Laying the groundwork to investigate scientists' structures and resources in their information environment

John D'Ignazio
School of Information Studies
Hinds Hall
Syracuse University
jadignaz@syr.edu

ABSTRACT

The set up to a case study that examined the interactions of a group of interdisciplinary social scientists conducting a systematic review of their research area examines why and how to investigate the information structures and resources relevant in science practice. Partners in other components of a national scale cyberinfrastructure project to develop systems that manage science data periodically perturbed this small group. Their resulting interactions provided an enhanced ability to explore socially versus locally contributed structures related to the resources that established their information environment. Coding and analysis of recorded progress meeting and interview transcripts and related artifacts created through the research process should reveal the structures attributable to information systems or disciplinary practice that these scientists attended to, engaged with, and created. The purpose of the coding is to give greater definition to resources from a practice viewpoint in order to support interdisciplinary eScience research via information systems. Also of interest in pursuing the case study is the impact of an informatician participating with the scientists as they conducted their project.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems – *Human Information Processing.*

General Terms

Human Factors

Keywords

Science data management; instrumental case study; metadata schema.

1. RESEARCH ORIENTATION

The large-scale investment in cyberinfrastructure to facilitate science investigations of complex phenomenon that require the participation of multiple disciplines has become a fixture of modern society. The development of cyberinfrastructure, a combination of hardware and software, has been receiving attention in the last decade, as collaborative tools that support interdisciplinary communication and the persistence and accumulation of data files affects the work practices of scientists. Adoptions of information and communication technology (ICT) in societal sectors, including cyberinfrastructure, have enabled productivity, but also have caused disruptions in this new form of science practice, or eScience. Information systems cause new roles and behavior to result in pursuit of science advancement and also result in altered responsibilities.

Such innovation and disruption leads to new terms and changing definitions, such as the terms "cyberinfrastructure" and "eScience." Because of their novelty, the terms are applied at different scales, in different disciplines, and may highlight changes to different facets of science work. For example, cyberinfrastructure could be considered on a scale of international system connections, such as astronomy's system of virtual observatories, or at a more local, discipline-specific manner, such as a new grid-computing system supported by the campus IT department for a professor's research. The functions that such ICT enables as part of an eScience endeavor might include collaboration, measurement, or analysis and visualization to generate results, or all of the above. Scale effects apply as well; eScience occurs with scientists involved on "big science" projects characterized by significant federal investment on centralized instrument platforms and discipline-level repositories, as well as in "little science" contexts where individual research groups depend on computer products and tools and share their work in an Internet-based collection devoted to progress on a comparatively narrow research agenda.

Applications of these terms have, however, resulted in consistently increased attention to the digital objects created by scientists in these ICT-enabled environments, usually classed under the term "data." The inclusive definition of data from the National Science Board provides a view to the range of what accumulates during the course of eScience: "any information that can be stored in digital form, including text, numbers, images, video or movies, audio, software, algorithms, equations, animations, models, simulations, etc" [NSF, 2005]. Whereas historically data may have been seen as a proprietary local, physical inscription [Latour & Woolgar, 1986], data as a digital object is increasingly a resource made available over the Internet to which research funders encourage access in order to enhance their investment and increase productivity [NSF, 2007]. What defines these digital objects is their tight relation to science work, or practice, versus the scholarly article, which although is inextricable from science, has a communication function as its reason for being. Practice-based artifacts such as data are defined by their materiality in terms of a work situation defined by understandings, skills, and activities [Palmer & Cragin, 2007].

Libraries have been long-standing stakeholders in science with capability and expertise applied to supplying publications the scientist uses in the research process. They, working alongside computer engineers and information system analysts, have helped build large-scale catalogs and scholarly-article-bases that are now the essential infrastructure supporting science communication [Brown, 2010]. Since the early 1990s, libraries have also participated in or hosted the construction of digital libraries and institutional repositories that have provided access to a wider range of digital objects than books and articles, with mixed success due to impact of this new media and related shifts in possible uses and needs [Saracevic, 2005; NRC, 1998].

Libraries are now stepping in to contribute their expertise in managing information and building collections to support eScience. For example, the Association of Research Libraries has recommended institutional programs be developed to accommodate “the emergent role of data curation, characteristics of virtual organizations, relevant policy for data and research dissemination, and tools and infrastructure systems” [ARL, 2007]. The research and practice sides of librarianship both have an active role, alongside domain scientists, in the two National Science Foundation DataNet projects designed to “provide reliable digital preservation, access, integration, and analysis capabilities for science and/or engineering data over a decades-long timeline” [NSF, 2007].

Management of data in collections is proving difficult for libraries at large research centers and universities because these digital objects lack much of the scholarly article’s formal elements, such as an object’s belonging to a genre, having a readable subject written in text, or existing in a tangible form [Wynholds, 2011]. Data as a practice-based artifact is bound up in the ongoing work of scientists—form and content mutate as part of workflows or levels of analytical or computing processes [Bose & Frew, 2005]. Scientists may resist handing over data to a library because of the corresponding associations with the publication of a finished product that such an action has in scientific communication [Borgman, 2007]. In addition, data may not have identifiable access points based on community-accepted attributes; for example, a data set’s “author” may be the principal investigator, or it may be one of a shifting array of graduate students in a research group [D’Ignazio & Qin, 2008]. Finally, the relation of data that is collected into repositories to scholarly article databases is an ongoing research and development area [Warner, Bekaert, et al., 2007].

People with new roles and responsibilities are taking on these information-oriented challenges related to science data with professional titles such as data scientist, data curator, or eScience librarian [Beagrie, 2006; Qin, D’Ignazio, et al. 2010]. Research into the contribution of these individuals, such as in astronomy and earth sciences, shows them facing a complicated environment of scientists working with technology that create a wide array of content which conforms to a variety of disciplinary or research-project-specific formats, types, and scales [Choudhury, DiLauro, et al., 2007; Karasti & Baker, 2006]. Their work is necessary to establish practices, mechanisms, and structures to build science data collections according to a typology of scale from local research support to large-scale community adoption as reference collections. Communication is required to adopt, extend, or create information structures, in some relation to pre-existing standards, and with connection to or backing of large research facilities and governing bodies [Borgman, Wallis, et al., 2007]. Such information-structure-building occurs in relation to investments in database systems and engineering, often in the case of “big science,” or according to the knowledge-base and willingness of research-group members who embrace open-source database programs to become nodes in a science network [Cragin & Shankar, 2006].

Because data has a different nature than the text-based resources that mature information retrieval systems are built to encode for community access, representation of datasets using metadata has loomed large as an issue in science data management [Qin & D’Ignazio, 2010]. The techniques in creating and applying metadata that are compiled as scheme-governed records can provide the means of consistent and meaningful access to objects in science data collections [Duval, E., W. Hodgins, et al., 2002; Dempsey & Heery, 1998]. Experience from digital library developments that relied on metadata creation and harvesting can aid in consistently describing science data across networked sites, but also enhance semantic operations in centralized systems [Arms, W. Y., N. Dushay, et al., 2003].

Metadata-level mechanisms such as crosswalks, application profiles, and registries help to bridge disparate databases and repositories, but require advanced technical knowledge and investment in either system or human resources for creation and management [Zeng & Qin, 2008]. Such effort can increase the return on investment of data repositories by facilitating interoperability, defined by the National Information Standards Organization as the ability of two or more information systems to exchange data with a minimal loss of content and functionality [NISO, 2004]. This is not, however, a strictly technical problem. Domain knowledge and work practice variation make a significant impact on data management and reuse by scientists from different disciplines and research groups, thus limiting the “social” interoperability between data repositories [D’Ignazio & Qin, 2008].

A dynamic, varied, technology-enabled environment thus characterizes cyberinfrastructure-enabled eScience. That environment includes work done by some type of information professional, henceforward dubbed “informatician,” to aid management of the resulting digital data objects as they persist in an ICT system. As required of infrastructure builders of the past, these new builders need assistance in understanding the elements and methods necessary to build reliable information systems that serve science, and that also leverage the infrastructure already in place. In order to clarify the best form of this assistance, an investigation is required from the viewpoint of scientists, guided by the following research questions related to data management and its representation via metadata systems:

1. What structures do scientists attend to, engage with, and create that defines their information environment?
2. How do scientists interact with data resources in their work processes, products, and systems?
3. What is the impact on structures of interest when an informatician is embedded with a research group?

2. LITERATURE-BASED MODEL

Justifying the new, context, and discipline-specific articles generated in the last five years on the topic of cyberinfrastructure, eScience, and science data management with the bulk of the scholarly literature rooted in the topics of librarianship, information retrieval, and information behavior, brings attention to the continued applicability of the core functions that are the focus of information studies. Also resulting was an insight into the central role of technology in advancing the character and complexity of information systems that facilitate human creation and communication of objects that are produced and contained or indexed by these systems. Previous eras of information system innovation and implementation have brought change in the types and scale of information needing to be managed in society, just as is now occurring again with eScience.

Starting with librarianship and moving through to today’s work in science data management, a phenomenon occurs when collections are formed and a resource model is applied to information artifacts such as books, scholarly articles, or even data sets. When something is recognized as a resource, its value is recognized for its ability to consistently provide what is needed to achieve some economically viable pursuit [Norton, 2000, p.83]. Despite its intangible and easily copied nature, information has been increasingly seen as an economic asset and managed per corporate and governmental strategy [Oppenheim, Stenson, et al., 2001]. When the form of information has stabilized so that people see the value in aggregating it into collections, the economic and functional value of providing a related service set becomes apparent through evaluation mechanisms set in the context of a societal domain [Lancaster, 1989, p.120].

The mechanisms that enable and surround a collection, and the phenomena that result and are enabled by it, are the main focus of information studies. This is made apparent through the consistent invocation of function words that establishes an expertise in management techniques and infrastructure creation, involving the gathering, organizing, storing, retrieving, and dissemination of information [Bates, 1999]. These management functions focus on information resources in terms of creating a stable, coherent collection that aggregates into its own resource, since it serves the diverse needs of members of a community.

At certain points in time, and with science as a driver, the information resources that information studies concerns itself with have changed, creating professional tension and generating new scholarly research. For example, a split occurred between "documentalists" and librarians in the early part of the prior century because of the growth in science and technical reports that did not fit the publication model of books and monographs but that scientists and industrialists needed for their work [Batty & Bearman, 1983, p. 366]. Again, beginning in the 1950s at Western Reserve University under the leadership of Jesse Shera, the need to have systems that enabled access to a large collection of scholarly articles in various science disciplines started the research and development of information retrieval [Wright, 1985].

While societal developments involving information resources led to new paradigms of information work, these trends were somewhat consistent in the need to manage externally generated, published artifacts, so that creation of these artifacts was less considered in the functional set. The need to manage business information, however, as it occurred in personal workstations accessing large-scale information systems in the last quarter of the 20th century resulted in the need to consider information creation and its value, internally and externally, to organizations and society [Levitan, 1982]. Information resource management, or its shortened form, information management, takes a strategic approach to the data-driven and technological environment of information systems in organizational contexts, so that the structures and design of these databases are built with an awareness of and support for the information flows of an organization [Lytle, 1988].

Information systems cannot operate unless the information environment is analyzed, data and metadata elements are established, and attributes for each are assigned according to best practices in all the functional areas. Such attention requires much human effort and social agreement, so it is generally applied to objects defined as resources. But how objects achieve their nature and consensus as socially valued resources, versus idiosyncratically created objects of individuals is a process worthy of attention. In science data management, currently, each instantiation of a data repository provides a distinct formulation of elements and attributes related to a set of objects, depending on the technology applied, the encoding language of description used, and the needs of the investigating scientists closest to the resource as reflected in the design [Baker & Yarmey, 2009; Borgman, Wallis, et al., 2007]. Mechanisms exist that enable users to bridge disparate systems and obtain resources outside their local control or association, using crosswalks, registries, and union catalogs [Heery, 2004], but these approaches add additional interface views based on yet more formulations and combinations of elements and attributes, often imposing a high cognitive effort for people obtaining resources, or reinforcing a narrow community of use. What is actually a resource in this complex, multi-level environment is a current puzzle.

Isolating three main, co-existing themes, and their interrelation, from the information studies literature defines a research area, science informatics, appropriate in the eScience context that helps to examine resources and their relations. Science informatics is composed of information resource management, information and communication technology, and disciplinary practices that drive growth in areas of science. The science informaticians who function in this environment, as opposed to other environments driven by different information systems, have as a focus the information resource management of the digital resources produced by ICT implementations that are governed by the needs and uses of particular disciplinary practices.

The relationships of these themes are portrayed in the following graphical representation:

Figure 1: The Science Informatics Model (arrows indicate interactions)

3. STUDY METHODOLOGY

The science informatics model was created looking through the meta-theoretical lens of structuration, introduced through the writings in the early 1970s of French social theorist Anthony Giddens, who maintained that there was a duality of human agency and social structure [Parker, 2000]. His writings explained the development and political operation of social classes, the manner in which it relates people, structures, interaction, rules, and resources through processes such as signification, legitimation, and intersubjective context creation [Cohen, 1989].

Structuration provides the means to investigate natural phenomenon whereby there is no part of information that is understandable by one human alone, and yet by nature is recognized and used by one human alone. This construction of the information environment of individuals is parallel to that of people acting alone while simultaneously composing their social environment. As presented in Figure 1, while the three components of the science informatics model occur simultaneously, they each are the work of no one individual, replete with social definitions, investment, and agreement, yet are sustained, reproduced, and created through the agency of individuals. Concretely, this structuration effect is exhibited by the assignment of the same word by both cognitive and information scientists. According to the former, people depend on their mental schemas, both learned and created, to relate concepts and objects in order to operate in their environment [Marshall, 1995, p. 39]. Informaticians, working with computer and domain scientists, have encouraged the establishment of data and metadata schemas that provide access to resource collections according to the operating philosophy that even this relatively intangible asset can be managed appropriately if it is well-modeled [Cronin & Davenport, 1991, p.9].

The structuration-based model of science informatics suggests a research methodology to pursue the three research questions listed previously that are driven by the appearance of cyberinfrastructure and eScience and answer some of the ambiguity created by the attention to science data managed as a resource in the form of a digital object. To conduct a case study at a field site via participant-observation in a science-domain-specific center of research strongly oriented to eScience systems and methods would explore the creation of and operation with structures that occur at both the individual and social levels simultaneously. These structures may be impacted by the participation of an informatician located at the site of an operating research group whose role it would be to facilitate their use and adoption of resources according to the requirements of communicating and operating in a discipline, and made capable by applications of ICT.

There have been many types of case studies identified, but all emphasize meticulous attention to highly contextualized activity in order to develop a rich description of particular issues and phenomena [Stake, 2005]. Approaching a research setting with established concerns as directed by the science informatics model and structuration theory, requires the application, according to Robert Stake's typology, of an instrumental case study methodology.

The field site was provided through their Advanced Study Program by the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. NCAR is one example of an organization that depends on cyberinfrastructure to further science—it's location as a node on the Teragrid, existence as a center of excellence in modeling and simulation, and maintenance of an array of advanced measurement devices and observation platforms both draws researchers to visit the component laboratories to conduct their work, and is a guiding force for remote researchers from various disciplines who monitor the organization's developments to be able to access and participate in the investigations of the moment. Participants in the NCAR cyberinfrastructure workshop held in 2002 "represented many of the intersecting disciplines within the broad environmental sciences" [NCAR, 2003]. The number of disciplines listed in the report as present, seven or more, reflect the impact of investments in this one, albeit major, research center. Through the efforts of its Earth Observing and Computational and Information System Laboratories, NCAR has established several massive data repositories to both archive data produced from the center's instruments and supply a community-oriented resource for the internationally distributed scientists studying the atmosphere. NCAR is therefore a good example of an institution creating and promulgating socially-defined and mediated structures of resources for use by constituents active in an interdisciplinary research area.

My hosts were members of the Integrated Information Services (IIS) department who run both the NCAR Library and the National Science Digital Library. I began an eight-month stay beginning mid-September 2010 hoping it would allow sufficient time to understand technical developments of NCAR's several large data repositories, access a variety of scientists representing research groups and domains, and incorporate a synthesized understanding of their cognitive representations of data resources from their work environment into a system design iteration. I confirmed Stake's experience that in case selection the researcher "leans toward those cases that seem to offer opportunity to learn" [Stake, 2005, p. 451].

In my research, the case that offered the most to learn was afforded by my sponsor, Mary Marlino, who as director of the IIS department is also one of the principal investigators on one of the two current NSF DataNet programs, the Data Conservancy (DC) project. NCAR, through Marlino's work and another of the DC PIs, is one among many constituent organizations of this large-scale project whose goal is to research and develop the future multi-institutional and multi-disciplinary digital data repositories for science. The NCAR role is to explore how social scientist data might be supported in repository designs that are, however, mostly specified to support physical science needs, such as biology, earth science and astronomy.

Amidst the hundreds of scientists at NCAR researching and modeling atmospheric science, and lately of international note, climate change due to global warming, there is a small, interdisciplinary department of social scientists. Because of the nature of the group, their situation in a hard science research center, and their sub-field's reliance on biophysical data for case study research, they represent in many dimensions the effects, promise, and constraints of interdisciplinary research [Haythornthwaite, 2006]. This group is attempting to master the output of a community of researchers studying urban vulnerability (UV) and resilience to climate change. This has risen to be a topic of great strategic import to NCAR and the wider community best represented by the Intergovernmental Panel on Climate Change. Recent output from this community indicate an increased focus on social aspects compared with strict physical science: a) humans as largely the source of recent climate change from global warming, and b) increasingly urbanized coastal populations are most at risk to this climate change.

This case is therefore a good way to reveal individual level adaptations, adoptions, creation, and use of structures scientists need in their information environment. I captured the progress of the UV research group as they conducted a systematic review and meta-analysis of case studies of urban areas around the globe. My own case study furthers recent work in information studies that revealed information "extractions" happening at the article level in a highly collaboratory manner during two research groups' systematic analyses of medicine and public health case studies [Blake & Pratt, 2006]. The thrust of the UV research project is to understand research lineages and coverage across the many individual studies of their subfield. Included in the case study are the effect on the UV research group of periodic injections and perturbations from the IIS and DC team members who, in partnership with the UV group are creating a "social science observatory" to support their community.

A main operational tenet of the DC project is the application of design process strategies such as participatory or contextual design, applied in the Human-Computer Interaction (HCI) community [Beyer & Holtzblatt, 1998]. The philosophy is to involve people from the community-of-practice as design progresses "early and often" in order to end up with a system, built from use-case-derived system requirements, that performs to meet the community's needs. While much of the HCI literature is related to tasks, by applying versions of methods used by expert or decision-support-system researchers known as knowledge elicitation techniques [Cooke, 1994], contextual design is serving to understand the more abstract, information environment of eScience.

Knowledge elicition techniques used in the course of the case study are meant to help establish data curation best practices through derivation of scientists' cognitive structures mentioned in recorded progress meetings of the group or described in interviews, or apparent in many artifacts produced in the course of the systematic review. This range of research artifacts is a common characteristic of case study research [Yin, 2003, p.85]. I plan to code the artifacts for structures and resources that may occur anywhere in the triad of the science informatics model. The coding and analysis would need to demonstrate a consensus of value of resources and composition of structures by my participants in the course of their work progress. This evidence will be used to disentangle what the participants' consider and rely on as data and how best a metadata scheme could provide access to these resources as an intersubjective construction [Walsham, 1995]. Specific data curation questions to be examined include how these scientists view provenance issues as to source and ownership of data, how they create and use schemes to handle data and information of various types, formats, and scales. Another important aspect on the semantic level would be their process of deciphering and contributing to definitions that would lead to controlled vocabularies [Mai, 2008].

One-time answers are unlikely to be satisfactory, since the system, interactants, and overarching research goals across various science disciplinary practices are mutable in multiple dimensions. Techniques for informaticians to conduct knowledge elicitation of scientists working in their information environment to drive data curation is another intended research outcome of this case study.

4. FUTURE DIRECTION

By gathering meeting transcripts and data-oriented artifacts as the research team conducted the systematic review and meta-analysis, I am anticipating exploring and validating the model of science informatics as graphically represented in Figure 1 according to the definitions of structure, interaction, and resource available in structuration theory and adapted for use in the information environment by the information studies literature. These definitions will drive the coding scheme applied to the data artifacts obtained through participant observation during an-intensive phase of activityby the UV research group.

Their strategies of processing a large number of research studies in their domain will reveal patterns of structure creation, adoption, and use in a science context more generally. It also reveals the impact of an informatician in helping to locate and work with resources of value to these scientists that should help indicate best practices of applying IRM functions to the structures of an eScience environment. The source and role of structures should be identified both across the triad and from either the societal or individual perspective as limited by the context of this case study.

5. ACKNOWLEDGMENTS

The author thanks the Institute of Museum & Library Services for support while conducting this research, the Data Conservancy project of the National Science Foundation DataNet program for the setting of the study, and the National Center for Atmospheric Research through the Advanced Study Program for hosting the student.

6. REFERENCES

[1] *Agenda for Developing E-Science in Research Libraries*. 2007. Association of Research Libraries, Washington, D.C.

[2] Arms, W. Y., Dushay, N., Fulker, D., et al. 2003. A case study in metadata harvesting: the NSDL. *Libr Hi Tech* 21, 2, 228-237.

[3] Baker, K. S. and Yarmey, L. 2009. Data Stewardship: Environmental Data Curation and a Web-of-Repositories. *Int J Digit Curation* 4, 2, 12-27.

[4] Bates, M. J. 1999. The invisible substrate of information science. *J Am Soc Inform Sci*. 50, 12, 1043-1050.

[5] Batty, D. and Bearman, T. C. 1983. Knowledge and Practice in Library and Information Services. In *The Study of Information: Interdisciplinary Messages*. F. Machlup and U. Mansfield, Ed. John Wiley & Sons, New York, NY**:** 365-369.

[6] Beagrie, N. 2006. Digital Curation for Science, Digital Libraries, and Individuals. *Int J Dig Curation* 1, 1, 3-16.

[7] Beyer, H. & Holtzblatt, K. 1998. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann Publishers, San Francisco.

[8] Blake, C. and Pratt, W. 2006. "Collaborative information synthesis I: A model of information behaviors of scientists in medicine and public health." *J Am Soc Inform Sci Tech* 57, 13, 1740-1749.

[9] Borgman, C. L. (2007). *Scholarship in the Digital Age*. MIT Press, Cambridge, Mass.

[10] Borgman, C. L., Wallis, J.C., Enyedy, N. 2007. Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries *Int J Digit Libr* 7, 1&2, 17-30.

[11] Bose, R. and Frew, J. 2005. Lineage Retrieval for Scientific Data Processing: A Survey. *ACM Comput Surv* 37, 1, 1-28.

[12] Brown, C. M. 2010. Communication in the Sciences. *Annu Rev Inform Sci* 44, 287-316.

[13] Choudhury, S., DiLauro, T., Szalay, A., et al. 2007. Digital Data Preservation for Scholarly Publications in Astronomy. *Int J Dig Curation* 2, 2, 20-30.

[14] Computer Science and Telecommunications Board of the National Research Council. 1998. Design and Evaluation: A Review of the State of the Art. *D-Lib Mag.* Corporation for National Research Initiatives, Reston, Va. 4.

[15] Cohen, I. J. 1989. *Structuration Theory: Anthony Giddens and the Constitution of Social Life.* St. Martin's Press, New York.

[16] Cooke, N. J. 1994. Varieties of Knowledge Elicitation Techniques. *Int J Hum-Comp St* 41, 6, 801-849.

[17] Cragin, M. H. and Shankar, K. 2006. Scientific Data Collections and Distributed Collective Practice. *Comp Support Coop W* 15, 2-3, 185-204.

[18] Cronin, B. and Davenport, E. 1991. *Elements of Information Management.* Scarecrow Press, Metuchen, N.J.

[19] *Cyberinfrastructure for environmental research and education*. 2003. National Center for Atmospheric Research**,** Boulder, Colo.

[20] *Cyberinfrastructure Vision for 21st Century Discovery*. 2007. Cyberinfrastructure Council, National Science Foundation**,** Washington, D.C.

[21] D'Ignazio, J. A. and Qin, J. 2008. Faculty data management practices: A campus-wide census of STEM departments. In *Proceedings of the 71st ASIS&T Annual Meeting: People Transforming Information - Information Transforming People*. (Columbus, Ohio, October 24-29) A. Grove and A. Rorissa, Ed. ASIS&T, 165-170.

[22] Dempsey, L. and Heery, R. 1998. Metadata: a current view of practice and issues. *J Doc* 54, 2, 145-172.

[23] Durrance, J. C. and Fisher, K. E. 2003. Determining How Libraries and Librarians Help. *Lib Trends* 51, 4, 541-570.

[24] Duval, E., Hodgins, W., Sutton, S., et al. 2002. Metadata Principles and Practicalities. *D-Lib Mag.* Corporation for National Research Initiatives, Reston, Va., 8.

[25] Haythornthwaite, C. 2006. Learning and knowledge networks in interdisciplinary collaborations. *J Am Soc Inform Sci Tech* 57, 8, 1079-1092.

[26] Heery, R. 2004. Metadata Futures: Steps Toward Semantic Interoperability. In *Metadata in practice*. D. I. Hillmann and E. L. Westbrooks, Ed. American Library Association**,** Chicago, Ill, 257-271.

[27] Karasti, H. and Baker, K. S. 2008. Digital Data Practices and the Long Term Ecological Research Program Growing Global. *Int J Dig Curation* 3, 2, 42-56.

[28] Lancaster, F. W. 1989. Measurement and Evaluation of Information Services. In *Principles and Applications of Information Science for Library Professionals.* J. N. Olsgaard, Ed. American Library Association**,** Chicago, 120-136.

[29] Latour, B. and Woolgar, S. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, Princeton, N.J.

[30] Levitan, K. B. 1982. Information Resource(s) Management. *Annu Rev Inform Sci* 17, 228-266.

[31] *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. 2005. National Science Board, National Science Foundation**,** Washington, D.C. 87.

[32] Lytle, R. H. 1988. Information Resource Management: Research, Education, and Practice. *J Am Soc Inform Sci* 39, 5, 337-339.

[33] Mai, J. E. 2008. Actors, domains, and constraints in the design and construction of controlled vocabularies. *Knowl Organ* 35, 1, 16-29.

[34] Norton, M. J. 2000. *Introductory Concepts in Information Science*. Information Today, Medford, N.J.

[35] Oppenheim, C., Stenson, J. Wilson, R. M. S. 2001. The attributes of information as an asset. *New Libr World* 102, 11/12, 458 - 464.

[36] Palmer, C. and Cragin, M. 2007. Scholarship and Disciplinary Practices. *Annu Rev Inform Sci* 42, 165-212.

[37] Parker, J. 2000. *Structuration*. Open University Press, Philadelphia.

[38] Qin, J. and D'Ignazio, J. A. 2010. The Central Role of Metadata in a Science Data Literacy Course. *J Libr Metadata* 10, 2&3, 188-204.

[39] Qin, J., D'Ignazio, J. A. Oakleaf, M., et al. 2010. Educating eScience Librarians. In *Proceedings of the 6th International Digital Curation Conference*. (Chicago, Ill., December 6-8).

[40] Saracevic, T. 2005. How were digital libraries evaluated? In *Proceedings of Libraries in the Digital Age* (Dubrovnik, Croatia, LIDA).

[41] Stake, R. E. 2005. Chapter 17: Qualitative Case Studies. In *The Sage Handbook of Qualitative Research.* N. K. Denzin and Y. S. Lincoln, Ed. Sage**,** Thousand Oaks, Calif., 443-466.

[42] *Sustainable Digital Data Preservation and Access Network Partners (DataNet).* Program Solicitation 07-601. 2007. National Science Foundation, Washington, D.C.

[43] *Understanding metadata*. 2004. N. Press, Ed. National Information Standards Organization, Washington, D.C.

[44] Walsham, G. 1995. Interpretive case studies in IS research: nature and method. *Eur J Inform Syst* 4, 1, 74-81.

[45] Warner, S. Bekaert, J. Lagoze, C., Liu, X. M., Van de Sompel, H. (2007). Pathways: augmenting interoperability across scholarly repositories. *Int J Digit Libr* 7, 1&2, 35-52.

[46] Wright, H. C. 1985. Shera as a Bridge between Librarianship and Information Science. *Journal Libr Hist Philos* 20, 2, 135-155.

[47] Wynholds, L. 2011. Linking to Scientific Data: Identity Problems of Unruly and Poorly Bounded Digital Objects. *Int J Digit Curation* 6, 1, 214-225.

[48] Yin, R. K. 2003. *Case Study Research: Design and Methods*. Sage, Thousand Oaks, Calif.

[49] Zeng, M. L. and Qin, J. 2008. Chapter 4: Schemas and Syntax. In *Metadata*. Neal-Schuman Publishers**,** New York, 131-147.