this 'Dyad Model' is successful, the data role becomes somewhat formalized in terms of working with more data, more public exposure, and perhaps expanding to include management of data for others in the laboratory or in a collaborative project. This Data Management Model morphs in response to a broadening of concern with local stakeholders and users of the data to include users of the data from off-site and larger communities. The Information Management Model includes more formalized management, organization, and delivery of the data. When data users include other communities and the public, the skills required to meet data needs are quite diversified – from server hardware and network security to information architecture and database design as well as data analysis and description. From this need comes the Informatics Team model that often includes not just a single system but multiple systems capable of handling highly structured data, highly complex data, as well as very large datasets. Figure 1 was developed as a conceptualization of the information management trajectory within the context of contemporary cyberinfrastructure developments. The OI Initiative has traveled through each of these stages with its 'Informatics Team Model' incorporating elements from each previous model and currently reconsidering the dynamics of the dyad arrangement nested within the team model.

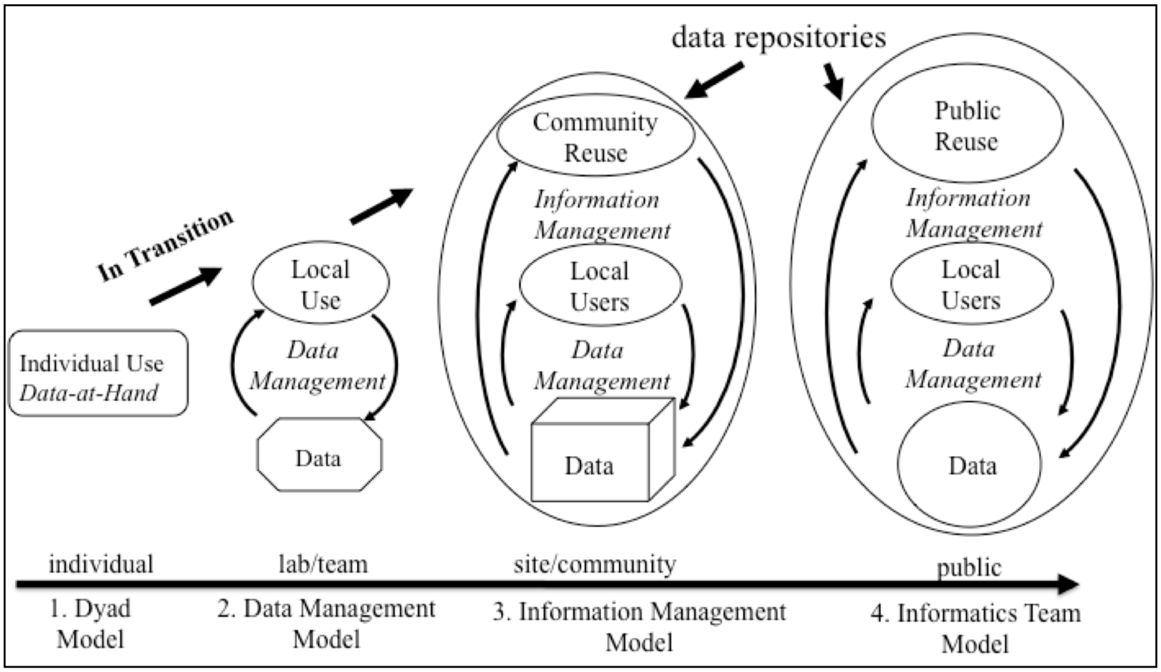

Figure 1. Evolution of information managers' roles and organizational arrangements. Local data and information management development is represented by four stages of organizational arrangements and roles. From: Baker and Millerand (2010).

Elements stimulating change are illustrated in Figure 2 that portrays the context within which the role of information management is developing. Over the years, the OI vision of information management broadened from that of serving science through data work to include managing of data over multiple scales and design involving new technologies and collaborative efforts (see: Baker and Karasti, 2004, Figure 3; Karasti and Baker, 2004, Figure 2). Figure 2 illustrates how the OI vision and practice have evolved since the period 1975-1991 when Baker worked within a dyad arrangement. In 1990 with the start of the Palmer LTER, data work expanded to include a number of scientific groups that comprised the Palmer components (birds, zooplankton, phytoplankton, bio-optics together with hydrography. With the advent of the internet, the accessibility of data transformed the role into one of information management involving digital data organization, semantics and metadata as well as data delivery and curation. In 2002, an exposure to Infrastructure Studies considered with a number of requests for a queriable information system at the PAL annual meeting. The two-faceted prompt enabled and shaped the Palmer information management vision of an Informatics Team Model. Expanded collaborative efforts with CCE and CalCOFI as well as with Science Studies participants have provided both resources and conceptual support for the team concept.

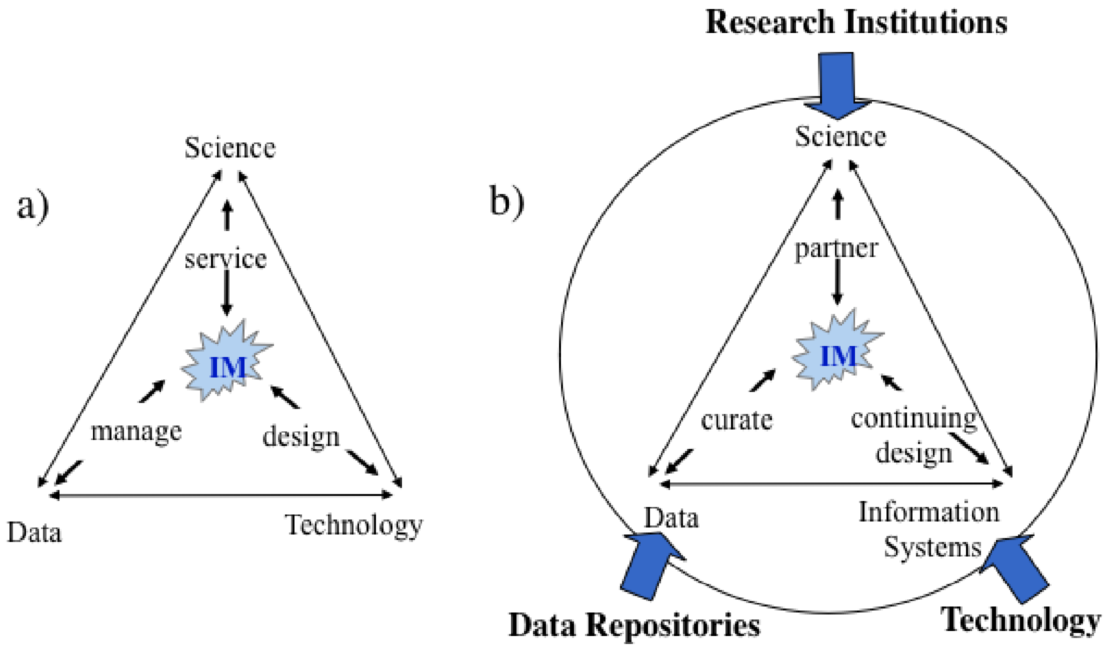

Figure 2. (a) Elements of information management (after Fig. 1, Karasti and Baker 2004; Karasti and Siikamaki, SAON 2007), and (b) elements of information infrastructure including research institution infrastructure arrangements, data repository configurations and technology as drivers of science, data, and information systems.

For the study of Ocean Informatics, we have built upon the earlier conceptualization to capture changes in internal relations as well as to include identification of what are often called larger-scale drivers. Figure 2b thus updates the 'service' label for the science-information management relationship to that of 'partner'. This shift indicates a realization of the need to recognize the needs of information management, the 'manage' label to include the repository-centric work of curation, and 'design' to include the long-term proactive aspects of technology use by updating the label to 'continuing design' (Karasti et al, 2010) to underscore the support of innovation and continuity. Note the 'technology' element is renamed as 'information systems' and 'technology' appears as a driver in Figure 2b. Other outside drivers include the various organizational infrastructure arrangements at research institutions that influence the science and information management alike as well as the duties of maintaining a data repository and interacting with other data repositories. As our understanding of the data/information work and its context develops so does our ability to discern new types of work and workflows as well as the roles associated with any number of work arrangements.

4.3 Building from and Scaling through Networked Collaboration

Ocean Informatics represents a local-scale effort where the local level is recognized as one element of a large-scale effort. The need for developing local-scale infrastructure - and a vocabulary for describing it - is less well recognized than large-scale efforts today. Here we specify 'local-scale' in particular, avoiding the term 'small-scale' as it brings to mind something at the other end of a continuum from 'large-scale'.

Though anchored locally, its involvement as the information management component for the geographically distributed components of the PAL LTER and as an active member of geographically distributed LTER research network, contribute to OI fitting loosely within the definition of a 'collaboratory' :

> *« A collaboratory is an organizational entity that spans distance, supports rich and recurring human interaction oriented to a common research area, and fosters contact between researchers who are both known and unknown to each other, and provides access to data sources, artifacts, and tools required to accomplish research tasks. » (Bos et al. 2007)*

Though developed to describe more traditionally conceived large-scale scientific undertakings, the definition of collaboratory provides an opportunity to consider various models of collaborative arrangements. The OI Initiative brings to the forefront a collaborative network model that supports coordination among quite a variety of networks and communities. Following a distributed rather than centralized approach, the OI information environment acts as a node for multiple entities within which it participates, e.g. entities include projects (PAL, CCE, CalCOFI-SIO, CalCOFI-SWFSC), ecological research communities (LTER, CaLCOFI), oceanographic networks (OOS, PACOOS), ecological networks (EcoInformatics), international programs (iLTER, IOOS), and so on. As such, OI emerges as a network within networks. Within the Bos et al. typology of collaboratories[13], the OI configuration represents a blend of three categories (Virtual Distributed Center, Virtual Learning Community, and Information management Community of Practice) but with the explicit addition of a local-scale level of collaboration.

The network-base of OI is recognized as providing an important blending of global context:

> *The balances we have to strike in bringing data together ...in packaging data, there are always constraints involved ... I'm hoping we are sidestepping some of those issues by defining what we do as local informatics. So Ocean Informatics ... its focus is on local scientific inquiry ... then the combining is done through the network. This keeps the local able to still focus in, draw in and elicit the particulars of the local site. ... the challenge then is to keep the balance between local and global, instead of ... a universal answer. (IM)*

Organized as a networked collaboration, OI evolves as a loose network that contributes to the framing of approaches, practices, and vision, without binding them. The focus is on enabling

[13] Bos et al. typology of collaboratories includes: Distributed Research Centers, Shared Instruments, Community Data Systems, Open Community Contribution Systems, Virtual Communities of Practice, Virtual Learning Communities, and Community Infrastructure Projects (Bos et al, 2007).

participants to learn and to respond to needed services with nimbleness and innovation. Largely inspired by the LTER model of network of networks, it recognizes the specificities of each of the independent entities to which it is tied, adopting from them what's useful, always looking for similarities that influence deign but respecting differences by resisting the urge to provide a unified 'one-size-fit-all' type of framework.

4.4 Taking Root Locally

Collaboration doesn't happen only at a distance with outside institutions but also locally among laboratories, departments, and divisions at a single location. Local or situated aspects of scientific work are rarely acknowledged within the large-scale collaboratory lens that prevails in contemporary debates on scientific challenges and cyberinfrastructure developments (see: Atkins, 2003; NSFCC, 2007). If recognized, they are not able to be accommodated. National efforts may be divided into regional center efforts but local efforts operate at a finer level of granularity. Such local information environments such as the one the OI Initiative has shaped provide critical services essential to daily scientists' work, innovative inputs and reality checks to community-level efforts, and a space to conduct information management.

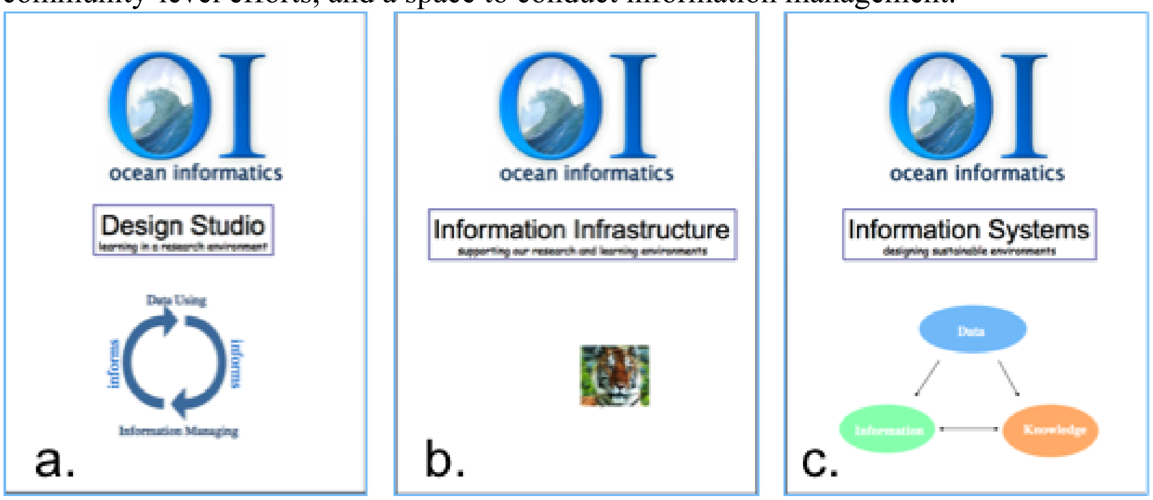

Figure 3. Ocean Informatics Office and Design Studio (SVH 2252, 2262) Posters showing: (a) the dynamic nature of information management as it co-evolves with the use of data, (b) the development of information infrastructure to support research and learning and (c) the need for information systems design to be done with a sensitivity to the complex relationships between data, knowledge, and information.

OI information environment is deeply rooted locally. It is a situated enterprise that emerged within an integrative oceanography division located within an oceanographic institution that is itself within a university. Anchoring its localness and its approach is the physical space in which it has developed over the years. Starting as an office space for programmers, 'Baker's Laboratory' was transformed into a Design Studio in two phases, first with the identification and use of a design table dedicated to collaborative design dialogue and second due to the opportunity created by a required move to a new room. A small budget from the director's office for refurbishing and furnishing the room allowed development of plans for a design studio that included a design table and a large wall monitor, individual work stations with associated side tables and a stool for brief consultations, and a number of mechanisms for hanging artifacts including an extensive magnet white board, a bulletin board, and a poster display space (Donovan and Baker, 2011; Figure 3).

Residing at or close to the data source and to the data originators represents a key element of OI situatedness. Proximity does facilitate the acquisition of data, but far more important, it also allows for the development of relations with field participants and data contributors. In contrast to remote data centers (e.g. national centers) where "t*here's a big disconnect between the center and the actual people that collected the data"* (in the words of one information manager), a local information environment such as OI recognizes the importance of the close ties between the people who manage the data and the scientists who collect it, as they work together, sometimes shoulder to shoulder. Being sensitive to 'the local' in terms of science needs, data management, and technological arrangements and constraints is then a priority in the OI vision. It is a sine qua non condition to be able to provide local representations of data needed by the scientists.

We suggest that the recent focus on data sharing in the scientific community, makes visible how data sampling, analysis, and description are in a difficult-to-transfer state that lacks an agreed upon vocabulary, suite of practices and sets of standards on which to build communication. This is particularly evident in the biological sciences given the extreme complexity of studying living entities. As reported by von Hippel (1998), a large portion of problem-solving associated with product customization involving application-specific tasks is carried out by users who themselves acquire "an inventory of knowledge regarding solutions that 'work' for the problem type at issue" (p. 629). This contradicts traditional notions of economies of specialization and economies-of-scale, suggesting the importance of taking into account 'complexities-of-scale' (Baker and Yarmey, 2009) as well as what von Hippel refers to as 'sticky information' where the measure of 'stickiness' is related to the effort required to transfer information to another site in a suitable form. von Hippel summarizes the findings of Szulanski (1996) regarding the most common barriers to information transfer: the absorptive capacity of the information receiver, inadequate encoding of information, and a distance between source and recipient that lacks alignment. Of course, stickiness can change, increasing based upon new arrangements and new skills with 'unsticking' information. On the one hand, the growing need to address institutional information management and multi-organization coordination within administrative, business as well as scientific realms seems to foreground a view of standardization as a quick solution to a pressing problem. Local designers, on the other hand, may have a broader scope in their approach to solving local problems so may take a more custom approach to design, tailoring a solution to a local issue that also takes into account the larger context of the domain and national standards.

The development of the role of information management from a catch-all set of tasks performed for production of a targeted bit of knowledge or product (i.e. a paper to be written by an individual scientist or a graphic to posted on a project web page) has opened up further as the work of information management is performed at multiple spatial, temporal, and organizational scales. Information management is still concerned with individual scientist needs but the work also involves interfaces with larger communities, centers, and networks. A participant in this role still responds to immediate data needs but is also dealing with new aspects relating to data repositories, networks, and change.

Making time for development of the role of information management through learning and leadership activities takes a concerted, well-planned and internally encouraged effort on a continuing basis. Despite feeding back immediately into an improved quality of information management services, the priority for such efforts is typically much less than that of meeting short-team scientific research project needs. The importance of expanding and diversifying the types of learning beyond the traditional options available to staff members is largely unrecognized within the university system. A centrally run, frequently outsourced training class or two does not meet the need for technology-savvy, information wise information managers. The result is visible in terms of the innovative individuals associated with scientific research projects rather than part of the centralized university services center who are effectively without a career trajectory so leave the university for expanded opportunities.

An aim of OI is to create an informatics set of services layered upon a foundational strata called a learning environment. OI has the special opportunity of being a workplace within a university and within the LTER network as well as a workplace where the purpose – as defined by research scientists – is to create information management support for data and information infrastructure to meet networking needs. A synergistic purpose comes from the OI focus on active participation in community and thereby in learning within various communities of practice. This is a purposeful redirection from traditional perceptions of advanced learning as taking place exclusively in graduate schools and with a highly structured organizational teaching element. This type of learning involves a rethinking of educational activities and the operationalism of learning in the form of situated learning via community participation (Lave, 1991; Brown and Duguid, 1991).

The educational focus is augmented also by comparative study at all levels and in any arena. For instance, application modules are purposefully designed with differing approaches and code libraries so that products and performance can be compared. Such methods augment existing educational arrangements by acknowledging and enabling learning-by-doing with guidance-upon-need available in the Ocean Informatics information environment where there is intentional guidance so that multiple perspectives and aspects of local and partnered information management are addressed. The aim is the development of social, organizational, and technical skills and knowledge to be established and then checked while carrying out locally situated tasks and activities.

5 The Importance of Local-Scale Contributions

In a time when the focus is on data aggregation, data curation, and large-scale infrastructure projects, Ocean Informatics focuses on *local data knowledge and organizational infrastructure as well as data repositories* and *the concept of a web-of-data repositories*. Thus OI contributes to contemporary developments in the realm of data work by providing an example that illustrates the potential of: (1) a local-scale information management design perspective of a multi-project effort in a site-network configuration, and (2) a new type of cyberinfrastructure initiative that represents a situated effort of an innovation community. Our investigation is an effort blending ethnography, co-learning, and action science. We report on the case of a keystone relationship existing with information managers embedded in the science and data work at the field-level, situated at ground-zero as part of the local information ecology (see Section 2.3.1) and forming an innovation community (see Section 2.3.2). Ethnography coupled with action science is demonstrated as adding to information management efforts both conceptually and in practice.

As a local community building effort, OI supports and contributes to ongoing field research programs involving a variety of heterogeneous data and data types. Ocean Informatics is unique in existing within a long-term organizational setting while being associated with several long-term research programs. Rather than being created as an individual, institutionalized center, it has grown as a program resource that holds potential as an institutional resource during a time when institutions are searching for ways of defining and developing information strategies useful with today's explosion of data. With a multi-faceted involvement in research – environmental and informatics – and an interest in new ways of collaborating and communicating, OI represents an information environment 'in formation'. Its development serves as an **exemplar** of one approach to organizing local information infrastructure in coordination with larger network-scale infrastructures.

Scientific information management efforts face a number of well-known dilemmas with data, that is, whether to preserve data close to its origin and its original form with its unique features or to bring some degree of standardization to the content as well as the format and documentation. This brings us to an appreciation of one of the strengths of OI's approach focusing on a 'web of data repositories' (Baker and Yarmey, 2009). Larger centralized data efforts aggregate data nationally and require a larger degree of standardization in order to enable data discovery, data delivery, and data integration. With increased scale and standardization, flexibility typically decreases in terms of a) design of information systems and b) handling of data anomalies. In the former case, data submission specifications become more stringent with a process more technologically driven than user-centered or participatory. In the latter case, non-standard data may not be submitted and if they are, work-arounds used to package anomalous data may mask 'the difference' they represent. Knowledge-making in science, however, moves forward both through macro analysis of large standardized datasets and through local analyses able to handle data anomalies. This type of understanding is what motivates the notion of 'a web of data repositories' where local level data repositories cater to local needs including data anomalies and network needs for standardized data while downstream one finds data curation efforts focusing more on standardization and large-scale delivery. In this way, a dilemma is addressed by cooperation. What appeared as a conflict of reality versus the ideal becomes instead a link-up of dichotomies into a multi-level web of repositories.

Further, in growing beyond the concept of individual technical systems, there is an emergent professionalization associated with informatics in general and information environments in particular. The hardware and information systems of Ocean Informatics are highly visible. Researchers, administrators and students see equipment made available, hardware and software being fixed, and data being delivered online. But the components of a contemporary information environment include a vast array of interdependent information elements (e.g. databases and metadata), data services (e.g. data organization and query interfaces), data comparability (e.g. dictionary development and standards use), and design aspects (e.g. continuing design and network configurations). Such a range of activities engenders the need for a wide ranging expertise. The traditional and familiar roles of technician and research assistant have blended in the past into that of data manager but now open up to include expertise in informatics, information architecture, data analysis, data synthesis as well as mediation and translation activities when managing highly interdisciplinary data in a scientific arena.

The OI Information environment defines a new type of informatics work arena for new types of collaborative arrangements among social scientists, domain scientists, and information managers. Taking seriously the definition of informatics as an applied approach taking place with the union of social sciences, information sciences, and ocean sciences, Ocean Informatics has been proactive in including social science participants as full, contributing partners in the evolution of the community and the information environment. Activating the participatory aspect of the environment, all OI members are challenged to find ways of defining their contributions within the work trajectory. A purposefully flattened workplace introduces its own challenges – need for increased communication and coordination while optimizing engagement, innovation, and learning.

OI represents a new-age network collaborator. Its 'networked collaboration' approach together with its long-term focus represent a new way of thinking and doing information management in situated, local environments. Its participatory, design-centric, learning-enhanced focus is an innovative, generative force at the local level. However, it undergoes continual change due to (re)balancing between a bounded set of project requirements as well as between short-term service-oriented efforts and long-term research-oriented undertaking. While exploring and developing the concept of an information environment, Ocean Informatics also works to establish an identity, an organizational placement. Differing views of Ocean informatics as an organization exist among information managers, researchers, and community participants. As a result, there is a constant ongoing pooling of insight, choice of techniques, assessing of design strategies, and balancing of constraints and benefits that largely goes unrecognized and yet is central to creating a viable information environment within an organizational setting.

Ocean Informatics may be seen as an exemplar of an approach for growing information infrastructure that provides a coordinated access to information management resources and a growing information expertise. Ocean Informatics represents an exploration in alternative ways of developing expertise, sharing resources, and managing information in a context where data sharing, integration, and synthesis involve high levels of complexity.

6 Acknowledgement

We offer special thanks to the Ocean Informatics community, its associates, and the graduate students who contributed (Steve Jackson, David Ribes, Brian Lindseth, and Claude Arsenault). This work is made possible by support from NSF (#0433369 Interoperability Strategies for Scientific Cyberinfrastructure: A Comparative Study) as well as by associated grants (PAL LTER OPP-02-17282, 08-23101, CCE LTER OCE-04-17616 & 10-26607, NOAA & CA OPC CalCOFI support) and Social Sciences and Humanities Research Council of Canada (SSHRC) (#756-2003-0099). Project and institutional support for cross-site visits (at San Diego and Montreal) facilitated collaboration given the longitudinal nature of the study.

7 References

Agre, P.E. (1997). Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. G.C. Bowker, S.L.Star, W.Turner, and L.Gasser (Eds.), Lawrence Erlbaum Associates, London.

Alverson, K. (2007). Why the world needs a global ocean observing system. *Marine Scientist* 21: 25-28.

Atkins, D. E. (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure*. Report of the NSF blue-ribbon Advisory Panel on Cyberinfrastructure: National Science Foundation. http://www.communitytechnology.org/nsf_ci_report.

Baker, K.S. (2005) *Informatics and the Environmental Sciences*. Scripps Institution of Oceanography (SIO) Technical Report Series, University of California San Diego. http://escholarship.org/uc/item/0179n650#page-1

Baker, K.S. (1998). Palmer LTER Information Management. In *Data and Information Management in the Ecological Sciences: a resource guide*. W. Michener, J. Porter and S.Stafford (Eds.), University of New Mexico, Albuquerque, New Mexico.

Baker, K.S., and C.L.Chandler (2008). Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics. *Deep Sea Research II* 55(18-19): 2132-2142. Special Series Palmer Long-Term Ecological Research. http://dx.doi.org/10.1016/j.dsr2.2008.05.009 Baker and Karasti, 2004

Baker, K. S., B. J. Benson, D. L. Henshaw, D. Blodgett, J. H. Porter and S. G. Stafford (2000). Evolution of a multisite network information system: the LTER information management paradigm. *BioScience*, 50(11): 963-978.

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, Hawaii

Baker, K.S., and H.Karasti (2004). *The Long-Term Information Management Trajectory: Working to Support Data, Science and Technology*. Scripps Institution of Oceanography (SIO) Technical Report Series, University of California San Diego. http://escholarship.org/uc/item/7d64x0bd

Baker, K.S., and F. Millerand (2010). Infrastructuring Ecology: New Challenges for Data Sharing. In Collaboration in the New Life Sciences. J. Parker, N. Vermeulen and B. Penders (Eds.) (pp. 111-138), Ashgate.

Baker, K.S., D.Ribes, F.Millerand, G.C.Bowker (2005). Interoperability Strategies for Scientific Cyberinfrastructure: Research Practice. *Proceedings American Society for Information Systems and Technology,* Charlotte, North Carolina, October 28-November 2.

Baker, K. S., and L. Yarmey (2009). Data Stewardship: Environmental Data Curation and a Web-of-Repositories. *International Journal of Digital Curation, 4*(2).

Baker, K.S., and J.Wanetick, 2010. SIO Ocean Informatics Update: Growing Infrastructure in Support of Scientific Research. *LTER Information Management Newsletter DataBits*, Spring 2010.

Berg, M. (1999). Patient Care Information Systems and Health Care Work: A Sociotechnical Approach. *International Journal of Medical Informatics* 55: 87-101.

Bijker, W. E., and Law, J. (1992). *Shaping Technology/Building Society. Studies in Sociotechnical Change*. Cambridge, MA: MIT Press.

Blomberg, J., L.Suchman, and R.H.Trigg, 1996. Reflections on a Work-Oriented Design Project. *Human-Computer Interaction* 11: 237-265.

Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., et al. (2007). From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication* 12(2).

Bowker, G. C., Baker, K. S., Millerand, F., & Ribes, D. (2010). Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In *International Handbook of Internet Research*. J. Hunsinger, J. M. Allen, and L. Klastrup (Eds.), Springer.

Brown, J. and P.Duguid (1991). Organizational Learning and Communities of Practice: Toward a unified view of working, learning, and innovation. *Organizational Science* 2(1): 40-57.

Callahan, J. T. (1984). Long-term ecological research. *BioScience* 34: 363-367.

Carusi, A., and M. Jirotka (2007). From data archive to ethical labyrinth. *3rd International Conference on e-Social Science*, Ann Arbor, Michigan, 7-9 October 2007.

Carusi, A. (2006). Data as representation and its ethical implications. *European Computing and Philosophy Conference*, Trondheim, Norway, 22-24 June 2006.

Donovan, J.M., and K.S.Baker, 2011. *The Shape of Information Management: Fostering Collaboration across Data, Science, and Technology in a Design Studio*. Scripps Institution of Oceanography (SIO) Technical Report Series, University of California San Diego.

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, (2007). Marine pelagic ecosystems: The West Antarctic Peninsula. *Philosophical Transactions of the Royal Society of London, Special Theme Issue, "Antarctic Ecology: From Genes to Ecosystems."*, *362*: 67-94.

Edwards, P. N., Jackson, S. J., Bowker, G. C., and C. P. Knobel (2007). *Understanding Infrastructure: Dynamics, Tensions, and Design*. Office of Cyberinfrastructure.

Edwards, P. N., Jackson, S. J., Bowker, G. C., & Williams, R. (2009). Introduction: An Agenda for Infrastructure Studies. *Journal of the Association for Information Systems* 10(5): 364-374.

Feng, P. (2000). Rethinking Technology, Revitalizing Ethics: Overcoming Barriers to Ethical Design. *Science and Engineering Ethics* 6:207-220.

Garfield, E. (1987) The Role of Information Scientists. *The Scientist* 1(22):9.

Geertz, C. (1973). Thick Description: Toward an Interpretive Theory of Culture. In *The Interpretation of Cultures: Selected Essays*. C. Geertz (Ed.), New York: Basic Books, p. 3-30

Glaser, B. G. (1978). *Theoretical sensitivity : advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Glasser, B.G., and A.L. Strauss (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago, Aldine Publishing Company.

Goguen, J.A. (1997). Toward a Social, Ethical Theory of Information. In *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. G.C. Bowker, S.L.Star, W.Turner, and L.Gasser (Eds.), Lawrence Erlbaum Associates, London.

Higgins, S. (2009). DCC DIFFUSE Standards Frameworks: Standards Path through the Curation Lifecycle. *International Journal of Digital Curation* 2(4).

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* 53:21-32.

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.

Karasti, H., K. S. Baker, and F. Millerand (2010). Infrastructure Time: Long-term Matters in Collaborative Development. *Journal of Computer Supported Collaborative Work* 19(3/4): 377-415.

Karasti, H., and K.S. Baker (2004). Infrastructuring for the Long-Term: Ecological Information Management. *Paper presented at the Hawaii International Conference on System Sciences* (HICSS'37), January 5-8 2004, Hawaii, USA.

Kling, R. (1996). *Computerization and controversy. Value conflict and social choices (2nd edition).* San Diego, CA: Academic Press.

Kling, R., and W. Scacchi (1982). The Web of Computing: Computer Technology as Social Organization. *Advances in Computer* 21: 1-89.

Lau, F. (1999). Toward a framework for action research in information systems studies. *Information Technology & People* 12(2):148-175.

Lave J. (1991). Situating learning in communities of practice. In *Perspectives on Socially Shared Cognition*. L. Resnick, J. Levine and S. Teasley (Eds.), Hyattsville, MD: American Psychological Association, p. 63-84.

Millerand, F., and K.S. Baker (2010). Who are the users? Who are the developers? Webs of users and developers in the development process of a technical standard. *Information Systems Journal* 20: 137-161.

Millerand, F., and G. C. Bowker (2009). Metadata Standards. Trajectories and Enactment in the Life of an Ontology. In *Standards and Their Stories*. S. L. Star and M. Lampland (Eds.), Cornell University Press.

Millerand, F., and G. C. Bowker (2008). Metadata, trajectoires et « énaction ». In *La cognition au prisme des sciences sociales.* C. Rosental (Ed.). Paris: Editions des Archives Contemporaines.

Mukerji, C. (1989). *A Fragile Power: Scientists and the State*. Princeton.

Nardi, B., and V. O'Day (1999). *Information Ecology: Using Technology with Heart.* Cambridge. Cambridge: MIT Press.

NOAA/NESDIS (2008). *Comprehensive Large Array-data Stewardship System (CLASS), Information Heterogeneity White Paper.* November 2008. U.S.Department of Commerce. NOAA/NESDIS CLASS-1215-CLS-WHT-IHET

NSFCC (2007). *National Science Foundation Cyberinfrastructure Council: Cyberinfrastructure Vision for 21st Century Discovery.* Washington, DC: National Science Foundation.

NSB (2005). *National Science Board: Long-Lived Data Collections: Enabling Research and Education in the 21^{st} Century*. NSF NSB-05-40.

NRC (2003). *National Research Council: Enabling Ocean Research in the 21st Century: Implementation of a Network of Ocean Observatories*. Committee on the Implementation of a Seafloor Observatory Network for Oceanographic Research, Washington, D.C., National Academy Press, 224 pp.

Ocean ITI Working Group (2004). *Trends in Information Technology Infrastructure in the Ocean Sciences*, 24 pp., www.geo-prose.com/oceans_iti_trends

Ocean Observations (2003). Special Issue Oceanography 16(4).

Ohman, M. D., and E. L. Venrick (2003). CalCOFI in a changing ocean. *Oceanography* 16: 76-85.

Powell, T.M., and J.H. Steele (1995) *Ecological Time Series*. Chapman and Hall, New York.

Raitt, H., and B.B.Moulton (1966). *Scripps Institution of Oceanography; first fifty years.* Ward Ritchie Press, http://scilib.ucsd.edu/sio/hist/caljsioa_gc29r2.pdf

Ribes, D., K. S., Baker, F. Millerand, and G. C. Bowker (2005). Comparative Interoperability Project: Configurations of Community, Technology, and Organization. *Paper presented at the Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*.

Ribes, D., and C. Lee (2010). Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories. *Computer Supported Cooperative Work - An International Journal* 19: 231-244.

Rogers, Y. (1997). Reconfiguring the Social Scientist: Shifting from Telling Designers What to Do to Getting More Involved. In *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. G.C. Bowker, S.L.Star, W.Turner, and L.Gasser (Eds.), Lawrence Erlbaum Associates, London.

Schuler, D., and A. Namiok (1993). *Participatory Design: Principles and Practices*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Segal, J. (2009). Software Development Cultures and Cooperation Problems: A Field Study of the Early Stages of Development of Software for a Scientific Community. *Computer Supported Cooperative Work* (CSCW) 18:581-606

Shor, E.N. (1978). *Scripps Institution of Oceanography: Probing the Oceans, 1936-1978*. San Diego, CA: Tofua Press.

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.

Star, L. (1991). The sociology of the invisible: the primacy of work in the writings of Anselm Strauss. In *Social organization and social process: essays in honor of Anselm Strauss*. D. R. Maines (Ed.). Aldine De Gruyter, New York, p. 265-284

Star, S. L., and A. L. Strauss (1999). Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *JCSCW, 8*(1/2) : 9-30.

Star, S. L., and K. Ruhleder (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, 7(1): 111-134.

Steele, J.H. (1989). The Ocean 'Landscape'. *Landscape Ecology,* 3(3/4): 185-192

Stocks, K.I., and K.S. Baker (2005). *CalCOFI Data Management White Paper 1*. http://calcofi.ucsd.edu/datamanagement2005.pdf

Stommel, H. (1963). Varieties of Oceanographic Experience: The ocean can be investigated as a hydrodynamical phenomenon as well as explored geographically. *Science* 139 (3555): 572-576.

Strauss, A., and J. Corbin (1998). *Basics of qualitative research : techniques and procedures for developing grounded theory (2nd ed.).* Thousand Oaks: Sage Publications.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge, MA: Cambridge University Press.

Suchman, L. A. (2002). Located accountabilities in technology production. *Scandinavian Journal of Information Systems* 14(2): 91-105.

Suchman, L. A. (1995). Making Work Visible. *Communications of the ACM: Special Issue on*

Representations of Work 38(9) : 56-64.
Suchman, L. A., and Jordan, B. (1989) Computerization and Women's Knowledge. In *Women, Work and Computerization*. K. Tijdens, M. Jennings, I. Wagner and M. Weggelaar (Eds). Amsterdam: North Holland, p. 153-160.
Swan, A., and S. Brown, 2008. *The Skills, Role and Career Structure of Data Scientists and Curators : An Assessment of Current Practice and Future Needs*. Report to the JISC. http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx
Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, 17(Winter Special Issue): 27-43.
Van Oost, E. S.Verhaegh, and N.Oudshoom (2009). From Innovation Community to Community Innovation: User-Initiated Innovation in Wireless Leiden. *Science, Technology & Human Values* 34: 182-2005.
Von Hippel, E. (1976). The dominant role of users in the scientific instrument innovation process. *Research Policy*, 5(3): 212-239. Doi: 210.1016/0048-7333(1076)90028-90027.
Von Hippel, E. (1986). Lead Users: A Source of Novel Product Concepts. *Management Science* 32(7) : 791-805.
Von Hippel, E. (2005). *Democratizing Innovation*. MIT Press.
Walters, T., and K. Skinner (2011). *New Roles for New Times : Digital Curation for Preservation*.
http://www.arl.org/rtl/plan/nrnt/
Wanetick, J., and D.Larsen (2010). *Data Zoo Design and Implementation of an online California Coastal Physical Oceanography Database* (http://zoo.ucsd.edu). (last checked July 2010).
Wenger, E. (2006). *Communities of practice: a brief introduction*. Web site: http://www.ewenger.com/theory/index.htm.
Williams, K. and J. Jaguscewski (forthcoming). *New Roles for New Times: Transforming Liaison Roles.* http://www.arl.org/rtl/plan/nrnt/nrntliaison.shtml
Winner, L. (1985). Do Artifacts Have Politics? In *The Social Shaping of Technology*. D. MacKenzie and J. Wajcman (Eds.). Philadelphia, PA: Open University Press.
Winner, L. (1986). *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. University of Chicago, Chicago.