THINKING CRITICALLY ABOUT DATA CONSUMPTION: CREATING THE DATA CREDIBILITY CHECKLIST

Lisa D. Zilinski, lzilins@purdue.edu
Megan Sapp Nelson, mrsapp@purdue.edu
Purdue University Libraries, West Lafayette, Indiana, Unites States

ABSTRACT

As STEM areas increasingly rely on pre-existing data, either to validate or extend the scientific body of knowledge, students who have baseline knowledge of how to find, evaluate, and access data will have an advantage. Accordingly, undergraduate STEM curricula is increasingly focused on research-based group projects that develop professional skills, building the professional portfolio needed for early career scientists, technologists, and engineers. This project works to develop new tools to implement basic data skills within the undergraduate disciplinary curricula. The first step in this process was to identify the competencies that are likely to be needed by those seeking data for reuse. Using competencies identified in the literature and via brainstorming, formatted similarly to the traditional model used by librarians to teach assessment of publication quality, a data credibility checklist was developed as the first tool in an emerging toolkit.

DEVELOPING DATA REUSE COMPETENCIES

Positive attributes of data objects include:

- Human readable/Machine readable
- Clear format, including easily identifiable units of measure and labeling
- Explicit date of creation or publication
- Documentation, content map or guide of some sort that outlines labels/relationships
- Quality control process identified
- Option for download, including the option to download in an open source format
- Clear process to download

Negative attributes of data objects include:

- Only human readable or machine readable
- Unclear authority including creator, creation date, and process by which data object was created
- No content guides to identify context
- No or poorly implement quality control, no process identified by which data object was checked for errors
- Unclear labels/ relationships with no documentation to explain
- Mixed formats or units of measure with no explanation
- No or an unclear process by which to download the data object
- Proprietary software formats as only option for download or no options for download

DATA CREDIBILITY CHECKLIST

Content map	Authoritative	Format expectations	Quality Control	Human Readable/ Machine Readable
What is covered?	Who created the data?	What units are used?	Who is in charge of checking for quality?	Can you open a file and understand what is in it?
What is not covered?	Who is managing the data?	What fields are present?	What process do they use?	Is the file available for download in an open source format?
Is it relevant to my research question?	Who paid for the data?	What naming conventions are used?	How is missing data handled?	Is there a clear process for download?
How is it relevant to my research question?	What bias might be implicit?	Are the dates of creation or update easy to find?		
Is there metadata included?	Is it currently maintained?			
	Has someone else used this dataset for reuse in the past? How?			
	Are there clear release versions, updates with release dates?			

COMPETENCIES & TASKS

The first set of competencies include:

- Ask a question and find a dataset that will have the data required
- Develop a research question based on the data in a dataset
- Understanding that there are fields within dataset
- Understanding what the fields within a dataset mean
- Understanding relationships between fields within a dataset
- Ability to read and interpret charts, graphs, and other data visualizations

Outline of specific tasks that could be completed in-class that would guide the students in defining the need(s) for specific data from a dataset:

- Ask a question and find a dataset that will have the data required
- Develop a question based on the data in a dataset
- Know what the databases contain

DISCUSSION

The tool that was created, the Data Credibility Checklist, can be used in a variety of courses in a number of different disciplines. Most recently, this checklist and corresponding class assignments and exercises have been used in engineering and technology courses (Zilinski et al., in press) with varying degrees of success. It will be important to find additional ways to test the checklist and incorporate it into the STEM curriculum through for-credit courses, one-shot library instruction sessions, workshops, and research programs.

Programs such as the Summer Undergraduate Research Fellowship (SURF) can offer opportunities to incorporate the checklist into a hands-on research experience for undergraduate engineering, science, and technology students. We can also see the Data Credibility Checklist used in research methods courses and undergraduate research fellowships in non-STEM disciplines, where students are exposed to datasets in the field.

CONNECTING COLLECTIONS, CULTURES, AND COMMUNITIES
77TH ASIS&T ANNUAL MEETING
October 31 – November 5, 2014
Sheraton Seattle Hotel, Seattle, WA

References

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. doi: 10.1515/libri-2013-0010

Carlson, J. & Fosmire, M. & Miller, 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.

Kreisberg, A, Frank, R., Faniel, I., & Yakel, E. (2013). The Role of Data Reuse in the Apprenticeship Process. In Proceedings of the American Society for Information Science and Technology, 50, 1-10. Montreal, Canada. doi: 10.1002/meet.14505001051

MacMillan, D. (2010). Sequencing Genetics Information: Integrating Data into Information Literacy for Undergraduate Biology Students. Issues in Science Technology Librarianship, 61. doi: 10.5062/F44F1NNK

NIH Data Sharing Policy and Implementation Guidance. (2003). National Institutes of Health.

NSF Proposal Processing and Review. (2011). National Science Foundation.

Qin, J., & D'Ignazio, J. (2010). Lessons Learned from a Two-Year Experience in Science Data Literacy Education. In International Association of Scientific and Technological University Libraries, 31st Annual Conference. West Lafayette, IN: Purdue University Libraries

Stephenson, E. & Schifter Caravello, P. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. Reference Services Review, 53(4), 525-540. doi: 10.1108/00907320710838354

Zilinski, L. D., Nelson, M. S., & Van Epps, A. S. (2014). Developing Professional Skills in STEM Students: Data Information Literacy. Issues in Science and Technology Librarianship, 77(Summer). Doi: 10.5062/F42V2D2Z

PURDUE UNIVERSITY LIBRARIES