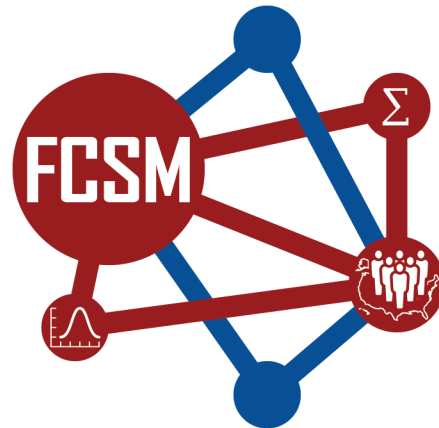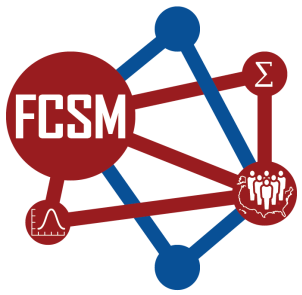# Who's Keeping the Score…on Data Quality!

Wednesday, October 25, 2023

Session F-1: Who's Keeping the Score….on Data Quality!

Organizer: Darius Singpurwalla, National Center for Science and Engineering Statistic

Agenda

1) Introduction to the FCSM Data Quality Framework.
2) Making the Case for Data Quality: The FCSM Data Quality Handbook of Case Studies
3) Rule-Based Data Validation and Reconciliation of Survey Responses
4) Continuing to Make the Case for Data Quality
5) A Data Quality Scorecard to Assess a Data Source's Fitness for Use
6) Still Making the Case for Data Quality
7) Panel Discussion w/ Audience Input

Overview of the FCSM Data Quality Framework

Figure ES 1. The FCSM Data Quality Framework
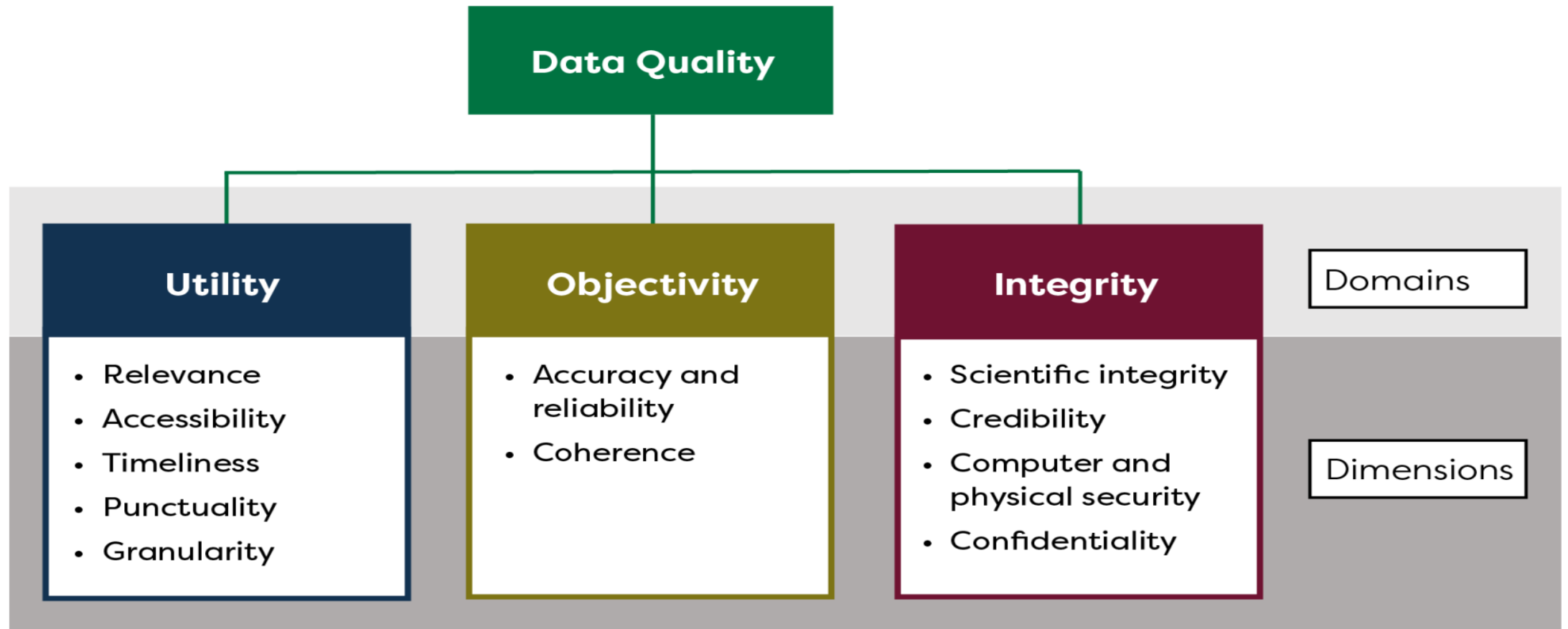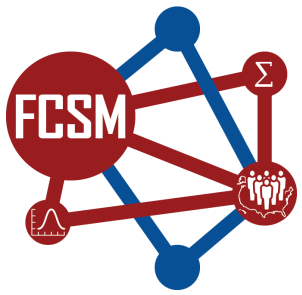
FCSM: Case Study Handbook

FCSM Data Quality Case Study Handbook
- Consists of seven case studies
- Follows this format:
  - Description of the data being assessed for quality.
  - How the FCSM data quality framework compares with previous efforts to assess data quality
  - Description of implementation including human capital, technology needed, and cost
  - Assessment of the source using the framework
    - Domains/Dimensions
  - Lessons Learned / Sustainability

Overview of the Data Quality Case Studies

Darius

NCSH Linked Mortality File

Overview: This case study uses the Framework for Data Quality to   assess the quality of National Center for Health Statistics' (NCHS) Linked Mortality Files (LMFs), which blends survey and mortality data.

Since **blended** data can increase the **disclosure** risk of a dataset, this case study describes the procedures and methods used by NCHS to systematically address the dimension of confidentiality ,in addition to other dimensions of the framework.

CPI – Crowdsourcing Gasoline Prices

Overview: This case study describes the data quality assessment of a new method for collecting Consumer Price Index (CPI) gasoline price data from retailers or data aggregators instead of a sample survey at BLS. While crowdsourcing data directly from retailers leads to efficiencies in both collection efforts and costs, the method also introduces the potential for increased errors in collection. This case study highlights how the BLS mitigates threats to the accuracy and reliability of these crowdsourced data and is using the framework to guide the expansion of alternative data into CPI estimation.

Overview of the Data Quality Case Studies

Darius

Motor Carrier Inspection Data

Overview: This case study evaluates data collected from the U.S. Department of Transportation (DOT)'s new roadside inspection tool, SafeSpect, highlighting the utility domain in the DQ framework and, more specifically, the relevance and timeliness dimensions. With respect to relevance, the tool continues to provide highly relevant data that support the Federal Motor Carrier Safety Administration's (FMCSA) mission. The case study describes in some detail how the new tool improves the timeliness of making the data available for review.

Physical Activity Monitoring from NHANES

Overview: This case study assesses the **utility** of physical activity monitor (PAM) data collected in the National Health and Nutrition Examination Survey (NHANES), which is conducted by the National Center for Health Statistics (NCHS) within the Centers for Disease Control and Prevention (CDC). This study focuses specifically on evaluating the accessibility of large data files that require subject matter expertise

to analyze.

Overview of the Data Quality Case Studies

Darius

Program Evaluation Data

Overview: Federal agencies interested in assessing the effectiveness of government programs and services often face limited resources. Thus, agencies often rely on administrative data that are already collected by the government for a different purpose and use the data to aggregate program outcomes and estimate policy impacts. Evaluators have become entrepreneurial in identifying data sets that will support answering research questions of interest through rigorous program evaluation, which may necessarily involve matching across multiple sources or require some supplemental data collection. It is imperative that researchers are transparent about the fitness for use of existing data for program evaluation, to strengthen credibility of study design and findings, and to account for other factors including data privacy. This case study examines how the FCSM Framework for Data Quality serves as a tool to determine the utility of administrative data in determining program effectiveness and provides considerations for the data's use that are paramount to upholding rigor and ethics as principles of program evaluation.

Conceptualizing a New Study

Overview: In this case study, the FCSM Framework for Data Quality was used to anticipate potential data quality threats during the planning stages for a new study that is designed to measure the nation's perceptions of the science and engineering (S&E) enterprise.

Panel Discussion Questions

1) **Data Quality Overall**
   1) What are the common challenges organizations face when it comes to maintaining high-quality data, and how can they address these challenges effectively?
   2) If you're a data producer, how does your organization currently report data quality to its users? If you are a data user, how would you like to see data quality assessed or otherwise reported?
   3) Can you share examples of how poor data quality has negatively impacted organizations or projects?
   4) How do you prioritize data quality initiatives in your organization, and how does it align with overall data strategy?
   5) Can you share success stories where improving data quality has led to tangible business benefits, such as increased revenue, cost savings, or enhanced customer satisfaction?

2) **Communication/Assessing/Reporting Data Quality**
   1) What technologies and tools can aid in maintaining and improving data quality, and how should organizations select the right ones?
      1) How does this tool differ from the FCSM data quality framework?
   2) What are the pros/cons of providing a numeric data quality score to a data product?
   3) How do you build a culture of data quality within an organization, and what role do employees play in this process?

3) **FCSM Framework Questions**
   1) How does the FCSM Data Quality Framework align with data governance and quality assurance practices within your organization?
   2) What strategies can organizations use to maintain and monitor data quality over time, ensuring that it remains consistent and reliable? Could the FCSM framework be a part of this strategy?
   3) What is the role of collaboration and coordination among federal agencies in ensuring consistent application of the FCSM Data Quality Framework across various data sources and programs?