The Data Quality Act: Developing IQ Standards in a Political Environment

Patrick Flanagan
Bureau of Transportation Statistics
U.S. Department of Transportation
pat.flanagan@bts.gov

Executive Summary/Abstract: The “Data Quality Act” is a recent law in the United States requiring every federal agency in the U.S. Government to produce information quality guidelines. It further implies that the agencies are required to follow those guidelines giving them the effect of standards.

At the U. S. Department of Transportation we have written guidelines to comply with the Act, improve data quality, and create a consistency across our data systems. Doing this in the political environment of the federal government adds additional challenges that had to be addressed in the development process. This presentation will describe how we simultaneously addressed data quality and the realities of implementation.

The “Data Quality Act”

- Federal agencies must issue guidelines:
  "ensuring and maximizing the quality, objectivity, utility, and integrity of information . . . ."
- Establish administrative public feedback mechanism:
  "allowing affected persons to seek and obtain correction of information ... that does not comply with the guidelines"
- Report annually on administrative mechanism

U.S. Department of Transportation (DOT)

- DOT establishes the United State’s overall transportation policy.
- Under its umbrella there are 11 “administrations.” Each is also a federal agency
- They cover highway planning, development, and construction; motor-carrier safety; urban mass transit; railroads; aviation; and the safety of waterways, ports, highways, and oil & gas pipelines.

DOT Data Systems - Examples

- Accident Reporting Systems (Air, Rail, Boats, Transit, Pipelines, Hazardous Materials)
- Fatality & Injury Data from States (Highway)
- Highway Performance Monitoring System
- National Transit Database
- Travel Surveys

Information Quality vs. Politics

- DOT data and analysis is used in rulemaking, congressional action, and funding distribution
- Information quality efforts strive for “good” data
- Groups affected by the data want it to support their interests
- If the data does not support their interests, they can attempt to discredit the data, support contrary analysis, or try to change the data
- Legislated data collection may have limitations

Political Concerns

- Special interests may use the “complaint system” to:
  - Fight data that they do not like
  - Challenge methods for same reason
  - Force priority changes to suit them
- Federal data may be withheld
- Hold feds responsible for third-party data (e.g., states)
The Politics - Impacts

- Will "complaints" bog down resources?
- Will compliance with the law be expensive to implement with no additional funds?
- Will requirements related to third-party data result in loss of third party data?
- If we give too much information will we be bogged down with challenges?

Method Step #1 - Develop Purpose List

- We began by brainstorming the general purposes to be served by the guidelines
  - Define minimum methodology in each phase (planning, collection design, collection, processing, analysis, data dissemination)
  - Maximize transparency of information to data users
- Then broke each general area into more detailed purposes
  - Use common data definitions
- Finally, we included constraints imposed by the Act

Method Step #2 - Political Issues

- Brainstorm issues affecting data quality that may challenge the success of implementing the guidelines
  - Do not withhold data simply to avoid public challenge
- Need Two Types of Guidelines – one under the Act and a second one meant as pure guidance on details
- Tailor guidelines to address third-party data like from states or industry sources
- Write guidelines to facilitate a data quality assessment program

Method Step #3 - Brainstorm Guidelines

- Taking one purpose at a time, brainstorm candidate guidelines, which, if followed, will achieve the purpose
- Create a matrix cross referencing guidelines to purposes
- Ensure adequate representation for all purposes & no guidelines without a purpose
- Note conflicts between guidelines and define balance in wording

The Matrix Tool

Some Results

- Guidelines will not establish a specific quality level; the planning phase should identify that
- Specific “minimum” methods identified in each phase
- Provide documentation of methods and other data quality information, even negative information
- Review data and supporting information prior to dissemination
- Gradually refine the guidelines as systems improve
Data Quality Level
- The needs of data quality must be linked back to data objectives derived from user needs
- All data has error – how much is decided in design and execution of collection
- Design the data collection process to meet quality needs, not “maximize quality”
- Contrary to popular belief, sometimes the users just need a “ballpark” estimate, not a precise one

Minimum Methods
- Identify processes that must be performed in each phase
  Example: “A minimum editing process should include range, validity, and consistency checks.”
- Phases include planning, collection, processing, dissemination, and evaluation.
- Wording of guidelines should be “go vs. no-go”

Documentation
- “Transparency” in dissemination is telling the user what we know about the data
- It is mostly about documentation which takes time
- Political Concern – The more detail you provide the more you can be “nitpicked”
- Reaction – Provide detail on what was done, but don’t go overboard on the “whys”

Documentation Impact
- Writing takes time, usually from an expert
- The documents have to be kept up: configuration management
- Not a problem if its in the budget; but it is often not
- Documentation does not pay for itself, except possibly in avoiding bad decisions

Documentation – Why?
- Planning documentation shows relevance between the data and the users
- Design documentation shows how collection will get data meeting the user needs
- Execution and evaluation documentation shows how it was carried out
- Analysis documentation shows how the data was used to draw a conclusion

Data Review – The Panacea
- Pre-dissemination review is touted as the solution to quality problems
- With documentation the data review is straightforward
- Without documentation it is usually questionable
- Is peer review necessary or even helpful?
Conclusions

- Improving IQ in the presence of constraints and political considerations
- Don’t fight politics; treat it along with IQ
- Mix purposes/objectives of both in design tools
- Set initial minimum IQ standards at attainable levels, and move the “bar” up sequentially
- Guidelines may be used to justify resources for IQ