



NCSES

National Center for Science and Engineering Statistics

Role of Contractors in NCSES Data Production, Archiving and Documentation

May Aydin

Panel on Transparency and Reproducibility for NCSES

May 21, 2019

National Center for Science and Engineering Statistics
Directorate for Social, Behavioral and Economic Sciences
National Science Foundation

www.ncses.nsf.gov

NCSES data systems and contractors

- NCSES obtains data files and documentation from data collection contractors/partners for inclusion in NCSES's central data repository.
 - Maintain 13 of 16 survey micro datasets in the central repository
 - Current total of 6 contractors and partners
 - Deliveries include micro-data, some metadata, and other documentation
 - Complex range of data from both demographic and establishment surveys with a variety of units of measures and variables.

Challenges

- Statements of work vary. Non-standard deliveries in content, format and scope
- Variability in data transfer methods
- Missing metadata, table specifications, change documentation
- Considerable resources within NCSES to identify differences, track-down info, transform data/metadata for usability and dissemination

Recommendations from Panel on Communicating S&E Data in the Information Age

Rec 3-1:

NCSES should incorporate provisions in the contracts with data providers for the receipt of versioned microdata, at the level of detail originally collected, in open machine-actionable formats.



Rec 3-3:

NCSES should require that data received from contractors be accompanied by machine-actionable metadata so as to allow for automated production of NCSES publications, comparability with previous analysis, and efficient access for third-party visualization, integration and analysis tools.

Process-driven metadata

**Data
Collector**

Prep

Collect

Deliver

NCSES

Receive

Transform

Disseminate

Archive

Data User

Access

Analyze

Success



Questions for the Panel

How and to what extent do we convey the role that data collection contractors and partners have for NCSES (and other federal statistical agencies) in achieving transparency and reproducibility?

Where is the best place to make efficiencies in the process?
(Not just the what, by the where and how)