Journal of eScience Librarianship
putting the pieces together: theory and practice

eScience in Action

A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs

Megan R. Sapp Nelson

Purdue University, West Lafayette, IN, USA

Abstract

Initial work in identifying data management or data information literacy skills generally went as far as identifying a list of proposed competencies without further differentiation between those competencies, whether by discipline, complexity, or use case. This article describes a significant innovation upon existing competencies by identifying a scaffolding (built upon existing competencies) that moves students progressively from undergraduate training through post graduate coursework and research to post-doctoral work and into the early years of data stewardship. The scaffolding ties together existing research that has been completed in research data management skills and data information literacy with research into the outcomes that are desirable for individuals to present in data management at each of the levels of education. Competencies are aligned according to application (personal, team, research enterprise) in such a way that the skills attained at the undergraduate level give students moving on to graduate work greater familiarity with data management and therefore greater likelihood of success at the graduate and then post graduate and data steward levels.

Introduction

What if we graduated data fluent students? The literature tells us that industrial partners do not consider our graduates to be data fluent (Van Tuyl and Whitmire 2016), faculty advisors do not consider graduate students to be prepared to manage data (Carlson and Stowell Bracke 2015), and early career faculty identify themselves as being unprepared to manage data in all areas of data stewardship (Sapp Nelson 2015a).

Therefore, what tools, strategies and approaches do the information professionals working to train the next generation of data managers need to be successful? This project began from this thought experiment. In order to graduate data fluent students, we, the faculty librarians at Purdue University have to define what it means for a student to be data fluent at graduation. In turn, we need a method of determining whether a student met that benchmark. A road map or plan to reach our students with those defined skills or competencies is necessary as well. Finally, we need tools to teach these skills or competencies to the students.

By examining the nascent discipline of data literacy (DL) or data information literacy (DIL) from the lens of this thought experiment, it is clear that the first step, to define what it means for a student to be data literate at graduation is addressed, at least in part by competencies (Qin and D'Ignazio 2010b, Carlson et al. 2011, Piorun et al. 2012, Calzada Prado and Marzal Miguel 2013, Schneider 2013). To date, five competencies have addressed the overarching vision for what students should know to be considered data literate. However, not one of the competency lists defines at what point all students must know the entirety of the data management competencies. Should it be at the end of their undergraduate career? Should students begin to learn data management skills in their graduate education and only be expected to master them once they were responsible for managing data as part of research projects? Should there be some competencies that are mastered at each level? This distinction of the timing of educational interventions and learning has not yet been addressed in a comprehensive way in the literature.

The tools to teach skills or competencies to students have been steadily emerging in the literature over the past 5-7 years. There is a wealth of creative pedagogy emerging from our field that addresses data management instruction in interesting and innovative ways (Athanases, Bennett, and Wahleithner 2013, Hogenboom, Phillips, and Hensley 2011, Johnston and Jeffryes 2014, Mandinach and Gummer 2013, Whitmire 2015, Wright and Andrews 2013).

The missing piece in this thought experiment is the road map. This "road map" or plan defining when we reach specific learners with specific competencies is lacking in the literature. The pilot scaffolding presented here seeks to fill in an important piece in the long progression toward an organized, structured educational strategy to ensure that all learners master data management skills.

Review of literature

As noted above, competencies were among the first piece of the data information literacy landscape to be offered to the community for critical reflection and the development of pedagogy. These competency proposals for the most part map to each other (See Appendix 1). A crosswalk of five competency proposals (Schneider 2013, Carlson et al. 2011, Qin and

D'Ignazio 2010a, Piorun et al. 2012, Calzada Prado and Marzal Miguel 2013) shows that the competencies mirror each other in the need to teach learners about the form and format of data and databases, the management of that data, the curation and reuse of data, and the metadata used to describe that data. At least three of the five agree in teaching the cultures of practice around data use, the preservation of data, the analysis of data, the visualization of data, and the ethics of data production and use. Overall, there is strong consensus in the literature of competencies regarding the major themes that should be addressed in data education.

Multiple authors mapped their competencies with the ACRL Information Literacy Competencies for Higher Education (Carlson et al. 2011, Calzada Prado and Marzal Miguel 2013). This effort shows that the overlap between existing information literacy efforts and data information literacy efforts are significant in areas such as basic data recognition and acquisition skills, and interpreting and presenting quantitative information (Calzada Prado and Marzal Miguel 2013). This indicates that, at least at the undergraduate level, it may be possible to "teach data information literacy" at the same time as information literacy is presented.

The Data Information Literacy Project called for the mapping of competencies across the curriculum and the levels of education (Sapp Nelson 2015b) but did not complete that portion of the work. Only Schneider went so far as to map which competencies should be taught to specific learner groups.

Schneider identified that undergraduates should be taught cultures of practice and documentation skills (Provide and Identify), that masters' students should learn those skills as well as Analysis and Metadata (Scope and Plan), and that data creators should have the same training as undergraduates (Schneider 2013). However, specific learning objectives and skills building for any of the levels were not identified within Schneider's competency proposal. It also didn't seem sufficient that a data steward (or data creator) should be given only the training that undergraduates receive in cultures of practice and documentation skills, when data stewards are legally responsible for the data. While this curricula matrix presented by Schneider was helpful, it didn't fully meet the needs of data educators.

Specifications for a new kind of competency matrix

Based upon an examination of the existing competencies, the DIL community had a basic framework at their disposal but there was a need for more explicit measurable, observable, and transferable learning objectives, similar to those that librarians had been used to working with (i.e. the ACRL Information Literacy Competency Standards (Libraries 2000) or the AACU Information Literacy VALUE Rubric (Universities 2009)). This format enables librarians to develop specific assessments for learners that are measurable and observable, and in turn allows for the tracking of skills mastery over time.

There is also a need for identifying the point when specific competencies should be introduced. There are multiple approaches to implementation for this aspect, including by undergraduate, graduate, and data steward, or competency by competency. Ultimately, competencies that move from a personal to a team to a research enterprise domain were selected over a specific learning time period (such as undergraduate, graduate, and data steward) based on feedback from librarians (See Figure 1).

Figure 1: Domains Represented in the Matrix
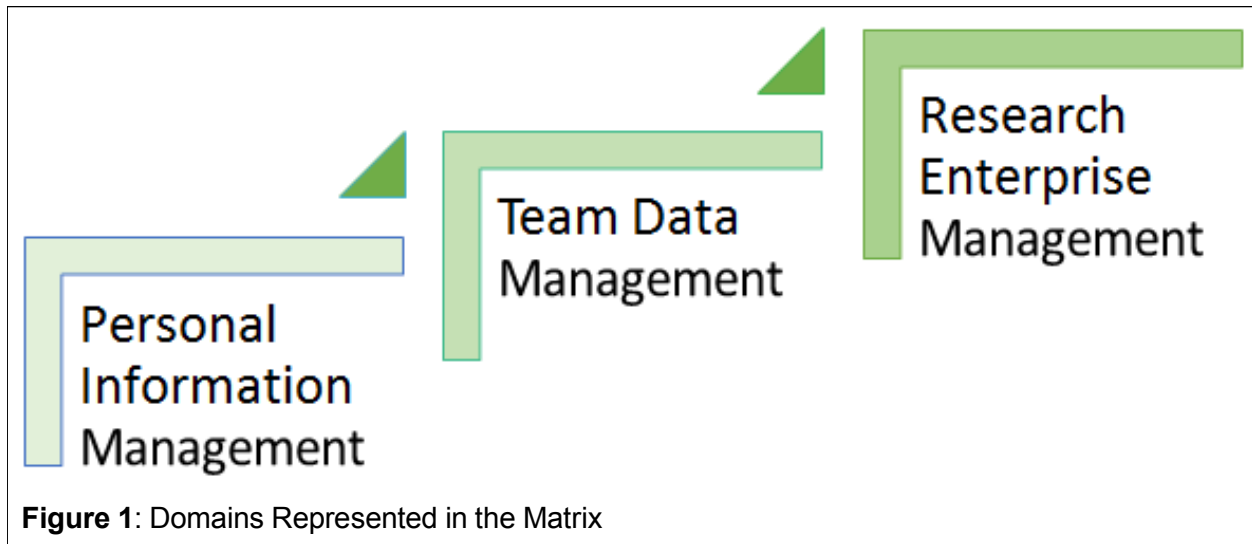
In the early years of the undergraduate level, learners are primarily interested in managing their own files successfully. The ruling domain is the personal and personal information management is the most meaningful type of data for these learners. A good goal statement for this group might be: When they graduate, can they manage their own collection of photos or tax documents? Can they transfer those skills to a desktop computer at their place of work? This domain continues to be important for lifelong learning for all data managers and will be an area where continued skills development occurs over time as new technologies emerge which allow for more sophisticated or efficient personal information management.

Learners at the later undergraduate (i.e. design and active learning classes) and graduate level are interested in successfully managing data within a research group or team, so team data management is a ruling interest for this group. A good goal statement might be: Can these learners manage data cooperatively to meet common goals? The skills built at this level are built around sharing internally with a team efficiently and effectively. Learners who master this skill may go into the workforce prepared to share data within the team environment of the workplace as well.

Learners at the data steward level are encountering a new level of complexity as they try to organize the data management of many individuals working on a common research endeavor. They are managing data at the research enterprise level. They have "The Buck Stops Here" levels of responsibility for the data, and so they not only need to know how to manage the data cooperatively, but how to teach and lead others to manage the data. A good goal statement for these learners might be: Can the learner make decisions about data knowledgably and efficiently? Can they teach members of the team a rationale and the routines resulting from those decisions? At this point, the learner has mastered most, if not all, data management skills to the point of being able to direct others to resources and find assistance with key tasks.

Some disciplines may primarily focus on building personal information management skills within their majors while other disciplines may require team or enterprise level skills building from their students early in their undergraduate careers. Due to the intended flexibility of the matrix, without a specified time range for when students should or must have specific skills,

disciplines can use this framework to determine within their disciplinary context what the term "data literate" or "data fluent" means for their students and their disciplinary pedagogy.

Learners should progress in their learning in ways appropriate to their current knowledge building situation, a learning philosophy termed "informed learning" (Bruce and Hughes 2010). Data education is ripe for informed learning (Maybee and Zilinski 2015) with the research based, hands on nature of many of the interactions with data. Learners frequently ask for exercises or case studies, but those are difficult to tailor to all learners at all domains and contexts. If the competencies are instead metered out across the domains, the cases and exercises become much more targeted and relevant to the learners consuming the educational content. Librarians have greater guidance and therefore a better ability to customize templates or cases to specific disciplinary needs if a discipline is able to identify which specific skills their students should have in order to be "data fluent".

Development of a pilot competency matrix

With the above goals for the competency list identified, a significant difficulty in the development of the competency matrix was simply to figure out what it would look like. The author was not familiar with any models that would allow for all of the requirements to be effectively displayed. After some searching, the author identified a vocational education overhaul that had mapped the educational requirements as competencies measured on Bloom's cognitive, psychomotor, and affective domains (Euler and Frank 2011). This allowed for "modularization" or as the author had been framing it "progression across domains", as well as a way of measuring educational attainment at the individual level. With that model in mind, the development of the pilot competency matrix began to take shape.

Because the Carlson et al. 2011 competencies were the most comprehensive in the crosswalk exercise, those competencies were used as the basic structure from which the matrix was built. The matrix is actually comprised of three matrices side by side, one each for personal, team, and research enterprise (Figure 2).
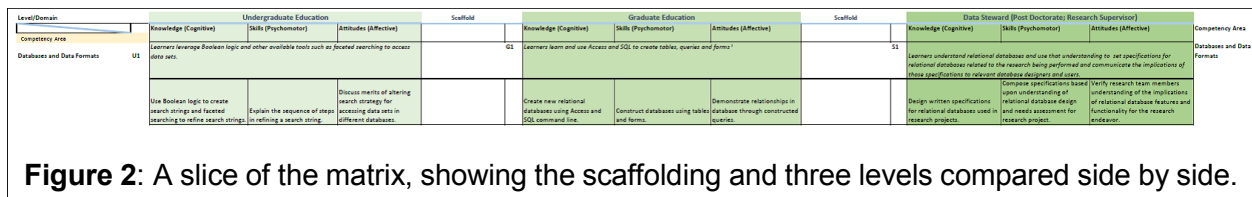
Figure 2: A slice of the matrix, showing the scaffolding and three levels compared side by side.

Each competency is articulated fully as an overarching goal and then has specific behaviors that are described on Bloom's cognitive (Knowledge), psychomotor (Doing), and affective (Attitudes) domains.

The knowledge, skills, abilities (KSAs) were gathered from various articles in the literature or created based upon the author's personal experience. The references are included in superscripts within the competencies and KSAs, as well as in a separate references tab in the spreadsheet.

These KSAs are measurable. This means that each competency can in theory be assessed in multiple ways. In a competency based framework, assessment is ideally based upon a yes or

no question answering "Does the learner meet the cognitive, psychomotor, or affective domains for that competency?"

Undergraduate Education		
Knowledge (Cognitive)	Skills (Psychomotor)	Attitudes (Affective)
U1 *Learners leverage Boolean logic and other available tools such as faceted searching to access data sets.*		
Use Boolean logic to create search strings and faceted searching to refine search strings.	Explain the sequence of steps in refining a search string.	Discuss merits of altering search strategy for accessing data sets in different databases.

Figure 3: Competency P1, including the competency as written and the behaviors on the cognitive, psychomotor, and affective domains.

Many of the competencies in the personal domain overlap with competencies that are traditionally considered to be those of information literacy/information fluency. In Figure 3 above, competency P1 significantly overlaps ACRL Information Literacy Competency Standards for Higher Education (Libraries 2000), competency 2.d. (Note: This manuscript and tool was drafted prior to the rescission of the ACRL Information Literacy Competency Standards).

The competencies as read from left to right build upon each other from the personal across the team and to the research enterprise levels (See Figure 4 for an example.). Notice how it builds

Graduate Education		
Knowledge (Cognitive)	Skills (Psychomotor)	Attitudes (Affective)
G1 *Learners learn and use Access and SQL to create tables, queries and forms [1]*		
Create new relational databases using Access and SQL command line.	Construct databases using tables and forms.	Demonstrate relationships in database through constructed queries.

Figure 4: Competency T1.

off of the basic skills found in P1 but doesn't repeat those skills. The skills may build in complexity, or may build in responsibility. Either is accurate and reflects the multiple domains (Personal, Team, Research Enterprise) that are being reflected in the competencies. The competencies are numbered from 1 to 36; each is also labelled as P (Personal), T (Team), or E (Research Enterprise), for an alphanumeric code that allows the user of the matrix to clearly communicate about which specific domain competency is being discussed.

Data Steward (Post Doctorate; Research Supervisor)		
Knowledge (Cognitive)	Skills (Psychomotor)	Attitudes (Affective)
S1 *Learners understand relational databases and use that understanding to set specifications for relational databases related to the research being performed and communicate the implications of those specifications to relevant database designers and users.*		
Design written specifications for relational databases used in research projects.	Compose specifications based upon understanding of relational database design and needs assessment for research project.	Verify research team members understanding of the implications of relational database features and functionality for the research endeavor.

Figure 5: Competency E1. The mastery level for this competency.

Between each matrix is a column labeled "scaffolding". This column is a placeholder indicating the actual work of many liaison and instruction librarians. As learners move from one phase to another, a learning scaffolding must be applied to build on existing skills and lift them to the next level of data skills (Dennen and Burner 2008). Scaffolding includes educational interventions such as learning objects, tutorials, modules, and in person or online instruction. Our discipline has been producing many learning objects and curricula, the most well-known of which include The Data Information Literacy project (www.datainfolit.org), New England Collaborative Data Management Curriculum (NECDMC) (http://library.umassmed.edu/necdmc/index), DataONE education modules (https://www.dataone.org/education-modules), and Research Data Mantra (http://datalib.edina.ac.uk/mantra/). Additionally, many librarians are creating individual courses and curricula for teaching data management. All of these education interventions will serve as the "lifting force" to bring learners from one competency level to the next as they progress throughout their career. The curricula just needs to be mapped to the competencies, to identify which learners will be addressed with the existing curricula.

The mastery levels of the matrix often change the skills to focus on teaching, instruction, supervision, strategic planning, or supervisory overview (See Figure 5 for an example). Notice how it refocuses for a different domain, as well as a different context, focusing on communicating with a research group and end users as well as performing research. This switch to *directive data management* frequently occurs in the mastery level of the data competency matrix.

The resulting competency matrix can be viewed and downloaded for reuse at: https://purr.purdue.edu/publications/2186/1.

Discussion

If the matrix were adopted by an institution as a guiding document for the development of an enhanced Data Information Literacy program, a number of potential advantages emerge. A primary advantage is having a document to start communications with faculty and departments about their current data management practices (both internal to the curricula and extracurricular), and what the department leaders and faculty would like to include through workshops or course integrated instruction. Even if the disciplinary faculty or department disagree with the content of the matrix, it will give them a concrete document to build on to craft their own version.

Relationships with specific stakeholders in areas of the institution such as Graduate Colleges or Honors Colleges can be enhanced by presenting competency based educational plans to meet the specific needs of their population of learners. Additionally, specific assessment needs for those stakeholder groups could be built into the partnerships from the beginning by communicating which Knowledge, Skills, Abilities (KSAs) should be measured during the learning interventions.

The document also provides opportunities for librarians to plan their own instruction more strategically. Online tutorials can be developed to facilitate the scaffolding between key transition points. By identifying specific competencies that are important to a broad range of stakeholders at an institution, library faculty and staff concentrate limited educational technology time and resources at maximal points of instructional effectiveness and can demonstrate impact based upon research and assessment.

Assessment can be built into DIL learning interventions by basing outcomes on concrete, observable tasks modeled on the KSAs language included in the matrix. This mode of assessment lends itself to badging and certificate program development, either within the libraries or in collaboration with disciplinary departments.

It is also possible for libraries to measure progress for specific learner groups along the matrix in collaboration with specific faculty or as long term assessment projects of the library. These allow the library to establish the impact of data information literacy instruction programs to the larger institution but also demonstrate the effectiveness of library instruction in a key institutional instruction area.

Existing data management services can identify gaps in available instructional objects and services that teach specific competencies. Existing learning objects and pedagogies can be compared to the matrix and indexed according to the competencies that they currently meet, without any further revision. All of the competencies that are currently addressed by instruction can then be indexed and gaps targeted for the creation of new instruction or for addition to current instruction modules. Additionally, if one domain (personal, team, or research enterprise) is significantly underrepresented in current instruction and outreach programs, those areas can be targeted for curricular development.

Additionally, liaisons or other librarians who were previously unengaged in data information literacy instruction may be introduced to the area by identifying concrete DIL skills that are similar to or overlap with information literacy skills (such as Boolean search strings or citation

of datasets), and then can be included within information literacy instruction sessions in the personal data management domain. Identifying skills that can be naturally integrated within undergraduate information literacy instruction gives liaisons who are less confident in teaching data management skills low hanging fruit to gain data management instruction experience.

At competency based institutions, learners may develop learning plans to self-direct their research data management skills acquisition by performing skills inventories based upon the matrix. Libraries can provide concrete, easily accessible educational interventions to pursue on their own time, leading to the acquisition of badges or certificates.

Finally, the matrix provides a lens that gives insight to specific, actionable intelligence on how to improve data services (i.e. Which domains are being currently served? Which domains are under- represented in current outreach? Which tools or instructional modules need updating or development?). This in turn gives administrations the evidence necessary to justify positions and resources for research data management. By showing/demonstrating the extent of key constituencies' DIL skills attainment and projecting where the addition of another position could meet those needs, library administrations would be in a better position to positively and proactively make the case for further data management support.

Next steps in the project

This pilot and tool is a necessary first step to develop a larger, structured data education program at an institution. The author presented this matrix to interested librarians at Purdue University. Significant editing to reflect the local situation is likely. The local situation will likely be specific to an institution or at the department level. This document doesn't indicate how the matrix should be used or with whom, and does encourage that the matrix be edited to reflect local nuances.

The author has a lens of a science and engineering librarian. While the author has been exposed to social science data work, the vast majority has been in technical disciplines. Therefore, these proposed competencies may have a science and engineering focus that though unintentional, will require translation to a humanities or social sciences data education program. Hence, changes are to be expected and encouraged.

The author expects that a matrix such as this would be a living document, undergoing revisions to reflect changes in data management practice, technology, scientific culture, and institutional and government mandates. Additionally, each institution could potentially have their own "flavor" of competencies that are important to address among their learners.

The matrix is intended to be a jumping off point for developing programs. However, it is not yet a program until librarians and data managers take the step of mapping curricula with the matrix competencies, identifying competencies not present in the curricula, inserting those competencies in the manner most appropriate to the local situation, and then teaching the curricula while assessing students for attainment of the learning outcomes specified in the curricula. This is a large endeavor that still has many steps to be accomplished before it is wholly implemented.

Conclusion

This matrix of data education competencies presents the opportunity to create a systematic approach to data management education. Facilitating the creation of data curricula enabling learning throughout the curriculum, tracking the growth of individuals across their academic careers, and grounding communication between instructors regarding which learners have been taught what, and by whom; and assessing the level at which those learners are successful in meeting set educational goals for data management education, will establish the library's commitment to DIL. With a more rigorous structure, it is the hope of the author that the data information literacy community will be better able to meet the needs of learners at all levels and in all domains. Please take this tool and use it to create rigorous data literacy programs appropriate to specific local situations in order to graduate data fluent researchers at all levels of education.

Supplemental Content

Appendix 1
An online supplement to this article can be found at http://dx.doi.org/10.7191/jeslib.2017.1096 under "Additional Files".

Disclosure

The author reports no conflict of interest.

References

Athanases, Steven Z., Lisa H. Bennett, and Juliet Michelsen Wahleithner. 2013. "Fostering Data Literacy Through Preservice Teacher Inquiry in English Language Arts." *The Teacher Educator* 48(1): 8-28. http://dx.doi.org/10.1080/08878730.2012.740151

Bruce, Christine and Hilary Hughes. 2010. "Informed learning: A pedagogical construct attending simultaneously to information use and learning." *Library & Information Science Research* 32(4): A2-A8. http://dx.doi.org/10.1016/j.lisr.2010.07.013

Calzada Prado, Javier and Ángel Marzal Miguel. 2013. "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents." *Libri* 63(2): 123-134. http://dx.doi.org/10.1515/libri-2013-0010

Carlson, Jake, Michael Fosmire, C.C. Miller, and Megan Sapp Nelson. 2011. "Determining data information literacy needs: A study of students and research faculty." *Portal: Libraries & the Academy* 11(2): 629-657. http://dx.doi.org/10.1353/pla.2011.0022

Carlson, Jake and Marianne Stowell Bracke. 2015. "Planting the seeds for data literacy: lessons learned from a student-centered education program." *International Journal of Digital Curation* 10(1): 95-110. http://dx.doi.org/10.2218/ijdc.v10i1.348

Dennen, Vanessa P. and Kerry J. Burner. 2008. "The Cognitive Apprenticeship Model in Educational Practice." In *Handbook of Research on Educational Communications and Technology* edited by J. Michael Spector, M. David Merrill, Jeroen van Merriënboer and Marcy P. Driscoll, 425-439. Abingdon: Routledge Handbooks Online. http://dx.doi.org/10.4324/9780203880869

Euler, Dieter and Irmgard Frank. 2011. "Courage or audacity? - Modularization and competence-orientation as key objectives of initial vocational education reform in Luxembourg." *Bonn, Germany Federal Institute for Vocational Education and Training (BIBB)*. https://www.bibb.de/en/23335.php

Hogenboom, Karen, Carissa M. Holler Phillips, and Merinda Kaye Hensley. 2011. "Show me the data! Partnering with instructors to teach data literacy." *Declaration of Independence: The Proceedings of the ACRL 2011 Conference, Philadelphia, PA.* http://www.ala.org/acrl/files/conferences/confsandpreconfs/national/2011/papers/show_me_the_data.pdf

Johnston, Lisa and Jon Jeffryes. 2014. "Data management skills needed by structural engineering students: Case study at the University of Minnesota." *Journal of Professional Issues in Engineering Education and Practice* 140(2). http://dx.doi.org/10.1061/(ASCE)EI.1943-5541.0000154

Libraries, Association of College and Research. 2000. "Information Literacy Competency Standards for Higher Education." Chicago: Association of College & Research Libraries. http://www.acrl.org/ala/mgrps/divs/acrl/standards/standards.pdf

Mandinach, Ellen B. and Edith S. Gummer. 2013. "A Systemic View of Implementing Data Literacy in Educator Preparation." *Educational Researcher* 42(1): 30-37. http://dx.doi.org/10.3102/0013189X12459803

Maybee, Clarence and Lisa Zilinski. 2015. "Data informed learning: A next phase data literacy framework for higher education." *Proceedings of the Association for Information Science and Technology* 52(1): 1-4. http://dx.doi.org/10.1002/pra2.2015.1450520100108

Piorun, Mary, Donna Kafel, Tracey Leger-Hornby, Siamak Najafi, Elaine Martin, Paul Colombo, and Nancy LaPelle. 2012. "Teaching Research Data Management: An Undergraduate/Graduate Curriculum." *Journal of eScience Librarianship* 1(1): e1003. http://dx.doi.org/10.7191/jeslib.2012.1003

Qin, Jian and John D'Ignazio. 2010a. "The Central Role of Metadata in a Science Data Literacy Course." *Journal of Library Metadata* 10(2/3): 188-204. http://dx.doi.org/10.1080/19386389.2010.506379

Qin, Jian and John D'Ignazio. 2010b. "Lessons learned from a two-year experience in science data literacy education." *International Association of Scientific and Technological University Libraries, 31st Annual Conference. Paper 5.* http://docs.lib.purdue.edu/iatul2010/conf/day2/5

Sapp Nelson, Megan. 2015a. "Data Management Outreach to Junior Faculty Members: A Case Study." *Journal of eScience Librarianship* 4(1): e1076. http://dx.doi.org/10.7191/jeslib.2015.1076

Sapp Nelson, Megan. 2015b. "Where do we go from here?: Further developing the Data Information Literacy Competencies." In *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, edited by Jake Carlson and Lisa R. Johnston, 231-245. West Lafayette, IN: Purdue University Press. http://www.thepress.purdue.edu/titles/format/9781612493527

Schneider, René. 2013. "Research data literacy." In *Worldwide Commonalities and Challenges in Information Literacy Research and Practice*, 134-140. Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-03919-0_16

Universities, Association of American Colleges and. 2009. "Information Literacy VALUE Rubric." *Association of American Colleges and Universities*. http://www.aacu.org/value/rubrics/pdf/InformationLiteracy.pdf

Van Tuyl, Steven, and Amanda Whitmire. 2016. "Investigation of private industry and non-academic data management practices." Presented at *Research Data Access & Preservation Summit, Atlanta, GA, May 4, 2016.*

Whitmire, Amanda. 2015. "Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned." *Journal of Librarianship and Scholarly Communication* 3(2): eP1246. http://doi.org/10.7710/2162-3309.1246

Wright, Sarah J. and Camille Andrews. 2015. "Natural Resources / Graduate Students / Wright & Andrews / Cornell University / 2013." *Data Information Literacy Case Study Directory* 2(1): Article 1. http://dx.doi.org/10.5703/1288284315476