Data Information Literacy Case Study Directory

Volume 3 *University of Minnesota*
Number 1 *Civil Engineering*

Article 1

2015

Civil Engineering/ Graduate Students/ Johnston & Jeffryes/ University of Minnesota/ 2012

Lisa Johnston
University of Minnesota - Twin Cities, ljohnsto@umn.edu

Jon Jeffryes
University of Minnesota - Twin Cities, jeffryes@umn.edu

Recommended Citation

Johnston, Lisa and Jeffryes, Jon (2015) "Civil Engineering/ Graduate Students/ Johnston & Jeffryes/ University of Minnesota/ 2012," *Data Information Literacy Case Study Directory*: Vol. 3: No. 1, Article 1.
http://dx.doi.org/10.5703/1288284315479

TEACHING CIVIL ENGINEERING DATA INFORMATION LITERACY SKILLS: An E-Learning Approach

Lisa R. Johnston, University of Minnesota

Jon Jeffryes, University of Minnesota

INTRODUCTION

The University of Minnesota (UMN) team collaborated with a civil engineering lab researching the structural integrity of bridges, experimentally and within the state of Minnesota, to identify the data information literacy (DIL) skills that graduate students in that discipline needed to be successful researchers. In-depth interviews with the civil engineering group found that graduate students lacked DIL skills, particularly metadata and data description, ethics and attribution, and digital preservation. The absence of these skills negatively impacted the students' abilities to effectively pass their data sets on to the next graduate student on the project.

Based on these findings, in the fall of 2012 the authors launched an instructional response to address the DIL skills absent from the curriculum. This instructional approach utilized a modularized e-learning format to reach busy graduate students (Brenton, 2008) through an extracurricular Data Management Course. The DIL team created a seven-module non-credit online course (http://z.umn.edu/datamgmt) using Google Sites, Screenflow, and YouTube. The self-paced course allowed students to complete the requirements outside of their formal course work and research activity. As a component of the course, each student wrote a draft data management plan (DMP) for creating, documenting, sharing, and preserving his or her data using a template offered by the instructors that aligned with each of the seven modules. The instructors offered this online course to all structural engineering graduate students in the fall of 2012 (11 students enrolled), giving students the whole semester to complete the requirements, and then opened up the course to any science, technology, engineering, or mathematics (STEM) graduate student in the spring of 2013. Forty-seven students enrolled in the spring semester (for a total of 58 students over- all). Five students from the fall semester completed the course (three out of these five choose to defer their participation to the spring semester when they expected to work with research data) and six additional students completed the course in the spring. The results of an assessment survey sent to students immediately after completing the course, iterative feedback on their completed DMP, and a follow-up survey on how they implemented the DMP 6 months after taking the course were positive. Results from this course informed the development of a "flipped classroom" version of the course in the fall of 2013.

DATA MANAGEMENT TRAINING AND PRACTICE IN THE CIVIL ENGINEERING DISCIPLINE

Currently civil engineering poorly defines its disciplinary expectations regarding teaching data management to its students. The topic of data literacy can only be inferred into existing learning outcomes or other standards that touch upon data tangentially, usually under outcomes that focus on the overall experimentation process.

The American Society of Civil Engineers' engineering curriculum, Civil Engineering Body of Knowledge for the 21st Century: Preparing the Civil Engineer for the Future (BOK 2) (ASCE, 2008), does not address data literacy explicitly. Currently the integration of these skills into the graduate-level curriculum remains completely voluntary. Students graduating have no guarantee of receiving formal education in the best practices of data management. Many students learn through informal instruction or address the problem when they suffer their own data loss.

A report produced between iterations of the BOK, Development of Civil Engineering Curricula Supporting the Body of Knowledge for Professional Practice, found room for improvement in the depth of students' engagement with data, citing one example where "students are not able to take an open-ended real world situation and design the experiments that would provide the necessary data to solve the problem" (American Society of Civil Engineers Curriculum Committee, 2006).

Data literacy skills can be inferred in many of the outcomes focused around its seventh outcome group, "Experiments." The relevant outcomes are

- Identify the procedures . . . to conduct civil engineering experiments
- Explain the purpose, procedures . . . of experiments
- Conduct experiments . . . according to established procedures
- Analyze the results of experiments (ASCE, 2008, p. 106)

Data literacy can also be inferred from the outcomes regarding communication (BOK 2, Outcome 16), which call for students to "use appropriate graphical standards in preparing engineering drawings" and "[o]rganize and deliver effective . . . graphical communications" (ASCE, 2008, p. 110). It can be read as part of Outcome 13: Project Management, if the new standard procedure for conducting experiments includes creating a plan to man- age data, including organization, security, and preservation (now mandated by some funding agencies).

The engineering field, more widely, shares this opacity of expectation with regard to data management. The outcomes suggested in the BOK 2 echo those already implemented by the Accreditation Board for Engineering and Technology (ABET) in their outcome, "an ability to design and conduct experiments, as well as to analyze and interpret the data" (ABET, 2012, General Criterion 3[b]).

Locally, UMN students and faculty receive somewhat varied and inconsistent DIL training. For example, the university requires all principal investigators (PIs) of grants to complete one of two Web-based instructional modules on the "best practices of research integrity" (University of Minnesota Research Education and Oversight, 2014). These modules cover some aspects of data control and intellectual property concerns. However, these responsible conduct of research (RCR) modules are only required for PIs and are not well described or discoverable to those looking for just-in-time data management education. Beginning in 2010, researchers could supplement that training with workshops taught by the libraries on “Creating a Data Management Plan for Your Grant Application” or “Introduction to Data Management for Scientists and Engineers,” available as drop-in library workshops and online video recordings (University of Minnesota Libraries, 2014). The former work- shop reached more than 300 faculty members and is offered for RCR continuing education credit (Johnston, Lafferty, & Petsan, 2012). However, both RCR training and library-led workshops were designed specifically for faculty PIs and therefore do not target the graduate student population.

It is possible that data management skills are being addressed, along with other information literacy competencies, in student research experiences such as undergraduate research opportunities programs, research assistantships, or cooperative educational programs, but the literature on information literacy has focused primarily on information retrieval skills (Jeffryes & Lafferty, 2012). One student in our study mentioned receiving some data management skills in an introductory research methods class, but considered it too early in her student career to be useful to her current research project. The current integration of data management skills into the graduate curriculum is neither constant nor at the point of need.

The DIL team also investigated the current data management best practices used by the discipline locally. One of the graduate student subjects worked in the Multi-Axial Subassemblage Testing (MAST) Laboratory, which provided explicit best practices for data management and support for data upload to the national NEEShub data warehouse, a National Science Foundation–funded data repository for earthquake engineering data. The other students in the study population did not receive documented support or management guidance during their research.

Data repositories, examples of curated data, and management protocols exist for some sub- disciplines relevant to the work conducted by the research population. The student working with the MAST Laboratory was required to post her data into NEEShub. Although the other researchers were not connected to a specific data repository, Table 7.1 provides examples of metadata schemas and requirements that researchers in structural engineering might encounter.

We discovered documentation and training opportunities provided by these bodies through Internet searches. Overall we found two disciplinary leaders within structural engineering, NEES and NISEE, both of which focus on the curation of earthquake engineering data (NEEShub, 2009; Thyagarajan, 2012; Van Den Einde et al., 2008; Wong & Stojadinovic, 2004).

METHODOLOGY

The UMN team interviewed the members of a structural engineering research group consisting of one faculty member and four graduate students ranging in experience from a first-year graduate student to a student in her final semester. The interview instrument, based on a modified version of the Data Curation Profiles Toolkit instrument (available for download at http://dx.doi.org/10.5703/1288284315510), allowed us to gather detailed information about the practices, limitations, needs, and opportunities for improving DIL practices from the perspective of both the faculty member and graduate students in the subject area. We collected and evaluated relevant documentation, including data set examples and supporting research practices.

The interviews took place between March 13, 2012, and April 20, 2012. These structured, 1- to 2-hour interviews took place in a library conference room using two audio recorders each producing a file that a graduate assistant transcribed for analysis. The interview comprised two components: a worksheet that participants filled out and a list of follow- up questions that were asked of interviewees based on their responses from the worksheet. The data we collected, including the sample of the research data provided by the research group, the interview transcripts and audio files, and the interview worksheets, were anonymized, compiled into a Microsoft Excel file, and analyzed.

TABLE 7.1 - Data Repositories Identified in the Disciplinary Environmental Scan of Civil Engineering

Repository	**Location**	**URL**
NEEShub (earthquake engineering)	Purdue University	http://nees.org
NISEE (earthquake engineering)	University of California, Berkeley	http://nisee2.berkeley.edu
DARPA Center for Seismic Studies	Arlington, Virginia	http://gcmd.nasa.gov/records/GCMD_EARTH_INT_SEIS_CSS_01.html

RESULTS OF THE NEEDS ASSESSMENT

The interviews provided a snapshot of the DIL skills needed for structural engineering graduate students at UMN. The analysis revealed several needs at various stages throughout the data life cycle. It was clear that the students had no formal training in DIL. Students reported collecting various types of data, but primarily data from sensors placed on the bridges they were evaluating, to study bridge integrity factors. The lab works with and receives funding from national and state agencies to conduct its research projects. These project partnerships have a noticeable effect on the treatment and handling of the data. The student working within NEES was expected to share data via the processes and standards for sharing and cu- rating data developed by the NEES repository. The state agency, on the other hand, claimed ownership over the data and required approval before the data could be shared. Although the work of the lab was influenced by the expectations of its external partners, no formal policies or procedures (for documenting, organizing, or maintaining data) existed in the lab itself. As a result, individual students approached data storage and management in different ways. The faculty researcher expressed concern about students' abilities to understand and track issues affecting the quality of the data, to transfer the data from their custody to the custody of the lab when they graduated, and to take steps to maintain the value and utility of the data over time:

> "The skills that they need are many, and they don't necessarily have it and they don't necessarily acquire it in the time of the project, especially if they're a Master's student, because they're here for such a short period of time."

We asked the participating faculty and students to indicate the importance for graduate students to become knowledgeable in each of the 12 competencies of DIL, by using a 5-point Likert scale, and then to explain their choices. Interviewees identified additional skill sets they saw as important for graduate students to acquire (see Figure 7.1).

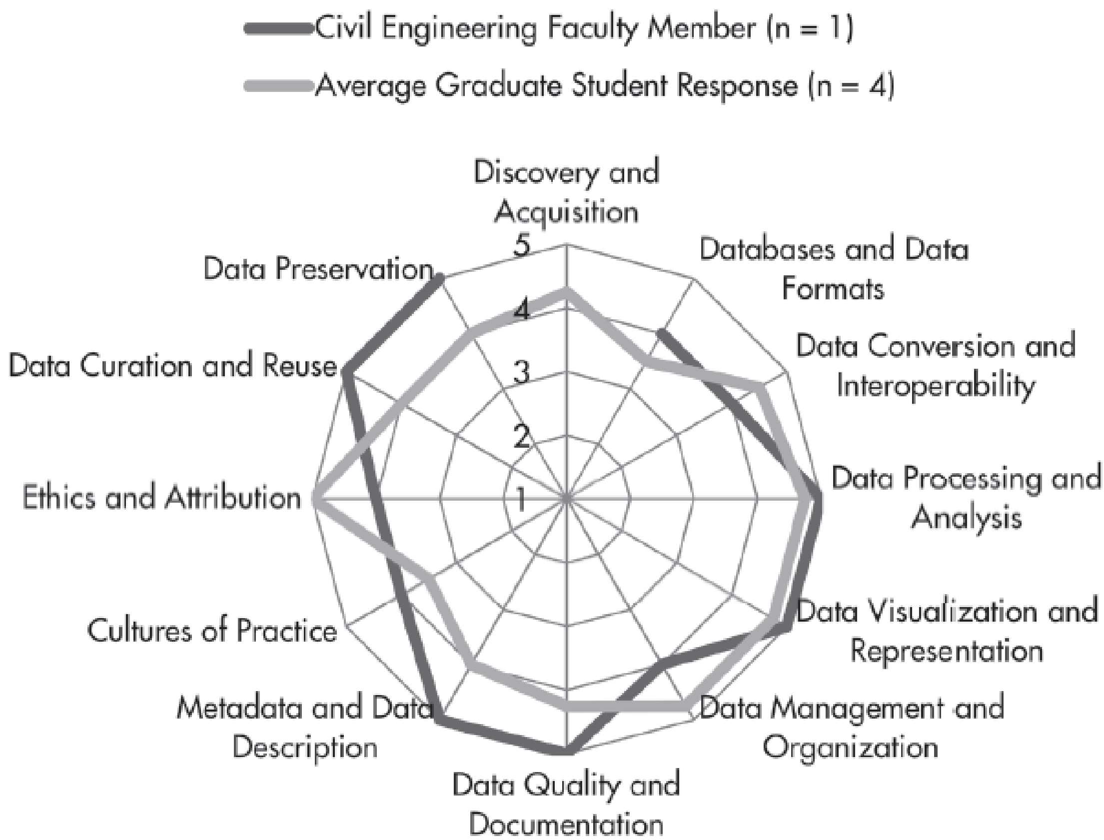


Figure 7.1 The rating of DIL skills by the UMN faculty member and the average graduate student response. Scale: 5 = essential; 4 = very important; 3 = important; 2 = somewhat important; 1 = not important. (NOTE: The faculty member did not rate Discovery and Acquisition.)

In the course of interviewing the graduate students, certain steps in the data life cycle were present regardless of the research project, though the students did not use a consistent vocabulary when describing these steps (see Table 7.2).

To analyze the skills and needs described in the interviews, we reviewed the results in the context of each of the stages of the data life cycle. Although the students did not explicitly identify preservation as a step in their data life cycle, they mentioned critical aspects of this topic throughout the results phase. These observations provided a foundation for a generalized approach to understanding the data interactions of structural engineering graduate students in a research group.

Stage 1: Raw Data

In the first module of the interview we asked the graduate students to describe the type of data with which they worked. All graduate students reported using sensor data as the crux of their research projects. Three out of the four graduate students collected data for projects that generated real-time sensor data to monitor the performance of local bridges, while one graduate student generated experimental data and simulations on concrete column performance in simulated earthquake conditions.

TABLE 7.2 - Data Life Cycle Stages as Described by the Case Study Graduate Students

Student Response					
Student	**Initial**	**Second**	**Third**	**Fourth**	**Fifth**
Grad #1	Raw Sensor Data	Processed Data	Processed with Figures	Comparison (with other research)	Share the Data (Stages 1 & 2)
Grad #2	Raw	Excel #1	Excel #2	Stress calculation / force and movement calculation	Final Excel file
Grad #3	Raw Numbers	Organization	Analysis and Conclusion		
Grad #4	Data Download from a Website	Organize Data into Test Folders and Regular Activity of Bridge Folder	Analyze Data	Create alarms to warn of potential problems on the bridge	
Data Stage					
	1. Raw Data	2. Collection and Organization	3. Processing and Analysis	4. Results	5. Sharing and Archiving

Although the expectations of their external partners influenced the work of the lab, the lab itself did not have formal policies or procedures in place for documenting, organizing, or maintaining their data. As a result, individual students approached data storage and management in different ways. The faculty researcher expressed concern about his students' abilities to understand and track issues affecting the quality of the data, to transfer the data from their custody to the custody of the lab upon graduation, and to take steps to maintain the value and utility of the data over time. For example, the faculty interview highlighted the need for students to understand the potential hazards of collecting "bad" data. The faculty member thought that having a better understanding of how sensors collect data might help. Several students mentioned knowing about potentially disruptive elements such as temperature conditions or scheduled construction/testing that might impact their data; however, their processes and documentation did not merge these events with the data they collected.

Stage 2: Collection and Organization

In discussions regarding data collection and organization, more trends emerged:

- Students used date-based file-naming structures, even when they weren't familiar with the concept of a file-naming structure. As one student remarked: "I've never even heard of a file naming system."
- Students did not consider data security an issue and felt that they had adequate protections in place.
- Backup of their data was often sporadic or nonexistent. Two of the students displayed some confusion about the concept of data backup versus data redundancy. For example, one student described her backup process as copying files to a separate folder on her desktop (which would not protect against theft or computer damage).
- Students agreed that they had no formal DIL instruction but had to rely on their peers, family, and previous experience for direction. As one student described: "I've had many projects with Excel files and stuff that I've needed to save, and I guess I learned [data management] just out of habit, mainly."

Students used formal and informal documentation practices to record the data collection process, and changes made to the data were ad hoc and varied. For example, while some students labeled columns in Excel, additional information, such as the bridge sensor locations, were in multiple locations and separate from the data files (e.g., in e-mail correspondence or schematic drawings). Most of the students did not have an understanding of the concept of metadata. Only one of the graduate students was familiar with the term, and when asked to define it the student replied, "It means data captured and saved during the test." The other students all responded negatively when asked if they were familiar with the term. Regardless, all of the students provided some level of metadata to the data they were working with, but the majority were not collecting or applying it in an intentional or formal manner.

When asked if they had any means of documenting the steps for someone else to repeat, the students described the inefficiencies of their own system. One student admitted, "I guess if I were to repeat [the research project], I would probably do it in a different way. I could probably document what I've done and I probably will do so, but then I'll also suggest maybe keeping things a little less complicated."

Stage 3: Processing/Analysis

Each of the graduate students described a process for analyzing, visualizing, and making conversions of the data beyond the original raw data stage. The majority of the graduate students spoke of a process of converting ASCII text files into Excel for further manipulation and sense making. One graduate student used a proprietary sensor program that allowed for data manipulation within her Web-based software. Regardless of format, they described a process of further manipulation of the data, such as re- moving "bad" data (i.e., bridge sensor readings contaminated due to noise during construction), synthesizing the rough data using equations, and creating graphical representations of the data ("plotting"), all to better communicate findings.

The faculty member held the graduate students' facility with Excel and MATLAB in high esteem, but had some concern that students weren't receiving all the support they needed in more advanced data analysis, saying:

> It's the relational databases . . . and their capabilities for statistical analysis that are a little weak. And there are courses they can take on campus for the statistical and the relational databases, so maybe it's something that we should be requiring. The problem is that if they're going to do a Master's thesis, they take only seven courses.

He echoed the sentiment for further development of student skills in this area by noting that students would benefit from further education on the strategy behind data plotting. His ideal would be for graduate students to demonstrate an "ability to take the data and come up with a way of conveying it so that the reader can pick it up very quickly." Indeed one student described his process of creating data visualizations in Excel as "mostly trial and error." The faculty member also specifically called out the need for students to be able to identify and track the quality of the data they were collecting when it may have been compromised by outside forces, such as with construction on the bridge where they collected sensor data. The professor commented that the students weren't currently tracking this aspect of their data analysis in the documentation, but "it would be nice, especially when they're collecting huge amounts of data, if we could some- how get measures of the quality of the data, statistically. And if we could use these measures to keep track of getting good data and when we're not getting good data."

Stage 4: Results

During discussions about ensuring long- term access to the data collected, numerous preservation concerns arose. Several issues were not addressed in the research group, such as physical storage (e.g., desktop computers used by graduate students would eventually be re- cycled) and file migration (e.g., use of a proprietary and future incompatible version of Excel) for data stored in the lab.

Students were unclear about whose responsibility it was to preserve the data for long-term access. Additionally, they were unclear about how to preserve data for 20 to 50 years, or the life of the bridge. For example, one student suggested that the contracting state agency held the responsibility for preserving the data and that the agency would keep the data "forever." When asked to identify the steps needed to preserve the data and if the state currently implemented those steps, the student responded: "I think that's just sort of what they do. . . . [B]ecause they've had issues in the past where people have completed projects and then others have wanted to repeat them or go more into depth with them and then haven't been able to find any of the original data for it, . . . I think that's kind of just their policy." When asked for steps to preserve the data set the graduate student responded, "Just putting [it] onto that hard drive and making sure it doesn't melt I guess."

In our conversation with the faculty member, the issue of data versioning for long-term access and preservation arose. Along with identifying and implementing steps to preserve and store data for the long term, researchers must choose which versions of their data should be preserved for future use and authenticity. The professor responded to the issue of versions:

> This is an interesting problem. There are actually multiple stages and multiple things that you do [to the data], and so how many data sets do you store? Clearly, you want the raw data. That's the purest form. And clearly you want the data that you think has been completely digested as you think it needs to be. But how many of the intermediate stages do you want to keep?

Stage 5: Sharing and Archiving

Each of the four students shared his or her data results in some way. One student shared her data in a formal process through the mandatory data archiving protocol of the NEEShub program, while the other students shared their data with state contractors, their advisor, and the graduate students continuing the project.

Although students had little to no experience with data citation, when asked their thoughts on its importance, they reported an understanding of the value of this practice. A student explained: "Because you need to know where this data is coming from, and obviously if it's not your own, then I feel like it's important to make other people aware that it is not data that you actually collected yourself."

As to the potential for other researchers to reuse their data, only one student felt that his analyzed data was unique and therefore of potential value. The other students had a harder time imagining how their data might be useful to researchers outside of their specific project. The graduate students demonstrated little to no knowledge of data repositories in their field or experience using another researcher's data from outside their lab. One student mentioned that looking at another researcher's data in the literature review led to his experiment, but he found the data by chance and the repository was not a standard destination.

The graduate students did not see the value in archiving similar data sets together in a subject-based repository structure. Referencing the Interstate 35W bridge in Minneapolis, which was rebuilt after the tragic 2007 collapse with sensors measuring strain in a similar way to the data obtained by our interviewee, the student noted, "Unless you could come up with some good way to compare the two sets of data, I don't know really what use it would be to collect them all into one place." The student did see the value of data repositories to save on space, however, so that "there aren't 50 external hard drives floating around."

Issues around privacy and confidentiality were a complex topic for students working on a state-contracted project analyzing bridge sensor data. Students knew to contact their advisor with requests to share the data owned by the state agency. One student described her caution with presenting the state-funded data results at a conference: "I had to get permission from [the state contractor] first before I could even do that." However, the reasons beyond "ownership" were unclear. The faculty member was able to explain the sensitive nature of the data when asked if the state agency had any specific interests in sharing this data beyond the agency. The professor replied:

> That's a really good question. They would like to share data, as long as they can protect their interests. And I don't mean any advantage in having that data. What they're afraid of is this data represents measurements that are taken off of real bridges, and that can very easily be misinterpreted and used to undermine a bridge that's actually not in bad shape, and then present a bloated and incorrect scenario about how bad the bridge problem is. Or the claim that a bridge is in great condition, when in fact it needs to be replaced. For that reason, they are very, very, very unwilling to have anything like open access.

All Stages

With our findings, the UMN team developed a list of skills needed by graduate students in this discipline. These are detailed in Appendix A to this chapter.

E-LEARNING APPROACH TO TEACHING DATA INFORMATION LITERACY SKILLS TO GRADUATE STUDENTS

The benefits of taking an e-learning approach to educating graduate students are enumerated in the literature reviews and discussions of many studies (Gikandi, Morrow, & Davis, 2011; Safar, 2012). The U.S. Department of Education (2010) in its meta-analysis of the literature found that "students in online conditions per- formed modestly better, on average, than those learning the same material through traditional face-to-face instruction" (p. xiv). Gikandi, Morrow, and Davis's review of formative assessment in online learning, citing the influence of Oosterhof, Conrad, & Ely (2008), posited that online learning benefitted students by providing instructors "many additional opportunities to dynamically interact with and assess learners" (p. 2333). Gruca (2010) nicely outlined benefits of libraries' adopting e-learning platforms to deliver their instruction. Most resonant with our experience was her assertion that "e-courses are equally accessible for full-time and remote students and may be a step towards inclusion for disabled students" (Gruca, 2010, p. 20). We wanted our instruction to be as accessible as possible to graduate students who carried a full course load as well as a time-intensive re-search schedule. Although Gruca (2010) never explicitly used the phrase, many of the benefits of e-learning she listed support the scalability of instruction inherent in an e-learning platform. Gruca stated that e-learning "saves teachers' and students' time" and "[o]nce published, an e-course may be improved and used many times" (p. 20). The ability to scale would be integral to ensuring expansion of our work at a university where we support tens of thousands of students.

Learning Objectives and Assessment Plan

Conceptualization and creation of the course took place over the summer of 2012. Table 7.3 shows the learning outcomes for each module of the course.

In the course design phase of the project, we met with the faculty partner to vet the learning outcomes and strategize on connecting students to our course content. Because the graduate-level curriculum was already quite full, the approach had to be a voluntary, extracurricular program for students. The online, e-learning format was clearly a good fit. In addition, modularized video lessons would be easy to download and watch on any device that matched the busy graduate student lifestyle. The syllabus is in Appendix B to this chapter.

TABLE 7.3 - Descriptions and Learning Outcomes of the Seven Modules in the UMN Data Management Course

Course Module	Brief Description	Learning Outcomes (Students will . . .)
1. Introduction to Data Management	In this module we introduce the concept of data management using an example from the academic discipline	Describe the benefits of data management to explicitly understand the benefits of participating in the course Articulate what they will get out of this program to reinforce the learning outcomes of the curriculum
2. Data to Be Managed	This module helps students define what information will be managed, document the data collection process, and create a plan to store, back up, and securely house these data	Create a data inventory for their research project (e.g., data, project files, documentation) to not overlook any aspects of their DMP Write a backup and storage plan to avoid potential loss of data
3. Organization and Documentation Methods	This module helps students plan for how to organize their data, track versions, create metadata, and document data collection for reuse	Plan an organizational structure for their data using a file naming system and directory structure that is well-documented and interoperable with other data sets to decrease versioning issues and data duplication Articulate a plan to collect and share the supplementary data points of their research to

		assist other researchers in making sense of their data Fill out a metadata schema example for their data to model ideal metadata practices
4. Data Access and Ownership	In this module we illustrate some of the intellectual property and access concerns that researchers face when sharing their data with others	Name the stakeholders of their data to understand the potential intellectual property and ownership concerns with releasing their data to a broader audience Report potential access concerns with their data to plan for the appropriate access controls Identify potential access controls to secure their data prior to release
5. Data Sharing and Reuse	In this module we describe the benefits of data sharing and potential for reuse as well as introduce students to the concept of data publishing and citation	Name the audience for whom the data will be shared to customize the documentation and format for potential reuse Explain an approach they will use to share the data to instill best practices for their future data sharing Cite their data in a properly structured format in accordance with emerging standards to prepare them to ethically reuse data in the future
6. Preservation Techniques	In this module we introduce the preservation and curation techniques used by information professionals who manage digital information for long-term access	Explain the life span of potential use for their data to recognize the long-term value of their data Identify the relevant preservation-friendly file format for their research data to ensure long-term access to their digital information
7. Complete Your DMP	This final module instructs students on how to complete and implement their DMP within their lab, research group, or future project	Map out an implementation plan to put their DMP into action. Identify the components of a DMP to repeat the process with future research activities

We thought the course needed a real-world application in which the students might demonstrate or test their newly acquired skills. Therefore, building on our earlier success offering data management training to researchers, we chose to use a DMP template as the framing device for course content delivery and evaluation. Each of the seven course modules mapped to a corresponding section of a DMP template where the student directly applied what he or she learned in the course. (See Appendix C to this chapter for a DMP template.) The resulting seven course modules became

1. Introduction to Data Management
2. Data to be Managed
3. Organization and Documentation Methods
4. Data Access and Ownership
5. Data Sharing and Reuse
6. Data Preservation Techniques
7. Completing Your DMP

Although data analysis and visualization skills came up in our interviews with faculty and students, we chose not to include them because the librarians did not have the expertise to teach them. As an alternative we added a page to our course website pointing students to local and freely available resources and training.

At the outset of our course design we decided that our guiding principle for creating online instructional modules would be to "utilize preexisting content." With that philosophy in mind our first step was to find content openly available for reuse, including video, images, and e-learning tools that covered any of our data management topics. A library science practicum student helped review relevant content. We discovered many sources labeled for reuse; including professional library-generated tutorials such as MANTRA (http://datalib.edina.ac.uk/mantra), a UK-based data management skills support initiative, as well as informal You-Tube videos and cartoons. We embedded several of these through the modules after receiving permission from the authors. In addition, we customized content from the in-person data management workshops that the UMN libraries have offered to focus on the particular needs of structural engineering graduate students.

To create the modules we wrote scripts, created slides, and recorded videos for each of the seven topics. The scripts were written to incorporate a logical flow of the information and to set up the student to respond to each learning outcome. Next, we built a slide deck in Micro- soft PowerPoint and then captured the screencast presentation with voiceover using ScreenFlow (http://www.telestream.net/screenflow/overview.htm), an Apple-based video recording software.

ScreenFlow was chosen because it allowed us to capture and edit existing YouTube videos that we embedded in PowerPoint presentations and included in our modules. ScreenFlow also presented a relatively easy-to-learn editing interface over alternative software such as Apple iMovie or Adobe Captivate. After creating the videos, we uploaded them to a YouTube channel to allow us to link or embed them into content platforms. YouTube also facilitated closed captioning of the videos, making them more accessible to a variety of learners.

The video content was organized on a Google Site as the course home page at http://z.umn.edu/datamgmt (see Figure 7.2). The Google Site allowed us to create separate Web pages for each module, which includes the following components:

- Text descriptions of each module's learning outcomes
- Instructional video (embedded from YouTube)
- Assignment (links to the student's DMP template)
- Links to additional resources (if applicable)
- Cartoon illustration of a relevant data management concept

The course site is open to the public. We choose Google Sites over other campus e-learning tools due to the ease of creation, discoverability, and potential for one-click "cloning" if the library adapts the course in future semesters or for disciplinary sections beyond civil engineering.

Beta testing of the e-course revealed several minor errors and inconsistencies with the video modules and website. The test users were primarily UMN librarians and members of the DIL grant project. ScreenFlow allowed for quick video edits and insertions while the written-out scripts proved easy to edit and rerecord.

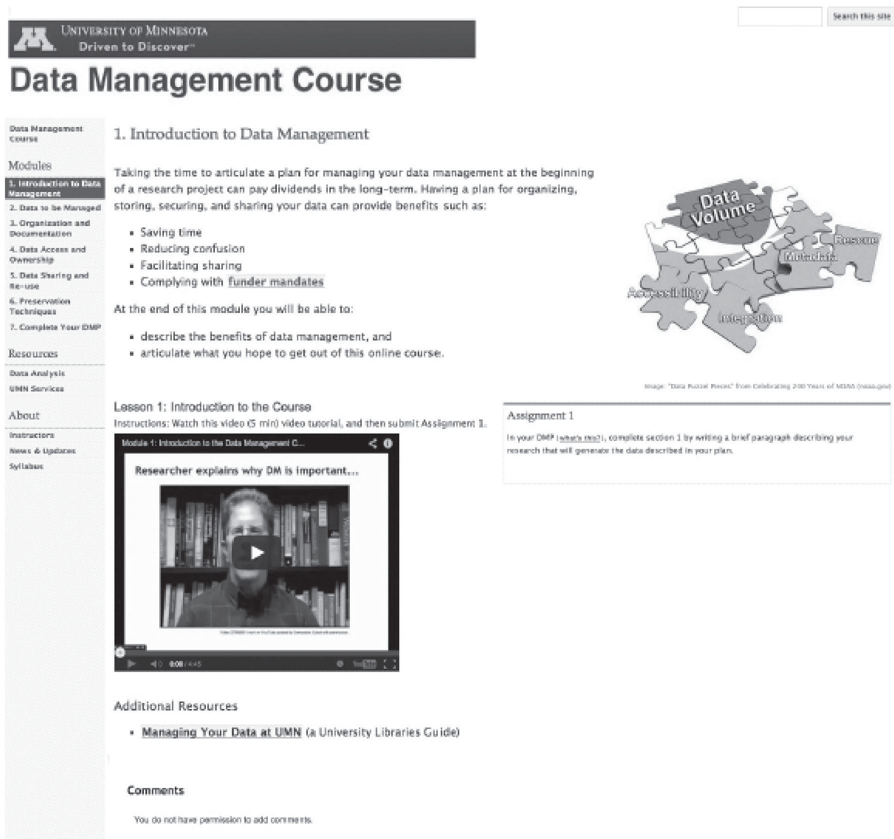
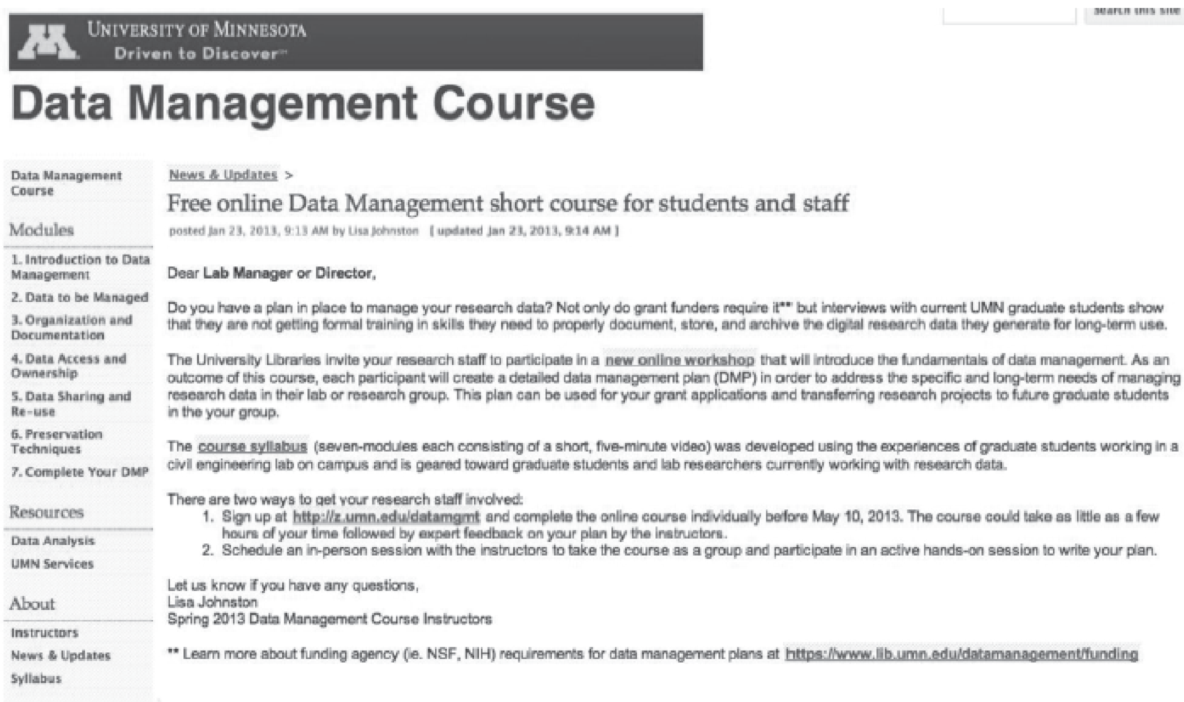


Figure 7.2
Screenshot of a module in the 2012 Data Management Course: Structures Section.

To assess the success of the instructional intervention we used a three-pronged assessment plan including formative and summative assessment techniques. Throughout the course students would take the information covered in the individual modules and apply it directly to their own research project through the creation of a DMP. The instructors created a unique copy of a DMP template that they shared with students via Google Drive (see Appendix C to this chapter) upon their enrollment in the course. We used the completion of the DMP template as a formative assessment throughout the course. Oosterhof, Conrad, and Ely (2008) described formative assessment as “those [assessments] that occur during learning,” analogous to “what a mentor does continuously when working with an apprentice” (p. 7). The different modules strategically mirrored the DMP template. This design made it easy for students to create a real-world application. Since the students’ DMP document was shared with the two instructors via Google Drive, we could check on the students’ understanding periodically and provide feedback via the “Comment” feature. This form of assessment allowed us to gauge student understanding in an organic way that would seem relevant to the students.

For the second prong of our assessment plan, we sent a course satisfaction survey immediately to students who had completed the course (see Appendix D to this chapter). These responses provided a summative view of each student’s experience in the course. The instructors learned which aspects of the instructional approach were effective, and which needed further improvement.



UNIVERSITY OF MINNESOTA
Driven to Discover

Data Management Course

Data Management Course

Modules

1. Introduction to Data Management
2. Data to be Managed
3. Organization and Documentation
4. Data Access and Ownership
5. Data Sharing and Re-use
6. Preservation Techniques
7. Complete Your DMP

Resources

Data Analysis
UMN Services

About

Instructors
News & Updates
Syllabus

News & Updates >

Free online Data Management short course for students and staff

posted Jan 23, 2013, 9:13 AM by Lisa Johnston [updated Jan 23, 2013, 9:14 AM]

Dear **Lab Manager or Director**,

Do you have a plan in place to manage your research data? Not only do grant funders require it** but interviews with current UMN graduate students show that they are not getting formal training in skills they need to properly document, store, and archive the digital research data they generate for long-term use.

The University Libraries invite your research staff to participate in a new online workshop that will introduce the fundamentals of data management. As an outcome of this course, each participant will create a detailed data management plan (DMP) in order to address the specific and long-term needs of managing research data in their lab or research group. This plan can be used for your grant applications and transferring research projects to future graduate students in the your group.

The course syllabus (seven-modules each consisting of a short, five-minute video) was developed using the experiences of graduate students working in a civil engineering lab on campus and is geared toward graduate students and lab researchers currently working with research data.

There are two ways to get your research staff involved:

1. Sign up at http://z.umn.edu/datamgmt and complete the online course individually before May 10, 2013. The course could take as little as a few hours of your time followed by expert feedback on your plan by the instructors.
2. Schedule an in-person session with the instructors to take the course as a group and participate in an active hands-on session to write your plan.

Let us know if you have any questions,
Lisa Johnston
Spring 2013 Data Management Course Instructors

** Learn more about funding agency (ie. NSF, NIH) requirements for data management plans at https://www.lib.umn.edu/datamanagement/funding

Figure 7.3 Announcement and e-mail invitation to participate in the 2013 spring Data Management Course.

The third prong measured the long-term impact of the course via an online survey that we sent out 6 months after the completion of the online course (see Appendix E to this chapter). This assessment was to show us whether completing the course impacted students' practice of managing research data. This form of assessment showed us whether the students successfully moved through the "hierarchical order of the different classes of objectives" found in Bloom's taxonomy, from knowledge, to comprehension, to application, to analysis, to synthesis (Bloom, 1956, p. 18). As Bransford, Brown, and Cocking (1999) stated in a report on the science of learning, "It is essential for a learner to develop a sense of when what has been learned can be used—the conditions of application" (p. xiii).

Results of the Fall 2012 and Spring 2013 Course

At the end of the first week of the fall 2012 semester the two library instructors discussed the data management course during the Civil Engineering Structures Seminar, a required course for all the graduate students in the "structures" track (around 20 students). We focused on why data management is important. At the end of the session the students completed a "1-minute paper" explaining how they thought a DMP would benefit their research. Subsequently, 11 students enrolled. The students controlled their own progress through the course. The instructors sent e-mails three times throughout the semester to nudge students to participate: once at the semester's midpoint, once a week before the course deadline (the last Friday of classes), and on the day of the dead- line of the course. The instructors periodically reviewed the DMPs of the enrolled students in Google Drive to provide feedback. There was no progress on the templates until late in the semester.

In the spring semester, we scaled the course to reach other researchers across our campus. We built the course so it would be relatively easy to replace the discipline-specific content with that of other research areas. In the spring of 2013, the instructors sought the help of 6 subject librarians, liaisons to the engineering and other science disciplines on campus. With their help, we opened the course to graduate students from other engineering and science disciplines (see Figure 7.3). There were 47 enrollees from 14 departments. No introductory session was offered in person as it had been in the fall due to the wide variety of students.

The spring course was similar to the fall semester course, except that liaison librarians, not the original course authors, sent periodic e-mail reminders to engage the students. Mid- way through the course, we offered an in- person 2-hour workshop that delivered all of the course material in a single, collaborative environment. Instead of working through the seven Web-based modules on their own, students could attend the workshop and ask questions and get feedback in class. They could learn from peers and discuss the practical application of data management with them. Thirteen students attended this session.

Course completion included not only watching the video modules (or attending the in-person session) but also completing a DMP. The plan had to be submitted to instructors for feedback before the course could be considered complete. At the end of the fall semester only 2 out of 11 students had completed the DMP template. Five students asked for extensions or permission to defer their enrollment into the next semester. The reasons for postponing included heavy workloads and lack of an actual data set to apply the principles covered in the videos. Three of those 5 students who chose to defer successfully completed the course in the spring, bringing the fall course completion rate to 5 students (a 45% completion rate). In the spring, 6 out of the 47 students who signed up successfully completed the course by turning in a written data management plan (13% completion rate). Overall, we ended the 2012–13 academic year with a total of 11 graduate students completing the course. This is a 19% completion rate for an online, non-required class—higher than that for most MOOCs (massive open online courses), which according to Parr (2013) is about 7%.

We sent a four-question survey to all 11 students once they finished the course, along with a certificate of completion for their UMN training history. Seven students (64%) completed the survey and demonstrated a high level of satisfaction. One student summed up the course:

> This course gave me good techniques which I will not only be able to implement in my current research in addition to what I have already been doing, but also use them in the rest of my career.

We received five (45%) responses to the 6-month follow-up survey. The questions mirrored the seven module topics of the course and the primary learning objects for each module. Overall the results and comments were very positive. Comments also demonstrated understanding of some of the primary learning objectives of the course—for example file naming and metadata schemas as illustrated by this comment:

> Some forethought on naming and metadata conventions goes a long way when managing data. This aspect of the course was very important and I have tried to employ it as often as possible. I sense that many students and possibly some researchers/professors don't commonly use a clear naming structure or metadata schema.

Comments also highlighted some surprising aspects of the course that students did not find relevant. For example, data ownership and access:

> This aspect of the class was also very thought provoking but isn't quite as relevant to my data. However, I am involved with many projects that have multiple organizations with interest in common data and so; some forethought on data ownership will help clarify who is in charge of this data and how to process/pass it along.

DISCUSSION: LESSONS LEARNED FROM THE E-LEARNING APPROACH

Our two semesters proved to be learning experiences in the presentation of this course. We applied key lessons from the first iterations of the e-learning approach, which included connecting to actual student data sets and providing generic simulations, as well as incentivizing the course to ensure completion.

Connection to Actual Data Sets

We attempted to make this course applicable by tying course content to the actual work students were doing in their labs. Therefore, students had to have their own research data to make the course useful. But many of the students interested in the course were not far enough along in their program to have started collecting data for their project. In the in person workshop we included an example of a completed DMP that provided students with a data set and a model they could follow when constructing their own plans. An approach to consider for students who do not have a research project is to provide a generic simulation to which students could apply the principles addressed in the video modules.

Ensuring Completion

Although a large number of students enrolled in the course, the completion rate was low. In the first iteration of the course a certificate of completion was used as a prompt for completion (on the advisement of our faculty partner), but only 2 of 11 students completed the course (though 5 more asked to defer their completion).

We are considering promoting the course through principal investigators and lab advisors.

We learned many lessons from implementing an online instruction model for teaching DIL. For example, we believed that our approach would allow busy graduate students to engage in supplementary materials on their own time. However, setting aside time to self-educate proved to be a major hurdle for students. The response to the optional workshop showed that students were willing to attend training in person because it provided a structure for completion. As one student stated: “I really liked the in-person lecture. Made it easy to set aside one block of time to go through all the information and have staff on-hand to answer questions.”

Therefore, in response to these findings we changed the pedagogy of the course in fall 2013 to a “flipped course.” Participants in the workshops met for 1-hour sessions once a week for 5 weeks. Students watched an online video before attending the corresponding hour-long hands-on workshop. In class we used fictional data scenarios from a wide range of disciplines to introduce students to practical aspects. To encourage completion, we offered participants who attended all five data management workshops a certificate of data management training for their UMN training records. Developing a written DMP was optional. The first offering of the flipped course was a success. To accommodate the number of students interested in attending, the library offered two classes for each of the five sessions. Eighty-three students enrolled in at least one of the five sessions. Attendance was a little over 50% on average for the series. Sixteen students (33% of attendees) completed all five sessions and received a certificate of data management in their UMN training history.

CONCLUSION

The results of this case study have been used to develop and implement several variations of online and flipped classroom instructional interventions. The UMN DIL team drafted a set of learning outcomes targeting the perceived greatest needs of graduate students that arose in the interviews. The partnering civil engineering faculty member vetted these out- comes and provided suggestions for involving students with the topic. Incorporating content from existing sources and tying instruction to federal requirements for data management, we developed a seven-module online course over three semesters.

The UMN librarians applied their expertise in organizing and managing information to the curation of research data. The civil engineering faculty member provided a reality check to en- sure that the skills would speak to the students’ experiences and fit within disciplinary norms. This partnership proved mutually beneficial, since the faculty could address a skill gap with- out creating the content to fill that gap. It gave the librarians a new way to engage with students and to introduce ourselves as resources for managing and sharing data.

This case study has been a starting point in the conversation of disciplinary norms. A replication or adaptation of this process ad- ministered more widely would gauge the DIL needs of students across institutions in the civil engineering field. Once the educational gaps have been identified, the ASCE’s BOK should be updated to address these skills.

Because the course lives online in a modular package, we were able to repurpose the pedagogy and teach the course in a way that better met student needs. Moreover, students can revisit the course material online and continue to develop their DMP through the openly accessible materials.

The course provides a framework for other librarians who hope to learn more about data management themselves or want to build learning objects for their institutions. Through the promotion of the DIL website, social media presence, and presentations at conferences, we have been in correspondence with librarians interested in examining what we are offering.

On our campus we've seen a hunger for guidance on these issues from both faculty and researchers. This is a natural extension of classic library services, including information classification and organization as well as information literacy instruction. DIL is a key component in the librarian's role on campus.

NOTES

Portions of this case study are reprinted with permission from:

Johnston, L., & Jeffryes, J. (2013, February 13). Data management skills needed by structural engineering students: A case study at the University of Minnesota. Journal of Professional Issues in Engineering Education and Practice. http://dx.doi.org/10.1061/(ASCE)EI.1943-5541.0000154;

and

Jeffryes, J., & Johnston, L. (2013). An e-learning approach to data information literacy education. Paper presented at the 2013 ASEE Annual Conference, Atlanta, GA. Available at http:// purl.umn.edu/156951.

This case study is available online at http://dx.doi.org.10.5703/1288284315479.

REFERENCES

Accreditation Board for Engineering and Technology (ABET). (2012). Criteria for accrediting engineering programs, 2012–2013. Retrieved from http://www.abet.org/DisplayTemplates/DocsHandbook.aspx?id=3143

American Society of Civil Engineers (ASCE). (2008). Civil engineering body of knowledge for the 21st century: Preparing the civil engineer for the future (2nd ed.). Reston, VA: American Society of Civil Engineers.

American Society of Civil Engineers Curriculum Committee. (2006). Development of civil engineering curricula supporting the body of knowledge for professional practice [Report]. Retrieved from http://www.asce.org/uploadedFiles/Committees_-_New/curriculacommreportdec2006.pdf

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals; Handbook 1: Cognitive domain. New York, NY: Longman, Green and Co.

Bransford, J., Brown, A. L., & Cocking, R. R. (Eds.). (1999). How people learn brain, mind, experience, and school. Washington, DC: National Academy Press.

Brenton, S. (2008). E-learning—An introduction. In H. Fry, S. Ketteridge, & S. Marshall, S. (Eds.). (2008). A handbook for teaching and learning in higher education: Enhancing academic practice. London, UK: Routledge.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011).Online formative assessment in higher education: A review of the literature. Computers & Education, 57(4), 2333–2351. http://dx.doi.org/10.1016/j.compedu.2011.06.004

Gruca, A. N. (2010). E-learning in academic libraries. New Review of Information Networking, 15(1), 16–28. http://dx.doi.org/10.1080/13614571003741395

Jeffryes, J., & Lafferty, M. (2012). Gauging work- place readiness: Assessing the information needs of engineering co-op students. Issues in Science and Technology Librarianship, (69). http://dx.doi.org/10.5062/F4X34VDR

Johnston, L., Lafferty, M., & Petsan, B. (2012).Training researchers on research data management: A scalable cross-disciplinary approach. Journal of eScience Librarianship, 1(2), Article 2. http://dx.doi.org/10.7191/jeslib.2012.1012

National Science Foundation (NSF) (2010). Scientists seeking NSF funding will soon be required to submit data management plans (Press release 10-077). Retrieved from http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928&org=NSF NEEShub (2009). NEES curation. Retrieved from http://nees.org/topics/NEESCuration

Oosterhof, A., Conrad, R., & Ely, D. P. (2008). Assessing learners online. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

Parr, C. (2013, May 10). Not staying the course. Inside Higher Ed. http://www.insidehighered.com/news/2013/05/10/new-study-low-mooc-completion-rates.

Safar, A. H. (2012). The students' perspectives of online training at Kuwait University. College Student Journal, 46(2), 436–458.

Stebbins, M. (2013). Expanding public access to the results of federally funded research. Retrieved from Office of Science and Technology Policy, the White House website: http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

Thyagarajan, R. (2012). NEEShub boot camp webinar—Data upload tools—January 2012. Retrieved from NEEShub website: http://nees.org/resources/4047

University of Minnesota Libraries. (2014). Managing your data. Retrieved from https://www.lib.umn.edu/datamanagement

University of Minnesota Research Education and Oversight. (2014). Core curriculum. Retrieved from http://research.umn.edu/reo/education/core.html#.Utr0zGTnYnU

U.S. Department of Education. (2010). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies [Report]. Retrieved from http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

Van Den Einde, L., Fowler, K., Krishnan, S., Rowley, J., Bhatia, K., Baru, C., & Elgamal, A. (2008). The NEES data model and NEES central data repository: A framework for data collaboration in support of earthquake engineering research. Paper presented at the Fourteenth World Conference on Earthquake Engineering, Beijing, China, October 12–17. Retrieved from http://www.iitk.ac.in/nicee/wcee/article/14_11-0168.pdf

Wong, J. M., & Stojadinovic, B. (2004, August). Structural sensor data repository: Metadata and Web-based user interface. Paper presented at the 13th World Conference on Earthquake Engineering, Vancouver, Canada. Retrieved from http://www.iitk.ac.in/nicee/wcee/article/13_956.pdf