

**First FCSM/WSS Workshop on Transparent Reporting on the Quality of Integrated Data:  
“Reporting on the Quality of Data Used as Inputs to Statistical Products”**

**Bureau of Labor Statistics Conference Center**

**December 1, 2017**

**Opening - 9:30-10:20**

9:30-9:45 – John Eltinge, Census Bureau – “Introduction to FCSM/WSS Workshops on the Quality of Integrated Data

9:45-10:00 – Chris Chapman, ED, National Center for Education Statistics – “Introduction of Speakers and Overview of the Input Workshop”

10:00-10:20 – Robert Groves, Provost for Georgetown University – “Advancing the Paradigm of Combining Data Sources – Considerations from the National Academy of Sciences”

**Break - 10:20-10:30**

**Session I - 10:30-12:10**

**Reporting on quality of administrative data.** Administrative data provide a potential source of data for mainline production of official statistics. They are often conceptually straightforward in principle, yet there remain important quality issues associated with administrative data such as conceptual ambiguity for data entry personnel, definition of the population for which the data apply, etc. What should an agency report regarding the quality of administrative data? How do these metrics relate to long standing survey-based quality metrics?

10:30-10:55 – Stephen B Cohen, RTI – “The Integration of Administrative and Survey Data in Support of Medical Expenditure Analyses: Utility and Challenges”

10:55-11:20 – Michael Berning and David Sheppard, U.S. Bureau of the Census. “Quality of Administrative Records as Source Data”

11:20-11:45 – Bonnie Murphy and Crystal Konny, Bureau of Labor Statistics – “Quality Considerations for Administrative Data Used for the Producer Price Index (PPI) and Consumer Price Index (CPI) Development”

11:45-12:10 – Mary Muth, RTI – “Assessment of Commercial Store and Household Scanner Data: Methods, Content, and Cautions”

**12:10-1:15 Lunch (on your own)**

**Session II - 1:15-2:55**

**Reporting on the quality of less structured data.** Less structured data are already used for a range of official purposes such as projecting weather conditions and monitoring transportation volume and safety. The data are collected through a wide range of processes, but for official statistics, they need to be processed in ways that statistical procedures can be applied to them. What should an agency report on metrics (or describe in the absence of a numerical measure) on the quality of less structured data? Are there important considerations for different kinds of data? How do these metrics relate to long standing survey-based quality metrics?

1:15-1:40 – Dr. Peter Elkin, SUNY Buffalo – “The Improvement in Sensitivity and Often Specificity when Adding Unstructured to Structured Data”

1:40-2:05 – David Johnson, USDA, National Agricultural Statistical Service – “Satellite Remote Sensing Imagery Quality and Timeliness: Considerations for Use in Regional Estimation of Crop Production”

2:05-2:30 – Subrat Mahapatra, Maryland Department of Transportation – “Sensing Data Quality in Sensor-Based Data”

2:30-2:55 – Roberto Rigobon, National Bureau of Economic Research – “Web-scraped Data, Consideration of Quality Issues for Federal Statistics

**2:55-3:15 Break**

**Session III - 3:15-4:15**

**Structured discussion**

Speakers and working group will engage with audience regarding integration of data collected specifically for statistical purposes (statistical or structured data) with administrative data (also often highly structured but for different applications than statistics), semi-structured data such as data from web-scraping processes, and unstructured data such as photos or satellite images for the production of official statistics. The discussions will largely consider extensions of existing statistical standards to include transparent reporting on the quality of data not originally collected for statistical purposes.

**Closing 4:15-4:25**