3-17-2015

Data Information Literacy and Undergraduates: A Critical Competency

Yasmeen Shorish
James Madison University, shorisyl@jmu.edu

Data Information Literacy and Undergraduates: a critical competency

Yasmeen Shorish[1]

Abstract

As a primer on data information literacy, this column will cover the background of the field and why it is relevant to college and university libraries serving undergraduate populations. This article includes how data information literacy (DIL) relates to information literacy, competencies associated with DIL, the relevance of DIL to undergraduates, DIL in library instruction, and the reasons for library engagement with DIL. Examining DIL within the larger framework of information literacy can help outreach and instruction librarians engage with a format that may be unfamiliar, but whose underlying foundation is well established.

Keywords: information literacy, data information literacy, data management, outreach, instruction, higher education, academic libraries, data curation

Introduction

Academic libraries serve to provide their communities with information resources that meet teaching, learning, and research needs. These information sources have changed format in response to technological advances: incunabula to printed text; print journal articles to PDFs. In addition to providing access to these sources, librarians serve to increase our communities' ability to effectively use this information. As libraries and librarians continue to support scholarship, it is critical to keep abreast of the changes in the scholarly communication landscape, including the role of data as a research source and output. Data traditionally have not

[1] Physical and Life Sciences Librarian; James Madison University, Rose Library Harrisonburg, VA 22807; shorisyl@jmu.edu

been considered a scholarly product to be communicated as a stand alone product. Researchers produce datasets as part of their own research process with little to no intention of sharing them with others, for a variety of reasons (Borgman, Wallis, and Enyedy 2007). However, as research has become increasingly collaborative and networked, data have become more valuable as a scholarly product with potential for reuse. For example, the National Science Foundation (NSF) replaced "publications" with "products" in the instructions for preparing a biographical sketch for a grant proposal, which can include citable and accessible data (National Science Foundation 2007, Chapter II.C.2f(c)). Efforts to standardize data citation, such as the Joint Declaration of Data Citation Principles (Force11 n.d.), illustrate how datasets are becoming recognized as standalone scholarly objects. This is especially true in the digital humanities, where the manipulation of existing datasets into new knowledge and new datasets forms the scholarly product of the field (Munoz 2013). As the value of data as a publicly-accessible research output increases, so does the demand for the skills required to make full use of this resource.

Considering that funding agencies require data management plans to be submitted with grant applications, and undergraduates are increasingly exposed to the research environment, it is critical that researchers (including students) become fluent in the description, organization, and overall management of research data, including its reuse. Science data literacy has been defined as the ability to collect, process, manage, evaluate, and use data (Qin and D'Ignazio 2010). These can be thought of as the actions of a "data consumer." As we enter more collaborative and interdisciplinary workspaces, it becomes equally important to address the issue of data sharing – the actions of a "data producer." For this reason, Carlson et al (2011) define data information literacy (DIL) as merging "the concepts of researcher-as-producer and researcher-as-consumer," building upon the foundation of science data literacy, as well as other established literacies that

focus on the consumer side, such as data, statistical, and information literacies (634). Employing these skills enables a researcher to fully engage with every step of the research lifecycle (Fig. 1)

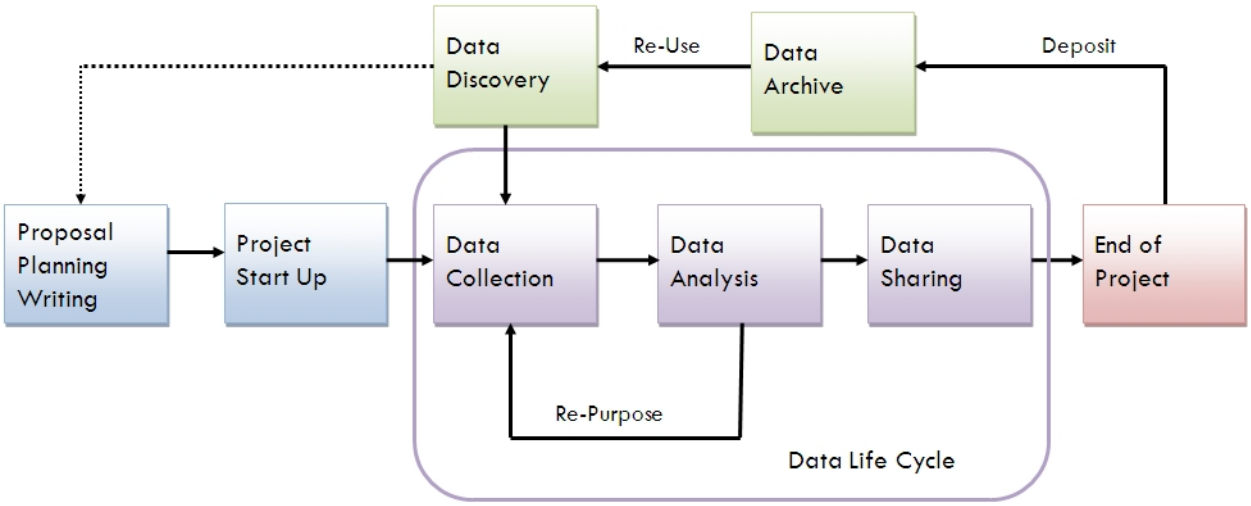

Figure 1: Research Lifecycle (University of Virginia Library n.d.)

Discussions around data sharing and reuse can appear to be limited to issues within "e-Science." While e-Science has been defined as a way of conducting scientific research in a collaborative and networked environment (Hey and Hey 2006), there is no reason to limit this methodology to science. Increasingly, research projects are interdisciplinary and data-driven projects are found in the social sciences and humanities (Association of College and Research Libraries: Working Group on Intersections of Scholarly Communication and Information Literacy 2013). Moreover, scientific research is increasingly team-based, distributed, and networked across institutions and nations. Perhaps it is time to drop the term e-Science altogether and acknowledge that technology has changed the way that research occurs. With these changes in technology, it is possible for unanticipated groups to take an interest in one's research data. For example, an ecologist's data on chemicals found in a waterway may be useful to an epidemiologist tracking environmental factors for a disease. The epidemiologist is not the primary audience for the ecologist, especially if the publication from the data does not highlight a chemical of interest. Nonetheless, the data gathered would be of value to this unintended

audience. With the awareness that data are created in all disciplines and can increasingly be shared electronically and re-used by unintended audiences, the importance of data documentation (Parsons and Duerr 2005) and DIL comes into focus.

The Evolution of Information Literacy and DIL

In order to equip students with the skills that they need to navigate the research landscape, and eventually their professional lives, libraries have engaged in information literacy training. Information literacy addresses how people learn, and gives students skills to "locate, evaluate, and effectively use information for any given need" (Association of College and Research Libraries 1989). Changes in information delivery formats have led to conversations around visual literacy and digital literacy, recognizing that the traditional application of information literacy has not been able to address the nuances of these associated competencies effectively.

Our profession recognizes these changes in the scholarly landscape, as evidenced in part by the decision to revise the Association of College and Research Libraries (ACRL) *Information Literacy Competency Standards for Higher Education*. Approved in 2000, these standards reflected the landscape of higher education at that time (Association of College and Research Libraries 2000). However, higher education is an evolving ecosystem, requiring constant attention. To that end, ACRL has begun to review the information literacy standards and create a new "framework," taking into account changes such as student population demographics, approaches towards student learning and team-based assignments, and the increasing importance of undergraduate research (Association of College and Research Libraries 2014). This new document, *Framework for Information Literacy for Higher Education*, attempts to take a more

holistic approach towards engagement with information, specifically noting data as a product with which students should gain proficiency.

This shift indicates a more intentional treatment of lifelong learning, as the *Framework* uses threshold concepts to couch activities from the points of view of both the information consumer and the producer. This is in line with the defining characteristic of data information literacy (Carlson et al. 2011), that it equips practitioners with the skills to make data reusable to others in a meaningful way. Data information literacy, with its characteristics of data documentation, preservation, and sharing, requires engagement with a wide array of information skills. To that end, DIL should be treated as any of the other literacy competencies and incorporated into the workflow of outreach librarians, with the acknowledgement that this may require input from other library units, such as metadata and scholarly communication (Association of College and Research Libraries: Working Group on Intersections of Scholarly Communication and Information Literacy 2013).

DIL Competencies

When thinking about the skills necessary to build data information literacy, it may be useful to consider the goal outlined in a draft of the forthcoming information literacy framework:

> Information literacy is a repertoire of understandings, practices, and dispositions focused on flexible engagement with the information ecosystem, underpinned by critical self-reflection. The repertoire involves finding, evaluating, interpreting, managing, and using information to answer questions and develop new ones; and creating new knowledge through ethical participation in communities of learning, scholarship, and practice. (Association of College and Research Libraries 2014, 4)

The information-consumer language of the original standards has been replaced to illustrate an active, information-producer role. Any DIL competencies should support this effort.

In the literature, one can find two approaches to DIL that initially appear quite different. The Data Information Literacy project funded by the Institute for Museum and Library Services involved the librarians at Purdue University, the University of Minnesota, Cornell University, and the University of Oregon. That project identified twelve core competencies that cover tool-based areas, such as data processing and analysis, databases, data discovery, data visualization, data quality, and data conversion and interoperability, as well as theory-based areas like data management, data preservation, data curation and reuse, metadata, cultures of practice, and ethics (Carlson et al. 2011). The approach of Calzada Prado and Marzal (2013) detailed five competencies couched in the familiar "understand, find, evaluate, and use" framework shared by ACRL's *Information Literacy Competency Standards for Higher Education*, with an additional entry for "managing data" (131).

Both approaches touch on areas where instruction should occur within the discipline, such as understanding data or analyzing data. These competencies are critical and are best taught by faculty within the context of the subject. The library community could capably address other competencies, such as ethics and preservation. Ethically using information has been foundational to information literacy instruction, often in the form of instruction about citation formatting. While there is currently no body of authority setting DIL standards, the two approaches outlined above provide the DIL community with a foundation upon which to build best practices.

DIL & Undergraduates

Observations of graduate student behavior at research-intensive universities are analogous to the behavior of undergraduates involved in research at institutions that do not have large graduate populations. While undergraduates may have less experience in research methods,

and may not be involved in as complex research studies (although this is not always true), the "freshness" of the undergraduate offers new opportunities. These students are just beginning to learn and understand research methodology; what better time to introduce data information literacy concepts to their workflow? Carlson and Bracke noted that graduate students at Purdue are involved in the collecting, processing, and analyzing of research data (2013). At colleges and universities with limited or no graduate student population, undergraduates fill this role.

Another concern that has been well documented in the data management and data curation literature is the issue of data inheritance (Carlson and Stowell-Bracke 2013; Doucette and Fyfe 2013; Carlson et al. 2013). Faculty direct multi-year research endeavors that can involve multiple cohorts of student research teams through the life of the experiment or project. As a result, work is built upon the data collected and documented by previous students. The amount of time that must be spent translating previous students' notes and processes could be considerable if there were no data management practices in place. Data sets may be opaque to a new student researcher and require mediation by the Principal Investigator (PI), or in the worst-case scenario may be completely unusable.

Institutions without a focus in undergraduate research still have a reason to engage with DIL. Carlson et al (2013) found that faculty expected graduate students to possess data management skills before working in their lab, either through prior experience or through their undergraduate education. However, other interviews indicated that faculty found graduate students ill-equipped in this area as the students lacked the skills and training necessary to effectively manage research data (Carlson et al. 2011). Obviously, there is a disconnect between the skills faculty in graduate research labs think students should learn in their undergraduate education and the competencies that those students actually have.

Data information literacy skills are relevant even if students do not go on to advanced degrees. The majority of individuals receiving post-secondary education in the United States seek a bachelor's degree as their terminal degree. In 2011-12, 1,791,046 bachelor degrees were conferred, more than twice the number of master's degrees and more than ten times the number of doctoral degrees conferred in that same period (Snyder and Dillow 2013). Moreover, these skills are critical to most aspects of business today. An analysis conducted by Gartner, a major information technology research company, found that business leaders, CIOs and compliance officers must adopt data management best practices in order to be cost-effective and agile (Dayley and Childs 2012). A subsequent trend report found that as more businesses rely on data manipulation, so-called "big data" will "become business as usual" (Buytendijk 2014, 1). Therefore, as one seeks to create a more informed and productive citizenry, one should seek to expose all college graduates to the skills required to effectively evaluate and use data.

DIL Instruction

As with information literacy instruction, there are many different ways to deliver data information literacy instruction to undergraduates. While library-based data management and DIL instruction is relatively recent, it has primarily focused on instruction to graduate students and faculty. This instruction has most often taken the form of seminars or workshops offered through the library as drop-in sessions or one-shot class instruction (Carlson et al. 2013). However, there has been some exploration of instruction at the undergraduate level.

In 2008 and 2009, Syracuse University offered an NSF Course, Curriculum, and Laboratory Improvement grant-supported course in science data management (Qin and D'Ignazio 2010). Focused on science data literacy, this course was open to both undergraduates and

graduate students. Qin and D'Ignazio found that the mixed audience class was challenging, as undergraduates had trouble fitting the class into their heavily proscribed schedules and it was difficult to deliver content at a useful level for all students (2010). The University of Massachusetts Medical School and Worcester Polytechnic Institute used funding from the Institute of Museum and Library Services (IMLS) to develop a curriculum framework that would address data management at the undergraduate and graduate levels. By creating modules, delivery to student populations could be tailored based on the experience level and need, since undergraduates may require more modules than graduate students (Piorun et al. 2012). In some environments, it may be most effective to target the instruction to research groups on campus instead of using class time to cover the material. Meeting with research teams on an individual or department level offers an opportunity to discuss in more detail the disciplinary nuances associated with data management. This may help to avoid some of the issues that arose in the Syracuse science data class due to disciplinary disparities.

Given the data management requirements of funding agencies, institutions are developing tools and resources. Many of these resources have been created with the PI or graduate student in mind, but there are several ways to adapt them to the undergraduate population. A proposed outcome from the aforementioned IMLS-funded DIL project is a model for librarians to use in developing DIL programs at their own institutions (Purdue University et al. 2013). Updates are posted to the project's website, where one can review handouts and recordings of the sessions from the 2013 DIL Symposium (Purdue University et al. 2013). The Symposium focused on exploring the DIL competencies with librarians and developing strategies for data management engagement with faculty and students.

Multiple-session DIL courses that can be studied and adapted for the undergraduate student include the University of Minnesota's flipped data management course (Johnston and Jeffryes 2014), the New England Collaborative Data Management Curriculum (Lamar Soutter Library, University of Massachusetts Medical School n.d.), the University of Edinburgh's MANTRA online course (EDINA and Data Library, University of Edinburgh n.d.), and the education modules available from DataONE (DataONE 2012). These are most often presented as sections covering a specific point in the research process, such as defining data, management planning, organizing, and sharing. As discreet units, it is possible to select the sections that are most applicable to an undergraduate student population of varying disciplines who are at different points in their research education.

An economics professor and a librarian at Haverford College have developed a protocol for the explicit purpose of teaching data management to undergraduates conducting empirical research (Ball and Medeiros 2012). This approach seeks to encourage students to integrate data management skills into their research practice. Ball and Medeiros suggested that teaching undergraduate students data management as a basic part of research helps assign responsibility and accountability to the findings (2012). This protocol is an exceptional example of how delivering DIL instruction can occur within a discipline's curriculum. Librarians can pair with faculty in their liaison departments to integrate DIL competencies into assignments in much the way that information literacy instruction currently takes place.

Conclusion

Emphasizing integrity and responsibility as a part of the research process is a critical component in any discipline. In the sciences, the tradition of meticulous documentation in lab

notebooks is considered basic to research training. Students are often told that they should document their notebooks thoroughly enough that anyone could continue work on an experiment from where they stopped, or replicate the steps that they already performed. As more research takes place in the digital realm, there appears to be a disconnect in translating these skills to digital data management.

Librarians are often tasked with providing students the training they require to build a practice of lifelong learning. Research and scholarship represent dynamic, evolving processes that occur on a continuum. Understanding that data have the potential to impact not only one's own research, but also the work of others – in fields that may appear unrelated – can help build an awareness of the diverse scholarship ecosystem. As academic librarians prepare students to engage with the world in a meaningful way, it is important that data information literacy not be overlooked. Empowering students to be responsible for the data that they generate, and instilling in them recognition that their data could be used to build further knowledge, should be an integral part of the research process. In order to support this research, libraries must engage with data information literacy for their constituents.

Acknowledgements

Thanks to Sharon Weiner, Carolyn Schubert, and Kelly Giles for their feedback on this column.

References

Bibliography

Association of College and Research Libraries. "Framework for Information Literacy for Higher Education." Association of College and Research Libraries, last modified February 20, 2014, accessed April 2, 2014, http://acrl.ala.org/ilstandards/wp-content/uploads/2014/02/Framework-for-IL-for-HE-Draft-1-Part-1.pdf.

———. "Information Literacy Competency Standards for Higher Education." Association of College and Research Libraries, accessed April 8, 2014. http://www.ala.org/acrl/standards/informationliteracycompetency.

———. "Presidential Committee on Information Literacy: Final Report." Association of College and Research Libraries, last modified January 10, 1989, accessed April 2, 2014. http://www.ala.org/acrl/publications/whitepapers/presidential.

Association of College and Research Libraries: Working Group on Intersections of Scholarly Communication and Information Literacy. 2013. *Intersections of Scholarly Communication and Information Literacy: Creating Strategic Collaborations for a Changing Academic Environment*. Chicago, IL: Association of College and Research Libraries.

Ball, Richard and Norm Medeiros. 2012. "Teaching Integrity in Empirical Research: A Protocol for Documenting Data Management and Analysis." *Journal of Economic Education* 43 (2): 182-189. doi:10.1080/00220485.2012.659647. http://www.tandfonline.com/doi/abs/10.1080/00220485.2012.659647.

Borgman, Christine L., Jillian C. Wallis, and Noel Enyedy. 2007. "Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries." *International Journal on Digital Libraries* 7 (1-2): 17-30.

Calzada Prado, Javier and Miguel Ángel Marzal. 2013. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri*. 63(2): 71-178. doi:10.1515/libri-2013-0010

Carlson, Jacob, 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://muse.jhu.edu/journals/portal_libraries_and_the_academy/v011/11.2.carlson.html.

Carlson, Jake, Lisa Johnston, Brian Westra, and Mason Nichols. 2013. "Developing an Approach for Data Management Education: A Report from the Data Information Literacy Project." *The International Journal of Digital Curation* 8 (1): 204.

Carlson, Jake and Marianne Stowell-Bracke. 2013. "Data Management and Sharing from the Perspective of Graduate Students: An Examination of the Culture and Practice at the Water Quality Field Station." *Portal: Libraries & the Academy* 13 (4): 343-361. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v013/13.4.carlson.html.

DataONE. "DataONE Education Modules." DataONE, accessed April 14, 2014, www.dataone.org/education-modules.

Doucette, Lise and Bruce Fyfe. 2013. "Drowning in Research Data: Addressing Data Management Literacy of Graduate Students." Indianapolis, IN, Assoc. of College and Research Libraries, April 10-13, 2013.

EDINA and Data Library, University of Edinburgh. "Research Data MANTRA [Online Course]." accessed April 14, 2014, http://datalib.edina.ac.uk/mantra.

289 Force11. "Joint Declaration of Data Citation Principles." accessed April 14, 2014,
290 https://www.force11.org/datacitation.

291 Hey, T. and J. Hey. 2006. "E-Science and its Implications for the Library Community." *Library Hi Tech*
292 24 (4): 515-528. doi:10.1108/07378830610715383.

293 Johnston, Lisa R. and Jeffryes, Jon. "Instructor Guide to the Flipped Data Management Course<*br* />."
294 University of Minnesota Libraries, accessed April 8, 2014, http://z.umn.edu/teachdatamgmt.

295 Lamar Soutter Library, University of Massachusetts Medical School. "New England Collaborative Data
296 Management Curriculum." accessed April 14, 2014, http://library.umassmed.edu/necdmc.

297 Munoz, Trevor. 2013. "Data Curation as Publishing for the Digital Humanities." *Journal of Digital*
298 *Humanities* 2 (3): 14-22.

299 National Science Foundation. "Proposal and Award Policies and Procedures Guide: Part 1 - Proposal
300 Preparation & Submission Guidelines (GPG)." National Science Foundation.

301 Parsons, M.,A. and R. Duerr. 2005. "Designating User Communities for Scientific Data: Challenges and
302 Solutions." *Data Science Journal* 4: 31-38.

303 Piorun, Mary, Donna Kafel, Tracey Leger-Hornby, Siamak Najafi, Elaine Martin, Paul Colombo, and
304 Nancy LaPelle. 2012. "Teaching Research Data Management: An Undergraduate/Graduate
305 Curriculum." *Journal of eScience Librarianship* 1 (1): 46-50. doi:10.7191/jeslib.2012.1003.
306 http://escholarship.umassmed.edu/jeslib/vol1/iss1/8.

307 Purdue University, University of Minnesota, University of Oregon and Cornell University. "Data
308 Information Literacy." accessed April 14, 2014, http://wiki.lib.purdue.edu/display/ste/Home.

309 Qin, Jian and John D'Ignazio. 2010. *Lessons Learned from a Two-Year Experience in Science Data*
310 *Literacy Education*. http://docs.lib.purdue.edu/iatul2010/conf/day2/5.

311 Snyder, T. D. and S. A. Dillow. 2013. *Digest of Education Statistics 2012 (NCES 2014-015)*.
312 Washington, DC.: National Center for Education Statistics, Institute of Education Sciences, U.S.
313 Department of Education.

314 University of Virginia Library. "Steps in the Research Life Cycle." accessed April 16, 2014,
315 http://dmconsult.library.virginia.edu/lifecycle/.

316

317