Evaluating Administrative Data Quality

Administrative Data Quality Dimensions

Administrative data, collected for operational, programmatic, or regulatory purposes, require a distinct quality evaluation approach compared to survey data. Federal statistical agencies broadly define data quality as "fitness for use," recognizing that different users of the same data may have different assessments of its quality. The following **quality dimensions** adapted from the Data Quality Assessment Tool for Administrative Data by the Federal Committee on Statistical Methodology can be utilized to aid the administrative data assessment:

- **Relevance**
  » Do the content, scope, level of measurement, coverage period, frequency, and timelines meet the user's needs?

- **Accessibility**
  » What are the administrative restrictions for accessing the data? How will the data be accessed by the user?
  » Was there any data collected but not included, due to confidentiality or other reasons?
  » What control methods or modifications were used to protect the confidentiality of the data?

- **Interpretability**
  » Are data variables and valid values clearly defined? Is the data dictionary available?
  » What methodology is used to recode original data or create a value for a new variable (e.g., assigning a reported age to an age range)?
  » Are data architecture and the relationship between key variables explained?
  » Does the dataset contain records for those who are denied eligibility to administrative programs and can those records be identified?

Does the value reflect only the "latest" version (refer to the concept of "Header Data" in **Data Quality Literacy Series 08: Evaluating Commercial Data Quality**)? Has the data been revised or updated? Have the superseded records been removed at a given time?

- **Coherence**
  » Are there any classification systems (e.g., race and ethnicity categories or NAICS) used for categorizing or classifying the data? Were there any changes to the system in the extracted data?
  » Were there changes or differences across geographic areas covered that would cause breaks in consistency such as different questions, revised questions, questions in different languages, or deleted questions?
  » Were there changes to the instructions that the data was collected or processed (e.g., instructions for completing the application form)?
  » Were there changes to geographical boundaries?
  » Were there substantial changes or differences across the geographical areas that influenced who participated in the program (e.g., legislative changes, eligibility changes, program expansions, or natural disasters impacting program participation)?

- **Accuracy**
  » What percentage of eligible participants are not included in the data file? What is known about their characteristics?
  » Are there duplicate records or missing values? What are the known sources of error?
  » What questions are most often misinterpreted?
  » Are there any revisions to the reported value and why are the changes made?

- **Institutional Environment**
  » Does the purpose of the administrative program align with the research purpose?
  » Who is the data collected from and how is source data collected?
  » Are there quality control standards and processes applied?
  » Will there be new records or revisions to existing records after data acquisition?

To learn more, refer to **Data Quality Literacy Series 05: Understanding Administrative 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

Iwig, W., Berning, M., Marck, P., & Prell, M. (2013). *Data Quality Assessment Tool for Administrative Data.* https://www.fcsm.gov/assets/files/docs/DataQualityAssessmentTool.pdf

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

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