

'Tis better to give than receive: considerations when sharing data

Melissa R. Pfeiffer, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

- Receiving data in other formats: trust but verify
 - Ask for data dictionaries/file layouts for field names, type, length, meaning, maybe even valid values/ranges
 - beware of incorrect or outdated information – make sure that data matches supplied documentation (example: file layout I received that indicated a variable was a date when it was actually \$1)
 - Importance of understanding what coded values mean (example: codes for race or ethnicity)
 - are there changes in coding, file structure over time? (examples: variables no longer available; changes in value meaning [marital status])
 - Ask for number of records in each file: aids in discovering truncated imports
 - Common import problems
 - embedded CRLF: solution. Not seen in xml files.
 - unknown lengths: checking for truncated values because the length was not set long enough.
 - Structurally sound
 - Relational tables make sense (example: no situations in which there is a driver-level record without a crash-level record)
 - All data expected are present (example: appear to be missing data for people born in a certain year, proportions of certain variable values make sense overall and when stratified temporally [e.g., have expectation of race/ethnicity proportion - any changes to this over time may indicate a change in coding], people who do not have transactions are permit only)
- Giving data to others
 - Ways to create different files (Excel, CSV) depending on whether
 - column headings are the variable name or label
 - values are unformatted or formatted
 - Example of multiple sheets in an Excel file
 - Supply a data dictionary – even a simple one.
 - Selecting just some output to create (ods select or creating data set with limited variables)
 - Consider adding key indicators, such as the % missing
 - Create a document with format information (using proc format cntlout)
 - Output selected formats