

Summary of Input Data Workshop

Chris Chapman
National Center for Education Statistics
January 25, 2017

- Organizing principle of workshop 1 was to identify data quality standards and issues for non-sample survey data
 - Statistical agencies have a wealth of guidance for quality metrics to report on survey data
 - Less guidance is available on quality metrics and reporting practices for administrative data, and less- and unstructured data
- A lot of terminology and guidance for the workshop was drawn from the CNSTAT report
 - “Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy”

- Presentations about administrative record data covered a range of subject domains
 - Health data
 - Price index data
 - Food purchase data
 - Overall integration processes
- Key factors to consider when evaluating and reporting on input data quality
 - Original purpose of collections
 - Level of documentation available
 - Issues with private data providers needing to protect data processes, and not being consistent over time

- Less structured and unstructured data discussions also covered a range of subjects
 - Health diagnoses
 - Agricultural production
 - Traffic
 - Pricing
- Common and new themes were discussed during the second session
 - Common was clarity about the purpose of the original collection
 - Technology used to collect data and limitations of the technology
 - How technology changes affect data comparability over time
 - Speed at which data are available

- Next steps
 - Complete summary report from workshop 1
 - Complete review of possible extant standards for administrative record data quality
 - Review applicability of existing broader data quality assessment frameworks
 - Present information on progress at the FCSM conference in early March

Thank you