Building Capacity of Academic Librarians in Understanding Quantitative Data, Data Quality Problems, and Evaluating Data Quality

A National Forum

INSTITUTE of **Museum** and **Library** SERVICES

Propose, Plan, Execute, and Deliver: Developing a National Forum and Digital Deliverables around Data Quality Evaluation

Grace Liu, Bobray Bordelon, Rashelle Nagar, Jordan Sarti, Uyen Nguyen
2023 Beyond the Numbers Conference

Propose: How the Idea is Developed?

Submit an article   Journal homepage

Research Article

Data quality problems troubling business and financial researchers: A literature review and synthetic analysis

Grace Liu ✉

❝ Download citation   🔗 https://doi.org/10.1080/08963568.2020.1847555   Check for updates

📄 Full Article   🖼 Figures & data   📄 References   ❝ Citations   📊 Metrics   🖨 Reprints & Permissions   **Get access**

Abstract

The data quality of commercial business and financial databases greatly affects research quality and reliability. The presence of data quality problems can not only distort research results, destroy a research effort but also seriously damage management decisions based upon such research. Although library literature rarely discusses data quality problems, business literature reports a wide range of data quality issues, many of which have been systematically tested with statistical methods. This article reviews a collection of the business literature that provides a critical analysis on the data quality of the most frequently used business and finance databases including the Center for Research in Security Prices (CRSP), Compustat, S&P Capital IQ, I/B/E/S, Datastream, Worldscope, Securities Data Company (SDC) Platinum, and Bureau van

Related research ℹ

People also read | Recommended articles | Cited by 6

Eight things to know about business research data ›

Alice Kalinowski et al.
Journal of Business & Finance Librarianship
Published online: 19 Nov 2020

Commercial data in academic business research: A study on use and access ›

2

Propose:
The Initial Idea and a Failed Proposal

NLG-LIBRARIES-FY21   West Chester University

The Role of the Academic Libraries in Promoting the Quality and Integrity of Business and Finance Research Data: A Pilot Study

West Chester University (WCU) seeks the Research in Service to Practice Grant of $149,950 to support a 2-year pilot qualitative study to address the high priority gaps of academic libraries in understanding business and finance research data quality and integrity issues and investigate the role of academic libraries in promoting the business research data quality and integrity. The study will develop new service models, best practices, and collaborative approaches for academic libraries and librarians and build the capacity of academic libraries to serve as catalysts to the scholars and researchers community.

Propose:
The Second Attempt to Make the Change

West Chester University

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

Propose:
Pre-to-full Proposal: Build a Stellar Team

- Ms. Grace Liu (Principal Investigator), West Chester University
- Ms. Rashelle Nagar (Co-PI), Stanford University
- Mr. Bobray Bordelon (Collaborative Consultant), Princeton University
  Ms. Marydee Ojala (Collaborative Consultant), Editor-in-chief of Online Searcher
- Dr. Jodi Schneider (Project Advisor), the University of Illinois at Urbana-Champaign
- Ms. Jordan Sarti (Graduate Assistant), the University of Illinois at Urbana-Champaign
- Ms. Uyen Nguyen (Graduate Assistant), the University of Illinois at Urbana-Champaign

Plan: 1st Year Schedule of Completion

			Schedule of Completion																		
			Baseline Time Estimate (hrs.)							**Timeline**											
		Activity (Year 1: August 2022- July 2023)	Project Lead	GL	BB	MO	RN	JSch	JS	UN	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
Phase I — Pre-Forum Planning	1	Bi-weekly project meetings & quarterly formal assessment	GL, JSch	60	24	24	24	8	18	18												
	2	Organize pre-forum informal focus group	BB, GL	12	10	2																
	3	Develop forum themes and panelist guidelines	GL, BB, MO	15	4	4		2	2				Extended									
	4	Identify 30+ expert panelists	BB, MO	8	30	10			4													
	5	Develop consent forms for recording, copyright, etc.	GL	12	2	2																
	6	Recruit 20+ expert panelists, request intro and slides	BB	5	40	5			10						Extended							
	7	Hire graduate student assistant	GL, JSch	20				6														
	8	Setup Zoom conferences, registration, and presenter prep	RN, JS	6	2		12		6													
	9	Advertise and promote the events	BB, RN, GL	8	12	3	8		4													
	10	Conduct literature review to complement expert opinions	GL	33					8													
Phase II — Execution	11	Host the national forms (7 sessions)	BB, GL	20	24	12	12		12	12												
	12	Conduct participant feedback survey and analysis	GL	12	2	2	2															
	13	Clean up and produce forum transcripts	MO, JS	5		5			32													
	14	Compile PPT, panelist intro, recordings, etc.	GL, JS	5					4													
	15	First-year project assessment	GL	10	2	2		4														
		First Year Total		231	152	71	58	20	100	30												

Plan: 2nd Year Schedule of Completion

			Schedule of Completion																		
			Baseline Time Estimate (hrs.)							Timeline											
	Activity (Year 2: August 2023 - July 2024)	Project Lead	GL	BB	MO	RN	JSch	JS	UN	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
1	Monthly project meetings & quarterly formal assessment	GL, JSch	32	12	12	12	8		12												
2	Edit transcripts for an adapted, edited collection	MO, GL	26	6	40				10												
3	Create intro and prepare other book content	MO, GL	14		15																
4	Confirm the manuscripts with panelists	MO, UN			5				5												
5	Design book cover, format, register ISBN and publish	GL, MO	20		10																
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			2												
8	Gather expert feedback, revise, finalize the strategies	GL, BB	16	10	2	2			5												
9	Create infographic guides	GL	28	2	2	2															
10	Build project website	GL, UN	20	2	2	2			30												
11	Disseminate project deliverables	GL, BB, MO	15	5	5	2			6												
12	Evaluate project outcome and submit IMLS report	GL, JSch	16	2	2	2	4														
	2nd Year Total		235	53	99	32	20		70												
	Two-year Summary		466	205	170	90	40	100	100												

Phase III: Post-Forum Deliverable Creation (Activities 1–10)
Phase IV: Dissem. (Activities 11–12)

Personnels: Yan (Grace) Liu (**GL**, Project Director); Bobray Bordelon (**BB**, Collaborative Consultant); Marydee Ojala (**MO**, Collaborative Consultant); Rashelle Nagar (**RN**, Project Member); Jodi Schneider (**JSch**, Project Advisor); Jordan Sarti (**JS**, Graduate Assistant); Uyen Nguyen (**UN**, Graduate Assistant)

Plan:
Seven Webinar Sessions

1. Evaluating Data Quality: Challenges and Competencies

2. Quality Assurance in Data Creation

3. Understanding and Evaluating Governmental Data (U.S. and International)

4. Commercial Data Quality Issues

5. Data Quality: Reproducibility and Preservation

6. Data Quality: Evolving Employer Expectations

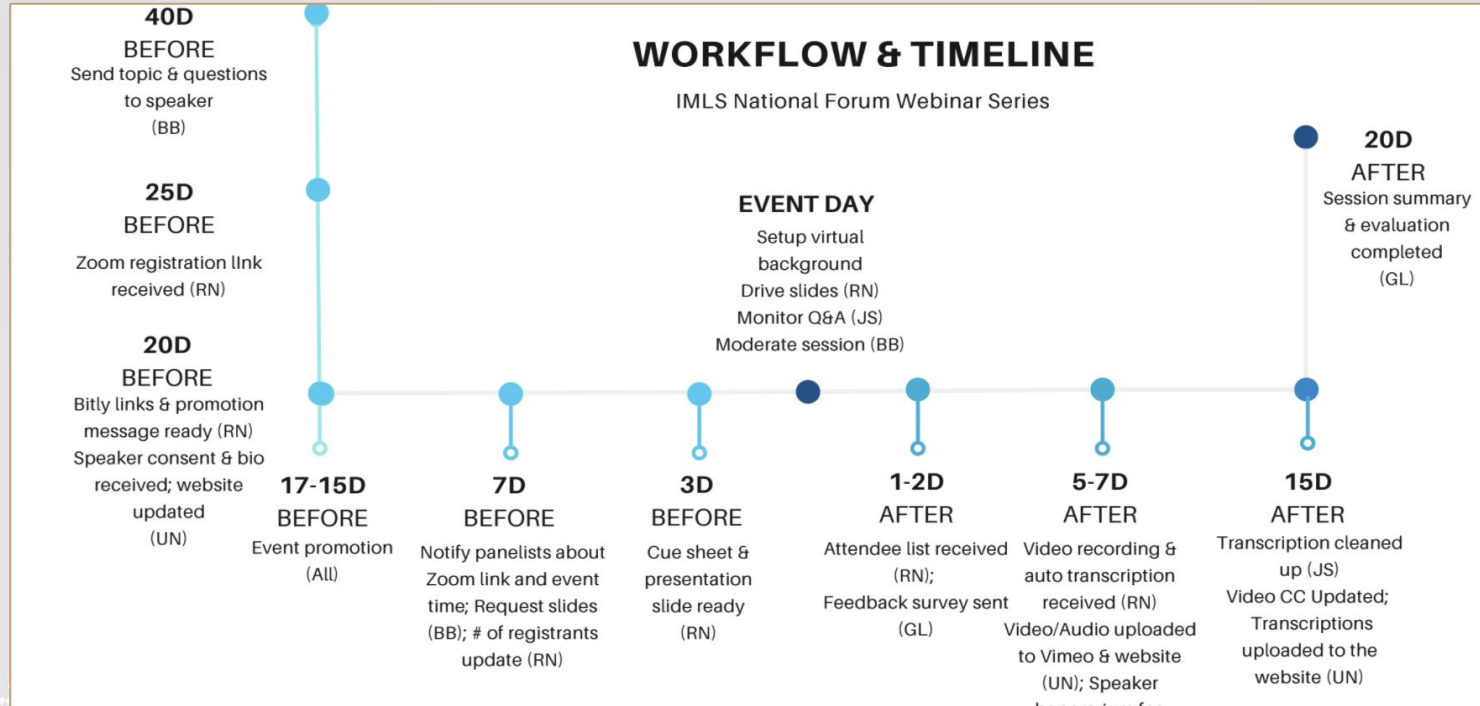7. Librarians' Role in Cultivating Data-Literate Citizens

Data Quality Evaluation
Awareness & Competencies

Event Schedule

01 Evaluating Data Quality — Challenges & Competencies
Thursday January 26, 2023
Registration is closed.

02 Quality Assurance in Data Creation
Thursday February 23, 2023
Registration is closed.

03 Understanding & Evaluating Government Data
Thursday March 30, 2023
Registration is closed.

04 Commercial Data Quality Issues
Thursday April 27, 2023
Registration is closed.

05 Data Quality Reproducibility & Preservation
Thursday May 25, 2023
Registration is closed.

06 Data Quality Evolving Employer Expectations
Thursday June 29, 2023
Registration is closed.

07 Librarians' Role in Cultivating Data-Literate Citizens
Thursday July 27, 2023

Plan: Event Workflow

WORKFLOW & TIMELINE

IMLS National Forum Webinar Series

40D BEFORE
Send topic & questions to speaker (BB)

25D BEFORE
Zoom registration lInk received (RN)

20D BEFORE
Bitly links & promotion message ready (RN)
Speaker consent & bio received; website updated (UN)

17-15D BEFORE
Event promotion (All)

7D BEFORE
Notify panelists about Zoom link and event time; Request slides (BB); # of registrants update (RN)

3D BEFORE
Cue sheet & presentation slide ready (RN)

EVENT DAY
Setup virtual background
Drive slides (RN)
Monitor Q&A (JS)
Moderate session (BB)

1-2D AFTER
Attendee list received (RN);
Feedback survey sent (GL)

5-7D AFTER
Video recording & auto transcription received (RN)
Video/Audio uploaded to Vimeo & website (UN); Speaker

15D AFTER
Transcription cleaned up (JS)
Video CC Updated; Transcriptions uploaded to the website (UN)

20D AFTER
Session summary & evaluation completed (GL)

9

Plan:
Event Promotion

- ALA Connect (ALA, ACRL, RUSA, RUSA-BRASS, IRRT, Core)
- ALA Affiliates
- State Library Associations
- ALA-accredited Library Schools
- Data Librarian List
- Academic Librarian List
- Special Library Association/IFLA
- Other Listservs: JESSE; BUSLIB-L; CODE4LIB
- Past Registrants

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Total	Unique
# of Registration	1615	995	967	526	582	712	771	6168	**3348**
# of Total Users	992	517	518	296	297	363	407	3390	**1931**

Execute: Speaker Recruitment
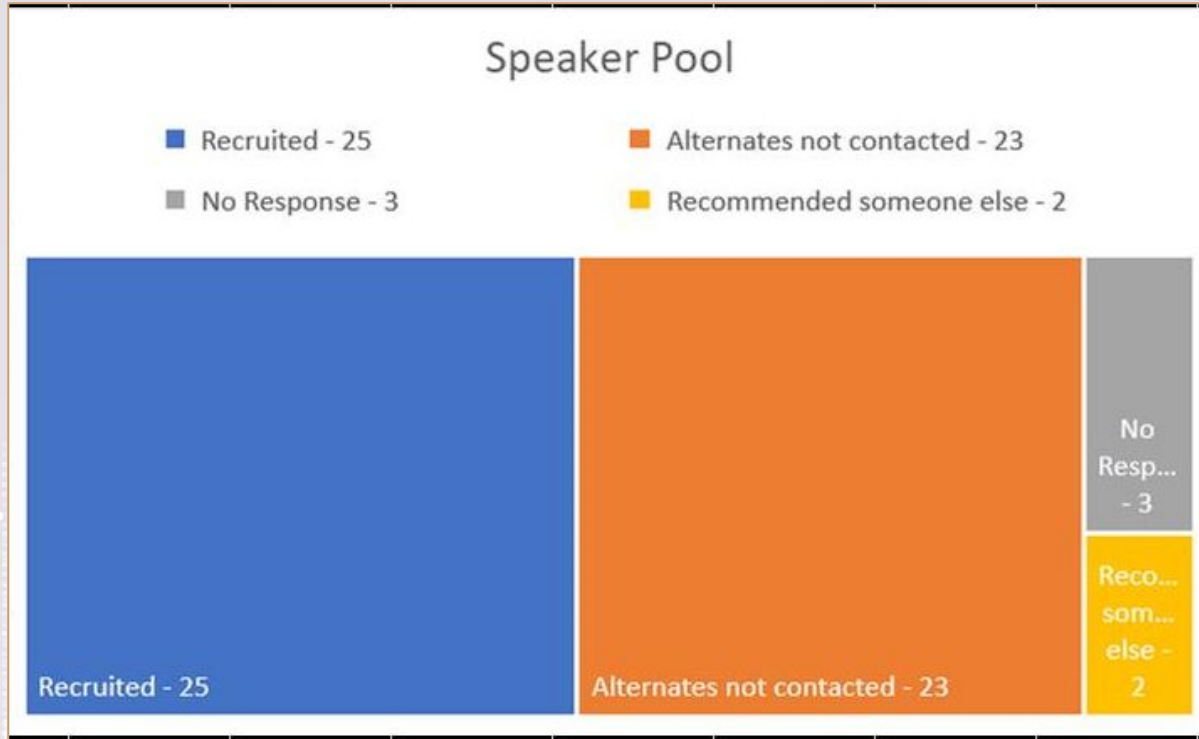
Execute: Speaker Recruitment

Execute: Speaker Recruitment
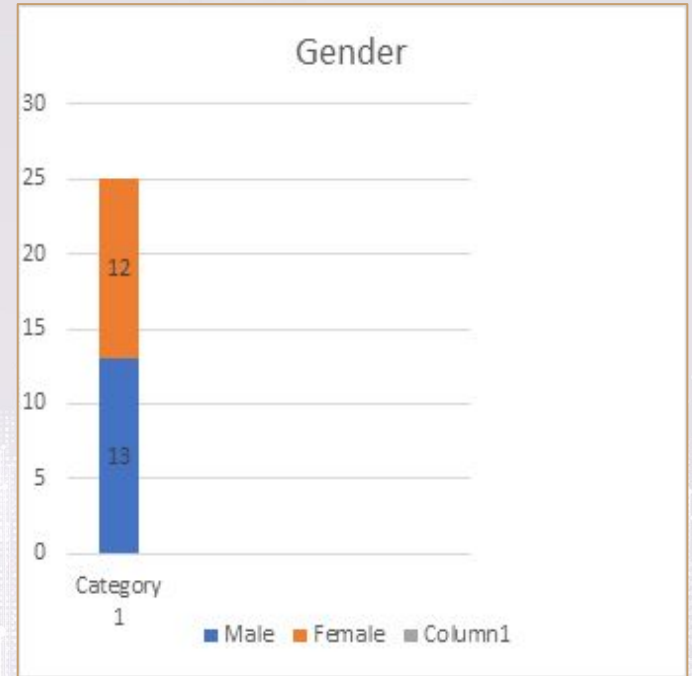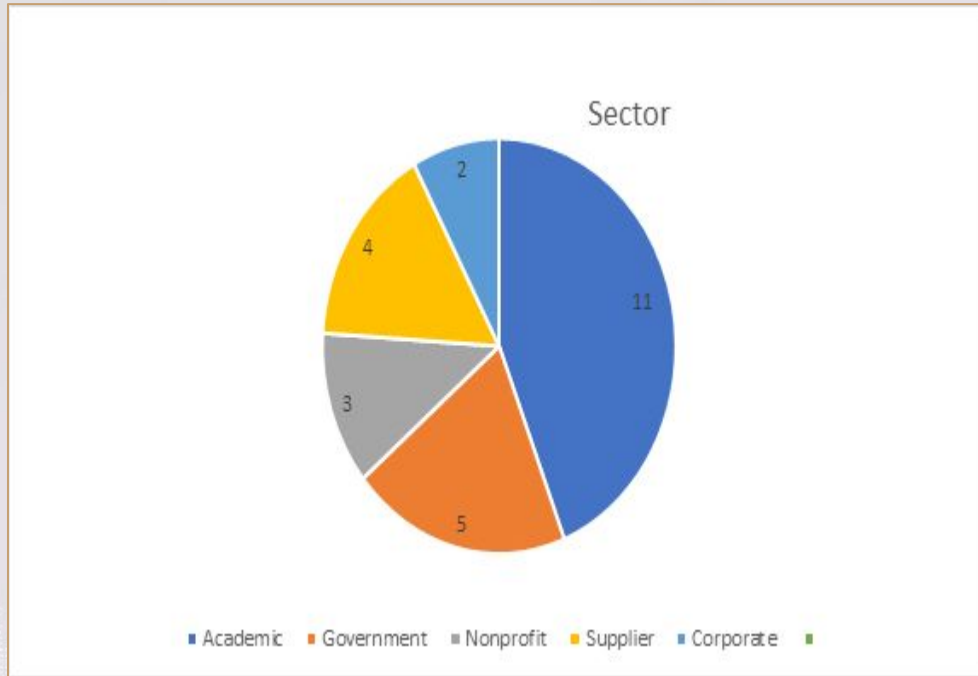
Execute: Speaker Recruitment

- Be open to suggestions from others
- Reach out to those you know
- Find a complimentary mix in terms of sectors, view points
- Do not assume No because seemingly out of scope
- Do not assume yes
- Be persistent and follow up

Execute: Moderating Webinar Sessions

For the Event:

- Do not try to do too much if moderating
- Have someone else handle technology
- Have someone else organize Q&A and remove technical & previously answered questions
- Let the audience know what to expect & NOT to expect (hands-on)

Q&A:

- Often the most interesting & practical part
- Moderator needs to know the subject area to be able to combine and prioritize
- Leave sufficient time

Execute: Branding

- Logo

- Website design elements

- Slide deck

- Zoom background

Execute: Technology

- Help from GSB Events

  & Operations Team

- Zoom Webinar

- Link tracking

Understanding & Evaluating
Governmental Data (U.S. & International)
with
James Church, Bill Sermons, and Jill Blaemers

March 30, 2023

This project was made possible in part by the Institute of Museum and Library Services [RE-252357-OLS-22].

Execute: Logistics

- Panelists

- Slides

- Cue Sheet

- Q&A Tracking document

Execute: Takeaways

- Plan, plan, plan

- But be flexible

- Accurately assess technology needs

Execute: Transcription & Webinar Q&A

- Cleaning up Zoom transcriptions

- Managing webinar Q&A behind the scenes

- Some takeaways

Deliver:
Build Project Website and Event Sessions

- Speakers confirmed

⬇

- Request materials (send reminder)

⬇

- Update website

⬇

- Speakers approval

Session Speakers

Kira Lillard
Technical Services Librarian, Federal Reserve Bank of Kansas City

Kira Lillard is a librarian in the Research Division at the Federal Reserve Bank of Kansas City where she is responsible for researchers' code and data releases, website administration, and is a member of the Data Governance Committee. She holds a BA from The Colorado College, an MA in Asian Studies from the University of Hawai'i, and an MLS from Emporia State University. Her research interests include web usability, research access and replicability, and universal design.

Lars Vilhuber
Executive Director, Labor Dynamics Institute, Cornell University School of Industrial and Labor Relations (ILR)

Lars Vilhuber holds a Ph.D. in Economics from Université de Montréal, Canada, and is currently on the faculty of the Cornell University Economics Department. He has interests in labor economics, statistical disclosure limitation and data dissemination, and reproducibility and replicability in the social sciences. He is the Data Editor of the American Economic Association, and Managing Editor of the Journal of Privacy and Confidentiality.

Kathleen Weldon
Director of Data Operations and Communications, Roper Center for Public Opinion Research, Cornell University

Kathleen has managed data collection and management efforts for the non-profit Roper Center for Public Opinion Research since 2016. During that time, she has led the implementation of a new acquisition policy based on a combination of methodological and transparency criteria, enactment of a Transparency Project for all data providers, and development of a bifurcated collection based on methodological evaluations, as well as enhanced presentation of methodological information to aid in the evaluation of historical data files. Most recently, she has directed the development of inclusion policies for building Roper Center-curated trends of polling questions.
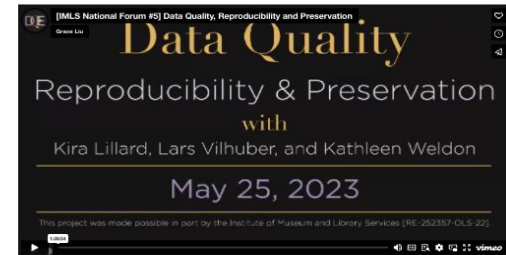
Deliver:
Prepare the Deliverables

- Google Sites:
  - Easier to use
  - Collaboration between team members
  - Sharing with others
  - Limited customizing capabilities
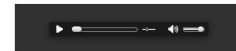 - Google domain and back-up storage

Assess:
Project Statistics

- 3,348 Total unique registrants
- 1,931 Total unique attendees
- Registrants from 55 countries
- Over 1,000 different organizations
- Diverse sections with academic focus
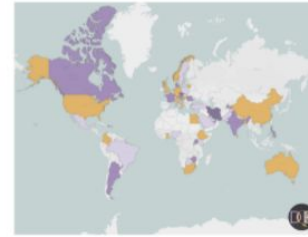- 10,000+ Website page views

10,674 Total Pageviews

3,000+

3,348 Total Unique Registrants

1,900+

1,931 Total Unique Web Attendees

Global Reach

The series attracted registrants from:

6 continents & 55 countries

Organizations

The series attracted registrants from:

Over 1,000 different organizations.

Sectors

The series attracted registrants from:

Academic institutions, public libraries, non-governmental organizations, central banks, government agencies, corporations, state libraries, national libraries, healthcare systems, and school libraries.

Questions & Answers

Visit our project website at:

https://sites.google.com/view/imls2022-data-quality