

# What can Studies of e-Learning Teach us about Collaboration in e-Research? Some Findings from Digital Library Studies

CHRISTINE L. BORGMAN

*Department of Information Studies, University of California, 235 GSE&IS Bldg, Box 951520, Los Angeles, CA, 90095, USA  
(E-mail: borgman@gseis.ucla.edu)*

**Abstract.** e-Research is intended to facilitate collaboration through distributed access to content, tools, and services. Lessons about collaboration are extracted from the findings of two large, long-term digital library research projects. Both the Alexandria Digital Earth Prototype Project (ADEPT) and the Center for Embedded Networked Sensing (CENS) project on data management leverage scientific research data for use in teaching. Two forms of collaboration were studied: (1) direct, in which faculty work together on research projects; and (2) indirect or serial, in which faculty use or contribute content to a common pool, such as teaching resources, concepts and relationships, or research data. Five aspects of collaboration in e-Research are discussed: (1) disciplinary factors, (2) incentives to adopt e-Learning and e-Research technologies, (3) user roles, (4) information sharing, and (5) technical requirements. Collaboration varied by research domain in both projects, and appears partly to be a function of the degree of instrumentation in data collection. Faculty members were more interested in tools to manage their own research data than in tools to facilitate teaching. They also were more reflective about their research than teaching activities. The availability of more content, tools, and services to incorporate primary data in teaching was only a minimal incentive to use these resources. Large investments in a knowledge base of scientific concepts and relationships for teaching did not result in re-use by other faculty during the course of the project. Metadata requirements for research and for teaching vary greatly, which further complicates the transfer of resources across applications. Personal digital libraries offer a middle ground between private control and public release of content, which is a promising direction for the design of digital libraries that will facilitate collaboration in e-Research.

**Key words:** collaboration, digital libraries, e-Research, e-Science, e-Learning, human-computer interaction, information seeking, information retrieval

## 1. Introduction

Cyberinfrastructure, as it is known in the U.S., or e-Research, as it is known in the U.K. and Europe, promises to facilitate scholarly collaboration by providing access to shared data and document repositories, tools, and services. Considerable progress is being made on building the technical framework, on establishing standards for interoperability, and on the construction of digital libraries to store scholarly content. However, relatively little

research has been done to determine how and whether these technologies will facilitate collaboration or enable access to new forms of knowledge. Evidence from prior social studies of science suggests that the adoption of information technologies such as e-Research is a complex and not always successful process (Kubicek and Dutton, 1997; Kline and Pinch, 1999; MacKenzie and Wajcman, 1999; Woolgar, 2003; Wouters, 2004; Foster and Gibbons, 2005).

One of the main drivers of e-Research is the “data deluge” (Hey and Trefethen, 2003). The volume of scientific data being generated by highly instrumented research projects (linear accelerators, sensor networks, satellites, seismographs, etc.) is so great that it can only be captured and managed using information technology. The amount of data produced far exceeds the capabilities of manual techniques for data management, and thus the need for control of these data is another essential driver of e-Research (Lord and Macdonald, 2003). Once these data are captured and curated, they can be shared over distributed networks. If these same data can be made available for other applications such as e-Learning, many opportunities arise for economic and political leverage of the investments in e-Research.

This article draws upon findings of two long-term digital library projects in the U.S. on the intersection between e-Research and e-Learning. In the Alexandria Digital Earth Prototype Project (ADEPT) (1999–2005), we studied the use of research-based digital libraries of primary scientific data for teaching at the undergraduate level. Within the Center for Embedded Networked Sensing (CENS), our project (2002–2008) has dual goals of developing an infrastructure for the management of research data by scientific teams and making these data useful for teaching at the middle school and high school levels. Our research in ADEPT and CENS confirms the need for studies of collaborative work in e-Research and e-Learning and for iterative design and evaluation of the technology. In this article I extract lessons about collaboration from our findings and reflect upon their implications for the design of e-Research infrastructure.

## **2. Users and uses of scientific data**

Sharing data is a core element of scientific collaboration. It is a complex social process involving trust, incentives, disincentives, risks, and intellectual property (Bishop et al., 2003; David, 2003b, 2004; David and Spence, 2003; Arzberger et al., 2004; Bowker, 2005). Data sharing between scientists is a complex and little-studied area (Borgman, in press; Hilgartner and Brandt-Rauf, 1994; Bowker, 2005). Data sharing between scientists, teachers, and students is even less well studied. CoVis is among the few projects to consider how students and teachers use scientific data, and that effort preceded present concerns about e-Science and cyberinfrastructure (Pea, 1993; Gomez et al., 1998).

Scientists who collaborate with each other tend to have similar disciplinary knowledge and analytical skills. Such similarities cannot be assumed when the same scientific data are shared with teachers and students (Enyedy, 2003). To serve these two communities with one set of resources, two potential conflicts must be addressed. One is that scientists and students collect and analyze data for different purposes. Scientists' primary goal is the production of knowledge, while students' primary goal is to learn the concepts and tools of science. For students, "doing science" is a means to learn new content and skills. In the ideal case, students also will generate data that contributes to knowledge in their classroom and to scientific knowledge.

The second conflict is that scientists, teachers, and students bring far different skill sets and epistemologies of science to the use of scientific data. As part of their graduate study and research training, scientists have learned practices to select, collect, organize, analyze, store, and disseminate data. Scientific practices reflect a tacit understanding about the nature of science, researchable questions, knowledge claims, and evidence necessary to support claims. By comparison, teachers and students at the middle and high school levels generally lack deep subject knowledge, research methods expertise, and knowledge of data management practices. Students in introductory university courses have only slightly more knowledge of scientific practices than do high school students. For scientific data from e-Research projects to be useful for learning, teachers and students need considerable assistance to bridge the gaps in knowledge and skills between them and the scientists who produced the data.

### **3. Background: ADEPT and CENS projects**

Some background on the goals of the ADEPT and CENS projects will help to explain the relationship of our findings to collaborative work.

#### **3.1. ALEXANDRIA DIGITAL EARTH PROTOTYPE PROJECT (ADEPT)**

The central goal of the ADEPT project (<http://www.is.gseis.ucla.edu/ADEPT/>) was to make geo-spatial information resources intended for research purposes usable for teaching and learning at the undergraduate level. The Alexandria Digital Library (ADL), constructed as part of the (U.S.) Digital Libraries Initiative Phase I (1994–1998), provides access to geo-spatial resources in many media via sophisticated searching mechanisms (Hill and Janeé, 2004). ADEPT is a set of services associated with the ADL intended to enable faculty (i.e., members of academic staff who teach and conduct research) to construct lectures and assignments using content from the ADL and other sources, enable teaching assistants in lab sessions to use the information resources assembled by the supervising faculty member, and enable students to explore the lecture resources and to perform interactive

assignments that utilize data, simulations, and other information resources assembled by their instructor. Thus ADEPT has three user communities (faculty, teaching assistants, students) and two purposes (research and teaching).

The research and development of ADEPT was conducted from 1999 to 2004; the education and evaluation component of the project continued to 2005. We studied the design, deployment, and adoption of prototype ADEPT learning environments at two university campuses. Research methods included classroom observations, interviews with faculty, students, teaching assistants, and developers; analysis of teaching materials (lectures, assignments, exams); and analysis of available metadata standards (Borgman et al., 2000, 2001, 2004a, b, 2005; Leazer et al., 2000; Mayer et al., 2002; Gazan et al., 2003; Champeny et al., 2004; Borgman, 2004; 2005a; D'Avolio et al., 2005).

Research on system design and architecture also is documented elsewhere (Hill et al., 1999, 2000; Ancona et al., 2002; Janee and Frew, 2002; Smith and Zheng, 2002; Hill and Freeston, 2003; Smith et al., 2003; Hill and Janee, 2004; Janee et al., 2004). Further work on the ADEPT software and on broader system deployment continues under a joint U.S. (National Science Foundation) and U.K. (JISC) funded project (*DialogPlus: Digital Libraries in Support of Innovative Approaches to Learning and Teaching in Geography*, 2003).

### 3.2. CENTER FOR EMBEDDED NETWORKED SENSING (CENS)

The Center for Embedded Networked Sensing (<http://www.cens.ucla.edu>) is a U.S. National Science Foundation Science and Technology Center based at UCLA that includes dozens of cooperating scientists, technologists, educators, and teachers (middle school and high school). CENS is a large, multidisciplinary research collaboration among multiple universities that involves sharing heterogeneous data collections, and thus fits the U.K. definition of e-Science (*e-Science Core Programme*, 2004). The Center was launched in August, 2002, with funding to 2007, and renewable to 2012. CENS investigators manage many additional grant projects through the Center.

CENS is developing embedded networked sensing systems and applying this technology to scientific applications. These are large-scale, distributed, systems composed of smart sensors and actuators embedded in the physical world. They monitor and collect information on such diverse subjects as plankton colonies, bird behavior, plant growth, contaminants in soil and water, and structural integrity of buildings, bridges, and other human-made structures. A central goal of embedded networked sensing systems is the ability to reveal previously unobservable phenomena. The researchers in CENS are investigating fundamental properties of these systems, developing new enabling technologies, and exploring novel scientific and educational

applications. Computer scientists, engineers, and scientists (e.g., biology, geology, seismology, environmental sciences, marine sciences) from multiple universities are collaborating to design and deploy these systems. As the Center has evolved, scholars in related fields have joined our projects, including participating faculty from statistics, law, architecture, design, and film.

The CENS education and data management teams have undertaken a broader scope of work than did the ADEPT education and evaluation team. Our scope in CENS includes design of infrastructure for managing scientific data, design of tools to make data useful for educational applications, and guidance on policy for sharing data. The content in CENS consists of real-time data generated by scientific research projects, rather than resources already collected into a digital library. The effort on this project was accelerated by an additional grant from the National Science Foundation specifically for data management and educational research (Sandoval and Borgman, 2004–2008).

Our early findings on data use are reported in several papers (Shankar, 2003; Borgman, 2004, 2005a, b; Borgman et al., 2006; Wallis et al., 2006). We have also presented findings in several unpublished talks at conferences and colloquia in the U.S. and Europe. Research on classroom deployments is reported elsewhere (Thadani et al., 2006).

CENS' sensor networks currently are deployed to study habitat biology, water quality, seismology, contaminant transport, marine microorganisms, and several other topics. The work reported here draws largely on data management and education studies of habitat biology research, complemented by observations and interviews across CENS' research areas. Habitat data generated by sensors at an ecological reserve in the mountains east of Los Angeles (*James San Jacinto Mountains Reserve*, 2004) can be monitored in real time or analyzed as datasets over selected time periods. For prototyping purposes, we deployed a similar set of habitat sensors near one of the participating schools. Scientists, teachers, and students (grades 7 through 12) have access to these data in real time and access to archives of previously generated data.

#### **4. Digital libraries and collaborative work**

Behavioral studies of digital libraries have focused largely on individual users, following the traditions of information-seeking research; only a few studies have examined collaborative aspects of these technologies (Ellis, 1989; Kuhlthau, 1991; Borgman, 2000a, b; 2003a; Case, 2002; Agre, 2003; Bishop et al., 2003; Lynch, 2003; Marchionini et al., 2003; Star, Bowker and Neumann, 2003; Van House, 2003; Van House et al., 2003). Our research questions in ADEPT focused initially on individual users and extended to

collaborative concerns as the project matured. The CENS research began about mid-way through the ADEPT project. Knowledge gained from ADEPT led to more emphasis on collaboration in the CENS research.

Lessons on collaboration in e-Research from ADEPT and CENS can be divided into five categories, which are used to organize the remainder of this article: (1) disciplinary factors, (2) incentives to adopt e-Learning and e-Research technologies, (3) user roles, (4) information sharing, and (5) technical requirements. Research supporting these findings is discussed in context rather than presented as a separate literature review.

#### 4.1. DISCIPLINARY FACTORS

It will come as no surprise to those schooled in collaborative work or social studies of science that the use of data and information varies by discipline or specialty (Meadows, 1998; Case, 2002). The degree of automation in data collection and analysis also is associated with the likelihood of sharing data (Pritchard et al., 2004). Thus research specialties that are more collaborative and make more use of instrumentation are more likely to use e-Research technologies.

##### 4.1.1. *ADEPT findings*

Due to the content of the introductory courses on geography that we studied, most of the research subjects in ADEPT were physical geographers; the rest were human or cultural geographers. In the one study that compared these two specialties, we found differences in degree of collaboration and in data sources. Physical geographers were more likely to participate in large, collaborative projects and their research was more data-driven. Several of them used data sets produced by agencies such as NCAR (*National Center for Atmospheric Research*). Human geographers in our sample tended to work on their own and to write sole-authored scholarly books; their research was more concept-driven than data-driven (Borgman et al., 2005). Fry (2003) also found that social/cultural geographers (comparable to the human geographers in our studies) tended to work individually rather than as part of collaborative teams.

Information resources used by geographers in both specialties were similar, however. All of the faculty we studied rely heavily on maps and spatial images, as would be expected for a field in which physical location and spatial orientation are organizing metaphors. When asked to define "primary data," faculty from all specialties responded (in varying ways) that primary data were anything raw and unprocessed, such as sensor data and field notes, while secondary data were analyzed or processed or interpreted in some way (Borgman et al., 2005).

#### 4.1.2. *CENS findings*

Our initial findings in CENS support the hypothesis that research specialties which are more collaborative and make more use of instrumentation are more likely to use e-Research technologies.

We spent the first year of the CENS project assessing the state of data management practices and available methods in each of the participating research groups (Shankar, 2003). We found a wide disparity across CENS in practices, tools, resources, and data archives. Habitat ecology and seismology were selected as initial topic areas due to their different histories of data management. Seismic data have been collected via automatic instruments since the early 1970s, and a common metadata standard has been in use since 1988 (*Standard for the Exchange of Earthquake Data (SEED)*, 2004), maintained by a global organization (*Federation of Digital Broad-Band Seismograph Networks (FDSN)*, 2004). The FDSN assigns network codes to provide uniqueness to seismological data streams. Many of these data are contributed to a community repository (*Incorporated Research Institutions for Seismology*, 2004), which has established policies and practices for the use of those data. Thus the seismology community has a long history of distributed, collaborative, and standardized management of its data.

Habitat ecologists, in contrast, tend to work alone or in small groups. These scientists have relied more on hand-sampling in the field than on highly instrumented data collection. Spreadsheets have been the preferred tool for data analysis, with data models specific to each project. These spreadsheets are used to produce graphs, charts, and tables for their research publications. Scientists who choose to participate in CENS presumably are those more amenable to collaborative work, as these projects involve partnerships with the computer scientists and engineers who design and deploy the sensor networks. As data collection becomes more automated and the volume of data increases, these scientists are likely to require more sophisticated tools for data management and analysis. We are exploring the degree to which their data practices change in parallel with the increasing automation of data collection. The CENS' habitat ecologists are aware of Knowledge Network for Biocomplexity (KNB), which is intended to serve as a community resource for data management in habitat ecology, but are not yet making much use of it for their own research. Software products and tools have been available from KNB since 2001 (*Ecological Metadata Language (EML)*, 2004; *Knowledge Network for Biocomplexity (KNB)*, 2004).

#### 4.2. INCENTIVES TO ADOPT E-LEARNING AND E-RESEARCH TECHNOLOGIES

Both the ADEPT and CENS projects aim to facilitate inquiry learning, which is a method of involving students in scientific practices so that they gain a

deeper epistemological understanding of science (Edelson et al., 1999; Sandoval and Reiser, 2003; Enyedy and Goldberg, 2004). While a laudable goal, scientists and teachers are very busy people. Few are interested in dabbling with technology for its own sake. Rather, they will adopt a technology only if it offers sufficient advantages to justify the investment in learning, in changing associated practices, and in costs of the technology itself (Rogers, 1995). Research is a collaborative activity in most scientific fields. Teaching is usually a solo activity. These differences appear to influence incentives to use e-Learning and e-Research technologies (Borgman, 2004).

#### 4.2.1. *ADEPT findings*

While the overall ADEPT project was based on the premise that faculty would incorporate more primary data in teaching undergraduate geography courses if they had better tools to mine digital libraries and to extract these data in forms useful for instruction, our team treated that premise as a research question. We assessed incentives and criteria for adoption of technology throughout the ADEPT research (D'Avolio et al., 2005).

Geography is a technology-intensive field, and we found participating ADEPT faculty to be sophisticated users of technology in their research. Although most expressed interest in experimenting with new instructional methods, few of them employed computer-based technology in their teaching. The most common reasons given were that too much advance planning was required for computer-based instruction and that too much assistance would be required to install and support the equipment (Borgman et al., 2000, 2001; Leazer et al., 2000).

Having anticipated these barriers, we built substantial resources into the ADEPT grant to lower the effort required to use the technology. Our hypothesis was that if we could provide sufficient assistance in course development to trade for the instructors' time in participating in the research, we could persuade faculty to teach with the ADEPT prototypes. Our investment of staff resources afforded the opportunity to conduct formative evaluation and to contribute iterative assessments to the design and implementation teams (Borgman et al., 2001; Gazan et al., 2003).

In the 2002–2003 academic year, we conducted an extensive assessment of a full prototype deployment of the ADEPT system in two sections of an introductory course in physical geography (Champeny et al., 2004). The course was taught twice by one instructor, once in the fall and once in the spring quarter; the system was refined during the interim winter term. The ADEPT “digital learning environment,” as implemented at that time, had three components that could be used in various combinations: the “object collection” of primary source modules (instructors could contribute their own or use those in the collection), the “lecture composer” to organize

lecture outlines with these objects embedded, and the “knowledge base” of concepts and relationships (referred to as an “ontology” in computer science). Ontologies are similar to thesauri, in that they establish terminology by consensus and map relationships between concepts, but more extensive in that they attempt to reflect the epistemology of a knowledge domain (Hovy, 2003; *OWL Web Ontology Language Guide: W3C Recommendation*, 2004; Ribes and Bowker, 2004; Bowker, 2005).

The instructor of this course used all three components, devoting most of his efforts to developing a knowledge base. He requested that approximately 1000 concepts be created for the 10-week course. At an estimated 1-h of labor per concept, the knowledge base for this one introductory course would have required the equivalent of 25 weeks of full-time effort (Champeny et al., 2004). As the course went on, the effort became unsustainable for the instructor and his graduate students and he did simplify his requirements.

When recruiting other faculty members to use ADEPT in their classrooms, we demonstrated the various options without offering judgments as to their usefulness. Two instructors implemented prototypes in 2003–2004 at a different university campus, and were presumed to be unfamiliar with the experience of the instructor discussed above. These two faculty members used the tools to construct lectures and embed primary data resources. Neither instructor chose to add concepts to the knowledge base nor to use concepts already developed. The ADEPT knowledge base was intended for “serial collaboration:” each instructor could use concepts contributed by prior instructors, add new concepts, and add or alter relationships in the database. Thus in the short time remaining in the project for other instructors to implement the system, no one else used the concepts and relationships laboriously constructed by one faculty member.

For the ADEPT design and development team, the massive investment in knowledge base construction was justified by the potential for re-use. We in the education and evaluation team were skeptical of the scale of this investment because classification by individuals can be very idiosyncratic. Knowledge organization mechanisms such as thesauri, ontologies, and taxonomies are most likely to succeed when they are constructed by communities that reach consensus on terminology and on relationships among concepts (Bowker and Star, 1999, 2001; Star et al., 2003; Bowker, 2005). Our team’s role was to evaluate each version of the system and to provide continuing input to the iterative design and development process. While the design and development team did not follow our advice in this instance, the deployment of a large knowledge base for an individual instructor provided a fruitful opportunity to study adoption and adaptation. This is not the first, nor likely the last, project to become mired in ontology construction (Goble and Wroe, 2005; Ribes and Bowker, 2006, in press).

#### 4.2.2. *CENS findings*

As expected, scientists' interest in data management appears to be growing as the amount of data from embedded sensor networks increases. The first 3 years of CENS were skewed heavily toward engineering work in sensor technology and networks. While scientists in application domains such as seismology, biology, habitat ecology, and environment have worked closely with the engineering teams since the beginning of the Center, only recently has the technology matured enough to produce a scientifically valid stream of data. The scope of the impending data deluge is now apparent, lending urgency to improving capture and management mechanisms. The methods and research questions in ecology were changed substantially by the introduction of remote sensing technology via satellites in the 1980s and 1990s (Kwa, 2005). Our continuing research in data management is assessing how these research areas will evolve with the introduction of embedded sensing networks, and whether the technology will help to shorten the time to discovery.

Middle and high school teachers are users, rather than producers, of scientific data and thus have different incentives to use e-Research technologies than do scientists. A goal of CENS is to teach with the same primary data used by participating scientists rather than to use "canned" datasets with pre-defined questions and answers, as is the case with most science learning projects. Few teacher education programs cover inquiry learning or the use of primary data sources in their core curriculum. Accordingly, these approaches rarely are employed in teaching at the middle and high school levels. Given the overhead of training teachers in these methods, we sought participation in CENS by innovative teachers who already are familiar with the approach.

Better pedagogy alone is insufficient incentive for teachers to participate. Content of science courses in California schools must be based on the California Science Standards. Teachers are required to teach to these standards and students are tested on them; no incentives exist to teach topics that are not included in these standards. Because the environment is a central topic in these standards, we had little difficulty selecting appropriate modules in habitat biology and water quality, however. The effectiveness of these modules is addressed by our research partners in education (Thadani et al., 2006).

#### 4.3. USER ROLES

An unexpected finding in the ADEPT project was differences in the use of information by role of researcher or teacher. We are making this comparison in CENS only indirectly, although other role comparisons appear to be interesting in this environment. Types of collaboration appear to vary by a person's role with respect to information resources.

#### 4.3.1. *ADEPT findings*

In initial stages of the project we treated geographers as one sample population. As differences emerged in their use of information when researching and when teaching, we realized that we had begun the project with the naïve assumption that sharing primary scientific data between research and teaching would be largely a matter of providing good tools, because the same people are both the researchers and the teachers. However, it was in the transfer of data between research and teaching that we encountered the greatest disconnect in the ADEPT project.

The researcher–teacher role differences were manifested in three ways. One was in their use of information technology. As noted above, even the most technologically sophisticated faculty members made little use of technology in their teaching. Their offices often contained two or more computer workstations, multiple hard drives, and other equipment such as scanners. Some of the faculty used supercomputers in their research. Yet most of them left these technologies behind in their offices, heading to class with chalk and overhead transparencies (Borgman et al., 2000, 2001; Leazer et al., 2000; Borgman, 2004). The second manifestation is in the use of data. These faculty members taught introductory courses from textbooks rather than from primary sources. Any research data used to illustrate their lectures usually came from their own research. These data usually were presented in synthesized forms (maps, images, tables) rather than as raw data for students to mine in course assignments (Leazer et al., 2000; Borgman et al., 2001, 2005; Gilliland-Swetland and Leazer, 2001; Gazan et al., 2003).

The third comparison, which we examined most closely, was between information-seeking for research and for teaching (Borgman et al., 2004a, b, 2005). Geographers sought information in support of their research in expected ways: they track the new literature in their fields, browse familiar sections of the library, bookmark favorite web sites, follow citation links, attend professional conferences, and receive sources and references from their scholarly peers. Searching for information in support of their teaching was more serendipitous. They might find useful items for teaching in the process of searching for research topics, and might also find research ideas or resources while gathering information for teaching.

#### 4.3.2. *CENS findings*

In our CENS studies of information and technology use, we have studied scientists only as researchers. Questions of how scientists use their research data in teaching undergraduate and graduate courses may be explored in later projects. In current research, we are asking questions about how sharing data for educational purposes differs from sharing data for scientific purposes. We also are comparing how scientists and technologists on the same team use data from embedded sensor networks.

Our education partners are studying middle and high school teachers' use of data in the context of classroom deployments of CENS' instructional modules. Compared to scientists, teachers have minimal skills in data management and far less access to information technology. Current CENS' instructional modules utilize only a small subset of the volume of data available to scientists, and include simple data analysis and visualization tools.

#### 4.4. INFORMATION SHARING

Collaboration is one of the factors that influences sharing of data and other information. Research collaborations frequently are based on sharing expensive instrumentation or resources (Olson and Olson, 2000; Finholt, 2002; Sonnenwald, 2006). Agreements about sharing data are central to establishing collaborations (David and Spence, 2003). Access to information and control of information are fundamental aspects of science. While "open science" drives the incentives to publish and to verify findings quickly through peer review (David, 2003a), many disincentives also exist to sharing information, especially research data. Disincentives fall into four categories: (1) rewards for publication rather than for data management; (2) the amount of effort required to document data for use by others; (3) concerns for priority, including rights to control results or sources until publication of research; and (4) intellectual property, both control and ownership of one's own information and access to information controlled or owned by others (Borgman, in press). Here I focus briefly on issues of priority and intellectual property that arose in the ADEPT and CENS projects, and how those factors influence collaboration in e-Research.

##### 4.4.1. *ADEPT findings*

We studied how faculty members select and use data and information for their research and for their teaching. We asked few specific questions about sharing information, but those issues arose throughout the project. As noted earlier, the premise of the ADEPT project was to add services to an extant digital library (the Alexandria DL) of geospatial resources to make it more useful for instructional purposes. The education and evaluation team treated this premise as a research question. Faculty were aware of the existence of ADL, but were making minimal use of it for either research or teaching purposes. It contains multiple collections of maps, satellite images, and other primary sources of geospatial data. The ADL is publicly available, currently containing about 15,000 items available for downloading. New features were added subsequent to the period of our research, including a Google Earth interface (<http://www.testbed.alexandria.ucsb.edu:8080/webclient/index.jsp>).

When offered the ADEPT system to construct classroom lectures, participating faculty preferred to create modules from their own resources rather than ADL resources, however. Although the ADEPT architecture initially was intended as an overlay onto ADL, the project was reframed to construct an independent digital library of instructional resources that could later be incorporated into the ADL. Faculty members wished to digitize materials they frequently use to illustrate concepts in teaching, including images from textbooks, journal articles, newspapers, and other copyrighted sources. When displayed in a classroom to a few dozen or a few hundred students, instructors commonly assume that use falls under the “fair use” educational provisions of U.S. copyright law (educational provisions of copyright laws vary widely by country; these studies focused on use within the U.S.). When the same resources are stored on a website in digital form, even if access is restricted to enrolled students, narrower fair use guidelines associated with digital resources usually apply (The Digital Millennium Copyright Act, 1998). If instructors wish to make those same resources publicly available via a digital library, their use definitely is subject to copyright permission.

Faculty members interviewed for ADEPT often felt that the effort required to manage the use of copyrighted resources in an e-Learning environment constrained their choice of teaching resources, compared to what they normally used with more traditional teaching methods (e.g., chalkboards, handouts, overhead displays, PowerPoint slides, maps, and objects such as rock samples). These constraints were a disincentive to using the ADEPT technology in teaching and to contributing resources to the collection. The availability of geospatial content in ADL was not sufficiently attractive to overcome their preferences for their own materials.

The education and evaluation team strongly encouraged the builders of the ADEPT teaching object collection to include metadata tags for copyright ownership on each object as it was created, but the builders were inconsistent in doing so. If ownership of items in the object collections were clearly identified, it would be possible to distribute the ADEPT software with the public objects and restrict access to copyrighted materials. The lack of clear distinction hampered the dissemination of the ADEPT collections.

Geography faculty members’ choices of data to use in teaching were influenced by intellectual property considerations, by availability, by familiarity, and by cost (Borgman et al., 2005). They draw upon their own data for a variety of reasons: these resources are familiar, available, usually do not require copyright permissions from others, and use does not incur additional monetary costs. Physical geographers often write grants to purchase datasets for their research. They may be able to use these same data in their teaching (or at least their own analyses of those data), but are unlikely to get instructional funds to purchase data for teaching alone.

#### 4.4.2. *CENS findings*

Issues of data sharing and ownership are explicit research questions in CENS. We are studying how scientists, teachers, and students determine their data requirements, their criteria for selecting and preserving data, their use of scientific data and how that use evolves over time, and incentives and disincentives to contribute data to repositories. Our research methods include attending workgroup meetings of scientific teams and analyzing their work products (datasets, websites, publications), interviewing individual faculty and research groups, visiting research sites, and identifying appropriate data repositories, metadata standards, and structures.

Data of interest to habitat ecology from embedded sensor networks fall into two general categories. The first category consists of continual measurements of the physical environment such as meteorological variables (e.g., temperature, humidity, rainfall, barometric pressure). These serve as baseline data that scientists can use in conjunction with their own observations. The physical monitoring data are considered non-proprietary and are made public immediately. The James Reserve, for example, posts data on local conditions every 15 min (<http://www.jamesreserve.edu/weather.html>). They also have webcams on multiple observation points (e.g., bird feeders, nest-boxes, moss growing) that Internet visitors can control remotely (<http://www.jamesreserve.edu/WebcamFiles/RoboCam.html>). The second category is data from investigators' research projects. Individual investigators are using small local arrays to study plants, animals (especially bird behavior), and soil and water conditions. Multiple investigators are obtaining ecological data from the James Reserve and elsewhere via three-dimensional robotic sensor data collection from the NIMS technology (Sutton, 2003; *NIMS: Networked Infomechanical Systems*, 2005; Batalin et al., 2004). Investigators control the data from their research projects. Some investigators may release data immediately. Others may not release data until they are cleaned, analyzed, and published (Borgman et al., 2006).

Seismology, with its long history of common standards, community repositories, and collaboration, has addressed access policies more explicitly. Data contributed to IRIS (*Incorporated Research Institutions for Seismology*, 2004) are embargoed from public use for 2 years from the date of experiment before being released, although requests can be made to individual investigators for data to be shared sooner (<http://www.iris.edu/services/faq.htm#iris>). The CENS seismology research team contributes its data to IRIS in the standard SEED format (*Standard for the Exchange of Earthquake Data (SEED)*, 2004). However, their embedded networked sensors are generating data at a higher rate (500 samples/s) than IRIS currently accepts, raising questions about how data management practices will keep pace with advances in data collection methods.

#### 4.5. TECHNICAL REQUIREMENTS

Differences in use based on research specialty and on role, varying incentives to use technologies, and varying practices in selecting and sharing information all have implications for the technical requirements of e-Research. In both ADEPT and CENS, the implications cluster into database and metadata issues. I discuss briefly the requirements identified in each of these projects, then return to implications for collaboration in the discussion and conclusions.

##### 4.5.1. *ADEPT findings*

In the initial design stages of ADEPT, we considered two extremes of architecture, both of which were rejected. A single shared collection was rejected due to the variance in roles and incentives. Independent collections for each faculty member were rejected lest the potential for re-use be undermined. A middle ground was agreed upon (Janeé and Frew, 2002) and its efficacy confirmed by findings of the education and evaluation team (Borgman et al., 2004a; Champeny et al., 2004; Borgman et al., 2005). The balance was achieved in ADEPT architecture by enabling each instructor to gather his or her own resources into a “personal digital library.” Faculty members can choose to share, or not to share, their personal digital libraries with others and to make items visible or not visible in the shared collection. This approach resolves some problems of intellectual property rights, and enables faculty members to use their research data for teaching without necessarily contributing them to the common pool (Borgman, 2003b).

Metadata is the second architectural issue arising from differences in the use of geographic resources for teaching and research. The Alexandria Digital Library, on which ADEPT is based, includes sophisticated metadata and a gazetteer. These capabilities were extended greatly in the ADEPT project by the teams at University of California, Santa Barbara (Janeé et al., 2004), providing extensive access by location (place name, latitude and longitude) and some access by concept. However, these access mechanisms proved insufficient for teaching purposes. Faculty, in their roles as teachers, asked for data, images, or simulations that would demonstrate concepts such as erosion or adiabatic processes, regardless of the physical location on Earth. Complicating matters further, an instructor might use an image or dataset to illustrate multiple points in one or more lectures (Borgman et al., 2005). Anticipating all the ways that a document or object might be described is one of the classic problems of knowledge organization, and it will not be solved by any single system (Svenonius, 2000).

##### 4.5.2. *CENS findings*

Similar divisions between public and private data are emerging in our CENS research but the origins are different. In ADEPT, we asked questions about

their use of public resources in the Alexandria Digital Library and about their use of data from their own research projects. In CENS, we are asking more generally about all their sources of data for research purposes. Thus far, we are seeing a distinction between publicly available baseline data on physical conditions and data collected by the investigator for specific projects. Data controlled by investigators may or may not be released later; this is the subject of research currently under way.

The personal digital library model should be useful in CENS, but “personal” libraries often will be controlled by collaborative teams rather than by individuals. Teams could manage their own data and set criteria for which data, at what time, will be released to the community repository. This approach also may allow access to scientific data for educational applications sooner, especially if the data are provided at a higher degree of granularity than is required for research publications.

The mismatch of metadata structures between research and teaching applications is even more striking in CENS. Metadata models for the habitat biology community (e.g., *Ecological Metadata Language (EML)*, 2004) describe the data (e.g., time, date, sensor location), while educational metadata models (e.g., *ADEPT/DLESE/NASA metadata (ADN)*, 2001; *IEEE Learning Object Metadata (LOM)*, 2004) describe the educational activity (e.g., grade, level, resources required for the activity, time to perform the activity, educational standards, etc.). We found no overlap in data elements between these scientific and educational metadata formats. The formats cannot be reconciled because they serve different purposes: metadata in scientific applications describe the data, while metadata in educational applications describe the pedagogy. A major thrust of our current work is to explore ways to bridge this gap via layered models, filtering tools, or other methods (Wallis et al., 2006).

## 5. Discussion and conclusions

e-Research is intended to facilitate collaboration through distributed access to content, tools, and services. *How* e-Research will facilitate collaboration is an open question. Neither of the two projects discussed here – the Alexandria Digital Earth Prototype Project and the Center for Embedded Networked Sensing research on data management – are specifically about collaborative work. Rather, each is about distributed access to scientific content, which is a key component of e-Research. In studying how faculty and teachers could use research resources for instruction, we delved into how those resources are produced and shared. We found it necessary to understand more about the conditions under which faculty would share content with each other as a prerequisite to understanding how they would share research content for instructional purposes.

Two types of collaboration arose in these projects. One is the more traditional model, in which faculty work together on research projects or to build new tools and services. The second is indirect or serial collaboration, in which faculty contribute or use contributed content. In the latter case, faculty members collaborate indirectly by building upon prior work of others whom they may or may not know. The shared content may be teaching modules, concepts and relationships (ADEPT), or research data (CENS).

While collaboration varied by research domain in both the ADEPT and CENS projects, the contrasts are more distinct in CENS. All of the people studied in ADEPT were geographers. Physical geography research is more data-driven than is research in human or cultural geography. Faculty in both areas implemented ADEPT prototypes in their classrooms but the samples were too small to draw conclusions based on specialty.

In CENS, faculty in seismology, which is the most data-driven and instrument-rich specialty studied, have the most standardized practices for data capture, preservation, and sharing. Researchers in habitat biology, which is the most hand-crafted, field-research oriented specialty studied, have the least standardized practices for data management. This comparison echoes the findings of Pritchard et al., (2004) that data sharing is associated more with the degree of automation of data collection and analysis than with research specialty, *per se*. Results of a pilot study also suggest that willingness to share is a function of the effort involved in gathering data (Borgman et al., 2006).

While the technology juggernaut of e-Science and cyberinfrastructure suggest that better tools will result in more use and re-use of scientific data, our results indicate that the relationship between tools, use, and re-use is very complex. Making data, teaching resources, or other forms of content easier to share does not mean that scientists *will* share. Geographers welcomed, in principle, better tools for teaching that would facilitate the use of primary content in their field. However, they made minimal use of the available digital library (the ADL) for teaching, preferring instead to transfer extant teaching materials and personal research data into the ADEPT software. In doing so, they created a new set of problems associated with the use of materials for which they did not control copyright.

Having begun with the assumption that ADEPT would add services to an already useful resource, the Alexandria Digital Library, we did not explore the value of ADL in sufficient depth. I can offer two explanations for why the ADL was not as popular as assumed. One is that it competed for attention among many other geospatial resources that were readily available online. In asking our subjects which resources they used for research and teaching, the ADL was rarely mentioned (Borgman et al., 2005). The other reason is the cumbersome user interface, which required considerable knowledge of the library collections. The ADL interface has improved since the time of our studies. The lesson learned here is that

people will use an information resource collaboratively only if the resource itself has sufficient value.

The component of ADEPT that was intended to have the greatest and longest-term collaborative value, the knowledge base of physical geography concepts, was not re-used during the course of our study. This was also the most laboriously constructed component of the ADEPT system. No matter how sophisticated the set of concepts and relationships built by an individual, communities may not use them if they lack a sense of ownership.

The mismatch of metadata for research and for teaching is an issue in both projects. Researchers search for content by concepts and, in geography, by location. Teachers search for examples that will illustrate concepts and processes. They often are less concerned with place or with specifics of the concept than with presentation criteria such as the clarity of the image and how well it will display in a classroom. Research materials rarely are described in these ways. Conversely, educational objects are described with pedagogical concepts and lack adequate description of the scientific content. Metadata standards for research and for teaching are based on different epistemologies. They cannot be reconciled through the usual technique of "crosswalks" (Godby et al., 2004) because minimal intersections exist. Collaboration via re-use of content for research and teaching is hampered severely by these epistemological differences, which in turn become embedded in system architecture.

Some promising directions for facilitating collaboration through e-Research have emerged from the ADEPT and CENS research projects. Personal digital libraries offer a middle ground between private control and public release of content. Better tools and services to manage content can improve capture, management, and preservation. Making content more shareable is a first step to facilitate sharing. Similarly, the differences we identified between uses of content for research and for teaching will inform the design of tools and services to bridge the gap. Research currently in progress addresses interactions between collaboration in research, uses of instrumentation for data collection and analysis, and sharing of data. Many of the factors influencing collaboration in e-Research are subtle, nuanced, and powerful. Much more research is needed to identify individual factors and the relationships between them. We are just beginning to understand *how* e-Research can facilitate collaboration. More research remains to understand *why*.

### **Acknowledgements**

ADEPT and CENS are both large collaborative projects. The education and evaluation research in ADEPT described here was conducted in 1999

to 2005 in collaboration with Anne Gilliland and Gregory Leazer, both of the Department of Information Studies, Graduate School of Education and Information Studies, UCLA, and Richard E. Mayer, Department of Psychology, UC-Santa Barbara. ADEPT was funded by National Science Foundation grant IIS-9817432, Terence R. Smith, University of California, Santa Barbara, Principal Investigator. We have been fortunate to have the participation of many talented students, all of whom are listed as co-authors on the ADEPT papers cited herein. The data management and education research in CENS is conducted in collaboration with William A. Sandoval and Noel Enyedy, both of the Department of Education, and Jonathan Furner of the Department of Information Studies; all three are faculty of the Graduate School of Education and Information Studies, UCLA. Students and post-doctoral fellows who participated on the data management team during the period of research reported here (2002–2005) are Stasa Milojevic, Jillian Wallis, Kalpana Shankar, and Eun Park. CENS is funded by National Science Foundation Cooperative Agreement #CCR-0120778, Deborah L. Estrin, UCLA, Principal Investigator; Christine L. Borgman is a co-Principal Investigator. Further research on data management described here is funded by NSF award number ESI-0352572, William A. Sandoval, Principal Investigator; Christine L. Borgman, co-Principal Investigator. We have enjoyed the extensive contributions to our work made by members of the education team, including (during the period covered in this paper) Sara Terheggen, Karen Kim, Joe Wise, Kathy Griffis, Andy Wu, Kelli Millwood, and Jeff Bockert. The James San Jacinto Mountains Reserve is part of the University of California Natural Reserve System; its director, Dr. Michael Hamilton, is a Co-Principal Investigator of CENS. The CENS seismology team is led by Prof. Paul Davis, Department of Earth and Space Sciences, UCLA, who is also a co-PI on CENS.

## References

- ADEPT/DLESE/NASA metadata (ADN). (2001): Retrieved from <http://www.dlese.org/Metadata/adn-item/index.htm> on 15 December 2004.
- Agre, P.E. (2003): Information and Institutional Change: The Case of Digital Libraries. In A.P. Bishop, N. Van House and B.P. Battenfield (eds.): *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press, pp. 219–240.
- Ancona, D., M. Freeston, T.R. Smith and S. Fabrikant (2002): Visual Explorations for the Alexandria Digital Earth Prototype. *Joint Conference on Digital Libraries*, Portland, Oregon: Springer-Verlag.
- Arzberger, P., P. Schroeder, A. Beaulieu, G. Bowker, K. Casey, L. Laaksonen, D. Moorman, P. Uhlir and P. Wouters (2004): An International Framework to Promote Access to Data. *Science*, vol. 303, no. 5665, pp. 1777–1778.

- Batalin, M.A., M. Rahimi, Y. Yu, D. Liu, A. Kansal, G.S. Sukhatme, W.J. Kaiser, M. Hansen, G. J. Pottie, M. Srivastava and D. Estrin (2004): *Call and Response: Experiments in Sampling the Environment*, Los Angeles. Retrieved from [http://www.ee.ucla.edu/%7Eka-iser/nims/Documents/Publications/SenSys\\_2004\\_Call\\_and\\_Response.pdf](http://www.ee.ucla.edu/%7Eka-iser/nims/Documents/Publications/SenSys_2004_Call_and_Response.pdf) on 6 December 2005.
- Bishop, A.P., N. Van House and B.P. Buttenfield (eds.) (2003): *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press.
- Borgman, C.L. (2000a): Digital Libraries and the Continuum of Scholarly Communication. *Journal of Documentation*, vol. 56,, 412–430.
- Borgman, C.L. (2000b): *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, MA: The MIT Press.
- Borgman, C.L. (2003a): Designing Digital Libraries for Usability. In A.P. Bishop, N. Van HouseVan and B.P. Buttenfield (eds.): *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press, pp. 85–118.
- Borgman, C.L. (2003b): Personal Digital Libraries: Creating Individual Spaces for Innovation. *NSF Workshop on Post-Digital Libraries Initiative Directions*, Chatham, MA. Retrieved from [http://www.sis.pitt.edu/~dlwshop/paper\\_borgman.html](http://www.sis.pitt.edu/~dlwshop/paper_borgman.html). on 19 January 2005.
- Borgman, C.L. (2004): The Interaction of Community and Individual Practices in the Design of a Digital Library. *International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society*, University of Tsukuba, Tsukuba, Ibaraki, Japan. Retrieved from <http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp9.pdf> on 10 April 2006.
- Borgman, C.L. (2005a): Disciplinary Differences in e-Research: An information Perspective. Keynote Talk. *First International Conference on e-Social Science*, Manchester, UK. National Center for e-Social Science. Retrieved from [http://www.ncess.ac.uk/events/conference/programme/presentations/ncess2005\\_borgman.pdf](http://www.ncess.ac.uk/events/conference/programme/presentations/ncess2005_borgman.pdf) on 10 August 2005.
- Borgman, C.L. (2005b): The Social Embeddness of Embedded Networked Sensing. *MIT-OII Workshop on New Approaches to Research on the Social Implications of Emerging Technologies*, University of Oxford. Oxford Internet Institute. Retrieved from <http://www.oii.ox.ac.uk/collaboration/?rq=specialevents/20050415> on 11 April 2005.
- Borgman, C.L. (in press): *Scholarship in the Digital Age: Information, Infrastructure and the Internet*. Cambridge, MA: The MIT Press.
- Borgman, C.L., A.J. Gilliland-Swetland, G.H. Leazer, R. Mayer, D. Gwynn, R. Gazan and P. Mautone (2000): Evaluating Digital Libraries for Teaching and Learning in Undergraduate Education: a Case Study of the Alexandria Digital Earth Prototype (ADEPT). *Library Trends*, vol. 49, no. 2, pp. 228–250.
- Borgman, C.L., G.H. Leazer, A.J. Gilliland-Swetland and R. Gazan (2001): Iterative Design and Evaluation of a Geographic Digital Library for University Students: A Case Study of the Alexandria Digital Earth Prototype (ADEPT). *Research and Advanced Technology for Digital Libraries: 5th European Conference, ECDL 2001, Darmstadt, Germany, September 2001: Proceedings*, New York, NY: Springer, Lecture Notes in Computer Science.
- Borgman, C.L., G.H. Leazer, A.J. Gilliland-Swetland, K.A. Millwood, L. Champeny, J.R. Finley and L.J. Smart (2004a): How Geography Professors Select Materials for Classroom Lectures: Implications for the Design of Digital Libraries. *Joint Conference on Digital Libraries*, Tucson, AZ: Association for Computing Machinery.
- Borgman, C.L., L.J. Smart, K.A. Millwood, J.R. Finley, L. Champeny, A.J. Gilliland and G.H. Leazer (2005): Comparing Faculty Information Seeking in Teaching and Research: Implications for the Design of Digital Libraries. *Journal of the American Society for Information Science and Technology*, vol. 56, no. 6, pp. 636–657.

- Borgman, C.L., L.J. Smart, K.A. Millwood, J.R. Finley, L. Champeny, A.J. Gilliland-Swetland and G.H. Leazer (2004b): *Getting Ready for Class: The Information Seeking Behavior of Geography Professors in Support of Teaching. Annual Meeting of the American Educational Research Association*. San Diego, CA: American Educational Research Association.
- Borgman, C.L., J.C. Wallis and N. Enyedy (2006): *Building Digital Libraries for Scientific Data: An Exploratory Study of Data Practices in Habitat Ecology*. Alicante, Spain: European Conference on Digital Libraries.
- Bowker, G.C. (2005): *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.
- Bowker, G.C. and S.L. Star (1999): *Sorting Things Out: Classification and its Consequences*. Cambridge, MA: The MIT Press.
- Bowker, G.C. and S.L. Star (2001): Social Theoretical Issues in the Design of Collaboratories: Customized Software for Community Support Versus Large-scale Infrastructure. *Coordination Theory and Collaboration Technology*. Mahwah, NJ: Lawrence Erlbaum Associates Inc, pp. 713–738.
- Case, D.O. (2002): *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. San Diego: Academic Press.
- Champeny, L., C.L. Borgman, G.H. Leazer, A.G. Gilliland-Swetland, K.A. Millwood, L. D'Avolio, J.R. Finley, and L.J. Smart (2004) Developing a Digital Learning Environment: An Evaluation of Design and Implementation Processes. *Joint Conference on Digital Libraries*. Tucson, AZ.
- D'Avolio, L., C.L. Borgman, L. Champeny, G.H. Leazer, A.J. Gilliland-Swetland and K.A. Millwood (2005): From Prototype to Deployable System: Framing the Adoption of Digital Library Services. *Annual Meeting of the American Society for Information Science and Technology*, Charlotte, NC.
- David, P.A. (2003a): The Economic Logic of 'Open Science' and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer. *The Role of the Public Domain in Scientific Data and Information*. Washington, DC: National Academy Press. Retrieved from <http://www.siepr.stanford.edu/papers/pdf/02-30.html> on 22 May 2006.
- David, P.A. (2003b): The Economic Logic of "Open Science" and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A primer. In J.M. Esanu and P.F. Uhler (eds.): *The Role of Scientific and Technical Data and Information in the Public Domain: Proceedings of a Symposium*. Washington, DC: The National Academies Press, pp. 19–34. Retrieved from <http://www.books.nap.edu/catalog/10785.html> on 13 April 2005.
- David, P.A. (2004): *Towards a Cyberinfrastructure for Enhanced Scientific Collaboration: Providing its 'Soft' Foundations may be the Hardest Part*. Oxford Internet Institute Research Reports: University of Oxford. 23 pp. Retrieved from <http://www.oii.ox.ac.uk> on 15 January 2005.
- David, P.A. and M. Spence (2003): *Towards Institutional Infrastructures for E-Science: The Scope of the Challenge*. Oxford Internet Institute Research Reports: University of Oxford. 92 pp. Retrieved from <http://www.oii.ox.ac.uk> on 15 September 2003.
- DialogPlus: Digital Libraries in Support of Innovative Approaches to Learning and Teaching in Geography* (2003): National Science Foundation (U.S.) and Joint Information Systems Committee (U.K.). Retrieved from <http://www.dialogplus.org/> on 19 November 2004.
- e-Science Core Programme* (2004): Engineering and Physical Sciences Research Council (EPSRC). Retrieved from <http://www.epsrc.ac.uk/ResearchFunding/Programmes/e-Science/default.htm> on 18 November 2004.

- Ecological Metadata Language (EML)* (2004): Retrieved from <http://www.knb.ecoinformatics.org/software/eml/> on 25 November 2004.
- Edelson, D.C., D.N. Gordin and R.D. Pea (1999): Addressing the Challenges of Inquiry-based Learning Through Technology and Curriculum Design. *Journal of the Learning Sciences*, vol. 8, no. 3–4, pp. 391–450.
- Ellis, D. (1989): A Behavioural Approach to Information Retrieval Design. *Journal of Documentation*, vol. 45,, 171–212.
- Enyedy, N. (2003): Knowledge Construction and Collective Practice: At the Intersection of Learning, Talk, and Social Configurations in a Computer-mediated Mathematics Classroom. *The Journal of the Learning Sciences*, vol. 12, no. 3, pp. 361–408.
- Enyedy, N. and J. Goldberg (2004): Inquiry in Interaction: Developing Classroom Communities for Understanding Through Social Interaction. *Journal for Research in Science Teaching*, vol. 41,, 862–893.
- Federation of Digital Broad-Band Seismograph Networks (FDSN)*. (2004): Retrieved from <http://www.fdsn.org/> on 25 November 2004.
- Finholt T.A. (2002): Collaboratories. In B. Cronin (ed.): *Annual Review of Information Science and Technology*, vol. 36. Medford, NJ: Information Today, pp. 73–107.
- Foster, N.F. and S. Gibbons. (2005): Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine*, vol. 11, no. 1. Retrieved from <http://www.dlib.org/dlib/january05/foster/01foster.html> on.
- Fry, J. (2003) The Cultural Shaping of Scholarly Communication with Academic Specialisms. Doctoral dissertation. University of Brighton, Brighton, UK.
- Gazan, R., G.H. Leazer, C.L. Borgman, A.J. Gilliland-Swetland, L.J. Smart, D. Ancona and R.M. Nilsson (2003) Use Scenarios in the Development of the Alexandria Digital Earth Prototype (ADEPT). *Annual Meeting of the American Society for Information Science and Technology*, Long Beach, CA.
- Gilliland-Swetland, A.J. and G.H. Leazer (2001): Iscapes: Digital Learning Environments for the Promotion of Scientific Thinking by Undergraduates in Geography. *First ACM/IEEE-CS Joint Conference on Digital Libraries*, Roanoke, Virginia.
- Goble, C. and C. Wroe (2005): The Montagues and the Capulets. *Comparative and Functional Genomics*, vol. 5, no. 8, pp. 623–632.
- Godby, C.J., J.A. Young and E. Childress (2004): A Repository of Metadata Crosswalks. *D-Lib Magazine*, vol. 10, no. 12. Retrieved from <http://www.dlib.org/dlib/december04/godby/12godby.html> on 22 May 2006.
- Gomez, L.M., B.J. Fishman and R.D. Pea (1998): The CoVis Project: Building a Large-scale Science Education Testbed. *Interactive Learning Environments*, vol. 6, no. 1–2, pp. 59–92.
- Hey, T. and A. Trefethen (2003): The Data Deluge: An e-Science Perspective. *Grid Computing—Making the Global Infrastructure a Reality*, Wiley. Retrieved from [http://www.rcuk.ac.uk/escience/documents/report\\_datadeluge.pdf](http://www.rcuk.ac.uk/escience/documents/report_datadeluge.pdf) on 20 January 2005.
- Hilgartner, S. and S.I. Brandt-Rauf (1994): Data Access, Ownership and Control: Toward Empirical Studies of Access Practices. *Knowledge*, vol. 15,, 355–372.
- Hill, L. and M. Freeston M. (2003): Introduction to Georeferencing in Digital Libraries, Tutorial. *Joint Conference on Digital Libraries*, Houston, TX. Retrieved from <http://www.rice.edu/jcdl03/tutorials.html#geo1> on 22 May 2006.
- Hill, L., G. Janec, R. Dolin, J. Frew and M. Larsgaard (1999): Collection Metadata Solutions for Digital Library Applications. *Journal of the American Society for Information Science*, vol. 50, no. 13, pp. 1169–1181.
- Hill, L.L., L. Carver, M. Larsgaard, R. Dolin, T.R. Smith, J. Frew and M.A. Rae (2000): Alexandria Digital Library: User Evaluation Studies and System Design. *Journal of the American Society for Information Science*, vol. 51, no. 3, pp. 246–259.

- Hill, L.L. and G. Janee (2004): The Alexandria Digital Library Project: Metadata Development and Use. In D. Hillman and E. Westbrook (eds.): *Metadata in Practice: a Work in Progress*. Chicago: American Library Association.
- Hovy, E. (2003): Using an Ontology to Simplify Data Access. *Communications of the ACM*, vol. 46, no. 1, pp. 47–49.
- IEEE Learning Object Metadata (LOM)*. (2004): Retrieved from <http://www.ltsc.ieee.org/wg12/> on 15 December 2004.
- Incorporated Research Institutions for Seismology*. (2004): Retrieved from <http://www.iris.edu> on 25 November 2004.
- James San Jacinto Mountains Reserve*. (2004): University of California. Retrieved from <http://www.jamesreserve.edu/> on 30 November 2004.
- Janee, G. and J. Frew (2002): The ADEPT Digital Library Architecture. *Second ACM/IEEE-CS Joint Conference on Digital Libraries*. Portland, OR: ACM Press, pp. 342–350.
- Janee, G., J. Frew and L.L. Hill (2004): Issues in georeferenced digital libraries. *D-Lib Magazine*, vol. 10, no 5. Retrieved from <http://www.dlib.org> on 19 January 2005.
- Kline, R. and T. Pinch (1999): The Social Construction of Technology. In D. MacKenzie and J. Wajcman (eds.): *The Social Shaping of Technology.*, Maidenhead: Open University Press.
- Knowledge Network for Biocomplexity (KNB)*. (2004): University of California, Santa Barbara. Retrieved from <http://www.knb.ecoinformatics.org/index.jsp> on 25 November 2004.
- Kubicek, H. and W.H. Dutton (1997): The Social Shaping of Information Superhighways: An Introduction. In H. Kubicek, W.H. Dutton and R. Williams (eds.): *The Social Shaping of Information Superhighways: European and American Roads to the Information Society*. New York, Campus Verlag: St. Martin's Press, pp. 9–44.
- Kuhlthau, C.C. (1991): Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science*, vol. 42, no. 5, pp. 361–371.
- Kwa, C. (2005): Local Ecologies and Global Science: Discourses and Strategies of the International Geosphere–Biosphere Programme. *Social Studies of Science*, vol. 35, no. 6, pp. 923–950.
- Leazer, G.H., A.J. Gilliland-Swetland, C.L. Borgman and R.E. Mayer (2000): Classroom Evaluation of the Alexandria Digital Earth Prototype (ADEPT). In D. Kraft (ed.): *ASIS 2000 Proceedings of the 63rd ASIS Annual Meeting, Chicago, IL, November 12–16, 2000*, vol. 37. Medford, NJ: Information Today, pp. 334–340.
- Lord, P. and A. Macdonald (2003): *e-Science Curation Report. Data Curation for e-Science in the UK: an Audit to Establish Requirements for Future Curation and Provision*. JISC Committee for the Support of Research. 85 pp. Retrieved from [http://www.jisc.ac.uk/uploaded\\_documents/e-scienceReportFinal.pdf](http://www.jisc.ac.uk/uploaded_documents/e-scienceReportFinal.pdf) on 1 November 2004.
- Lynch, C.A. (2003): Colliding with the Real World: Heresies and Unexplored Questions about Audience, Economics, and Control of Digital Libraries. In A.P. Bishop, N. Van HouseVan and B.P. Buttenfield (eds.): *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press, pp. 191–216.
- MacKenzie, D. and J. Wajcman (eds.) (1999): *The Social Shaping of Technology*. 2. Maidenhead: Open University Press.
- Marchionini, G., C. Plaisant and A. Komlodi (2003): The People in Digital Libraries: Multifaceted Approaches to Assessing Needs and Impact. In A.P. Bishop, N. Van HouseVan and B.P. Buttenfield (eds.): *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press, pp. 119–160.
- Mayer, R.E., T.R. Smith, C.L. Borgman and L.J. Smart (2002): Digital Libraries as Instructional Aids for Knowledge Construction. *Educational Technology*, vol. 42, no. 6, pp. 38–42.

- Meadows, A.J. (1998): *Communicating Research*. San Diego: Academic Press.
- National Center for Atmospheric Research. (2004): Retrieved from <http://www.ncar.ucar.edu/ncar/> on 2 December 2005.
- NIMS: *Networked Infomechanical Systems*. (2005): Retrieved from <http://www.cens.ucla.edu/portal/nims/> on 3 March 2005.
- Olson, G.M. and J.S. Olson (2000): Distance Matters. *Human-Computer Interaction*, vol. 15, no. 2-3, pp. 139-178.
- OWL Web Ontology Language Guide: *W3C Recommendation*. (2004): Retrieved from <http://www.w3.org/TR/owl-guide/> on 22 May 2006.
- Pea, R.D. (1993): The Collaborative Visualization Project. *Communications of the ACM*, vol. 36, no. 5, pp. 60-63.
- Pritchard, S.M., L. Carver and S. Anand (2004): *Collaboration for knowledge management and campus informatics*. University of California, Santa Barbara. 38. Retrieved from [http://www.library.ucsb.edu/informatics/informatics/documents/UCSB\\_Campus\\_Informatics\\_Project\\_Report.pdf](http://www.library.ucsb.edu/informatics/informatics/documents/UCSB_Campus_Informatics_Project_Report.pdf) on 14 November 2005.
- Ribes, D. and G.C. Bowker (2004): *Ontologies and the Machinery of Difference: Towards a Sociology of Knowledge Representation*. Paris: Society for the Social Studies of Science.
- Ribes, D. and G.C. Bowker (2006, in press): Ontologies and the machinery of difference: Towards a sociology of knowledge representation. *Journal of the Association of Information Systems*.
- Rogers, E.M. (1995): *Diffusion of Innovations*. 4. New York: The Free Press.
- Sandoval, W.A. and C.L. Borgman (2004-2008): *CENS Education cyberInfrastructure (CENSEI), NSF Award ESI-0352572*. National Science Foundation, Directorate for Education and Human Resources. Retrieved from <http://www.cens.ucla.edu> on 20 May 2005.
- Sandoval, W.A. and B.J. Reiser (2003): Explanation-driven Inquiry: Integrating Conceptual and Epistemic Supports for Science Inquiry. *Science Education*, vol. 87., 1-29.
- Shankar, K. (2003): *Scientific Data Archiving: the State of the Art in Information, Data, and Metadata Management*. Retrieved from <http://www.cens.ucla.edu/Education/index.html> on 19 January 2005.
- Smith, T.R., D. Ancona, O. Buchel, M. Freeston, W. Heller, R. Nottrott, T. Tierney, A. Ushakov (2003): The ADEPT Concept-based Digital Learning Environment. *European Conference on Digital Libraries*, Trondheim, Norway: Springer.
- Smith, T.R. and M. Zheng (2002): ADEPT Knowledge Organization Team. Structured Models of Scientific Concepts as a Basis for Organizing, Accessing and Using Learning Materials. *Joint Conference on Digital Libraries*, Portland, Oregon: ACM Press.
- Sonnenwald, D.H. (2006): Scientific Collaboration: Challenges and Solutions. In B. Cronin (ed.): *Annual Review of Information Science and Technology*, vol. 40. Medford, NJ: Information Today, in press.
- Standard for the Exchange of Earthquake Data (SEED)*. (2004): Federation of Digital Seismographic Networks (FDSN). Retrieved from [http://www.iris.edu/manuals/SEED\\_chpt1.htm](http://www.iris.edu/manuals/SEED_chpt1.htm) on 25 November 2004.
- Star, S.L., G.C. Bowker and L.J. Neumann (2003): Transparency Beyond the Individual Level of Scale: Convergence Between Information Artifacts and Communities of Practice. In A.P. Bishop, N. Van House and B.P. Buitendijk (eds.): *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press, pp. 241-270.
- Sutton, C. (2003): *UCLA Develops Mobile Sensing System for Enriched Monitoring of the Environment*. UCLA. Retrieved from <http://www.engineer.ucla.edu/stories/2003/nims.htm> on 30 November 2004.

- Svenonius, E. (2000): *The Intellectual Foundation of Information Organization*. Cambridge, MA: The MIT Press.
- Thadani, V., M. Cook, K. Millwood, A. Harven, D. Fields, K. Griffis, J. Wise, K. Kim and W.A. Sandoval (2006): *Eyes on the Prize: Considering how Design Research can Lead to Sustainable Innovation*. Paper presented at the Annual Meeting of the American Educational Research Assn., San Francisco, April 7–12 Visited.
- The Digital Millennium Copyright Act (1998): Pub. L. No. 105–304, 112 Stat. 2860, Oct. 28, 1998. Retrieved from <http://www.copyright.gov/legislation/dmca.pdf> on.
- Van House, N.A. (2003): Digital Libraries and Collaborative Knowledge Construction. In A.P. Bishop, N. Van House and B.P. Buttenfield: *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA, MIT Press, pp. 271–296.
- House, N.A. Van, A.P. Bishop and B.P. Buttenfield (2003): Introduction: Digital Libraries as Sociotechnical Systems. In A.P. Bishop, N. Van House Van and B.P. Buttenfield (eds.): *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press, pp. 1–21.
- Wallis, J.C., S. Milojevic, C.L. Borgman and W.A. Sandoval (2006): The Special Case of Scientific Data Sharing with Education. *American Society for Information Science and Technology*, Austin, TX: Information Today.
- Woolgar, S. (2003): *Social Shaping Perspectives on e-Science and e-Social Science: the Case for Research Support : A Consultative Study for the Economic and Social Research Council (ESRC)*. Economic and Social Research Council (ESRC). Retrieved from [http://www.sbs.ox.ac.uk/html/faculty\\_area\\_sci\\_tech.asp](http://www.sbs.ox.ac.uk/html/faculty_area_sci_tech.asp); <http://www.sbs.ox.ac.uk/downloads/e-SocialScience.pdf> on.
- Wouters, P. (2004): What is the Matter with e-Science? – Thinking Aloud about Informatisation in Knowledge Creation. *Society for Social Studies of Science*, Paris. Retrieved from <http://www.nerdi.knaw.nl> on 29 October 2004.