Building Capacity of Academic Librarians in Data Quality Evaluation: A National Forum
West Chester University and collaborative consultants Mr. Bobray Bordelon and Ms. Marydee Ojala will request $149,995 funding to organize a national forum of six expert panel discussion webinars on building the capacity of librarians in evaluating data quality. This project ties closely to the **Laura Bush 21st Century Librarian Grant Goal 3 and Objective 3.3**. The targeted audiences include academic and research librarians, data librarians, business librarians at universities, colleges, research institutions, etc. We will bring together data quality experts including experienced researchers, data users, database providers, journal editors and reviewers, librarians, statistical agency officials, data analysts, and information industry professionals to share their insights on this topic. An open-access edited collection based on the webinar will be published. Training materials such as library guides, infographic guides, lesson plans, instruction handouts, data reference workbooks will be developed for librarians to facilitate their instruction, consultation, workshops, or data literacy and information literacy programs on data quality evaluation. The project will fill librarians' knowledge gap in data evaluation and will allow them to provide value-added service to benefit the research community and enhance library users' data literacy, critical thinking, and lifelong learning.

Project Justification
Data is ubiquitous; so are data quality problems. According to Gartner's Data Quality Market Survey, the average annual financial impact of poor data quality on an organization is $15 million per year (Gartner, 2018). IBM estimated that the yearly cost of poor data quality is $3.1 trillion in the US alone (Redman, 2016). The ability to understand data quality issues and critically evaluate data quality is crucial for today's working professionals. Critical thinking over data quality is an integral part of data literacy and is an essential skill for college students to have especially in the age of algorithms and big data (Head et al., 2020). Academic and research librarians have played an important role in educating college students in information literacy and data literacy. Among all data literacy competencies, data quality evaluation is the core competency since each of the other competencies includes some variation of evaluation of data, and 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, data quality evaluation is far more complicated than general information evaluation. The common information evaluation strategies such as the CRAAP test (Currency, Relevance, Authority, Accuracy, and Purpose) are not sufficient to understand complex data quality issues. The principal investigator's recent study has identified 11 categories of data quality problems in reputable library-subscribed business research databases, which include 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).

There is a lack of in-depth research and discussion in the library field on data quality evaluation. Previous library projects such as data information literacy focused on data management and believed that librarians have the capability to teach evaluating data quality for its authority, relevance, compatibility, and metadata accuracy (Carlson et al., 2011). But in general, most academic librarians have limited knowledge on how reliability and validity are ensured in the research design process; how data is aggregated and data quality is controlled in subscribed databases; how research data is evaluated in the peer review process and replication studies; what special data evaluation skills are needed in business settings when dealing with sales, customer, and credit data, etc. and even what data quality evaluation means in the context of big data, machine learning, and quantum computing. This project intends to fill librarians' current knowledge gap in this regard.

This project will target librarians whose job involves locating data, understanding data, using data, and evaluating data, including but not limited to academic librarians, research librarians, data librarians, business librarians at universities, colleges, research institutions, federal agencies, and international organizations. Any librarian who is interested in information literacy and data literacy will find this topic relevant and valuable. The project will ultimately benefit researchers and students who are tomorrow's working professionals and leaders. The knowledge of data quality evaluation will build their critical thinking skills, data literacy, and information literacy skills for life-long learning. Through a series of expert panel discussions, we hope to approach data quality evaluation more deeply, comprehensively, and more critically. We will bring together data experts such as experienced researchers, data users, database providers, journal editors, peer-reviewers, librarians, statistical agency officials, data analysts, and information

industry professionals who possess deep insights into data quality and evaluation issues to fill this knowledge gap for librarians. This project ties closely to the **Laura Bush 21st Century Librarian Grant** Goal 3 and objective 3.3.

Project Work Plan

Project team includes Grace Liu (principal investigator), Assistant Professor, Business Librarian at West Chester University; Bobray Bordelon, Economics & Finance Librarian/Data Services Librarian at Princeton University; and Marydee Ojala, Editor-in-chief of Online Searcher. We will bring our extensive experience and 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. Several data quality experts from Princeton University, Cornell Federal Statistical Research Data Center, Census Bureau Quality Program, Bureau of Labor Statistics, CRSP, Standard & Poor's, Federal Reserve Economic Data, American Economic Association, ACRL, RUSA BRASS have been identified and expert identification will be carried out further through research, outreach, and referrals. **August 1, 2022-December 2022**: PI (Grace Liu) will lead the creation of a detailed project plan, develop themed topics, and brainstorm detailed strategies for recruiting 20 speakers for six national forum webinar series. Bi-weekly meetings will be organized to ensure the progress of the project and initial efforts will be made by Mr. Bordelon to recruit speakers identified. **January 2023 - July 2023**: Mr. Bordelon will further lead the recruitment of experts with the help of the team and lead to organize six virtual expert panel webinars (1.5 hrs. for each session). The webinars will be hosted via West Chester University's Zoom Conference Platform. WCU Conference Services will assist with the registration. The Webinars will be transcribed by a student assistant based on automatic zoom transcription and reviewed by Ms. Ojala as we prepare the edited collection. **August 2023 to July 2024**: Ms. Ojala will lead the creation of the edited collection and the eBook publication. Ms. Grace Liu will lead a literature synthesis on the data quality evaluation beyond the library field and will synthesize the expert insights from the webinars and the research into actionable data evaluation strategies and develop training materials for librarians to easily integrate data quality evaluation into their research consultation, instruction, data literacy workshops, and information literacy programs. The handbook and training materials will be reviewed and finalized by the group. Grace Liu will work with the student assistant to build the website to host all the content and promote the website to targeted audiences.

Diversity Plan

The project will engage experts and audiences from diverse backgrounds. Efforts will be made to reach out to minority librarians through list-serves such as The Black Caucus of the American Library Association, The National Association to Promote Library & Information Services to Latinos and the Spanish Speaking (Reforma), The American Indian Library Association, The Asian/Pacific Library Association, and the Chinese-American Librarians Association.

Project Results

Through a series of expert panel discussions and engaging conversations, we will raise wide awareness of data quality issues and will build librarians capability in evaluating data quality. But the more sustainable and far-reaching impact will be achieved through the further creation of other deliverables. These deliverables include recorded webinars, an open access eBook (edited collection), and a training/instructional material package (including library guides, lesson plans, case studies, slides, infographic guides, instruction handouts, data reference workbooks) that will facilitate easy integration of data evaluation into the library instruction, research consultations, data literacy workshops, and information literacy programs. A website will be created to host the recording (including expert introduction and presentation slides), the eBook, and training materials. We will assign a Creative Commons License to these materials and allow download, adaptation, and distribution of these resources widely. Promotion efforts will be made to distribute the materials to targeted audiences via listservs and the webinar registration list. All the materials will still be available after the grant period and the website will be sustained for at least five years and the updating effort will be made as needed.

Budget

Faculty salary ($54,952.07) and fringe benefits ($9,308.88), student assistant ($1,032), eBook Design and Publication ($1,000), Collaborative Consulting/Professional Services Agreements ($51,984), speaker honorarium fees ($10,000), Webinar Hosting & Technology ($2,000), ISBN and Web Registration ($495), Indirect Cost ($19,223.48), in total $149,995.43.