Building Capacity of Academic Librarians in Understanding Quantitative Data, Data Quality Problems, and Evaluating Data Quality: A National Forum

Summary

West Chester University in collaboration with Stanford University, the University of Illinois at Urbana-Champaign, collaborative consultants Mr. Bobray Bordelon (Economics & Finance Librarian/Data Services Librarian at Princeton University) and Marydee Ojala (Editor-in-chief of Online Searcher, formerly with Research Library & Information Center at the BankAmerica Corp.), will request $149,992 funding to organize a **national open forum** (one prep session and six expert panel discussion Zoom conference sessions) on building the capacity of academic librarians in understanding quantitative data, data quality problems, and evaluating data quality. This project ties closely to the Laura Bush 21st Century Librarian Grant **Goal 3** and **Objective 3.3**, especially for building library workforce skills and expertise in **cultivating critical and creative thinking skills**. The **target group** of this project is academic librarians including subject librarians (in both science and social science subjects), research librarians, data librarians, information specialists at universities, colleges, and other research institutions. The project will engage the target groups through pre-forum and post-forum focus group discussions and participant feedback surveys to ensure the forum meets their needs and expectations. The national forum sessions will be open to any interested librarians from the targeted groups and beyond. We will invite data experts from diverse backgrounds and industries to share their perspectives and answer audiences' questions regarding data quality issues and data evaluation topics. Based on the forum, the project will create an open-access ebook (edited collection), infographic guides of a synthesized data evaluation strategy, and a project website to share the forum recordings, presentation slides, and other deliverables with the public. All project **deliverables** will be assigned CC-BY-NC Creative Commons License to allow sustainable access and broad impact. This project will help raise academic librarians' awareness of data quality problems, fill their knowledge gaps, and build their capacity and confidence in teaching data evaluation, critical thinking about data, and data literacy. As a long-term **impact**, the project will demonstrate the library's impact on educating data-literate citizens, improving the quality and integrity of scholarly output, reducing the social and economic costs caused by the data quality issues, and improving the quality of data-driven decision-making.

Project Justification

This national forum project addresses fundamental needs to fill the knowledge gap of academic librarians in understanding quantitative data, data quality problems and build their capacity in evaluating data quality and teaching data literacy. Data is ubiquitous in today's data-driven world; so are data quality problems. These problems incur high social and economic costs. In many ways, they can introduce errors and biases into research, pollute data analysis, distort research results, destroy research efforts, and mislead decisions based on such research. According to Gartner's Data Quality Market Survey, in 2017, poor data quality resulted in $15 million of annual financial cost on average to an organization (Moore, 2018)[1]. IBM estimated that the yearly cost of poor data quality is $3.1 trillion in the U.S. alone (Redman, 2016). Librarians often suppose that with millions of dollars paid to data providers each year, libraries have access to high-quality proprietary research data. Unfortunately, many data quality problems exist in the most reputable and costly research databases such as Compustat, CRSP, Datastream, I/B/E/S, Worldscope, SDC Platinum, BvD Orbis, Capital IQ. Rossi (2011) criticized Datastream - "the data is so bad and flawed that statistical inferences driven without a thorough review and correction exercise are, at best, totally unreliable" (p. 3). Bratten et al. (2016) found Compustat has an overall data error rate of 13% for financial statement footnote entries. Ma et al. (2011) estimated that the company

---

[1] For detailed citations, see the Supporting Document 1 References (Supportingdoc1.pdf)

profiles from Mergent miss 75% of all competitor pairs. McGuire et al. (2016) articulated the concept of "database effect" - "researchers would likely come to a different conclusion based on the database used" (p. 186). Many researchers have overturned prior research results after correcting specific data problems (Beedles & Simkowitz, 1978; Guenther & Rosman, 1994; Payne & Thomas, 2003; Brown & Larocque, 2011). The principal investigator's evidence synthesis research has identified **98** articles discussing such data quality problems in library databases and categorized 11 types of data quality problems: missing values, data errors, discrepancies, biases, inconsistencies, static header data, standardization, changes in historic data, lack of transparency, reporting time issues, and misuse of data (Liu, 2020).

The capability of understanding data quality issues and critically evaluating data quality is crucial for college students, scholars, researchers, knowledge workers, and professional employees, especially in the age of algorithms and big data (Head et al., 2020). Critical thinking over data quality is an integral part of data literacy. Academic librarians have played an indispensable role in educating college students and researchers about information literacy skills (Association of College & Research Libraries, 2016). And recent years have witnessed a rising demand for academic librarians to contribute to data literacy education (Prado & Marzal, 2013). The ability to evaluate data quality is the core of data literacy competency because it sets the "foundation for interpreting data, data-driven decision making, and data ethics, all of which rely on high quality, trustworthy data" (Pothier & Condon, 2020, p. 16). However, evaluating data quality is far more complicated than assessing general information (Batini et al., 2009). The typical information evaluation strategies such as the CRAAP (Currency, Relevance, Authority, Accuracy, and Purpose) Test are insufficient to understand complex data quality issues (Blakeslee, 2004; Batini et al., 2009; Liu, 2020). There are significant knowledge gaps in quantitative data, data quality problems, and data evaluation that need to be filled for academic librarians, so they can confidently and comfortably engage in data literacy education.

Previous library literature and IMLS-funded projects have contributed to building a framework for data literacy competencies, and all included evaluating data quality as a core competency (Carlson et al., 2011; Prado & Marzal, 2013; Pothier & Condon, 2020; Shields, 2005). For example, a previous IMLS-funded library project, Data Information Literacy, discussed evaluating data authority, relevance, compatibility, and metadata accuracy (Carlson et al., 2011). Our preliminary research also identified several data evaluation strategies/frameworks (Batini et al., 2009; Coughlan, Cronin & Ryan, 2007; Lee, 2006). However, library research and projects mostly focused on data management and have not cast enough light on data quality problems and data quality evaluation. And current data evaluation frameworks are often developed outside of the library field. They are either focusing on disciplinary research or on evaluating complex information systems (Batini et al., 2009; Coughlan, Cronin & Ryan, 2007; Lee, 2006).

Building upon prior research and results, this project seeks to deepen academic librarians' understanding of the topic through a series of expert panel discussions in the format of a national open forum. We will invite experts on data quality and data evaluation from diverse backgrounds and industries to share their insights and directly interact with the academic librarian audience through Q&A sessions. Based on the discussions, the project will create an edited collection ebook to allow the knowledge to be shared widely with audiences outside of the forum. Incorporating expert insights and comments, the project will develop a synthesized data evaluation strategy and create infographic guides for the strategy, which will be more suitable and adaptable for librarians' workshops, instruction sessions, and consultations. Hopefully, this project will help librarians build their capability and confidence in teaching data evaluation, data literacy, and critical thinking about data.

The national forum will be open to any interested librarians but will target <u>academic librarians</u> whose jobs involve locating data, understanding data, using data, evaluating data, and/or educating data users, including but not limited to subject librarians (including both science and social science subjects), research librarians, data librarians, information specialists at universities, colleges, and other research institutions. According to the data from the National Center for Education Statistics (NCES) and the Institute of Museum and Library Services (IMLS) that is cited by the American Library Association (ALA) (2018), there are **26,606** librarians working in academic libraries, and as one of the largest association for academic librarians, the Association of College and Research Libraries (ACRL) has **8,662** members (ACRL, 2021). The project will engage this target group through various channels and Listservs at ALA, ACRL, and beyond. Besides, the project will also reach out to <u>potential academic librarians</u> in Graduate Library and Information Sciences programs. Although this project targets academic librarians, any librarian who is interested in information literacy and data literacy will find this topic relevant and valuable. The **ultimate beneficiaries** of the project will be students, scholars, researchers, knowledge workers, professional employees, and decision-makers. The project will demonstrate the library's impact on educating data-literate citizens, improving the quality and integrity of scholarly output, reducing the social and economic costs caused by data quality issues, and improving the quality of data-driven decision-making.

Project Work Plan

1. Guiding Principles

The project work plan follows the project management best practices from the project management institute (Salapatas, 2000). The project will be conducted based on the following guiding principles to ensure effectiveness, efficiency, quality, and timeliness (See the *Performance Measurement Plan* for details):

1) A <u>work breakdown structure</u> and <u>baseline time estimate </u>(see the *Schedule of Completion*) will be used to track the status and progress of project tasks and ensure efficient and timely delivery of project tasks.
2) <u>Assessment</u> will be embedded in the whole project process. Periodical assessment (monthly, quarterly, and yearly) and task-related assessments will be discussed at regular group meetings.
3) <u>Target groups' perspectives</u> will be integrated into the design of the forums and deliverables through pre-forum and post-forum focus group discussions and forum participant feedback surveys.
4) <u>Diverse external expert perspectives</u> will lead the national forums and will be widely shared through a series of deliverables (forum recordings, ebook, website, and infographic guides). Their input will be validated, and consensus will be built through the feedback process after the forums.
5) <u>Event distribution and dissemination efforts</u> will be made to ensure the project reaches its target group widely and provides a lasting impact.
6) <u>Diversity, equity, inclusion, and accessibility considerations</u> will be foundational to this project, from recruiting focus group members, expert panelists, graduate student assistants, designing forum topics, to creating, promoting, and distributing event announcements and product deliverables.

2. Personnel

The project team (see the **List of Key Project Staff and Consultants**[2] for details) will bring our experience in project management, understanding of data quality issues, the capability of outreaching to data quality experts in both academia and industry, and the expertise in publication and instructional material design to ensure the success of the project. Each of us will have a project lead role based on our expertise to ensure the success of specific activities and maximize the value of the project investment and output.

---

[2] For details, please refer to Supporting Document 2 - Letters of commitment from consultants and project advisors (Supportingdoc2.pdf).

- Ms. Yan (Grace) Liu (**GL**, Project Director), Assistant Professor, Business Librarian at West Chester University. Ms. Liu will bring her extensive experience in project management, her knowledge on data quality issues, and successful practices in instructional design to this project. She will direct this project, lead pre-forum and post-forum focus group discussions, and develop a synthesized strategy for data evaluation.
- Ms. Rashelle Nagar (**RN**, Project Member), Business Research & Collections Librarian at Stanford University. Ms. Nagar will bring her diverse experience and knowledge working with data, data products, and data users to this project. She will lead the Zoom conference setup, promote the forum, and bring essential perspectives for the synthesized strategy development.
- Mr. Bobray Bordelon (**BB**, Collaborative Consultant), Economics & Finance Librarian/Data Services Librarian at Princeton University. Mr. Bordelon will bring his extensive experience with data, his knowledge of data quality issues, and his capability of reaching out to data quality experts in both academia and industry to lead the panelist recruitment, host the forum, and assist in developing a synthesized data evaluation strategy.
- Ms. Marydee Ojala (**MO**, Collaborative Consultant), Editor-in-chief of Online Searcher, formerly with BankAmerica Corp. Research Library & Information Center. Ms. Ojala will bring her industry expertise and extraordinary experience with editing, conference organizing, and webinar production to lead the ebook publication and assist panelist recruitment.
- Dr. Jodi Schneider (**JS**, Project Advisor), Assistant Professor at the School of Information Sciences, the University of Illinois at Urbana-Champaign. Dr. Schneider will bring expertise in library and bibliometric applications of data science and data quality, experiences in directing research and grant projects, as well as in continuing education of library professionals. She will provide project feedback and formal evaluations to enhance the assessment of this project and assist in hiring a graduate student in Library and Information Sciences.
- Graduate Assistant (Hourly) (**GA**, to be hired). The project will provide an opportunity for graduate students who are interested in data librarianship or related academic librarianship. The GA will assist with project logistics, collating forum recording and content, editing transcripts, preparing ebooks, and assisting with building the project website.

In addition, the project team will recruit 20+ expert panelists and will work with the Stanford University Information Technology department to host the national forum Zoom conferences.

3. Project Schedule

A. Phase I: Pre-Forum Planning (August - December 2022)

1) Conduct Project Meetings & Assessment

The project team will meet every other week in the first year and every month in the second year to make sure the project is on track. Project performance will be evaluated at regular meetings. The project director and project lead for each specific task (see **the Schedule of Completion**) will closely monitor the progress of the task at hand to identify early signs of bottlenecks and convene the group to make timely adjustments to ensure project milestones are completed according to schedule. Assessment will be conducted throughout the project on a monthly, quarterly, and yearly basis. Task-related assessments will be discussed at regular group meetings. The external project advisor, Dr. Jodi Schneider will attend quarterly project assessment meetings to provide feedback to the project and will provide an independent formal assessment for the first year and at end of the project (Project Lead: **GL; JS**)

2) Organize Informal Focus Group Discussion with Targeted Group

The project team will organize an informal focus group discussion with targeted group members at the beginning of the project in August-September 2022. We will select academic librarians from diverse backgrounds, subject areas, and educational settings to discuss their interests, needs, knowledge gaps, and expectations for the forum. The discussion will help develop forum themes and panelist topic guidelines. (Project Lead: **BB, GL**)

3) *Develop Forum Themes and Panelist Topic Guidelines*

Forum Themes and Panelist Topic Guidelines will ensure the forum discussion meets the target group's needs and expectations. We will organize a prep session to build up the target group's knowledge to understand quantitative data and prepare librarians for understanding more complex data quality problems. We will then host six expert panel discussion sessions that allow data experts to engage with academic librarians to help them understand data quality problems in different scenarios and strategies for evaluating data quality. The project will recruit panelists from varied backgrounds and industries and allow librarians to gain diverse perspectives on this issue. (Project Lead: **GL, BB, MO**)

Tentative Forum Themes

[Prep Session] Understanding Quantitative Data

This session will set the stage and prepare academic librarians with basic knowledge of numerical data, data collections, quantitative research design (reliability and validity), and basic statistical inferences. It will prepare the audience to understand deeper data quality problems and further develop critical thinking about data.

[Forum Session #1] Data Quality Problems and Data Evaluation: Perspectives from Academic Researchers

This session will invite academic researchers with extensive experience in primary quantitative research to help librarians understand how data quality is controlled in research design and is evaluated during data collection.

[Forum Session #2] Data Quality Problems and Data Evaluation: Perspectives from Data Users

This session will invite a) researchers who have conducted critical analysis on secondary data sources; b) leading journal editors/researchers who have engaged in replication studies; and c) librarians who have provided extensive support for data users to understand data quality. This session will help librarians understand data quality problems encountered when using library data sources and researchers' strategies in evaluating data quality and handling data problems.

[Forum Session #3] Data Quality Problems and Data Evaluation: Perspectives from Data Providers

This session will invite content managers or data quality personnel from the leading data providers to help librarians understand how data providers manage data challenges and control data quality. This session will allow librarians to understand data normalization and quality issues from a vendor perspective and provide insight into what data challenges can be remedied, which can be not.

[Forum Session #4] Data Quality Problems and Data Evaluation: Perspectives from Statistical Agencies

This session will invite expert panelists from statistical/data collection agencies to discuss the data quality issues they encounter when conducting major national surveys (such as the American Community Survey and the Consumer Expenditure Survey). This session will help librarians understand how statistical agencies control and evaluate data quality, create representationally diverse sample populations, and disclose potential limitations for using national statistics and data.

[Forum Session #5] Data Quality Problems and Data Evaluation: Perspectives from Employers

This session will invite company data analysts to discuss data quality problems they encountered at the workplace, what strategies they use to evaluate the data quality. The panelists will also provide insight into the expected skillset of data-literate college graduates when entering the job market. It will help librarians understand the data quality problems and data evaluation at the workplace and inform data literacy competencies for college students.

[Forum Session #6] Data Quality Problems and Data Evaluation: What Future Awaits

This session will put the data quality problems and evaluation in the context of artificial intelligence, the Internet of Things, quantum computing, and big data. We will invite information industry professionals to help us understand what data quality problems will be resolved by further technologies, what will remain, and what new data quality challenges we will face.

West Chester University

Tentative Panelist Topic Guidelines

The topic guidelines will be revised based on the target group's focus group discussion and will be tailored to individual sessions.

- What is the specific context that the data quality problems occur?
- What is the typical object of quantitative data products to be evaluated (data points, datasheets, datasets, database, a combination, or other?)
- What are the data quality problems encountered by the panelist?
- Can you give specific case studies or examples to demonstrate the problems?
- In the instances of data creation, what steps/procedures/approaches/strategies are used/can be used to control data quality?
- In the instances of using secondary data, what steps/procedures/approaches/strategies are used/can be used to evaluate data quality?
- What cognitive and critical thinking skills do researchers need to develop for data evaluation?
- What affective skills and mental habits do researchers need to develop for data evaluation?
- How Diversity, Equity, Inclusion, and Accessibility (DEIA) issues are considered in research design, data collection, and data evaluation?
- How to potentially solve the data quality problems identified through evaluation?
- How to teach students and researchers critical data evaluation skills?
- What data quality evaluation knowledge and competencies do college students need to develop to get career-ready?
- What can librarians do to help students and researchers develop data evaluation skills?
- How do data literacy, information literacy, and critical thinking correlate, and how the responsibilities can be shared among librarians and educators to cultivate data-literate citizens?

4) Identify and Recruit Expert Panelists

We have developed a list with examples of potential panelists (see panelist examples below). Extensive team efforts will be put into identifying 30+ potential expert panelists through research, outreach, and referrals and recruiting 20+ Panelists. The potential panelist information will be organized in a shared excel file and details of the recruitment process and stages will be recorded and tracked. We will code and index the potential panelists' expertise to ensure the diversity, breadth, and depth of the topics and better manage the recruitment efforts. Mr. Bordelon will lead the effort of recruiting panelists, Ms. Ojala will assist in reaching out to industry experts and Ms. Liu will provide research support for identifying panelists. After identifying the expert panelists, consent forms will be prepared and signed by panelists to clarify copyright issues and get permission for content distribution on the web and to target groups. Panelist introduction will be requested for event promotion. Zoom meetings may be scheduled with the panelists to explain the purposes, goals, and expectations of the forum and to answer any questions they may have. (Project Lead: **BB, MO**)

Panelist Examples

- Academic researchers with extensive expertise in quantitative data collection, such as Researchers from Kahneman-Treisman Center for Behavioral Science and Public Policy at Princeton University, NORC at the University of Chicago, Climate Change Institute from the University of Maine.
- Data specialists from data curation leaders such as Inter-university Consortium for Political and Social Research (ICPSR), Roper Center for Public Opinion Research (Roper), and UK Data Archive (UKDA).
- Librarians/administrators from research data centers such as Indiana State Data Center, and Cornell Federal Statistical Research Data Center.

- Scholars and researchers who have conducted critical analysis on using secondary data sources, such as authors of the article: McGuire, J. B., James, B. E., & Papadopoulos, A. (2016). Do your findings depend on your data (base)? A comparative analysis and replication study using the three most widely used databases in international business research. *Journal of International Management*, 22(2), 186-206.
- Statisticians or data specialists from statistical/data collection agencies such as the U.S. Census Bureau, Bureau of Labor Statistics (BLS), National Center for Environmental Information (NCEI), who engaged in major national surveys and data collection, such as the American Community Survey (ACS), the Consumer Expenditure Survey (CEX); National Longitudinal Survey of Youth (NLSY) and the Atmospheric, Coastal, Geophysical, and Oceanic research.
- Content managers, or data quality control/assessment personnel from the leading data providers and aggregators such as Center for Research in Security Prices (CRSP), Refinitiv, Standard & Poor's, WRDS Database, Simply Analytics, Statista, DataPlanet, etc.
- Journal editors from the American Economic Association, who have conducted replication studies for submitted scholarly articles; researchers from the Center for Open Science, who have engaged in the high-profile Reproducibility Project: Cancer Biology (RPCB) initiative or scholars who published replication reports in the Advances in Methods and Practices in Psychological Science.
- Academic Librarians, data librarians, information specialists, who have provided extensive support for data users to understand data quality.
- Data analysts from top employers in the nation such as Walmart, UPS, IBM, etc.,
- Researchers or data analysts from top marketing research firms (e.g. Nielsen), investment banks (e.g. JP Morgan Chase), data analytics companies (e.g. IQVIA), etc.
- Information or technology industry professionals (such as experts from Google, Microsoft, Zillow, etc.) who have extensive knowledge about the data issues with artificial intelligence, Internet of Things (IOT), quantum computing, big data, etc.

5) Setup Zoom Conference and Technology Support

The project team will work with the Stanford University Information Technology department to set up the Zoom Conference. The forum will be organized using Stanford University's Zoom Webinar platform and license. The registration form will be set up via the Zoom webinar platform. (Project Lead: **RN**)

6) Prepare for Event Announcements and Promotion

The project team will distribute the event announcements widely through diverse channels, social media, and listservs, including but not limited to **ALA**, ALA Affiliates, ALA State and Regional Chapters, Student Chapters, **ACRL**, Reference and User Services Association (**RUSA**), Special Library Association (**SLA**), etc. A special effort will be made to reach out to underrepresented groups, such as the Black Caucus of the American Library Association (**BCALA**), the National Association to Promote Library & Information Services to Latinos and the Spanish Speaking (**Reforma**), the American Indian Library Association (**AILA**), the Asian/Pacific Library Association (**APALA**), the Chinese-American Librarians Association (**CALA**). Additional efforts will be made to distribute the event information directly to **graduate library students and educators** via the Association for Library and Information Science Education (**ALISE**) and the offices of Graduate Library and Information Sciences programs and encourage graduate library students' participation, spark interest in academic and data librarianship and build their capacity for data evaluation and data literacy early on. (Project Lead: **BB, RN**)

7) Conduct Literature Reviews to Complement Expert Opinions

One of the project goals is to develop a synthesized data quality evaluation strategy (see detailed explanation in Phase III) for easy adoption and adaptation for library workshops, instruction sessions, and consultations. Our preliminary

research has identified several references, but a further review of existing literature on the topics of data quality evaluation frameworks and strategies will be conducted to complement the expert panel discussions. (Project Lead: **GL**)

B. Phase II: National Forum Execution (January 2023 - July 2023)

1) Host the Prep Session and Six Expert Panel Discussion Sessions

The project team will host the national forum (one prep session and six expert-panel discussion sessions) between January 2023-July 2023. The forum sessions will be moderated by Mr. Bordelon. Each session will run for about 1.5 hrs and allow the audience to engage with panelists through a Q&A session. The Zoom sessions will be enhanced with auto-generated captions for accessibility purposes and will be recorded and transcribed. The graduate student will further help clean up the Zoom transcription, upload the recordings to Youtube, and organize panelist introductions and presentation materials for building the project website. (Project Lead: **BB**)

2) Conduct Participant Feedback Survey

A Participant Feedback Survey will be created and distributed to registered participants after each discussion session. The feedback survey will incorporate the questions to help assess the effectiveness and quality of forum sessions. The survey results will be analyzed and feedback will be incorporated into the planning of future sessions. (Project Lead: **GL**)

3) First-year Project Assessment

After the delivery of all forum sessions, a major first-year project assessment will be conducted to evaluate the project outcome. Dr. Jodi Schneider will provide an independent assessment for the project outcome with a short written report. The project team will assess the data collected for the performance measurements and refer to Dr. Schneider's report to make adjustments and improvements (See the ***Performance Measurement Plan*** for details). (Project Lead: **GL, JS**)

C. Phase III: Post-Forum Deliverable Creation (August 2023 - June 2024)

1) Create an Ebook (Edited Collection)

Based on the transcripts of the forum recordings, we will create an edited collection ebook. The ebook will be organized into sections/chapters based on the forum themes and each panelist's perspectives will be edited into an individual article with a structured format. The individual ebook manuscript will then be confirmed with the panelist and will be reviewed and proofread by project team members. The ebook will be distributed with the Creative Commons License Attribution-NonCommercial (CC BY-NC). Compared with white papers, an ebook with an ISBN will reach a much wider audience. We will catalog the ebook into West Chester University (WCU) Library Catalog, which will be included in WorldCat, and we will promote it to be included in other library catalogs as well. The ebook will be deposited into WCU digital commons and will be discovered by Google Scholar. We will also share the ebook via Open Educational Resources (OER) platforms such as ACRL Sandbox, OER Commons, MERLOT, OpenStax, and on the project website to allow free download by the public. (Project Lead: **MO, GL**)

2) Create a Synthesized Data Evaluation Strategy and Infographic Guides

In order to make the data evaluation approaches shared by the experts more easily adopted by librarians, the project team will make extra effort to develop a synthesized data evaluation strategy and infographic guides. The project team members will bring their expertise on the topic and Dr. Schneider will bring her expertise in educating library professionals to develop a synthesized data evaluation strategy based on transcript content synthesis, literature review, expert comments, and input from the targeted groups. We will send the draft strategies to data experts for their comments and will gather the input from the targeted group through a focus group discussion. Several infographic guides will be created via Canva.com or Adobe Illustrator to display the synthesized strategy. The infographics will be designed in different sizes, formats, and file types (PDF, PNG, JPG, GIF, SVG, MP4, etc.) for easy adaptation, so they can be

printed as large posters or small bookmarks, used or shared digitally for varied purposes. Studies show that the human visual system can process and assign meaning to imagery much faster than reading through a text explanation, so infographics can be an effective tool to optimize the learning process (Gareau, et al., 2015; Bystrova, 2020). The project director has been using synthesized infographic guides for several years in teaching and distributing complex information literacy concepts. Her recent contribution to the ACRL Sandbox of the infographic guide "Master Literature Review Process" in Public Administration has been downloaded 11,834 times in less than a year. Her "4-step Strategy on Evaluating Online Sources" infographic guide can be used as a model for creating the synthesized data evaluation strategy. The infographic guides will be shared via ACRL Sandbox, WCU digital commons, and the project website to allow free download by the public. (Project Lead: **GL, BB, RN**)

3) Organize Post-forum Informal Focus Group Discussion

A post-forum focus group discussion will be organized to inform the finalization of the synthesized data evaluation strategy and infographic guides. The participants will be recruited from the pre-forum focus groups and via the forum participant feedback survey. A draft synthesized data evaluation strategy and infographic guides will be shared with the focus group to get their input for changes, adaptation, and wide distribution. This effort will ensure the data evaluation strategy meets the target group's expectations and is applicable and adaptable to facilitate academic librarians' workshops, instruction sessions, and consultations. (**Project Lead: GL**)

4) Build Project Website

We will build the project website under **wcupa.edu** domain name and utilize the West Chester University's Web Team Services to help maintain the project website for long-term access. The Web Team Services will provide support for webpage search engine optimization, data analytics, and accessibility check. The website will display all the content created for the national forum, including the forum youtube videos, expert introductions, PowerPoint slides, the ebook, and infographics. The accessibility of all digital products will be assessed with the help of the Web Team Services via WAVE Web Accessibility Evaluation Tool, with reference to BCCampus' Accessibility Statement. All content will be assigned Attribution-NonCommercial (CC BY-NC) Creative Commons License and allow free download by the public. (Project Lead: **GL, GA**)

D. Phase IV: Dissemination (July 2024 and onwards)

The project deliverables will be distributed widely under the CC BY-NC license for broad impact. They will be **(1)** shared with national forum participants and panelists; **(2)** shared in Open Educational Resources platforms such as ACRL Sandbox, OER Commons, MERLOT, OpenStax; **(3)** saved in library catalogs and institutional repositories at the WCU and will be discovered by Google Scholar and WorldCat; **(4)** shared with librarians (in the target group and beyond) via listservs, social media at ALA, ALA Affiliates, ALA State and Regional Chapters, Student Chapters, ACRL, RUSA, SLA, and related groups at the International Federation of Library Associations and Institutions (IFLA); **(5)** special efforts will be made to distribute to underrepresented groups such as BCALA, AILA, APALA, Reforma, CALA; **(6)** additional efforts will be made to distribute to graduate library students and educators via ALISE and offices of the Graduate Library and Information Sciences programs; **(7)** The ebook will be promoted to be included in university library catalogs to allow wide discovery. **(8)** the project website will be search-engine optimized for easy discovery online. The project team will continue the distribution effort on the data evaluation strategy and the project after the project completion through **(9)** publications (possible venues include *American Libraries*, *College & Research Libraries News*, *Online Searcher*, etc.) and **(10)** presentations (possible venues include ALA, SLA, ACRL Annual Conference, International Association for Social Science Information Service and Technology (IASSIST) Conference, LibLearnX, Loex Conference, Information Literacy Symposiums and Summits and conferences outside of library field such as Federal Reserve Bank of St. Louis' Beyond the Numbers Conference).

Diversity Plan

The Forum will strengthen the library's commitment to diversity, equity, inclusion, and accessibility practices. The National Forum will engage data experts from diverse backgrounds and industries. The DEIA topic will also be integrated into the expert panel discussion and help librarians understand how the DEIA issues are considered during data collection and data evaluation. The project will invite librarians from diverse racial/ethnic, socioeconomic, gender, geographical, national, institutional, and disciplinary backgrounds to participate in the national forum sessions. Besides regular communication channels with ALA, ACRL, RUSA, and SLA, participant invitations and deliverables will be directly sent to underrepresented groups through channels such as BCALA, AILA, APALA, Reforma, CALA, etc. Participants from diverse subject areas and institutional backgrounds will be invited to participate in the pre-forum and post-forum focus group discussion and collaborate in defining forum themes, topic guidelines, and developing a synthesized data evaluation strategy. The project deliverables will be checked for their accessibility to individuals with visual disabilities. This project provides a paid graduate assistant position for an MLIS student and will make deliberate efforts to prompt the forum and disseminate project deliverables to graduate library students via ALISE and offices of Graduate Library and Information Sciences programs to help fulfill IMLS's commitment to developing future professionals in the library fields.

Project Results

The project will advance knowledge and understanding of data quality problems and data quality evaluation through a series of project deliverables, which include: **(1)** the national forum expert-panel discussions, presentation slides, and recordings, **(2)** an open-access ebook collection, **(3)** a synthesized data evaluation strategy, **(4)** infographic guides, and **(5)** the search-engine optimized project website. These project outputs will build greater knowledge, skills, awareness, and abilities of library professionals in understanding data quality problems and data evaluation strategies. These project deliverables will provide sustained benefits to academic librarians in facilitating their workshops, instruction sessions, consultations, and in-depth conversations with students, scholars, and researchers on data quality problems, data quality evaluation, and critical thinking about data. All project deliverables will be made readily adaptable, generalizable, and usable by other institutions and communities nationally. The deliverables will be open to the public under the CC BY-NC license and allow users to remix, adapt, and build upon this work non-commercially. The ebook will be posted to digital commons (discovered by Google Scholar), OER Platforms, library catalogs, and WorldCat for long-term preservation and easy access for students, researchers, educators, and the public. All project deliverables will be shared with the public via the project website under wcupa.edu domain. The recordings will be made available via YouTube and embedded into the project website. All of these efforts will allow project output to be sustainably preserved for long-term access after the project completion. The project members will make continued efforts to promote the deliverables through publication and presentation at the venues in and outside of the library field during and after the project completion.

As a short-term outcome, the project will help raise academic librarians' awareness on data quality problems and help fill librarians' knowledge gap, bolster their confidence in in-depth conversations with scholars and researchers on data quality issues, and motivate librarians to teach data evaluation, data literacy and critical thinking about data. As a long-term impact, this effort will eventually help raise educators', researchers', and students' awareness of data quality issues and help them develop well-rounded data evaluation skills and critical thinking skills about data. The raised awareness and improved critical thinking skills about data quality will help improve the overall quality of data collection and data creation over time. It will help demonstrate the library's impact on educating data-literate citizens, improving the quality and integrity of scholarly output, reducing the social and economic costs caused by the data quality issues, and improving the quality of data-driven decision-making.

Schedule of Completion

Phase	#	Activity (Year 1: August 2022- July 2023)	Project Lead	GL	BB	MO	RN	JS	GA	Task
Phase I Pre-Forum Planning	1	Bi-weekly project meetings & quarterly formal assessment	GL, JS	60	24	24	24	8	18	**Admin.**
	2	Organize pre-forum informal focus group	BB, GL	12	10	2				**Preparation**
	3	Develop forum themes and panelist guidelines	GL, BB, MO	15	4	4		2		Forum
	4	Identify 30+ expert panelists	BB, MO	8	30	10			10	Ebook
	5	Develop consent forms for recording, copyright, etc.	GL	12	2	2			2	Website
	6	Recruit 20+ expert panelists, request intro and slides	BB	5	40	5			20	Infographics
	7	Hire graduate student assistant	GL, JS	20					6	**Deliverable**
	8	Setup Zoom conferences, registration, and presenter prep	RN, GA	6	2		12		10	Forum
	9	Advertise and promote the events	BB, RN, GL	8	12	3	8		6	Ebook
	10	Conduct literature review to complement expert opinions	GL	33					10	Website
Phase II Execution	11	Host the national forms (7 sessions)	BB, GL	20	24	12	12		12	Infographics
	12	Conduct participant feedback survey and analysis	GL	12	2	2	2			**Evaluation**
	13	Clean up and produce forum transcripts	MO, GA	5		5			30	**Dissenm.**
	14	Compile PPT, panelist intro, recordings, etc.	GL, GA	5					10	
	15	First-year project assessment	GL	10	2	2		4		
		First Year Total		231	152	71	58	20	128	

Baseline Time Estimate (hrs.) columns: GL, BB, MO, RN, JS, GA

Timeline (Aug, Sep, Oct, Nov, Dec, Jan, Feb, Mar, Apr, May, Jun, Jul)

Schedule of Completion

	Activity (Year 2: August 2023 - July 2024)	Project Lead	GL	BB	MO	RN	JS	GA	Timeline (Aug–Jul)	Milstones	Deadline
Phase III — Post-Forum Deliverable Creation			*Baseline Time Estimate (hrs.)*								
1	Monthly project meetings & quarterly formal assessment	GL, JS	32	12	12	12	8	12		Forum	7/31/2023
2	Edit transcripts for an adapted, edited collection	MO, GL	26	6	40			8		Ebook	6/30/2024
3	Create intro and prepare other book content	MO, GL	14		15			5		Website	12/31/2023
4	Confirm the manuscripts with panelists	MO, GA			5			10		Infographics	6/30/2024
5	Design book cover, format, register ISBN and publish	GL, MO	20		10					Final Report	7/31/2024
6	Develop synthesized data evaluation strategies	GL, BB, RN	30	12	2	6	8				
7	Organize post-forum informal focus group	GL	18	2	2	4		5			
8	Gather expert feedback, revise, finalize the strategies	GL, BB	16	10	2	2		5			
9	Create infographic guides	GL	28	2	2	2		5			
10	Build project website	GL, GA	20	2	2	2		20			
Phase IV — Dissem.											
11	Dissenminate project deliverables	GL, BB, MO	15	5	5	2					
12	Evaluate project outcome and submit IMLS report	GL, JS	16	2	2	2	4	2			
	2nd Year Total		235	53	99	32	20	72			
	Two-year Summary		466	205	170	90	40	200			

Personnels: Yan (Grace) Liu (**GL**, Project Director); Bobray Bordelon (**BB**, Collaborative Consultant); Marydee Ojala (**MO**, Collaborative Consultant); Rashelle Nagar (**RN**, Project Member); Graduate Student Assistant (**GA**, To be hired); Jodi Schneider (**JS**, Project Advisor)

Applicant Name: West Chester University, Stanford University, and the University of Illinois at Urbana-Champaign

Project Title: Building Capacity of Academic Librarians in Understanding Quantitative Data, Data Quality Problems, and Evaluating Data Quality: A National Forum

Digital Products Plan

Type

The project will create the following digital products:

1. **Recordings** for one prep session and six expert-panel discussion sessions. Zoom webinar software will be used to host and record the national forum sessions. Recordings will be saved in **MP4** format. Estimated 1.5 hrs for each session and altogether **10.5** hrs of total recording. Recording files will be uploaded to YouTube and embedded into the project website.

2. **Transcripts** of the seven recordings will be created in word format and be edited into an ebook.

3. **Powerpoint Slides** (20+) from the panelists will be shared in **PDF** format.

4. **Ebook:** one ebook (edited collection) will be published (approx six sections and 3-4 manuscripts for each section, all together approx. 20 articles) in **PDF** format.

5. **Infographics**: several infographics will be created and produced in **JPG, PNG**, **PDF,** and other available formats.

6. **Website**: The project website will be built under **wcupa.edu domain** and display all the digital content in **HTML** format.

Availability

The digital products will be made widely available to <u>librarians and the public</u>, through the following efforts:

1. All digital products (forum recordings, transcripts, PowerPoint slides, an ebook, and infographics) will be **publicly available online** via the project website. The website will be built under wcupa.edu domain and will be search-engine optimized to increase discovery.

2. The digital products (including ebook, infographics, and project website links) will also be shared in **Open Educational Resources** platforms such as Association of College & Research Libraries (ACRL) Sandbox, OER Commons, MERLOT, OpenStax.

3. The ebook will be saved in **institutional repositories** at West Chester University and will be discovered by Google Scholar and will be included in West Chester University's **library catalog** and **WorldCat** to allow wide discovery. The ebook will be promoted to be included in other university libraries' catalogs as well.

4. The digital products (including ebook, infographics, and project website links) will be distributed to librarians via various channels, listservs, social media at ALA, [ALA Affiliates](#), [ALA State and Regional Chapters](#), [Student Chapters](#), ACRL, RUSA, SLA, and IFLA.

5. Special efforts will be made to distribute to underrepresented groups such as BCALA, AILA, APALA, Reforma, CALA, etc.

6. Additional efforts will be made to distribute to graduate library students and educators via ALISE and offices of the [Graduate Library and Information Sciences programs](#).

7. The project will continue to prompt the project and project results via publications after the project completion (possible venues include *American Libraries*, *College & Research Libraries News*, *Online Searcher*, etc.). In the case that publication is not freely available, the author will self-archive the pre-print copy in an institutional repository in compliance with the publication's copyright policies.

8. The project will continue to prompt the project and project results via presentations (possible venues include ALA, SLA, ACRL Annual Conference, International Association for Social Science Information Service and Technology (IASSIST) Conference, LibLearnX, Loex Conference, [Information Literacy Symposiums and Summits](#) and conferences outside of library field such as Federal Reserve Bank of St. Louis' [Beyond the Numbers Conference](#)).

Access

1. **Permissions** will be obtained from panelists to record the forum sessions and distribute the recording and presentation slides under Attribution-NonCommercial (CC BY-NC) **Creative Commons License**.

2. All digital products created by the project members (ebook and infographics) will be publicly available online via the project website under the **CC BY-NC License** and allow users to **download**, **remix, adapt** and **build upon** this work non-commercially.

3. The **accessibility** of all digital products and project web pages will be checked with the assistance of West Chester University's Web Team Services via [WAVE Web Accessibility Evaluation Tool](#), with reference to [BCCampus' Accessibility Statement](#) before dissemination.

Sustainability

1. The digital products created by this project will be **freely and readily available** for use and reuse by **the public**.

2. The public-facing website will be built under the wcupa.edu domain name and will be maintained by the West Chester University's Web Team Services for an indefinite period of time with the commitment for at least 10 years.

3. The recordings will be made **permanently available** via YouTube and embedded into the project website.

4. The ebook will be **permanently preserved** via library catalogs, open educational resource platforms, and institutional repositories mentioned above.

5. The infographics will be **permanently preserved** via open educational resource platforms and institutional repositories mentioned above. The infographic guide will continue to be **updated** and the updated version will be shared via the educational resource platform and institutional repositories as well.

West Chester University

Organizational Profile

Organization's Mission
Founded in 1871, West Chester University is a community of educators dedicated to developing graduates who succeed personally and professionally and contribute to the common good. WCU, one of the top regional comprehensive public universities in the nation, serves more than 17,700 students at the undergraduate and graduate levels. It is the largest state-owned university in the Commonwealth of Pennsylvania. We offer more than 100 undergraduate academic opportunities, more than 80 certificate and master's degree programs and four doctoral degrees.

To fulfill its mission as a public institution serving the Commonwealth, West Chester University will focus on student success and institutional improvement as measured by the degree to which WCU demonstrates: Access to learning; Community and cultural engagement; Continuous improvement; Critical thinking; Inclusion; Scholarly and creative activities; and, Sustainability.

West Chester University Libraries provides the best possible support for West Chester University's mission of educational excellence by offering the highest quality collections, services, and facilities for learning, teaching, and research. The libraries have 759,000 print resources, over 7000 journal subscriptions and 40 full time staff.

Governance Structure
WCU is a part of the Pennsylvania's State System of Higher Education, established in 1983. A Board of Governors is responsible for oversight of the State System. The Board establishes broad educational, fiscal, and personnel policies. Among other tasks, the Board appoints the chancellor and each university president. Additionally, each university has a Council of Trustees that serves as a policy-setting board for the university. At the University level, the University Libraries is under the Academic Affairs division along with six other Colleges and School and there is shared governance between the administration and faculty.

Organizational Service Area
WCU has a student population of 17,719 graduate and undergraduate. The student body is 61% female and 39% male. The majority of the population is white at 74%, additional groups include 11% African American, 6% Latino, 3% Asian and the remaining population are Other. While most undergraduates are recent high school graduates preparing for career objectives, many others are older individuals, including veterans and homemakers, who either never before had the opportunity for a college education or whose schooling was interrupted. Most students are residents of Pennsylvania but students from other states and foreign countries are welcomed. The student body at WCU represents a cross section of many ethnic, racial, and religious groups and includes students from all socioeconomic levels.

The main campus located in West Chester, PA has two libraries: The Francis Harvey Green (FHG) Library and the Presser Music Library. In addition, to its physical space, the University Libraries provides service to all distance education students.

Brief History of the FHG Library
The first library on campus was housed in Old Main Hall. Student contributions to the collection and growing demand led to the construction in 1902 of what is today known as Old Library. In 1947, the library was named for beloved English professor Francis Harvey Green. Two decades later, a new building was constructed; a later five-story addition completed today's FHG Library. The library is currently a keystone of the campus quad, a central hub of student learning, and a home to millions of accessible educational resources.