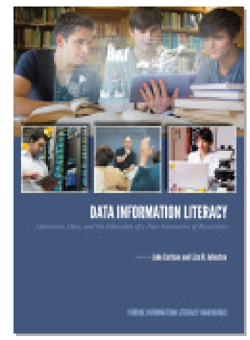
Data Information Literacy

Carlson, Jake , Johnston, Lisa



Published by Purdue University Press

Carlson, Jake and Lisa Johnston.
Data Information Literacy: Librarians, Data and the Education of a New Generation of Researchers .

Purdue University Press, 2014.
Project MUSE.muse.jhu.edu/book/42542.

CHAPTER 1

DETERMINING DATA INFORMATION LITERACY NEEDS

A Study of Students and Research Faculty

Jake Carlson, University of Michigan

Michael Fosmire, Purdue University

C. C. Miller, Purdue University

Megan Sapp Nelson, Purdue University

INTRODUCTION

The nature and practice of research and scholarship is undergoing dramatic change with the advent of ready access to high-bandwidth networks, the capacity to store massive amounts of data, and a robust and growing suite of advanced informational and computational data analysis and visualization tools. The practice of technology-driven research, known as *e-science,* or more broadly as *e-research,* has had a transformative effect in the science and engineering fields. E-research applications are growing within the humanities and social science disciplines as well, where e-research is poised to have similar effects on the nature and practice of research.

The complexity and scale of e-research in turn requires an evolution of traditional models of scholarly communication, library services, and the role of librarians themselves. In response, librarians are initiating discussions and projects to situate themselves in those areas of e-research most in need of library science expertise (Jones, Lougee, Rambo, & Celeste, 2008). In light of the federal expectation that grant proposals have a data management plan (DMP; NSF, 2011), libraries are starting conversations in their universities to negotiate a role in the management of research outputs.

Data management skills also provide the opportunity for an evolution of instruction in libraries. Academic libraries offer information literacy courses and programs as part of the educational mission of the institution. Extending information literacy to include programs on data management and curation provides a logical entry point into increasing the role of libraries in supporting e-research. A successful education program, however, must be based on a firm understanding of current practice and standards as well as the needs of the target audience. There is a lack of research on the needs of both the researchers and the students grappling with these issues in the classroom and in the laboratory. The authors attempted to address this knowledge gap by gathering data from interviews with faculty researchers and from the authors' own Geoinformatics course. With this information, the authors proposed a model set of outcomes for data information literacy (DIL).

BACKGROUND

E-Research and Implications for Libraries

E-research has had a tremendous impact on a number of fields, increasing the capabilities of researchers to ask new questions and reduce the barriers of time and geography to form new collaborations. In astronomy for example, the National Virtual Observatory (NVO) makes it possible for anyone from professional astronomers to the general public to find, retrieve, and analyze vast quantities of data collected from telescopes all over the world (Gray, Szalay, Thakar, Stoughton, & vandenBerg, 2002; National Virtual Observatory, 2010). For scholars of literature, the HathiTrust Digital Library not only provides a tremendous collection of scanned and digitized texts, but also its Research Center provides tools and computational access to scholars seeking to apply data mining, visualization, and other techniques toward the discovery of new patterns and insights (HathiTrust Research Center, n.d.). It should be no surprise, of course, that such projects simultaneously produce and feed upon large amounts of data. The capture, dissemination, stewardship, and preservation of digital data are critical issues in the development and sustainability of e-research.

Funding organizations and professional societies identified a need for educational initiatives to support a workforce capable of e-research initiatives. The National Science Foundation (NSF) first described the connection between e-research and education. The 2003 Atkins Report highlighted the need for coordinated, large-scale investments in several areas, including developing skilled personnel and facilities to provide operational support and services (Atkins et al., 2003). In 2005 the National Science Board produced a report that articulated existing and needed roles and responsibilities required for stewarding data collections, followed by a series of recommendations for technical, financial, and policy strategies to guide the continued development and use of data collections (National Science Board, 2005). The American Council of Learned Societies issued a report in 2006 calling for similar attention and investments in developing infrastructure and services for e-research in the humanities fields (Welshons, 2006). More recently, the National Academy of Sciences issued a report advocating the stewardship of research data in ways that ensured research integrity and data accessibility. The recommendations issued in the report included the creation of systems for the documentation and peer review of data, data management training for all researchers, and the development of standards and policies regarding the dissemination and management of data (National Research Council, 2009).

While the rich, collaborative, and challenging paradigm of e-research promises to produce important, even priceless, cultural and scientific data, librarians are determining their role in the curation, preservation, and dissemination of these assets. In examining how e-research may affect libraries, Hey and Hey argued that e-research "is intended to empower scientists to do their research in faster, better and different ways," (Hey & Hey, 2006, para. 10). They particularly emphasized that information and social technologies made e-research a more communal and participatory exercise, one that will see scientists, information technology (IT) staff, and librarians working more closely together. A particular challenge looming with the rise of e-research is the "data deluge"—that is, the need to store, describe, organize, track, preserve, and interoperate data generated by a multitude of researchers to make the data accessible and usable by others for the long term. The sheer quantity of data being generated and our current lack of tools, infrastructure, standardized processes, shared workflows, and personnel who are skilled in managing and curating these data pose a real threat to the continued development of e-research.

Gold (2007) provided an outline of the issues and opportunities for librarians in e-science. Starting from the familiar ground of GIS (geographic information systems), bioinformatics, and social science data, Gold argued that librarians working in e-science will develop relationships—both upstream and downstream of data generation—and the effort may be "both revitalizing and transformative for librarianship" (Sec. 2.2, para. 6). Similarly, the *Agenda for Developing E-Science in Research Libraries* outlined five main outcomes that focused on capacity building and service development in libraries for supporting e-science (Lougee et al., 2007). Walters (2009) further asserted that libraries taking "entrepreneurial steps" toward becoming data curation centers are on the right track, reasoning that "a profound role for the university research library in research data curation is possible. If the role is not developed, then a significant opportunity and responsibility to care for unique research information is being lost" (p. 85). In other words, the academic library community seems reasonably sure that supporting e-research is not so novel that it falls outside of the mission and founding principles under which libraries operate.

Educational Preparation for E-Research

Ogburn (2010) predicted that e-science will quite certainly fail if future generations of scholars are not savvy with *both* the consumption and production of data and tools. "To prepare the next generation of scholars the knowledge and skills for managing data should become part of an education process that includes opportunities for students to contribute to the creation and the preservation of research in their fields" (p. 244). It is not enough to teach students about handling incoming data, they must also know, and practice, how to develop and manage their own data with an eye toward the next scientist down the line. The Association of Research Libraries reported to the NSF in 2006 that because

> many scientists continue to use traditional approaches to data, i.e., developing custom datasets for their own use with little attention to long-term reuse, dissemination, and curation, a change of behavior is in order. . . . [This change] will require a range of efforts, including . . . perhaps most important of all, concerted efforts to educate current and future scientists to adopt better practices. (Friedlander & Adler, 2006, p. 122)

The inspiration for the authors' own work on instructional components to e-science comes from the NSF's *Cyberinfrastructure Vision for 21st Century Discovery,* in which the dramatic rhetoric of revolution and recreation does indeed trickle down to education:

> Curricula must also be reinvented to exploit emerging cyberinfrastructure capabilities. The full engagement of students is vitally important since they are in a special position to inspire future students with the excitement and understanding of cyberinfrastructure-enabled scientific inquiry and learning. Ongoing attention must be paid to the education of the professionals who will support, deploy, develop, and design current and emerging cyberinfrastructure. (National Science Foundation Cyberinfrastructure Council, 2007, p. 38)

Although many articulated the need for educating a workforce that understands the importance of managing and curating data in ways that support broad dissemination, use by others, and preservation beyond the life of its original research project, there has been very little examination of what such a program would contain. We believe that librarians have a role in developing these education programs and will need to actively engage in these discussions.

Gabridge (2009) notes that institutions experience

> a constantly revolving community of students who arrive with . . . uneven skills in data management. . . . Librarian subject liaisons already teach students how to be self-sufficient, independent information consumers. This role can be easily extended to include instruction on data management and planning. (p. 17)

With the respectful elision of "easily," we argue in the remainder of this chapter that there are indeed gaps in the knowledge of current e-researching faculty and students (both as producers and consumers of data) that librarians may address by developing DIL curricula.

Environmental Scan of Related Literacies

For the sake of clarity, it is important to distinguish DIL from other literacies such as data literacy, statistical literacy, and information literacy. Typically, data literacy involves understanding what data mean, including how to read graphs and charts appropriately, draw correct conclusions from data, and recognize when data are being used in misleading or inappropriate ways (Hunt, 2004). Statistical literacy is "the ability to read and interpret summary statistics in the everyday media: in graphs, tables, statements, surveys and studies," (Schield, 2010, p. 135). Schield finds common ground in data, statistical, and information literacy, stating that information literate students must be able to "think critically about concepts, claims, and arguments: to read, interpret and evaluate information." Furthermore, statistically literate students must be able to "think critically about basic descriptive statistics, analyzing, interpreting and evaluating statistics as evidence." Data literate students must "be able to access, assess, manipulate, summarize, and present data." In this way, Schield (2004, p. 8) creates a hierarchy of critical thinking skills: data literacy is a requisite for statistical literacy, and, in turn, statistical literacy is required for information literacy. Stephenson and Caravello (2007) extol the importance of data and statistical literacies as components of information literacy in the social sciences, arguing that the ability to evaluate information essentially requires that one understand the data and statistics used in an information resource.

Qin and D'Ignazio (2010) developed a model, Science Data Literacy, to address the production aspect of data management. SDL refers to "the ability to understand, use, and manage science data" (p. 2) and an SDL education

> serves two different, though related, purposes: one is for students to become e-science data literate so that they can be effective science workers, and the other is for students to become e-science data management professionals. Although there are similarities in information literacy and digital literacy, science data literacy specifically focuses less on literature-based attributes and more on functional ability in data collection, processing, management, evaluation, and use. (p. 3)

Whereas definitions of *data, statistical,* and *information literacy* focus on the consumption and analysis of information, the production of information is often overlooked in literacy instruction. E-research is, by definition, a social process, and contributing to—not just extracting from—the community's knowledge base is crucial. DIL, then, merges the concepts of researcher-as-producer and researcher-as-consumer of data products. It builds upon and reintegrates statistical, information, and science data literacy into an emerging skill set.

Prior Instructional Efforts in Data Information Literacy

Several libraries have developed programs or prototypes to address those needs. The Massachusetts Institute of Technology Libraries created a robust "Manage Your Data" subject guide/tutorial, supplemented by seminars such as Managing Research Data 101 (Graham, McNeill, & Stout, 2011). Both resources include data planning checklists that include the following topics:

- Documentation and metadata
- Security and backups
- Directory structures and naming conventions
- Data sharing and citation
- Data integration
- Good file formats for long-term access
- Best practices for data retention and archiving

The University of Virginia Library created the Scholars' Lab and Research Computing Lab. These projects, collaborative ventures between IT and library departments, created a new service model that included traditional roles for IT (software support and training) and librarians (subject knowledge and departmental interactions), as well as services that bridged those disciplines such as data management and analysis, computational software support, and knowledge of emerging technologies. Librarians from the University of Virginia explained: "We chose to promote the service areas of software support, current awareness, data, collaboration, and research communication. . . . Collectively, we view these as being supportive pieces to the entire research lifecycle, rather than just a single point" (Hunter, Lake, Lee, & Sallans, 2010, p. 341). While the University of Virginia model focused primarily on reference and project-based services, the Scholars' Lab also provided workshops and seminars on special topics in data management such as GIS, Web application development, and text digitization.

The Science Data Literacy project at Syracuse University developed a program "to train students with the knowledge and skills in collecting, processing, managing, evaluating, and using data for scientific inquiry" (Qin & D'Ignazio, 2010, p. 2). As part of the project, Qin developed a credit-bearing course, Science Data Management, covering the fundamentals of scientific data and its description, manipulation, visualization, and curation. Project SDL made its syllabus for the course, with lecture notes, available online (Science Data Literacy Project, 2010).

The Purdue University Libraries are active in this area as well. Two of the authors of this chapter developed a Geoinformatics course with a faculty member in the Department of Earth, Atmospheric, and Planetary Sciences (Miller & Fosmire, 2008). The instructors designed Geoinformatics for beginning graduate and advanced undergraduate students. The course provided a holistic approach to GIS and spatial data, encompassing the full cycle of data, from discovery and acquisition to conversion and manipulation, analysis, and finally visualization, metadata, and re-sharing. The syllabi are online (Miller, 2010).

ASSESSMENTS OF FACULTY AND STUDENT NEEDS IN DATA INFORMATION LITERACY

Like e-research, DIL is not new, but rather compiles expertise and portions of existing research methods, information and other literacies, and computing curricula to offer more holistic, communal, and participatory perspectives and techniques for e-researchers. Just as e-research encourages researchers from a variety of disciplines to collaborate to advance scientific knowledge, disciplinary and library faculty must work together to determine the skill sets that a data literate student should demonstrate and to develop best practices for imparting those skills to the students. Both faculty members and students have perspectives on the necessary data management skill sets in their fields. Grounded in these perspectives are their real-world perceptions and practices and a first-hand knowledge of how one conducts research in his or her respective discipline. Any attempt to define a DIL program must be aligned with current disciplinary practices and cultures if it is to be relevant to and accepted by its intended audience(s). The authors compiled the perspectives of both faculty and students from two different research projects, one based on interviews with faculty members and the other on surveys of students and an analysis of their course work. In the next two sections, the authors report on the DIL priorities articulated by both faculty and students as discovered through our assessments.

Assessment of Faculty Needs: A Reexamination of the Data Curation Profiles Project

In the fall of 2007, the Purdue University Libraries and the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign (UIUC) received funding from the Institute of Museum and Library Services (IMLS) to carry out the Data Curation Profiles (DCP) project. The goals of the DCP project were to better understand the willingness of research faculty to share their data with others—including the conditions necessary for data sharing to take place—and to investigate possible roles for librarians in facilitating data sharing and curation activities.

The investigators interviewed participating faculty at Purdue and UIUC, focusing on three broad areas: the nature and life cycle of one of the data sets generated by researchers; their data management practices; and their needs for making their data available to others and curating their data for long-term access. These interviews resulted in the creation of "data curation profiles," each of which summarized the information gathered from the interview under a common framework that enabled comparisons to be made among the researchers' responses (Witt, Carlson, Brandt, and Cragin, 2009).

The first round of interviews for the DCP project took place at Purdue and UIUC in the summer and early fall of 2008. A convenience sample of faculty participants was recruited from a broad selection of departments in the sciences and engineering on the basis of prior relationships with project personnel or liaison librarians. The semi-structured interviews asked broad, open-ended questions to allow participants to control the direction of the discussion and identify the most important issues related to sharing and curating their data. The investigators then extracted common themes from the transcripts using grounded theory.

The DIL project was predicated in part by the Data Curation Profiles project, which explored the willingness of research faculty to share their data with others—including the conditions necessary for data sharing to take place—and to investigate possible roles for librarians in facilitating data sharing and curation activities.

One of the common themes emerging from the interviews concerned the skills, knowledge, and training needed by graduate students to effectively manage and curate research data. Graduate students actively generated and curated data in support of their own research. Many also oversaw the management of data generated by the entire research group. A few of the faculty noted that their graduate students had been asked to share their data with individuals not affiliated with the research and therefore had to consider similar issues of whether or not to share and what conditions to place on sharing. Typically, faculty reported that graduate students were unprepared to manage or curate the data effectively. While acknowledging that this was an area of concern, they often could not provide adequate guidance or instruction because it was not an area that they knew well or fully understood.

The investigators conducted a second round of interviews in the spring of 2009 to gather additional details from faculty and address gaps from the first interview. Investigators asked the faculty participants at Purdue whether there was a need for a data management and curation training program for graduate students, and what such an educational program should contain. Responses from these second interviews were coded and analyzed with the information from the first interviews. A total of 19 faculty from both schools completed both interviews.

The overwhelming majority of researchers in this study felt that their students needed some form of DIL education.

Faculty Assessment: Results

Generally, faculty in this study expected their graduate students to carry out data management and handling activities. However, the extent of data management responsibilities varied among the faculty interviewed. Some took an active, hands-on role in managing their data with minimal student involvement, while others delegated most data management tasks to their students. Typical responsibilities of graduate students included processing or cleaning the data to enable use or analysis, assuring the quality of the data, compiling data from different sources, and organizing the data for access and use by project personnel. In addition, faculty often considered data management duties as distinct from other research responsibilities.

Analysis of the interviews revealed that the training graduate students received and the training methods varied widely. Some of the researchers taught their graduate students data management tasks, such as how to develop and assign metadata to the data files. Other researchers reported that their graduate students had not received much, if any, formal training in data management and were left to figure things out on their own.

Given the variance in the range of responsibilities and training in data management received by graduate students, it is not surprising that faculty presented a mixed picture in assessing the work of their students in this area. Several faculty expressed frustration with their inability to understand or make use of the data their students had been working on, especially after they graduated. Other comments provided a positive statement of individual students' skills, which they generally acquired without formal training.

The overwhelming majority of researchers in this study felt that their students needed some form of DIL education. However, even in stating a need for such a program, several respondents expressed an uncertainty or a reluctance to teach data management skills to their students themselves. Some faculty expressed a concern about getting too involved in telling students what to do in what should be the students' own work, or in making their work more difficult by introducing new software or formats to work with. Furthermore, although faculty identified the lack of data management skills in their graduate students as a strong concern and described broad themes that should be addressed, they often could not articulate precisely what skills should be taught to remedy the situation.

> *Interviewer:* Is there a need for education in data management or curation for graduate students?
>
> *Faculty:* Absolutely, God yes . . . I mean we're . . . We have the ability to accumulate huge datasets now[,] especially with the new tools that we have.
>
> *Interviewer:* So, what would that education program look like, what would it consist of? What kind of things would be taught?
>
> *Faculty:* Um, I would say, um, and I don't really know actually, just how do you manage data? I mean, where do you put it? Um, how secret does it need to be? Or you know, confidentiality things, ethics, probably um . . . I'm just throwing things out because I hadn't really thought that out very well. (Soil Scientist)

After coding and analysis, several major themes emerged from the faculty's observations of graduate students' deficiencies in data management. These themes are metadata, standardizing documentation processes, maintaining relationships among data, ethics, quality assurance, basic database skills, and preservation.

Metadata

An understanding of metadata and how to apply it were frequently mentioned as areas of need, although the term *metadata* was not used often. More often, researchers said their students needed to know how to annotate and describe data. In most cases, references to "annotations" included both a need to provide information about a data file as well as information about individual components of the data (such as a cell in a spreadsheet). The main reasons for providing metadata include assuring that data can be understood by others (both within the lab and by external audiences), enabling its continued usability over time, and fostering use of the data beyond its original purpose.

Researchers also expressed the need to apply and conform to metadata standards. One researcher stated that not only must students be taught "how to approach the idea of metadata," but also they must develop an awareness of standardized disciplinary ontologies and how to apply them to their own work.

Standardizing Documentation Processes

Standardizing documentation processes is a rather broad theme that applies to both high-level organization as well as to specific, local needs. Researchers frequently reported a need for students to be able to organize data by documenting it in a systematic and logical fashion. Explanations given for the need for rich documentation often extended beyond the immediate needs of the researcher's lab and included such high-level needs as enabling the sharing of data outside the research team, submission to repositories, reuse by external audiences, and preservation beyond the research life cycle. At the local level, this category addresses folder and file naming conventions, data sharing among the lab/project team(s), and assigning staff responsibilities for managing data, communication, and workflow.

Several major themes emerged from the faculty's observations of graduate students' deficiencies in data management: metadata, standardizing documentation processes, maintaining relationships among data, ethics, quality assurance, basic database skills, and preservation.

Researchers expected their graduate students to share responsibility for documenting the lab or project's data, as well as the student's own interactions with it. Documenting data focuses on what needs to be recorded and provided while generating, processing, analyzing, and/or publishing the data to later validate and verify it. This includes such tasks as generating and maintaining data dictionaries, glossaries, or definitions of variables; maintaining lab notebooks or their equivalent; and capturing the provenance of the data. Overall, researchers expressed that students' documentation needs to stand the test of time.

Researchers in this study acknowledged the problem of data documentation, not only for their students but for themselves as well. Difficulties in documenting data contributed to a larger concern: the lack of standardization and consistency in how the data are organized. Faculty repeatedly mentioned that every student employs different methods of documenting his or her data. The lack of standardized and shared data management protocols and practices across a research group often led to a "tower of Babel" situation, where it is difficult to understand what was done, by whom, and for what reason. This further led to difficulties in correlating and relating one data file with another or with the data collection as a whole. The inevitable turnover of students exacerbated this problem. Although most of the researchers in this study required their students to document their work with the data, actual documentation practices followed by the students varied from one to the next. Moreover, they often did not provide complete or detailed enough documentation to enable others to understand their work.

Several researchers suggested creating a standard operating procedure for data formatting and management. One faculty member noted that he created standard operating procedures for most equipment and procedures in the lab and proposed that a similar standard operating procedure be developed for handling and managing his data. When asked to describe an ideal situation for organizing data, several of the faculty members noted the need for students to develop and use a standardized set of best practices.

Maintaining Relationships Among Data: Master Files and Versioning

Many interviewees described the challenge of relating data files to each other. This includes issues related to taking data generated at a particular time or for a particular purpose and enabling its integration with other data to create a new data set. This category also includes the converse action, generating a subset of the data from a larger data set or file.

Several researchers specifically mentioned the need for the creation of an official record of the data (a "master file") to ensure the authority and integrity of this record compared to the working copies of data sets or files created and used for specific purposes by subsets of lab or project personnel.

Many researchers desired that the master file bring a number of disparate files together into a searchable database that engenders question development and helps assure quality control for research. A lack of standardization in data management practices, a high learning curve, and a perceived lack of support for the advanced database utilities and programs required to create such files hindered the ability of researchers to achieve these goals.

Researchers expressed the need to balance the requirements for a particular research project with those for making the data accessible and useful to the larger research community.

This focus on the specific research needs of the student (or the faculty sponsor in some cases) often led to situations in which the faculty member could not retrace the steps taken in processing the data and relate the graduate student's work back to the larger data set to which it belonged.

Akin to these issues of compiling or merging data, researchers frequently brought up versioning as an often neglected but very important concept for students to learn. In this study, researchers clearly reported the importance of maintaining documentation of different versions of their data. They wanted to know which data files were used for what analysis, what file contained the current version being used by the research group, and how these versions differed from each other. However, several faculty members admitted that they themselves had a difficult time in maintaining adequate documentation and struggled to consistently generate the needed documentation in a timely manner.

Ethics

Faculty members in this study identified "data ethics" as another area where most students need assistance. Data ethics includes intellectual property rights and ownership of data, issues of confidentiality/privacy and human subjects, implications and obligations of sharing (or not sharing) data with others (including open access), and assigning attribution and gaining recognition of one's work. Although faculty clearly stated ethics as a needed area of instruction, they generally did not provide much description as to what the curriculum of such an ethics program would include. In one case, the professor tied ethics to an understanding of ownership of data.

Basic Database Skills

Several researchers expressed the expectation that students be able to understand and develop relational databases and use database tools effectively. Frequently, students' lack of basic understanding of database development and usage frustrated the interviewees. However, the expectations of student skills differed among the researchers. A civil engineering professor acknowledged that students needed some basic understanding of relational databases, normalization of data, database tools, and documentation techniques.

Quality Assurance

Researchers expected their graduate students to review or check their data and evaluate its quality. Interviewees mentioned the difficulty of knowing exactly what their students had done to compile and analyze the data. Thus the provenance of the data was unknown. One professor stated that she could not understand the work done by her students.

Quality assurance is in some ways a blend of technical skills (familiarity with equipment), disciplinary knowledge (whether the result is even theoretically possible), and a metacognitive process that requires synthesis on the part of the students. Primarily, quality assurance is the ability to recognize a pattern or consistency in the data. Quality assurance may also facilitate or impede the quality of documentation (annotation/metadata) produced, and the organizational schema, of a given data set.

Faculty repeatedly mentioned that every student employs different methods of documenting his or her data.

Preservation

Researchers expect their students to know how to preserve their data and document the processing of the data. Much like the discussion of metadata, faculty members generally understood the term *preservation* in a broad and loose sense of the word, often conflating it with the simple backing up of files. They were unaware of or unacculturated to preservation from a library perspective, instead focusing much more on the immediate issues and procedures surrounding backing up their data.

Although researchers recognized the need for backups, the methods and timing of performing backups differed considerably among research groups. Some, having learned the hard way through lab disasters, kept geographically dispersed backups. Others relied largely on graduate students to create backups on departmental servers. Still others had no real-time backup system in place. A common problem expressed with backups was tracking versioning.

Faculty Assessment: Lessons Learned

The design of any DIL program requires an understanding of the real-world needs of research groups, where research either progresses or is impeded by their ability to handle data in the ways described here. The faculty supervisors are no doubt acutely aware of the deficiencies in their students' abilities to properly care for their research input and output. The interviews analyzed for this study provide a window into the ground-level interaction with data and in fact become a magnifying glass through which we can spot the deficiencies and gaps in knowledge that a DIL curriculum might target.

Although faculty clearly stated ethics as a needed area of instruction, they generally did not provide much description as to what the curriculum of such an ethics program would include.

We would be remiss, however, not to account for the gaps in faculty responses on data practices, as these interviews also expose faculty interaction with data. Many faculty admitted or otherwise revealed that they themselves lacked expertise or experience with data management, even as they critiqued their students' abilities. We must assume their critiques of their students' (and their own) facility with any or all aspects of data management may be somewhat shallow. In other words, they may not know what they don't know about data management and curation. Therefore, a program based entirely on faculty self-report risks incompleteness, and other viewpoints on what should constitute the objectives for DIL must be taken into account.

As a complement, then, the next section will draw conclusions that help to complete our DIL core objectives from a direct source, a course taught at Purdue University that broached some of these exact topics, including data source evaluation, metadata, databases, preservation, and sharing.

This course allowed us to examine the DIL of students directly and learn from firsthand observation. Because we gained insight into what the students do not know, our own evaluation of student performance in a (classroom-simulated) research environment can serve as an important second front in developing a richer and more comprehensive list of core DIL objectives.

ASSESSMENT OF STUDENT NEEDS

Enrollees in the 2008 and 2010 offerings of the course Geoinformatics provided the sample population for our student assessment. The combined number of students enrolled totaled 27: 12 in 2008 and 15 in 2010. Most of these were students majoring in earth, atmospheric, and planetary sciences, but other majors represented in this course included civil engineering, agricultural and biological engineering, and forestry and natural resources. In 2008, the core course content revolved around a "whodunit" concept. Students were asked to track down, over the course of several laboratory exercises, the location of a fictitious chemical spill by gathering data (both spill data and underlying geology) and using various geospatial analysis and visualization techniques. Student projects provided the rest of the context for learning DIL skills. The 2010 course dropped the "whodunit" mechanism to shift more attention toward a longer, more involved semester project.

To improve and tailor the course, the authors used several methods to probe students' interests, their perceived needs, and their abilities to carry out data management tasks. Among these were a pre-course assessment to inventory the students' technology and information skills and a post-course survey to determine their perceptions of how important different topics were to their research. The instructors also analyzed student semester projects to determine how well they demonstrated mastery of DIL skills.

We administered the pre-course survey in both offerings of Geoinformatics. It contained short-answer questions, mainly probing the students' background in databases, GIS, and programming, such as "What computer programming languages do you know (for example, Fortran, C)?" and "What geospatial software do you use?" The instructors then tailored the course content to address the ability levels of the students. The post-course survey was given only to students in 2008. For each course topic, students rated, on a 5-point Likert scale, the lectures, the lab, and the importance of the topic to the course and to their own research. They also recommended improvements to the course labs.

These instruments probed students' attitudes toward various topics related to DIL. However, there were disconnects between student perceptions and their performance. As Grimes and Boening (2001), among others, have observed, novices tend to overstate their expertise, in large part because they don't know what they don't know. To provide a check of the degree to which the students actually demonstrated DIL skills, the instructors analyzed the students' projects. The project required students to identify a problem or research question with geospatial components and use the skills and techniques discussed in class to advance that research and present the results of their work. It required both the acquisition of original data and the use of external, "published" data. And it involved analysis and visualization and required a summary of how the research answered or at least clarified the question or problem. It should be noted that this course did not teach research methods or disciplinary content knowledge: the students needed to get content assistance from their own research group.

Student Assessment: Results

Although in both course offerings several students indicated they had a rudimentary understanding of the technologies identified in the pre-course survey, none indicated that they felt able to command the tools to accomplish their own ideas and solutions. The survey, in fact, revealed low levels of exposure to most of the course content. Students reported little experience with GIS at all, and the experience they had was limited to a handful of data types and rather turnkey operations. Both offerings of the course required the instructors to cover fundamental concepts before moving on to a higher order agenda. These lessons included an introduction to databases and data formats, basic use of GIS and GPS tools, rudimentary visualization and analysis techniques, and metadata and presentation skills. The instructors decided

TABLE 1.1 *Results of the 2008 Post-Course Survey, on a 5-Point Likert Scale, of the Importance of Different Topics to the Course and to the Students' Research (n = 5)*

Topic	Importance to Course	Importance to Research
Databases	4.8	5.0
Data formats	5.0	4.8
Data gateways/portals	4.6	4.6
Introduction to GIS	4.8	4.8
GIS analysis	5.0	5.0
GIS data conversion	5.0	5.0
Workflow management	4.6	4.6
Metadata	5.0	5.0
Statistics	4.6	4.4
GPS	4.6	4.2
Data visualization	5.0	5.0
Ontologies	4.0	3.6
Data preservation	4.2	4.2

against using some technologies because, for example, students had no experience working in Unix/Linux systems or using low-level programming languages.

Students indicated a high level of interest in all the topics covered in the class and had an appreciation for DIL skills. In the standard end-of-course evaluations to which all students (n = 12) responded, the course received an overall rating of 4.8 out of 5.0, and several students remarked that after taking the course they finally understood what they were doing and now could contribute new procedures for analyzing data to their research groups. Of the 12 enrolled students, 5 completed the 2008 post-course survey, with the results summarized in Table 1.1.

The high level of interest in basic topics such as data formats and an introduction to databases indicate the relative lack of preparation in the core technology skills necessary to work in an e-research environment. All but one topic (ontologies) received a rating of at least 4.0 (very important) as important to research.

In addition to extracting information from course surveys, the instructors also carefully examined students' completed course work to determine which concepts, skills, or ideas students still lacked. For example, the authors found that most students had ready access to the primary data used by their research groups and that these data often formed the basis for their semester project analysis. A focus of the course was on students' abilities to identify and synthesize supplementary data, such topographic, political, or land-use data to overlay on the data collected by the research group. Analysis of the student semester projects indicated that students indeed could find, identify, and incorporate external data sources into their analysis and/or visualization.

However, the analysis of the students' semester projects from both years revealed recurring shortcomings. While students did apply external data appropriately to their work, frequently these data were not cited properly. Although students correctly documented traditional published literature, they might not consider data to be a valid, citable scholarly source or have a clear understanding of how to cite a data set.

Students also struggled to fully comprehend the importance and complexity of data sharing, though the course was geared toward pushing this point explicitly. The following issues appeared multiple times over the two separate semesters:

1. *Preservation/archiving.* The students' final task in 2008 was to submit their data to the GEON Portal (www.geongrid.org) for safekeeping and redistribution. In 2010, GEON was merely a suggestion and students were encouraged to identify a repository in their domain to which they could submit their project data. Although many students attempted these submissions in good faith (despite some technical difficulties with GEON both years), several students shared the sentiments of one in particular, who argued that a department-run website that "everybody in the [domain] community knows about" was a better ultimate destination for their data than any more formal data repository.
2. *Metadata.* Although the time allocated for metadata was limited, the instructors managed to include the concepts of schema, authoritative terminology, XML, indexing, and searchability. Each course offering had a metadata unit during which instructors introduced students to several proper examples of metadata. The students then completed a lab in which they wrote their own simple metadata documents. While some students did write good accompanying metadata for their final project materials, most did not. One deficit seemed to arise from students creating metadata from the perspective of "how I did it," rather than striving to make the data more discoverable by the next scientist down the line.
3. *The technologies and workflows of data sharing.* Students (despite instructor warnings) expected to accomplish far more than they were able during a single semester. This was an outcome of students' expectations that, once analyzed, their data could be visualized fairly easily and shared online. The complexity of building data-driven, interactive Web applications was not apparent until it was too late.

DISCUSSION

The authors sought to triangulate the needs related to DIL through interviews with research faculty and analysis of the results of our own geoinformatics-themed DIL course. We found a substantial amount of overlap between the needs identified: databases, metadata, data sharing, preservation and curation of data, and formatting and documentation of data.

The assessments also uncovered differences that were more clearly a focus for one group than the other. For example, the interviews with faculty members primarily focused on data they created themselves, while a significant portion of the Geoinformatics course involved locating data from external sources. An analysis of course work showed that students needed to learn "the basics" of much of information technology, even before broaching data issues. Additionally, to manipulate the data, students had to learn how to use analysis and visualization tools, use workflow management tools, and develop a minimum computing background to take advantage of the available cyberinfrastructure. On the other hand, the production- and publication-focused faculty researchers described the need for data curation and management, such as good versioning, documentation, and quality assurance and the merging of data. In addition, the faculty surfaced the concept of data ethics: when to share data, who owns data, and how to appropriately acknowledge data. To that extent, these two investigations provide complementary information about perceived DIL needs.

We have argued that an understanding of either faculty or student practices and needs alone is insufficient to develop the foundational objectives necessary for a DIL program. Instead, both faculty and student perspectives must be understood and analyzed in tandem to inform a more complete understanding of what is needed in DIL. We now reintroduce another foundational component toward developing objectives for a DIL program: the perspective of the librarian. The organization, description, dissemination, curation, and preservation of information resources, which increasingly includes research data, are the hallmark of librarians. Although DIL must be grounded in real-world needs as expressed by students and faculty, the librarian brings the broader perspective and a connection to the larger "information ecology" that exists beyond the single research project or classroom. This connection can ensure that holistic best practices strengthen current practices.

Comparison of Data Information Literacy With ACRL IL Standards

To help articulate and ground our core DIL objectives, we found it useful to examine these topics through the prism of the ACRL (Association of College and Research Libraries) information literacy competency standards (2000), which have been widely adopted by many institutions and accreditation agencies and guide many library instruction initiatives. To that end, the next section first lists the ACRL standards, then briefly examines each standard for its relevance to these DIL objectives.

One readily identifiable gap in applying the ACRL information literacy standards to a DIL program is the difference in focus. The ACRL standards focus on educating information consumers—people seeking information to satisfy an information need. Although faculty and students do consume research data, our analysis of faculty and students indicates a strong need to address their roles as data producers as well. Therefore, the underlying objectives for any DIL program need to accommodate both the data producer's viewpoint as well as that of the data consumer.

The ACRL standards state that information literate individuals are able to:

1. Determine the extent of information need.
2. Access needed information efficiently and effectively.
3. Evaluate information and its sources critically and incorporate selected information into one's knowledge base and value system.

4. Use information effectively to accomplish a specific purpose.
5. Understand the economic, legal, and social issues surrounding the use of information, and access and use information ethically and legally. (ACRL, 2000, pp. 2–3)

ACRL Standard One: Determining Nature and Extent of Information Need

When gathering information, one often skips the research question formulation stage that is the foundation of the information search process (Kuhlthau, 2004). However, without articulating and understanding the question deeply, one cannot arrive at a relevant answer. The instructors addressed this concept in the semester project for the Geoinformatics course—for example, the overall assignment asked students to identify their research question and determine what data they needed to address that question. In the case of geospatial data, students needed to determine whether to use raster or vector data, because each type has its own strengths and weaknesses for analysis and presentation. Thus, the authors' curricular topic of *databases and data formats* fit best into this competency standard, as it is fundamental to understanding the nature of the information needed. In fact, Standard One already explicitly addresses data, stating that a student "realizes that information may need to be constructed with raw data from primary sources."

From the data producer's standpoint, identifying the nature and extent of the potential needs and uses of the data being generated provides the foundation for effectively sharing, reusing, curating, and preserving data. The cultural practices and norms of the producer's discipline, including being aware of any existing community resources, standards, or tools, inform these data functions.

ACRL Standard Two: Access Needed Information Efficiently and Effectively

Students need to consult common disciplinary and general data repositories as well as understand the formats and services through which data can be accessed in order to access information efficiently and effectively. In the Geoinformatics course, students investigated several data sources and were required to use external data extensively to supplement their own data. In addition to finding data relevant to their research question, the variety and complexity of data formats made the process of locating supplementary data challenging for the students. Several students needed assistance converting data from one format to another and understanding how to merge data sets with different resolutions or timescales.

Standard Two addresses these issues, as an information literate student "extracts, records, and manages the information and its sources," including using "various technologies to manage information selected and organized" (ACRL, 2000, pp. 10–11). Not only will DIL students need to know where data exist, but they also must harvest, convert, possibly merge, and ultimately feed it into analysis or visualization tools that may or may not require still other formats. Although a direct graft of classic information literacy competency standards to DIL would focus on the process of bringing data *into* one's research, as the faculty interviews revealed, these concepts are similar for publishing data to the world. Thus, DIL concepts related to this competency standard include data repositories, data conversion, data organization, sharing data, and interoperability.

ACRL Standard Three: Evaluate Information Critically

When evaluating data, students understand and critically evaluate the source. Students must determine whether the research group that provided the data is known to be reliable and/or if the data repository or its members provide a level of quality control for its content. Users also need to evaluate the data for relevancy and compatibility with their own research. As part of the quality assurance component of data evaluation, students need to evaluate associated metadata. Among other attributes, metadata specifies the details of the experiment or data product, including the following: the conditions under which the data were collected or created; the apparatus or procedures used to generate the data; distribution information and access rights; and spatial and temporal resolution, units, and parent sources. It is a vital tool in the evaluation of the quality and authority of the resource. While the ACRL standards would approach this from a data user perspective, the faculty interviewed made it clear that data producers need to provide quality assurance for data and metadata as well.

ACRL Standard Four: Use Information to Accomplish a Specific Purpose

In this standard, students carry out a project and need to "communicate the product or performance effectively to others." As such, students should use a format and employ information technologies that best support the purpose of the work. Here, in the expansive verb "communicate" and phrase "appropriate information technologies," one can assume the concepts of data sharing, reuse, and curation, as well as connections to analysis and visualization tools.

In addition, this standard includes the application of information toward the planning and creation of a product, revising the development process as appropriate along the way. These components parallel the statements made by faculty on the importance of documenting the processes used to develop research data (the "product" in this case). Researchers also identified the careful management and organization of data as essential in enabling its eventual transfer "from their original locations and formats to a new context" (as stated in Standard Four) for internal use by others in the project, or for reuse by others.

ACRL Standard Five: Understand Economic, Legal, and Social Issues and Use Information Ethically

Data ethics are certainly an important component of a well-rounded DIL program, especially since intellectual property issues concerning data are much less defined than, for example, those concerning traditional textual works. Students need to not only determine when and how to share data, which varies among disciplines, but also document their own sources of data. We found students struggled with the latter in the Geoinformatics course, as exhibited primarily by a failure to acknowledge those parties responsible for the data they consumed and reused. The ethical issues surrounding students as data producers and publishers, a concern raised by research faculty, appears to be entirely absent from the ACRL standards and would be a largely novel component of a DIL curriculum.

CORE COMPETENCIES FOR DATA INFORMATION LITERACY

With information gleaned from the faculty interviews, the Geoinformatics course, and the ACRL Information Literacy Competency Standards, the authors propose the following educational objectives for a DIL program. Disciplinary implementation of these outcomes would naturally incorporate technologies or techniques specific to that discipline. The following are the proposed core competencies, organized by major theme.

Introduction to databases and data formats. Understands the concept of relational databases and how to query those databases, and becomes familiar with standard data formats and types for the discipline. Understands which formats and data types are appropriate for different research questions.

Discovery and acquisition of data. Locates and utilizes disciplinary data repositories. Identifies appropriate data sources and can import data and convert it when necessary, so that it can be used by downstream processing tools.

Data management and organization. Understands the life cycle of data, develops DMPs, and records the relationship of subsets or processed data to the original data sets. Creates standard operating procedures for data management and documentation.

Data conversion and interoperability. Proficient in migrating data from one format to another. Understands the risks and potential loss or corruption of information caused by changing data formats. Understands the benefits of making data available in standard formats to facilitate downstream use.

Quality assurance. Recognizes and resolves any apparent artifacts, incompletion, or corruption of data sets. Utilizes metadata to anticipate potential problems with data sets.

Metadata. Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by members of the work group and external users. Develops the ability to read and interpret metadata from external disciplinary sources. Understands the structure and purpose of ontologies in facilitating better sharing of data.

Data curation and reuse. Recognizes that data may have value beyond the original purpose, (i.e., to validate research or for use by others). Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community-driven e-research. Recognizes that data must be prepared for its eventual curation at its creation and throughout its life cycle. Articulates the planning and actions needed to enable data curation.

Cultures of practice. Recognizes the practices, values, and norms of the chosen field, discipline, or subdiscipline as they relate to managing, sharing, curating, and preserving data. Recognizes relevant data standards of a field (metadata, quality, formatting, and so forth) and understands how these standards are applied.

Data preservation. Recognizes the benefits and costs of data preservation. Understands the technology, resource, and organizational components of preserving data. Utilizes best practices in preservation appropriate to the value and reproducibility of data.

Data analysis. Becomes familiar with the basic analysis tools of the discipline. Uses appropriate workflow management tools to automate repetitive analysis of data.

Data visualization. Proficiently uses basic visualization tools of the discipline and avoids misleading or ambiguous representations. Understands the advantages of different types of visualization—for example, maps, graphs, animations, or videos—for different purposes.

Ethics, including citation of data. Understands intellectual property, privacy, and confidentiality issues and the ethos of the discipline related to sharing data. Appropriately acknowledges data from external sources.

The authors compared the DIL core objectives with the course syllabus from the Science Data Literacy curriculum of Qin and D'Ignazio (2010) and found similarities between the two formulations. The chief difference appeared to be the depth of treatment of different topics. While the SDL course concentrated on metadata, for example, our approach focuses as much on the consumption of data (tools) as it does on documenting and annotating data. The Geoinformatics course perhaps had too little coverage of metadata, but we found that students and faculty both needed as much help with data manipulation as they did with enhancing the documentation of their data. Naturally, instructors must balance using tools and creating interoperable infrastructure in teaching this type of course.

We have alluded to the notion that a comprehensive DIL program may not be entirely the responsibility of librarians. However, librarians who have the skills to teach database management and data analysis, for example, could teach those concepts. Indeed, learning those skills supports the educational mission of the university. However, the authors recommend collaboration between disciplinary faculty and librarians as the best practice for teaching DIL skills. DIL needs to be grounded in the culture of the discipline in which it is embedded, and also imbued with the greater, communal perspective possessed by a librarian.

CONCLUSION

> Thirty years ago, it was good laboratory practice [that] you had a bound paper manual, you took good notes, you took fifteen or twenty data points, maybe a hundred, and you had a nice little lab book. But we've scaled now to getting this mega amount of information and we haven't scaled our laboratory management practices. . . . It makes perfect sense to me that . . . you get this [data management skills] in people's consciousness, make them aware it's a problem early on in their careers as graduate students, before they go on and do all the other things and get too set in their ways. . . . And . . . that takes a fair amount of education . . . and training. (Civil Engineer)

The authors uncovered a growing need among research faculty and students for DIL skills. As a result, the authors brought together data from different audiences to propose a suite of core DIL skills that future e-researchers need to fully actualize the promise of the evolving cyberinfrastructure.

DIL represents an opportunity to expand information literacy from the library into the laboratory. In much the same way that libraries' information literacy programs have gone beyond the "one-size-fits-all" approach, librarians will need to go beyond a "one-size-fits-all" approach to data management and curation literacy. The Data Curation Profiles project (Cragin, Palmer, Carlson, & Witt, 2010; Witt, Carlson, Brandt, & Cragin, 2009) indicated that different disciplines and subdisciplines have different norms and practices for conducting their research and working with data. These differences are manifest in the myriad ways they manage (or don't manage), share (or don't share), curate, and preserve their research data. While we have provided a general summary of common themes from these interviews, we understand that any DIL program focused on a specific discipline needs to identify, incorporate, and address these specific differences in the curriculum. Models will help ascertain the educational needs of subdisciplines with regard to their data and then design DIL programs that will address these needs. These results serve to start a conversation and propose general concepts, rather than to provide a final, detailed curriculum.

Upon examination of the ACRL standards for information literacy, it is clear that DIL falls within the scope of standard library practice. The conceptual overlap between the ACRL standards and the DIL objectives indicates that these skills are very much aligned with librarianship. With some exceptions, the ACRL standards are written generally enough to accommodate DIL skills, and indeed the standards do have several specific outcomes related to data. Still, given the ballooning interest in data management for e-research, the new iteration of the standards should incorporate more data-related outcomes, especially from the perspective of the user as publisher of information.

Additional research should be done to identify the skill sets librarians need to support the DIL objectives, either as stated here or as they develop in practice. This will not only speed the development of DIL curricula, but also push the library community to work to adapt the collective DIL practice to trends in e-research.

ACKNOWLEDGMENTS

The authors wish to acknowledge support from the Institute of Museum and Library Services, Grant #IMLS LG-06-07-0032-07, for the faculty assessment portion of this chapter, and the work of Anupma Prakash, University of Alaska Fairbanks, whose Geoinformatics course provided insight into the development of our own.

NOTE

This chapter was originally published in 2011 (Carlson, J., Fosmire, M., Miller, C., & Sapp Nelson, M. [2011]. Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy, 11*[2], 629–657. http://dx.doi.org/10.1353/pla.2011.0022) as an article describing the needs assessment research done by the Purdue University Libraries. The authors' articulation of DIL competencies and how they relate to information literacy served as the springboard for the Data Information Literacy project. The editors feel that this work serves as an important milestone in DIL history, and as such it is reprinted here with minimal revision.

REFERENCES

Association of College and Research Libraries (ACRL). (2000). *Information literacy competency standards for higher education.* Retrieved from http://www.ala.org/acrl/files/standards/standards.pdf

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., . . . Wright, M. H. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.* Retrieved from National Science Foundation website: http://www.nsf.gov/cise/sci/reports/atkins.pdf

Cragin, M. H., Palmer, C., Carlson, J., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A, 368*(1926), 4023–4038. http://dx.doi.org/10.1098/rsta.2010.0165

Friedlander, A., & Adler, P. (2006). *To stand the test of time: Long-term stewardship of digital data sets in science and engineering.* Retrieved from Association of Research Libraries website: http://www.arl.org/storage/documents/publications/digital-data-report-2006.pdf

Gabridge, T. (2009). The last mile: Liaison roles in curating science and engineering research data. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC, 265,* 15–21. Retrieved from http://old.arl.org/bm~doc/rli-265-gabridge.pdf

Gold, A. K. (2007). Cyberinfrastructure, data, and libraries, part 2: Libraries and the data challenge: Roles and actions for libraries. *D-Lib Magazine, 13*(9/10). http://dx.doi.org/10.1045/september2007-gold-pt2

Graham, A., McNeill, K., & Stout, A. (2011). *Managing research data 101.* Retrieved from MIT Libraries website: http://libraries.mit.edu/guides/subjects/data-management/Managing_Research_Data_101_IAP_2011.pdf

Gray, J., Szalay, A. S., Thakar, A. R., Stoughton, C., & vandenBerg, J. (2002). *Online scientific data curation, publication, and archiving* [Arxiv preprint cs/0208012]. Retrieved from http://arxiv.org/abs/cs/0208012

Grimes, D. J., & Boening, C. H. (2001). Worries with the Web: A look at student use of Web resources. *College & Research Libraries, 62*(1), 11–22.

HathiTrust Research Center. (n.d.). Our Research Center. Retrieved from http://www.hathitrust.org/htrc/

Hey, T., & Hey, J. (2006). E-science and its implications for the library community. *Library Hi Tech, 24*(4), 515–528. http://dx.doi.org/10.1108/07378830610715383

Hunter, C., Lake, S., Lee, C., & Sallans, A. (2010). A case study in the evolution of digital services for science and engineering libraries. *Journal of Library Administration, 50*(4), 335–347. http://dx.doi.org/10.1080/01930821003667005

Hunt, K. (2004). The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly* (Summer/Fall), 12–15. Retrieved from http://www.iassistdata.org/downloads/iqvol282_3hunt.pdf

Jones, E., Lougee, W., Rambo, N., & Celeste, E. (2008). *E-science talking points for ARL deans and directors.* Retrieved from Association of Research Libraries website: http://www.arl.org/storage/documents/publications/e-science-talking-points.pdf

Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services* (2nd ed.). Westport, CT: Libraries Unlimited.

Lougee, W., Choudhury, S., Gold, A., Humphrey, C., Humphreys, B., Luce, R., . . ., & Young, P. (2007). *Agenda for developing e-science in research libraries.* Retrieved from Association of Research Libraries website: http://old.arl.org/bm~doc/ARL_EScience_final.pdf

Miller, C. C. (2010). Geoinformatics course site. Retrieved from Purdue University Libraries website: http://oldsite.lib.purdue.edu/gis/geoinformatics

Miller, C., & Fosmire, M. (2008). *Creating a culture of data integration and interoperability: Librarians collaborate on a geoinformatics course.* Paper presented at the International Association of University Libraries 29th Annual IATUL Conference, Auckland, New Zealand. Retrieved from http://www.iatul.org/conferences/pastconferences/2008proceedings.asp

National Research Council. (2009). *Ensuring the integrity, accessibility, and stewardship of research data in the digital age* [The National Academies Press Openbook version]. Retrieved from http://www.nap.edu/catalog.php?record_id=12615

National Science Board. (2005). *Long-lived digital data collections enabling research and education in the 21st century.* Retrieved from http://www.nsf.gov/pubs/2005/nsb0540/

National Science Foundation Cyberinfrastructure

Council. (2007). *Cyberinfrastructure vision for 21st century discovery.* Retrieved from http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

National Science Foundation. (2011). Chapter II—Proposal preparation instructions. In *Grant proposal guide.* Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp

National Virtual Observatory. (2010). What is the NVO? Retrieved from http://www.us-vo.org/what.cfm

Ogburn, J. L. (2010). The imperative for data curation. *portal: Libraries and the Academy, 10*(2), 241–246. http://dx.doi.org/10.1353/pla.0.0100

Qin, J., & D'Ignazio, J. (2010, June). *Lessons learned from a two-year experience in science data literacy education.* Paper presented at the International Association of Scientific and Technological University Libraries, 31st Annual Conference, West Lafayette, IN. Retrieved from http://docs.lib.purdue.edu/iatul2010/conf/day2/5

Schield, M. (2004). Information literacy, statistical literacy and data literacy. *IASSIST Quarterly, 28*(2/3), 6–11.

Schield, M. (2010). Assessing statistical literacy: Take CARE. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 133–152). Chichester, England: John Wiley & Sons.

Science Data Literacy Project (2010). Syllabus. Retrieved from http://sdl.syr.edu/?page_id=23

Stephenson, E., & Caravello, P. S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *RSR: Reference Services Review, 35*(4), 525–540. http://dx.doi.org/10.1108/00907320710838354

Walters, T. O. (2009). Data curation program development in U.S. universities: The Georgia Institute of Technology example. *International Journal of Digital Curation, 4*(3), 83–92. http://dx.doi.org/10.2218/ijdc.v4i3.116

Welshons, M. (Ed.). (2006). *Our cultural commonwealth: The report of the American Council of Learned Societies Commission on cyberinfrastructure for the humanities and social sciences.* Washington, DC: American Council of Learned Societies.

Witt, M., Carlson, J., Brandt, D. S., & Cragin, M. H. (2009). Constructing data curation profiles. *International Journal of Digital Curation, 4*(3), 93–103. http://dx.doi.org/10.2218/ijdc.v4i3.117