Data processing flow

The diagram below depicts how raw data was processed for the production of the Measuring Broadband America 2014 report.

Raw data

Raw data for the chosen period is collected from the measurement database. The ISPs and products that panellists were on are exported to a ‘unit profile’ file, and those that changed during the period are flagged.

Data is cleaned. This includes removing measurements when a user changed ISP/product during the period. Anomalies and significant outliers are also removed at this point. See the data cleansing document (<http://data.fcc.gov/download/measuring-broadband-america/2016/validated-data-cleansing-sept2015.docx>) for more information.

Clean data

Per-unit results are generated for each metric. It is at this point that the 7-11pm averages are computed and the trimmed mean calculated for each metric. The SQL scripts used here are contained in [http://data.fcc.gov/download/measuring-broadband-america/2016/sql- processing-scripts-sept2015.sql.gz](http://data.fcc.gov/download/measuring-broadband-america/2016/sql-%20processing-scripts-sept2015.sql.gz)

SPSS processing

Per-unit results (CSV)

The per-unit CSV data is processed by SPSS scripts (available at: [http://data.fcc.gov/download/measuring-broadband-america/2016/spss- processing-scripts-sept2015.zip](http://data.fcc.gov/download/measuring-broadband-america/2016/spss-%20processing-scripts-sept2015.zip)), coupled with the unit profile data. This process removes ISPs/products with low sample sizes and computes statistical averages for the remainder that can be used in the report.

Summary data tables and charts in Excel are produced from the statistical averages. These are used directly in the report.

Excel tables / charts