Traffic Data Quality Measurement
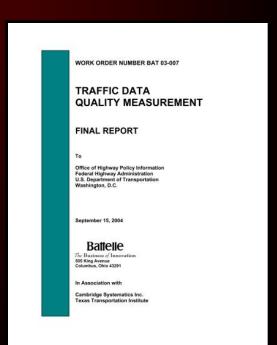
Shawn Turner
Texas Transportation Institute

USDOT Data Quality Webinar

November 2, 2009

Texas Transportation Institute

2004 FHWA Report

- "Traffic Data Quality Measurement: Final Report"
  - EDL #14058, September 2004
  - Data quality measures
  - Calculation procedures
  - Data quality target values
  - Monitoring costs

WORK ORDER NUMBER BAT 03-007

TRAFFIC DATA QUALITY MEASUREMENT

FINAL REPORT

To

Office of Highway Policy Information
Federal Highway Administration
U.S. Department of Transportation
Washington, D.C.

September 15, 2004

Battelle
The Business of Innovation
505 King Avenue
Columbus, Ohio 43201

In Association with

Cambridge Systematics Inc.
Texas Transportation Institute

Texas
Transportation
Institute

Traffic Data Quality Measures

1. Accuracy
2. Availability
3. Timeliness
4. Coverage
5. Validity
6. Accessibility

Section 1201

Accuracy

- Degree of agreement between data values and a source assumed to be correct

- Most difficult to measure

- Requires a benchmark/reference

Texas
Transportation
Institute

Issues with Accuracy

- No consensus on how to establish benchmark/reference
  - Single floating car/GPS trace
  - Reidentification (Bluetooth, license plate, etc.)
  - Fixed-point sensors?
- No consensus on measures
  - mean absolute percent error
  - root mean squared error
  - mean absolute error (I-95 Corridor Coalition)
  - seconds per mile

Texas
Transportation
Institute

Developing Consensus for Accuracy

- ITS American Working Group – NATWG

- VDOT Pooled Fund Study

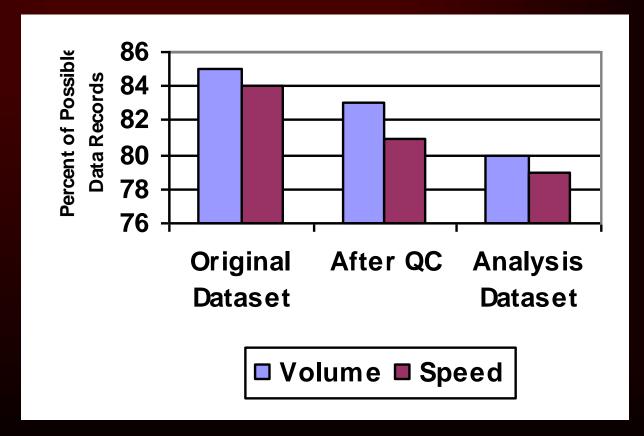**Texas Transportation Institute**

Availability

- Degree to which data values are present in the attributes that require them
  - % available
  - % available for measured vs. estimated data
  - % available for real-time vs. historical
- Simple accounting, BUT
  - Where in the data stream do you count?

Texas Transportation Institute

Availability at Different Steps

- Real-time ➔ Intermediate ➔ Archive

Texas Transportation Institute

Timeliness

- Degree to which data values are provided at the time required or specified
  - % timely data
  - Average delay for late data
  - Update frequency (Section 1201)
- Similar to latency – if data are latent or not being updated, then it shows up in the accuracy results

Texas
Transportation
Institute

Coverage

- Degree to which data values in a sample accurately represent the whole of that which is to be measured.
  - % coverage with description of universe and sampling
- Section 1201 NPRM
  - 100% of Interstate highways in 2 years
  - XX% of non-Interstate metro highways in 4 years

Texas
Transportation
Institute

Details, Details

- How often to assess accuracy and availability?
- How much of data is within accuracy limits?
- What are time/space parameters for availability?
  - Each link at 90%? System at 90%?
  - Every day at 90%? Sample of days?
- Where in data stream to measure availability?
  - Real-time feed or archive?

For More Information

- "Traffic Data Quality Measurement: Final Report"
  - http://www.its.dot.gov/library.htm
  - Search EDL #14058
  - September 2004