Grace Liu, Bobray J. Bordelon, and Rashelle Nagar

Data Quality Literacy

Empowering Academic Librarians to Teach Data Quality Evaluation

Data quality literacy lies at the intersection of information evaluation and data literacy. It focuses on developing literacy skills in evaluating data quality and developing critical thinking about data. Although the ACRL guidelines for teaching information literacy shifted from the Information Literacy Competency Standards for Higher Education (2000) to the Framework for Information Literacy for Higher Education (2016), the expectation for students to develop information evaluation skills has remained consistent. The Standards emphasized the importance of students learning to evaluate information and its sources critically.[1] Similarly, the Framework asks students to demonstrate the ability to evaluate a wide range of information sources, as in the frame Searching as Strategic Exploration. It also expects students to recognize that information reflects its creators' expertise and credibility and should be evaluated based on the information needed and the context in which the information will be used, as in the frame Authority Is Constructed and Contextual.[2]

In recent years, data literacy has gained significant attention from academic librarians, with growing research exploring academic librarians' role in teaching data literacy skills. Efforts have been made to develop data literacy competencies and align them with the Framework for Information Literacy for Higher Education.[3,]) Many of the data literacy competency frameworks developed in recent decades highlight "quality assurance," "evaluating data," or "evaluating and ensuring quality of data sources," as one of the core data literacy competencies.[3] However, these frameworks often address data quality evaluation only at a surface level. In this article, we share our collaborative efforts in expanding academic librarians' understanding of data quality issues and providing resources to empower academic librarians in teaching the evaluation of data quality.

IMLS Grant Project

In 2020, West Chester University, in collaboration with Stanford University and the University of Illinois at Urbana-Champaign, received grant funding from the Institute of Museum and Library Services (IMLS) for the project "Building Capacity of Academic Librarians in Understanding Quantitative Data, Data Quality Problems, and Evaluating Data Quality."[5] As a National Forum Project, it engaged data experts from diverse backgrounds and industries to share their perspectives on data quality issues and data evaluation. The project aimed to raise academic librarians' awareness of data quality problems, fill our knowledge gaps, and build our capacity and confidence in teaching data evaluation, critical thinking about data, and data literacy.

Grace Liu is the business librarian and associate professor at the West Chester University FHG Library, email: yliu@wcupa.edu. Bobray J. Bordelon is the economics and finance librarian at Princeton University, email: bordelon@princeton.edu. Rashelle Nagar is the business research and collections librarian at the Stanford Graduate School of Business Library, email: rmnagar@stanford.edu.

Between January and July 2023, the project organized seven online national Forum webinar sessions:

1. "Evaluating Data Quality: Challenges and Competencies"
2. "Quality Assurance in Data Creation"
3. "Understanding Commercial Data Quality Issues"
4. "Evaluating Governmental Data (U.S. and International)"
5. "Data Quality: Reproducibility and Preservation"
6. "Data Quality: Evolving Employer Expectations"
7. "Librarians' Role in Cultivating Data-Literate Citizens"

The sessions featured a motley group of speakers, including data librarians from Princeton, Stanford, Yale, and the University of California, Berkeley; the chief scientist of the US Census Bureau; former chief statistician of the United States; experts from the US Department of Agriculture, Catholic Charities USA, Federal Reserve Banks of Cleveland and Kansas City, Center for Health & Wellbeing at Princeton University, Labor Dynamics Institute at Cornell University, Meta Platforms, and Microsoft Research; as well as leaders from the Inter-University Consortium for Political and Social Research, the Roper Center, Ithaka S+R, and the American Statistical Association.

The project team edited the National Forum transcriptions into an open access ebook and developed knowledge briefs outlining a synthesized data evaluation strategy (see details on the project website).[6] In the long term, the project aims to demonstrate the library's role in fostering data-literate citizens, enhancing the quality and integrity of scholarly output, reducing the social and economic costs of data quality issues, and improving data-driven decision-making.

Resources for Teaching Data Quality Literacy

The ebook and knowledge brief lay some groundwork for supporting academic librarians in teaching data quality literacy and engaging researchers for deeper conversations on data quality issues.

The knowledge briefs cover 13 topics:

1. Data Reference Interview
2. Evaluating Data Documentation
3. Evaluating Dataset for Research Needs
4. Using and Evaluating U.S. Federal Statistics
5. Understanding Administrative Data
6. Evaluating Administrative Data Quality
7. Understanding Commercial Data
8. Evaluating Commercial Data Quality
9. Commercial Data Quality: Conversation with the Vendors
10. Commercial Data Quality: Conversation with Researchers
11. Evaluating International Government Data Quality
12. Understanding Survey Data and Public Poll
13. Evaluating Survey Data Quality

These knowledge briefs focus on the quality aspect of data literacy competencies and aim to help academic librarians gain a deeper understanding of different types of data products to help researchers develop critical thinking skills to evaluate data quality. Some examples from the knowledge brief include the following.

Evaluating Data Documentation

Data documentation provides the contextual information needed to discover, understand, access, and reuse data. The knowledge brief covers the quality indicators and the characteristics of good data documentation. It also includes ideas to fill the data documentation gaps.

Administrative Data Evaluation

Administrative data refers to data collected for operational, programmatic, or regulatory purposes rather than statistical or research purposes. The evaluating administrative data knowledge brief follows a series of data quality dimensions to guide researchers in assessing the relevance, accessibility, interpretability, coherence, accuracy, and institutional environment to evaluate the fitness for use of the administrative data product.

International Government Data

International government data presents some unique data quality challenges, which require some specific evaluation strategies. Researchers need to pay attention to missing geographies and values, temporal limitations, adjustments, changes in methodologies, changes in historical data, discrepancies, data manipulation, and concept-measurement gaps.

Commercial Data Evaluation

The knowledge brief covers a list of common data quality issues that need researchers' attention. The issues include missing values, data errors, biases, inconsistencies, discrepancies, header data, standardization, superseded data, actual versus estimated data, reporting time issues, misuse of data, and lack of transparency. It also provides guidance for academic librarians to engage in conversations with data providers and researchers to identify potential data quality issues.

Survey Data Evaluation

Survey data evaluation follows the Total Survey Error (TSE) paradigm, which is a framework widely used to assess and enhance the quality of survey data. It guides researchers to evaluate the measurement aspect of a survey to assess the survey data's potential specification error (validity), measurement error, processing error, and analytical error and to evaluate the survey's representation aspect to evaluate potential coverage, sampling error, nonresponse error, and adjustment error. Further, total survey quality is also dependent on other nonstatistical quality dimensions such as credibility, comparability, usability/interpretability, relevance, accessibility, timeliness and punctuality, completeness, and coherence.

Conclusion

Unlike evaluating information quality based on currency, relevance, authority, accuracy, and purpose, evaluating data quality is a more complex and challenging process. The evaluation

framework can vary depending on the specific type of data product, such as data documentation, federal statistics, administrative data, commercial data, or survey data. The project provides foundational resources to help academic librarians enhance their awareness and competencies in evaluating data quality, develop critical thinking skills around library data products and sources, and build their ability to teach these skills in the classroom or through research consultations.

Notes

1. ACRL, *Information Literacy Competency Standards for Higher Education* (Chicago, IL: ACRL, 2000), http://hdl.handle.net/11213/7668.

2. ACRL, *Framework for Information Literacy for Higher Education* (Chicago, IL: ACRL 2016), http://www.ala.org/acrl/standards/ilframework.

3. Wendy Girven Pothier and Patricia B. Condon, "Towards Data Literacy Competencies: Business Students, Workforce Needs, and the Role of the Librarian," *Journal of Business & Finance Librarianship* 25, nos. 3–4 (2020): 123–46, https://doi.org/10.1080/0896356 8.2019.1680189.

4. P. B. Condon and W. G. Pothier, "Advancing Data Literacy: Mapping Business Data Literacy Competencies to the ACRL Framework," *Journal of Business & Finance Librarianship* 27, no. 2 (2022): 104–26, https://doi.org/10.1080/08963568.2022.2048168.

5. Institute of Museum and Library Services, "RE-252357-OLS-22: West Chester University (FHG Library)," Laura Bush 21st Century Librarian Program Grant, 2022, https://www.imls.gov/grants/awarded/re-252357-ols-22.

6. All session videos, transcripts, the ebook, and knowledge briefs are freely available on the project website (https://www.dataqualityliteracy.org).