International Association of Scientific and
Technological University Libraries, 31st Annual
Conference

31st Annual IATUL Conference

Jun 22nd, 10:30 AM - 11:30 AM

Lessons learned from a two-year experience in science data literacy education

Jian Qin
Syracuse University, jqin@syr.edu

John D'Ignazio
Syracuse University, jadignaz@syr.edu

Follow this and additional works at: http://docs.lib.purdue.edu/iatul2010

LESSONS LEARNED FROM A TWO-YEAR EXPERIENCE IN SCIENCE DATA LITERACY EDUCATION

Jian Qin & John D'Ignazio

School of Information Studies

Syracuse University

Syracuse, NY 13244, USA

{jqin, jadignaz}@syr.edu

Abstract

In spring 2008 and 2009, we offered a course "Scientific Data Management" to undergraduate and graduate students from science and technology majors with support from an NSF CCLI (award #0633447) grant. We actively advertised the course on campus and conducted outreach efforts to individual classes and faculty representing a wide range of science and technology disciplines. Each time the course was offered, we conducted pre- and post-course surveys to assess the instructional effectiveness and learning outcomes. This paper will describe our experience as well as lessons learned in three areas: science faculty perceptions on data management, effect of science curriculum structures on data literacy education, and changes of students' perceptions and aptitudes, and discuss how the lessons learned will help shape our e-science Librarianship curriculum development, which is being funded by IMLS.

Introduction

The Science Data Literacy (SDL) Project was conducted at Syracuse University's (SU) School of Information Studies (iSchool) from May 2007 to May 2009. The project's funding came from a U.S. National Science Foundation (NSF) Course, Curriculum, and Laboratory Improvement grant (Award #0633447) and allowed the Principal Investigator (PI), Jian Qin, an associate professor at the iSchool, and research assistant John D'Ignazio, to execute an ambitious plan to accomplish a two-part, outcome-based goal. The goal was to first assess the needs for scientific data literacy education by engaging with the science, technical, engineering, and mathematics (STEM) faculty about their attitudes and practices and then use that information to inform the design of an undergraduate course. Learning strategies, techniques, and materials on science data and their lifecycle were created according to a designed instructional framework for a 14-week course, that was then evaluated using outcome-based evidence regarding student uptake of science data literacy concepts and progress developing related projects.

Reflecting back on the SDL project's efforts, successes, and failures provides a means to comment on and generalize about this fairly new problem space faced not only by STEM faculty and researchers, but also by students who wish to participate in cutting-edge investigations and the legions of science and technical librarians and information specialists who seek to support the research efforts occurring across the departments, domains, and disciplines at their institutions. Our experience in this area is instructive in the manner in which both librarians and educators can deliver data literacy support and education to students and work to address this issue in this complex academic and research environment.

In particular, the contribution of the SDL project was to:

1) define the need for and work to enhance "data literacy."
2) develop or apply pre-existing strategies to raise STEM faculty awareness about this topic.
3) develop or apply pre-existing strategies for outcome-based learning and assessment to prepare students for an environment of changing scientific work practice.
4) assess the efficacy of one approach to enhancing data literacy in a campus environment.

Defining Data Literacy

Science data literacy (SDL) emphasizes the ability to understand, use, and manage science data. The 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. In either case, the focal point is to train students with the knowledge and skills in collecting, processing, managing, evaluating, and using data for scientific inquiry. 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 (Table 1). This emphasis on operational skills coincides with the practice-based production, operation, and use of digital datasets during scientific research.

Table 1. Differences between different types of literacy
[ACRL, 2010; Finn, 2004; Qin & D'Ignazio, in review]

Characteristics	Information literacy	Digital literacy	Science data literacy
Ability and Skills	Finding, retrieving, analyzing, and using information	Locate, organize, understand, evaluate, and *create* information using digital technology	Collecting, processing, managing, evaluating, and using data for scientific inquiry
Object / Resource	Information	Information	Data

In the middle of the last decade, the SDL Project PI recognized two main gaps that were occurring in the area of science data literacy where the iSchool could make a significant contribution. The first gap was between what science needs from the next generation workforce to manage a new kind of information resource, given the development and increasing prevalence of cyberinfrastructure, a combination of hardware and software, that results in the persistence and accumulation of data files occurring as digital resources. This has become known as eScience, and although associated with "big science" efforts at large, federally funded and centralized research facilities for large-scale investigations, is also occurring at the "little science" level of individual scientists and small research groups in labs and out in the field.

The second gap was how much the STEM faculty had become aware of their own shifting expectations of the skill set required by individuals to assist them in both research conduct, as incoming graduate students and research assistants, and information support from institutions like the academic library or specialized information centers. Signs of both gaps were a lack of response reflected in the science curricula to accommodate the changing e-science/e-research environment, and few information workers available to step in and engage in effective resource management of their science data.

Addressing the first gap required learning about the changing nature of science to design an effective course offering that trains students to function in the new environment. STEM fields vary in terms of data scope, type, format, subject, and size. Even within the same field on a macro scale, and also within an individual scientist or lab group on a micro scale, the varying nature of purposes and kinds of research can result in very different sets of data attributes. The creation of a course to raise student literacy about data in this complex environment centered on student immersion in case studies and activity on related group projects. These case studies and projects concentrated on the nature of this digital resource and the role it plays in science investigations. This strategy developed through an increase in the project staff's awareness about the science environment in which these digital resources through designing a survey and administering this vehicle as a census of science data management practices and attitudes at all STEM departments in both SU and SUNY-ESF departments. Detail about both the design and conduct of the SDL SU/ESF Faculty Survey, and the resulting course design it influenced is available in the article "Teaching Metadata as a Key Ingredient in Developing Science Data Literacy" in the *Journal of Library Metadata* [Qin, J. & J. D'Ignazio, under review].

The remaining portion of this article covers the response to the second gap about faculty awareness regarding one means to address their data management problem, lessons learned in addressing this gap, and lessons learned in assessing student progress through two instances of what became known as IST 400/600 Science Data Management.

Lessons Learned 1: Raising Awareness

To surmount the problem of a low awareness about data literacy we engaged in a comprehensive communications effort that sought to both 1) raise awareness about this issue and to 2) raise awareness and participation in our response in the form of our planned course offering. The participation of an interdisciplinary advisory board of professors representing several science, technology, and education disciplines, as well as a range of professional experience and position, helped us come up with or adjust strategies to reach faculty and students.

We tried various venues and channels to promote the special topics course not listed as part of any program or specialty. The course had, however, an open enrollment policy making it available to any student interested from across Syracuse University or even from the co-located campus of the State University of New York, Environmental Science & Forestry (SUNY-ESF). The outreach efforts included an array of both "push" strategies designed to take the message to both students and faculty where they worked in their various departments and campus buildings, as well as "pull" strategies designed to get their attention to the SDL project and the iSchool.

- Pull Strategies
  - an expert panel who presented on science data management and use in the real world,
  - a website containing course information, project events and highlights, and science data management and curation job listings.
- Push Strategies:

- flyers posted in campus buildings and distributed to advisory board members to hand out in their home departments,
- email notice to science, technical, engineering, and mathematics (STEM) faculty members from both campuses targeted as possible participants in a survey on their data management practices and attitudes,
- a campus mailing of flyers to survey participants, department chairs, and student advisors in STEM departments, and
- visits to science classes for a short presentation and Q&A session.

An Example "Pull" Strategy

The iSchool had fairly recently occupied a renovated, more centrally located campus building so the SDL project staff thought to host a high-profile event in this venue to draw attention to digital data and its impact on science and technology in both academic and business environments. We held an early evening panel discussion, The Big Bang of Science Data—New Opportunities/New Skills, as the first major communication effort. The panel featured Dr. Robert J. Corona, vice president of clinical, medical, and scientific affairs and chief medical officer at a nearby medical device manufacturer, Welch Allyn Inc., and Paul Gandel, who was serving at that time as SU's Chief Information Officer (CIO).

Corona spoke from his position as a licensed physician and surgeon who is board-certified in three medical specialties and from his corporate management position responsible for biostatistics, new product development, and clinical testing. He spoke about the data management strategies being developed and used at Welch Allyn to design and certify the latest medical devices used in surgery and rehabilitation to gather and present key information to medical teams during comprehensive patient care. Gandel spoke from his capacity as CIO, responsible for all aspects of information technology and information services at SU. His comments focused on advances in the IT side of data management to support cutting-edge research on campus (lately known as cyberinfrastructure), why this matters, and how students can prepare themselves for career possibilities in this area. The event was, except for iSchool graduate students, rather poorly attended, and while motivating to the attendees and the SDL staff regarding the scale and reach of the data management issue in the increasingly digital environment of academic and corporate research and development, proved to be a largely unsuccessful method to raise campus awareness for either students or faculty.

An Example "Push" Strategy

After more than a year into the project, the SDL Project staff processed the still very low level of awareness of data management issues and the iSchool response to them among STEM faculty and students. So we made plans in summer 2008 to increase our outreach efforts during the next fall semester with an aggressive "push" style campaign. The strategy included identifying potential classes and faculty members that had a need for science data literacy as well as management from course listings in the SU and SUNY-ESF catalog and online schedule, as well as individual department websites. A library science graduate student with a zoology background joined the project and helped conduct this outreach activity. The student assistant represented to STEM students her perspective as a recent graduate with a baccalaureate degree who had participated in a zoological research project that resulted in a commitment to working in the science information field. After this testimonial in 16 sections of science and

technology courses, the project research assistant concluded the 10-minute presentation with general information about science data management, the need for people with skills in this area, and details about the course. The classes visited, where staff also distributed course brochures are presented in Table 2.

Table 2. Classes visited in fall 2008 with estimated student attendance

Course Code	Course Name	Sections	Approximate Number Students Attending	Total
EFB307	Principles of Genetics	1	75	75
GOL314	Mineralogy	1	18	18
IST 769	Advanced Database	1	20	20
IST 659	Data Admin. Concepts and Database Mgmt.	2	20	40
IST 511	Intro. to the Library and Info. Profession	1	30	30
BIO121	General Biology	2	200	400
BEN 467/667	Advanced Biomechanics Lab	2	10	20
PHY 211	General Physics I	1	125	125
PHY 451	Problems of Contemporary Physics	1	10	10
PHY 300	Waves and Optics	1	18	18
GEO 595	Geography & The Internet	1	5	5
PSY 322	Cognitive Psychology	1	30	30
PSY 252	Statistical Methods II	1	35	35
Total				**826**

Awareness-raising Results

One large measure of the overall results from the communications effort could be measured in student enrollment for the two instances of the science data management course. The yield from these combined efforts was rather low: we managed to retain ten students for spring 2008 and eight students for spring 2009. To successfully offer the course, enrollment was opened to graduate students who were interested in this emerging field. In spring 2009, five students from the iSchool's Cyberinfrastructure Facilitator program, also funded by an NSF grant, were required to take the science data management course as one of their core requirements, which allowed the course to be offered a second year in a row. In fact, all but two students were enrolled in iSchool undergraduate or graduate programs, which do count as technology-oriented under the STEM umbrella. A science orientation was represented through the undergraduate degrees obtained by a number of iSchool graduate students taking the course, including meteorology, zoology, environmental science, biology, and psychology.

Despite this somewhat poor performance in getting students to take the course, it may be useful to indicate that push strategies were more effective and influential in the course of the project than pull strategies. Getting out of the iSchool and interacting with the broader STEM faculty enabled the project staff to experience directly comments from both professors and students as well as picking up an environmental awareness from the structure of classrooms, lectures, laboratories, and notices posted on hallway bulletin boards. Two examples are worth noting. The first is that the SDL SU/ESF Faculty survey turned into a significant push strategy in its own right, despite the lack of that particular intention. The survey included a couple months of

interaction with the roster of STEM faculty at two campuses, starting with the analysis to decide which departments fall into this category. Multiple emails to encourage participation generated personal responses from faculty with clarifying or correcting comments, and the survey's open-ended responses generated meaningful statements to analyze. In addition, the final amount of participants, nearly a third of the eligible total of about 300 research and teaching STEM faculty, could be counted as people who became more aware of the issue of science data literacy by taking the survey, although the full set of email solicitations and reminders repeatedly introduced the issue and the course.

The second push strategy to elaborate is the presentations to undergraduate STEM classes. Faculty members who welcomed the 10-minute presentation and Q&A about this course provided positive feedback on the necessity of science data literacy education. In more than a few occasions, after our presentation the hosting professor contributed in the question and answer period with personal stories of their struggles dealing with their own data and cited a need for students to contribute in their labs or field observations with science data awareness and effective management skills. This effort landed the only two undergraduate STEM students who enrolled in the project's course offerings—a biology major and an aeronautical engineering major.

Lessons Learned 2: Outcome-Based Learning and Assessment

As previously indicated, in-depth discussion of the course design can be found in a *Journal of Library Metadata* article currently under review. This section will include details of the pedagogical experience of conducting the two course instances, including assessment of the students' experience with outcome-based learning about science data management. The P.I. applied her 15-year career in library and information science research and teaching when establishing the pedagogical framework in this new area of e-science. Through clearly differentiated modules and tiered skill development in related exercises, students gained skills and awareness to manage data resources in the e-science environment. The research assistant apprenticed in the course development and introduction by serving as the teaching assistant in the first instance, and then contributed design improvements based on that experience to teach the second instance of the course a year later under the oversight of the P.I.

Project staff observed that among the students enrolled in this course, graduate students had much better comprehension of the material and intellectual engagement in the exercises than undergraduate students. This prompted us to pair undergraduate and graduate students together in project groups to make sure each group had a good balance of levels of intellectual maturity and types of skills and subject background. We tried to make the course interesting to students representing various disciplines using data examples from a number of different science research fields and also reiterating fundamentals of science data management that crossed disciplinary character or research group practice. The course's active learning style included not only instructor, but also student, presentations of case studies about data types, formats, use, and management in different disciplines. Other student activity included quizzes on lectures and readings, interview practice, exercise completion, guided group work, and the delivery of a multiple stage, authentic project based on a scientist's or technologist's work practice.

The students of the course over two semesters profited from the life experience and expertise of

over a half-dozen people, not counting the researchers they worked with to complete their separate final projects. This type of team teaching is suitable to the subject of science data literacy given its blend of topics and skills in the application of library and information science techniques to treat data, which can be ephemera in the flux of the technical computing environment of e-science, as dependable digital information resources necessary for the successful conduct of scientific research. Two invited guests lectured to the class: one serving in the SU information technology group as a cyberinfrastructure specialist presented about grid and high performance computing environments, and the other, an advisory board member and professor of biophysics spoke about types of data accumulated and used in the course of several decades long research endeavors. The scientific data literacy project involved three doctoral students, who contributed their backgrounds in large-scale data analysis, database construction, systems engineering, and design, and communication to course planning and execution.

Pedagogical Results

At the beginning and end of each semester, we conducted an evaluative survey targeting student attitudes and reaction to the course experience. Students enrolled in this course in both years felt a closer affinity with science disciplines after its conclusion, but their relative comfort with technology and experience in databases and teamwork varied by the year (Table 3). This is reflective of the fact that more of the students enrolled in this course came from a science discipline in 2009; a majority in 2008 majored in a technology field. The drop in perceived competence in 2009 may reflect the relative inexperience of the research assistant as a teacher, but it also may reflect his insistence that they work with data or metadata from a target science domain versus staying in comfort zones such as web design or writing papers.

Table 3. Comparison of student attitudes about science and being a data manager

	INITIAL SURVEY 2008 MEAN (STD DEV)	FINAL SURVEY 2008	INITIAL SURVEY 2009	FINAL SURVEY 2009
N	14 (6UG+8G)	9 (4UG+5G)	11 (4UG+7G)	8 (2UG+6G)
Science Affinity (up to 5)	2.32 (1.08)	3.30 (1.02)	3.27 (1.13)	3.75 (.46)
Comfort with Databases and Computers (up to 5)	3.45 (.90)	3.96 (.75)	2.91 (.94)	3.19 (.75)
Comfort with Teamwork (up to 5)	3.29 (.89)	4.11 (.93)	2.82 (.90)	3.0 (1.31)
Perceived Competence Scale (up to 7)	5.61 (1.25)	5.83 (1.15)	5.68 (1.07)	4.18 (1.04)

Despite the change of teachers and adjustments to the course, reaction to the pace of the class stayed roughly the same (Table 4). Students generally considered the assignments/project the most interesting and useful part of the course. These were broken apart for greater granularity in the second year survey, where the results match the positive comments about the value of the project experience. The high marks for the exercises indicated the students felt they received necessary skills by doing them. Lectures were found lacking in 2009, once again reflecting the teacher's comparative lack of experience in the classroom, although these students were also more negative about the guest speaker. Perhaps the higher proportion of graduate students in 2009 shows they are more demanding if spending time in class than undergraduate students.

Table 4. Comparison of student attitudes about course aspects

	FINAL SURVEY 2008 MEAN (STD DEV)	FINAL SURVEY 2009 MEAN (STD DEV)
N	9 (4UG+5G)	8 (2UG+6G)
Pace of Class (up to 5(fast))	3.67 (.87)	3.50 (.535)
Lectures were Interesting (up to 5)		2.50 (1.07)
Lectures were Useful (up to 5)	4.0 (1.0)	2.50 (1.41)
Assignments were Interesting (up to 5)		3.13 (.84)
Assignments were Useful (up to 5)		3.25 (1.28)
Projects were Interesting (up to 5)		4.13 (.64)
Projects were Useful (up to 5)	4.44 (.73)	3.75 (1.04)
Speakers were Interesting (up to 5)		2.38 (.52)
Speakers were Useful (up to 5)	3.44 (1.01)	2.38 (.92)

We collected student comments as responses to open-ended questions asked in the post-course survey in both years. These can be compared with the Likert-scale responses of the rest of the surveys and also be appreciated in their own right as honest reactions to the course, to science, and to students' newly found appreciation for the role played in science by effective data management. It was rewarding to note a lack of comments about the readings in 2009 compared with 2008, given the amount of time to identify and assign more topical, incisive readings and better orient them to each lecture. The next such major refinement to the course should target the exercises, which got a similar amount of negative reaction comparing 2009 responses to 2008 survey results. More precise wording regarding the choice of activity between metadata or data level may help the students and perhaps these should be separate, parallel units in the future. Example solutions were again called for, although preparing these "canned" illustrative examples takes time and needs to apply across the science domains the students might choose, unless the science domains are restricted to allow more cogent teaching.

Table 5. Student responses about course mastery

Social and Technical Skill	- Importance of organizing information in a way that is useful and neat - XML Schema - How to interview someone - What a relational database is - Exercises 1-4: helped out with database skills, E-R diagram creation.
Science Data Awareness	- Complexity of datasets - Science data life cycle - How IT and science can mix - Scientific research process - Different science disciplines alter metadata schemas to fulfill the discipline's needs - How to develop a data management/use scenario and recommendations for a real project - Different repositories for the sciences - Learning about metadata: It makes intuitive sense that context needs to be provided for datasets

Open-ended comments solicited from the students in the final survey indicated successes of the project where their comments coincided with the goals of the SDL project. Student comments can be organized according to the social and technical skills needed and science awareness required participating in the practice of e-science. Example responses to the questions: "three things I learned from this course" and appear in Table 5. Additionally, student responses to "what is your (least) favorite part of the course?" questions point to successful features and areas that need work about the course design and implementation (Table 6).

Table 6. Student responses about favorite and least favorite course aspects

Positive	• Making database and website: it's really fun to create something useful for other people that make their jobs easier. • Fundamentals about science data and data management with case studies: easier to learn. • Project work: It gave me real life experience. • Talking with scientists to see what they cared about and why. • I liked learning about the different standards
Negative	• Highly technical terms sometimes, especially related to specific scientific fields: It requires deep background knowledge to fully comprehend some concepts. • Reading the course packet: I did not always find the readings useful. I felt like the lectures were a lot more useful. Just looking at the packet, makes everything seem overwhelming. • Exercises: Were not well established, too much ambiguity. • My least favorite part of the course was my lack of knowledge in databases. This class has no prerequisites & claims to not be a database [course], but I felt very hindered in progressing due to not having previous experience in creating databases.

Lessons Learned 3: Assessing efficacy of one solution

The SDL Project's solution to raising data literacy in future science researchers and information workers was a one-size-fits-all approach delivered by a single school that sits among many in the STEM academic environment. Not all universities have a library or information school to host such a course as we designed, and even if they do, it proved difficult to draw students to take such a course. Early in the project, the advisory board members expressed the concern that each of their undergraduate science curriculum structures would likely deter students who learned of and wanted to take the course. Many of the science curricula are structured in a way where each undergraduate is mandated to complete a fixed number of required courses and electives. Working with faculty advisors to create a schedule containing a full slate of these discipline-based courses before graduation leaves little room for students to take cutting-edge and somewhat demanding courses such as this one, particularly for the juniors and seniors who were the best candidates to take our course. Students in the later years of their program have more experience in their discipline and data-gathering for research and thus are more likely to be able to apply the lessons from our course.

In addition to the curricular boundaries, there were also many disciplinary boundaries the students were required to repeatedly negotiate when trying to process the case studies. The one-size-fits-all approach of our course required a variety of case studies in different research domains, which included all the individual discipline or domain's terminology, methods, perspectives and paradigms. One method to address this constraint is to expend effort to generalize methods in resource management of data resources that occur across disciplines in e-science. Another method is for more involvement with STEM faculty to raise data literacy

within their courses and departments, with involvement from library personnel as presently occurs with information literacy training. Our science data literacy course content could be incorporated into science curricula so that the data literacy topics and practices can be put in context, perhaps as a unit within already existing labs or courses. We did consider the possibility of working with a few STEM faculty members on creating tailored versions of our course, but due to the time and resource constraints, we decided not to pursue it during the project lifetime. Through our advisory board, we heard of one effort at the department level to educate students about data management, but given the number of STEM departments on the campuses, each department separately developing pedagogical solutions to the same, increasingly common problem seems an inefficient approach

Conclusion

We learned several lessons from our experience in the past two years in science data literacy education. While science data literacy skills are becoming increasingly important as the e-science/eResearch environment evolves, administrative support at the university and college levels is critical for the success. This includes, among other things, the awareness about science data literacy among university and college administrators and accompanying support for experimental and interdisciplinary projects that incorporate data literacy training for undergraduate STEM students. These observations led us to believe that science data literacy training needs to be provided at different levels via different venues, and that the training needs to adapt to science disciplinary context, terminology, and workflow. At the undergraduate level, the goal would be to train future science workforce with a solid understanding and skill set in data management and use issues.

It is worth mentioning that our experience in this project led to a successful grant proposal for building an e-science librarianship curriculum (eSLib), funded by the Institute for Museum and Library Services (IMLS), which is a three-year project that started in 2009. The eSLib proposal's success was built on the lessons learned from the SDL project, one of which was that one powerful way to supply people capable of working in e-science resource management of digital data requires a special type of master of science in library and information science (MSLIS) degree. We recruited people with at least an undergraduate degree in science as well as experience working on science research projects and are designing an educational program adapted from the MSLIS program and including a series of courses, including Science Data Management, that should prepare them to operate in the future, technologically mediated, interdisciplinary, collaborative environment of tomorrow's science investigations.

References

Finn, Charles, W.P. (2004). The mandate of digital literacy. Retrieved April 20, 2010, from http://www.techlearning.com/article/2602

ACRL. (2010). Introduction to information literacy. Retrieved April 20, 2010, from http://www.ala.org/ala/mgrps/divs/acrl/issues/infolit/overview/intro/index.cfm

Qin, Jian & J. D'Ignazio (under review). Teaching Metadata as a Key Ingredient in Developing Science Data Literacy. *Journal of Library Metadata.*