See Also:

Data literacy instruction in academic libraries: best practices for librarians

Amanda Wanner - amandathinks@gmail.com

Amanda is an MLIS student at the University of British Columbia. She is interested in health librarianship and open knowledge initiatives.

Abstract:

This paper discusses the challenges and opportunities of bringing data literacy instruction to academic libraries. Information literacy and digital literacy in libraries has been widely discussed in the library sciences and education literature, but until recently very little focus has been given to data literacy. However, new e-government and open data initiatives over the past decade have created widely available public data that is of great interest to students and academics. Increased technological capabilities to process "big data" have also created new opportunities for the layperson and researcher alike. One popular article claims, "Ensuring that big data creates big value calls for a reskilling effort that is at least as much about fostering a data-driven mindset and analytical culture as it is about adopting new technology" (Harris, 2012). The influx of available data presents unique challenges for librarians. How can libraries play a role, for example, in this "reskilling effort" to develop a "data-driven mindset"?

Copyright:

Data literacy instruction in academic libraries: best practices for librarians

Amanda Wanner - amandathinks@gmail.com

Introduction

In this paper, I will discuss the challenges and opportunities of bringing data literacy instruction to academic libraries. Information literacy and digital literacy in libraries has been widely discussed in the library sciences and education literature, but until recently very little focus has been given to data literacy. However, new e-government and open data initiatives over the past decade have created widely available public data that is of great interest to students and academics. Increased technological capabilities to process "big data" have also created new opportunities for the layperson and researcher alike. One popular article claims, "Ensuring that big data creates big value calls for a reskilling effort that is at least as much about fostering a data-driven mindset and analytical culture as it is about adopting new technology" (Harris, 2012). The influx of available data presents unique challenges for librarians. How can libraries play a role, for example, in this "reskilling effort" to develop a "data-driven mindset"?

What is data literacy and how does it differ from its counterpart, information literacy? What might a data literacy curriculum look like in post-secondary institutions? This paper seeks to address these questions. Part I of this paper critically examines the concept of data literacy – how is it different, or similar, to other kinds of literacies, and why is it important? Part II examines data literacy instruction in academia, including a short literature review of recent instructional interventions. The paper

concludes with a set of best practices for librarians wishing to pursue data literacy instruction at their institutions and recommendations for future research.

Part I: Conceptualizing Data Literacy

Information Literacy

The broad conception of "information literacy" is well known in librarianship. Many universities offer instruction in information literacy, either embedded throughout the curriculum or as a workshop series through the library system. Much has been written about the importance of teaching students the ability to find and evaluate sources of information. When students are able to "determine their own information needs, to use some information retrieval tools efficiently, to evaluate the retrieved information, and to use that information to answer their needs" (Julien and Boon, 2004, p. 122), students feel more confident in their abilities and less anxious, make better use of the library resources available to them, and learn transferable literacy skills (ibid, p. 138).

Information literacy instruction in universities is generally guided by the Association of College & Research Libraries (ACRL) Information Literacy Best Practices Committee, which has produced two major standards:
- The Information Literacy Competency Standards for Higher Education, adopted by the committee in 2000, was the standard for 15 years, and has been recently supplemented by the Framework for Information Literacy for Higher Education (Association of College and Research Libraries [ACRL], 2015)
- Characteristics of Programs of Information Literacy that Illustrate Best Practices: A Guideline, last revised in 2012.

ACRL defines information literacy as "a set of abilities requiring individuals to

recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information." (2000) However, the relationship between information literacy and data literacy is still unclear. "Substitute 'data literacy' for 'information literacy' and are we on the way to defining the best practices for data literacy programs?" Hunt asks rhetorically (2004). Indeed, the relationship between the two concepts is nebulous; conceptions of data literacy in the literature vary in their association to information with literacy. Some suggest, for example, that the two belong on a sort of continuum (Calzada & Marzal, 2013, p. 126), while others suggest that data literacy is a smaller component of a larger information literacy concept.

Data literacy is a relatively new concept to librarianship, although the skills required to be data literate, including evaluating, curating, analyzing, and managing information are not. What has changed in the information landscape is the scope and depth of data available. Increases in technological capabilities and large "open" datasets spurred by e-government initiatives have created the catalyst for data-specific instruction.

Defining Data Literacy

As of yet, there are no unified standards for data literacy (Mandinach & Gummer, 2012), although many researchers have proposed core competencies and curriculum guidelines based on evidence-based practice. Indeed, the term 'data literacy' itself is not fixed. Other terminology used by professionals in the literature – either as a synonym or a related concept – includes quantitative literacy, quantitative reasoning, statistical literacy, data information literacy, numeracy, information visualization, data research management, science data literacy, eResearch, and eScience.

On a basic level, data literacy means the ability to work with and understand data.

Most accepted definitions revolve around this core concept, as in Carlson et. al.'s basic overview: "Data literacy involves understanding what data mean, including how to read graphs and charts appropriately, draw correct conclusions from data, and recognize when data are being used in misleading or inappropriate ways." (Carlson, 2011, p. 633). D'ignazio and Qin define data literacy as "the ability to examine multiple measures and multiple levels of data, to consider the research, and to draw sound inferences" (2010, p. 189) Stevenson and Caravello succinctly note that data literacy is "also known in the literature as statistical literacy, quantitative literacy, and numeracy" (2007, p. 525). These basic definitions all revolve around a functional use of data in practice.

Some researchers take a conservative approach, linking the similarities between information literacy and data literacy, noting that they both require skills such as finding and using information and evaluating sources. However, there is disagreement as to the extent that these concepts overlap and the specific definition, with some conflating the two concepts and others making the case that they are conceptually and in practice very different.

One common approach is to view information literacy as focused on the findability and usability of information, with data literacy dealing with more technical aspects of data production, curation, and management. Within this framework, information literacy is seen as a broader concept that doesn't account for new responsibilities in research data management. Qin and D'ignazio, for example, discuss the tension between data literacy programs that are "pushing" rather than "pulling" data to their students; that is, focusing on usability of data versus processing and managing data (2010, p. 190). They note that the latter is more commonly seen within more technical programs such as computer science and use the terminology "science data literacy" to encompass "pulling" data with "an emphasis on scientific inquiry

through collecting, transforming, managing, and using data" (ibid, p. 189).

Carlson et. al. makes a similar argument: "Whereas definitions of data, statistical, and information literacy focus on the consumption and analysis of information… data information literacy merges the concepts of researcher-as-producer and researcher-as-consumer of data products" (2011, p. 634). Under this framework, data literacy is a "social process" (ibid) in which researchers not only require basic information literacy skills but the ability to share data and understand the upstream and downstream effects of their data management practices for other researchers. Calzada and Marzal define this relationship as a "continuum", noting, "Data literacy can be viewed both as a whole and as an integrated assemblage of other competencies… Data literacy can be defined, then, as the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle and ethically use data" (2013, p. 125-126).

Calzada and Marzal suggest that there are two main trends in conceptualizing data literacy: one based around functional use of data in practice (i.e. a consumer-based perspective), and the other relating to research data management (i.e. a producer-based perspective). This tidy perspective accounts for widely varying perspectives between, for example, teaching undergraduate students the basics of geospatial data and working with faculty on data management and curation practices. Thus, one common conceptualization of data literacy is that it contains core aspects of previous definitions of information literacy, but adds crucial critical thinking and evaluative skills that are needed for students to produce as well as consume data.

What is increasingly clear, however, is that while there is some agreement on the basic components of data literacy, there is no consistent, established definition or core competencies for instruction. One white paper, which analyzed and

summarized the findings of a group of stakeholders over a day and a half deliberating over a definition of data literacy, concluded simply, "a formal, simple definition remains elusive" (Mandinach & Gummer, ii).

The Importance of Data Literacy

On the surface, data literacy and information literacy seem to be similar concepts. Both conceptually have their roots in information use and practice. The rise in e-government, big data, and open data movements of the past decade have precipitated a skills crisis in using data effectively. What use is data, after all, if it cannot be understood, manipulated, and shared? Stephenson and Caravello note that many people lack everyday critical thinking skills to evaluate data: "It is not clear whether the average reader actually understands the statistical material presented in the daily newspaper on yet another medical study, public opinion poll, or even the NBA playoffs." (2007, p. 526). Carlson et. al. go one step further: "The sheer quantity of data being generated and our current lack of tools, infrastructure, standardized processes, shared workflows, and personnel who are skilled in managing and curating these data pose a real threat to the continued development of e-research." (2011, p. 631)

Finally, data literacy skill development is not limited to students. A report from the Association of Research Libraries to the National Science Foundation in 2006 found: "Many scientists continue to use traditional approaches to data, i.e., developing custom datasets for their own use with little attention to long-term reuse, dissemination, and curation… even modest collaborative projects are inconsistent in their attention to data management and few individual scientists think beyond posting selected results and data on the Internet or submitting a final data product to a data archive if required to do so." (Friedlander & Adler, 2006, p. 122). Lack of data information literacy skills, indeed, could be at crisis levels.

7

Three major trends emerge for the importance of data literacy instruction: developing skills in the next generation workforce, where data management is becoming increasingly important, using best practices in research data management, and more broadly, developing critical thinking skills in university students that are transferable to a wide range of applications, including the workforce, academia, and everyday life.

For Gunter, data literacy instruction is primarily a tool for creating more well-informed and educated citizens. This broad conception is related to general critical thinking and problem-solving skills to be used in everyday life, and transferable to a variety of situations. Gunter claims that data literacy represented a large conceptual shift from how things have been done previously, even going so far as to broadly claim that today's learners are fundamentally different from previous generations in the ways that they communicate and use information (2007).

For others, data literacy is deeply tied to specific skill sets required for the workplace or the research lab (or both). In interviews, for example, Carlson et. al. found that faculty enthusiastically expressed importance of data literacy skills, but often failed to define precisely what data literacy is and acknowledged that their own data management practices were lacking. In fact, in many cases, faculty expressed that data literacy training was spotty, or that students were expected to come to the lab with data literacy skills already developed. One faculty member expressed, "The way we [teach these skills], is generally speaking, we just say, 'Well, go learn it' and [the graduate students] just figure it out" (Carlson, 2011, p. 637). In this study, faculty were often unprepared to teach their students and research assistants the conceptual framework and tools required for data literacy understanding.

If faculty are failing to lead by example, there is opportunity for libraries to become a

centralized hub for data literacy instruction. Indeed, Carlson et. al. conclude, "We believe that librarians have a role in developing these [data literacy] education programs and will need to actively engage in these discussions" (Carlson, 2011, p. 633).

Part II: Data Literacy Instruction

Given the rise of big data and open data initiatives, there appears to be a data literacy gap in university classrooms and labs. This section discussed how libraries and classrooms in post-secondary education are addressing the data literacy gap. First, a short literature review discusses current data literacy initiatives in the classroom, followed by a guideline of best practices derived from evidence-based practice.

Literature review

Data literacy instruction is still a new area for exploration in libraries. However, a few prominent initiatives have emerged in recent years documenting their data literacy programs and results. The following studies are all data literacy initiatives taking place in post-secondary education, either embedded within a classroom or initiated by the university library. Each of the studies are summarized and the main course outcomes outlined to demonstrate the range of initiatives and diversity of practices.

Information and statistical literacy in sociology (Stephenson and Caravello)

Stephenson and Caravello (a librarian and data archivist) co-taught a one-credit information literacy and statistical literacy sociology course. The goals of the course were to build skills in critical thinking and information evaluation. The data literacy module contained four specific learning outcomes:

- Develop the ability to read and critically evaluate simple 2×2- or 3-way tables
- Produce accurate bibliographic citations for data tables
- Use American Factfinder to create a table, which they could describe and

cite correctly

- Read an article containing a graphical representation of data and discuss the table in relation to the article content (2007, p. 530-531)

Data literacy in geoinformatics (Carlson, Fosmire, Miller, & Nelson)

Carlson et. al.'s extensive study focused on a data literacy needs assessment of faculty and assessed student outcomes from a geoinformatics course offered in 2008 and 2010. Interviewed faculty came from science and engineering departments at Purdue and the University of Illinois at Urbana-Champaign. The results from student and faculty assessments were then "filtered through the perspective of ACRL's information literacy competency standards" (2011, p. 629) to create a set of learning goals for data literacy instruction. By integrating information from existing information literacy standards and faculty and student assessments, Carlson et. al's study is one of the strongest yet for a proposed data literacy competencies standard.

Their proposed core competencies for data information literacy include:

- Introduction to databases and data formats
- Discovery and Acquisition of data
- Data management and organization
- Data conversion and interoperability
- Quality assurance
- Metadata
- Data curation and reuse
- Cultures of practice
- Data Preservation
- Data Analysis
- Data visualization
- Ethics, including citation of data (ibid, p. 652-653)

Science data literacy (Qin & D'ignazio)

Like Carlson et. al., Qin and D'ignazio surveyed science faculty on their data management practices and built a science data literacy course based on their findings. The course was developed for a mix of undergraduate and graduate students from a variety of subject departments. The course utilized a variety of pedagogical strategies to engage students, including: three strong course modules based on faculty needs assessment; an emphasis of the role of metadata in science data management (SDM) from multiple perspectives; teaching by case studies from a variety of STEM disciplines (with an emphasis on geoinformatics data for its relative accessibility); and, "authentic" activities to get involved in data management projects, including "reverse engineering" a data repository and in-class presentations (2010, p. 201).

The course was broadly based around three modules:

- Fundamentals of science data and data management
- Managing data sets in aggregation
- Broader issues in science data management (ibid, p. 197).

Geospatial and Data Curation in geoinformatics (Miller & Fosmire)

Miller and Fosmire discuss a geoinformatics course offered to beginning graduate and advanced undergraduate students in 2008. The course was intended to be a holistic overview of geospatial and data curation topics, while also providing practical hands-on experience working with data and data management tools. The first half of the course, in which theoretical concepts were introduced, uses a case study scenario with an ongoing story arc. The second half of the course focused on student projects and more advanced lecture and lab modules. Throughout, the course emphasized data research management principles, including a librarian-led curriculum with data preservation, ontology, and metadata modules.

Course topics included:

- Basic computing environments

- Geospatial data

- Geographic Information Systems (GIS)

- Data gathering

- Independent statistical procedures

- Scientific workflow tools (2008)

Data literacy instruction framework (Calzada & Marzal)

Calzada and Marzal reviewed the literature of information literacy and data literacy instruction and proposed a common reference framework for data literacy instruction. The framework builds on trends in data literacy instruction in both data use and data management. The framework translates competencies identified within the literature into discrete modules that can be used as the foundation for course instruction:

1. Understanding data
2. Finding and/or obtaining data
3. Reading, interpreting, and evaluating data
4. Managing data
5. Using data

Data Literacy Instruction Best Practices

Although the research on data literacy instruction is still emerging, several best practices have emerged from a review of the literature. The following section identifies several common themes from data literacy initiatives within higher education. These best practices represent a comprehensive set of recommendations, derived from evidence-based practice, for institutions wishing to pursue a data literacy curriculum.

Critical thinking skills

One of the most widely discussed topics in data literacy instruction is the importance of critical thinking skills. At the core of data literacy is the ability to transfer practical skills in the classroom into other areas of life, including the everyday, workplace, and research lab environment. Gunter emphasizes the importance of "critical and problem-solving skills" (2007) for general well-rounded 21st century citizens. Stephenson and Caravallo (among others) emphasize critical thinking skills for the workplace and in academia. In interviews with faculty, they found that "critical thinking and evaluation of materials" (2007, p. 530) was the second most mentioned theme.

Collaboration between library and departments

Carlson et. al.'s article concludes, "the authors do recommend a collaborative venture between disciplinary faculty and librarians as the best practice for teaching data information literacy skills" (2011, p. 653). Most data literacy initiatives reviewed involved a strong partnership between the university library (or librarians) and faculty. These partnerships allow for broad centralized data literacy modules to be developed to meet the needs of the students while creating a feedback loop so that the curriculum can be tailored or refined as needed and in collaboration with faculty.

Librarian involvement

Similarly, there appears to be great need for librarians to initiate data literacy programming. Carlson et. al. surprisingly found that although faculty enthusiastically identified a need for data literacy instruction and skill-building, "several respondents expressed an uncertainty or a reluctance to teach data management skills to their students themselves" (2011, p. 638) and "could not articulate precisely what skills should be taught to remedy the situation" (ibid). Many of the faculty interviewed themselves lacked data literacy skills or, worse, didn't realize how little they knew.

This "ignorance loop" (ibid, p. 644) reveals that there may be a need for faculty and staff education in addition to student data initiatives. Additionally, a centralized data literacy program or library staff support could help ease this pressure from faculty. Many are quick to identify the library as a key place to provide this kind of support given that many information literacy initiatives already stem from university libraries.

Ongoing data literacy instruction at all levels of schooling

One of the most important themes to emerge from the literature is the importance of ongoing data literacy instruction to be embedded into the curriculum at all levels of schooling. That is, students need repeated exposure to data literacy concepts so that they can build a mental framework and learn to apply the concepts to new and unique situations. "One shot" library instruction is not enough to impart a thorough and critical understanding of data use and management. "Just as writing skills are more easily developed as part of various learning and curricular experiences, the variety of data literacy skills can be obtained more comprehensively when incorporated across the curriculum, in core theory or topical seminars as well as methods courses," Stephenson and Caravello eloquently note (2007, p. 527), before recommending a number of ideas for embedding data literacy modules into other courses at their institution.

This can also be accomplished through a curriculum with varied activities. Stephenson and Caravello, for example, suggest a multifaceted approach, incorporating individual consultations, class orientations, reference desk instruction, incorporating data literacy activities into other library instruction initiatives, using hands on activities and so forth (2007). Others use creative, practical pedagogical techniques, such as Miller and Fosmire's use of an ongoing case study over a period of 7 weeks (2011).

Of all the themes in the literature, the need for ongoing instruction is the most widely and emphatically discussed. Whether the skills are developed for ongoing research in academia or the workplace, there is general consensus that "if students are expected to use data skills as they move into higher level, evaluative, and interpretive research and coursework, they will need multiple opportunities in which to develop data literacy" (Stephenson & Caravallo, 2007, p. 535).

To some degree, best practices of data literacy instruction reflect what is already known about instruction generally in library settings. Students need repeated contact with material, over a period of several courses (to several years) with progressively more complex material that is presented in different ways.

Assessment

There is little difference between data literacy program assessment and other kinds of library instruction assessment; however, a few components are worth mentioning. First, it is widely understood that many students and to some degree faculty[1] overestimate their informational literacy skills, known as the Dunning-Kruger effect. For this reason, it is of limited usefulness to rely on student self-assessment of data literacy skills, as results will not reflect that students may not know what they don't know. Thus, more recent assessments of data literacy initiatives tend to utilize two strategies. First, faculty assessments are commonly used to see what's being done in the classroom and what skills students are perceived to be lacking. This kind of needs assessment is used to ensure that the initiative will be useful to students and faculty. Second, a pre- and post- skills test is commonly used to gauge student learning from classroom interventions to see whether the initiative was effective.

---

1   See Carlson et. al., 2011

Conclusion

Libraries are well positioned to become leaders in providing data literacy services for students and researchers. Libraries can fill this need by ensuring that researchers have the knowledge to find, evaluate, clean, use, and attribute data properly. Data literacy skills go well beyond simply searching for a government dataset and downloading the file or using a data analysis computer program, although these skills are important. Data literacy also requires critical-thinking and problem-solving skills to learn how to manipulate and interpret data, assess its quality, and recognize patterns.

It's not immediately clear that data literacy instruction belongs in the library. For many, the jump between current workshop offerings in data-related topics and data literacy training is an obvious one, but others are more hesitant to ascribe yet another responsibility to spaces that are increasingly seeing budgetary cuts and resource reductions. How can libraries provide yet another service when the resources for current services are already so tight in many places? To be most effective, data literacy instruction requires collaborative partnerships throughout the institution, staff training or the appointment of a "data librarian", and a comprehensive plan of action to provide instruction throughout the student lifecycle. And yet, as Merrill eloquently states: "the message is that data literacy instruction needs a home. Why can't that home be the library?" (2011, p. 147).

References:

ACRL Information Literacy Best Practices Committee. (2012). Characteristics of programs of information literacy that illustrate best practices: A guideline Approved by the ACRL Board of Directors, June 2003, revised January 2012. College & Research Libraries News, 73(6), 355-359.

Association of College and Research Libraries (2015). Framework for Information Literacy for Higher Education. Chicago: American Library Association. Retrieved from http://www.ala.org/acrl/standards/ilframework.

Association of College and Research Libraries (2000). Information literacy competency standards for higher education. Chicago: American Library Association. Retrieved from http://www.ala.org/acrl/standards/informationliteracycompetency.

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. Libri: International Journal of Libraries & Information Services,63(2), 123-134.

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. Portal: Libraries and the Academy, 11(2), 629-657.

D'ignazio, J., & Qin, J. (2010). The central role of metadata in a science data literacy course. Journal of Library Metadata, 10(2), 188-204.

Friedlander, A., & Adler, P. (2006). To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering. A Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships--The Role of Academic Libraries in the Digital Data Universe (Arlington, Virginia, September 26-27, 2006). Association of Research Libraries.

Gunter, G. A. (2007). Building student data literacy: An essential critical-thinking skill for the 21st century. MultiMedia & Internet@Schools [H.W.Wilson - EDUC], 14(3), 24.

Harris, J. (2012). Data is Useless Without the Skills to Analyze It. Harvard Business Review. Retrieved from https://hbr.org/2012/09/data-is-useless-without-the-skills.

Hogenboom, K., Phillips, C. M. H., & Hensley, M. (2011, March). Show me the data! Partnering with instructors to teach data literacy. In Declaration of Interdependence: The Proceedings of the ACRL 2011 Conference, March (pp. 410-417).

Hunt, K. (2004, April). The challenges of integrating data literacy into the curriculum in an undergraduate institution. Paper presented at Iassist conference, Madison, WI. Retrieved from http://www.iassistdata.org/downloads/iqvol282_3hunt.pdf.

Julien, H., & Boon, S. (2004). Assessing instructional outcomes in Canadian academic libraries. Library and Information Science Research, 26(2), 121.

Mandinach, E., & Gummer, E. (2012). Navigating the landscape of data literacy: It IS complex. (White Paper). WestEd. Retrieved from http://www.wested.org/online_pubs/resource1304.pdf

Merrill, A. (2011). Library+. Public Services Quarterly, 7(3-4), 144-148.

Miller, C., & Fosmire, M. (2008). Creating a culture of data integration and interoperability: Librarians collaborate on a geoinformatics course. In Proceedings of the 29th Annual IATUL Conference.

Qin, J., & D'ignazio, J. (2010, June 22). Lessons learned from a two-year experience in science data literacy education. Paper presented at International Association of Scientific and Technological University Libraries, 31st Annual Conference. Accessed from http://docs.lib.purdue.edu/iatul2010/conf/day2/5.

Stephenson, E., & Caravello, S. P. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences. Reference Services Review, 35(4), 525-540.