Using and Evaluating U.S. Federal Statistics

U.S. Federal Statistics and its Quality: The Basics

- The U.S. federal statistical system is a decentralized, interconnected network of 13 principal statistical agencies with about 100 additional federal statistical programs.
- Legislation such as the **Information Quality Act** ensures the quality of information disseminated by federal statistical agencies.
- The Office of Management and Budget issued Information Quality Guidelines and statistical policy directives such as Standards and Guidelines for Statistical Surveys to specify quality requirements regarding:
  - » Utility: the usefulness of the information to the intended users.
  - » Objectivity: the information is accurate, reliable, and unbiased and is presented in an accurate, clear, complete, and unbiased manner.
  - » Integrity: protecting information from unauthorized access or revision and not compromised through corruption or falsification.
- Agencies including the U.S. Census Bureau created their own Statistical Quality Standards to ensure statistical information quality from planning programs, acquiring data, producing estimates, analyzing data, reporting results, and releasing information to documentation.

Evaluating Federal Statistics for Research Needs

- Despite the high quality of federal statistics, researchers still need to assess the fit between the dataset and their research needs (refer to **Data Quality Literacy Series 03: Evaluating Dataset for Research Needs**).
- Researchers also need to be aware of the potential

dataset structure quality issues in federal statistics:

- » **Changes in the dataset schema and structure.** Data collection can also be discontinued due to funding cuts or changes in mandates. This change can affect data processing and the comparability of data across time.
- » **Inconsistent use of variable codes.** Do not assume variable codes are the same between data collections. For example, if age 12 was coded as XYZ in a 2020 survey, it can be coded as XYZ2 in the next release. There can be underlying differences in how data was collected or tabulated.
- » **Changes in survey questions or variable codes or labels between surveys.** For example, a new survey collection instrument is used in one year but not the next. Then, the trend can become discontinuous.
- » **Latency or source release schedule changes.** For example, the Internal Revenue Service (IRS) can be late in publishing an updated statistical release.
- » **Changes in reference materials**. For example, updates to North American Industry Classification System (NAICS) codes every five years can affect data comparability over time, potentially disrupting continuity in series.
- » **Lack of documentation.** Especially for administrative records, the data documentation may be insufficient for researchers to understand how to use the data.

References

Liu, G., Bordelon, B., Nagar, R., Sarti, J., Nguyen, U., & Boettcher, J. (2024). *Data Quality Literacy: A Guidebook*. Institute of Museum and Library Services (IMLS) Grant Project. https://doi.org/10.31219/osf.io/ruawm

Office of Management and Budget. (2022). *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information.* https://www.federalregister.gov/documents/2002/02/22/R2-59/guidelines-for-ensuring-and-maximizing-the-quality-objectivity-utility-and-integrity-of-information

Visit the project website to learn more!
https://www.dataqualityliteracy.org

INSTITUTE of Museum and Library SERVICES