Data Quality and Why We Should Care

2103 FHWA HIS – NHI Arlington, VA

Tianjia Tang, PE. Ph.D.
Chief, Travel Monitoring and Surveys Division
Office of Highway Policy Information
Federal Highway Administration

Outline

1. Data Quality Dimensions
2. Data Accuracy Dimensions
3. Quality Data Assurance Framework

Data Quality

What does quality data mean to you?

Data Quality Dimensions

1. Relevance
2. Believability
3. Accuracy
4. Consistency/Coherence
5. Timeliness
6. Accessibility
7. Interpretability
8. Traceability

Data Quality Dimensions

1 - Relevance

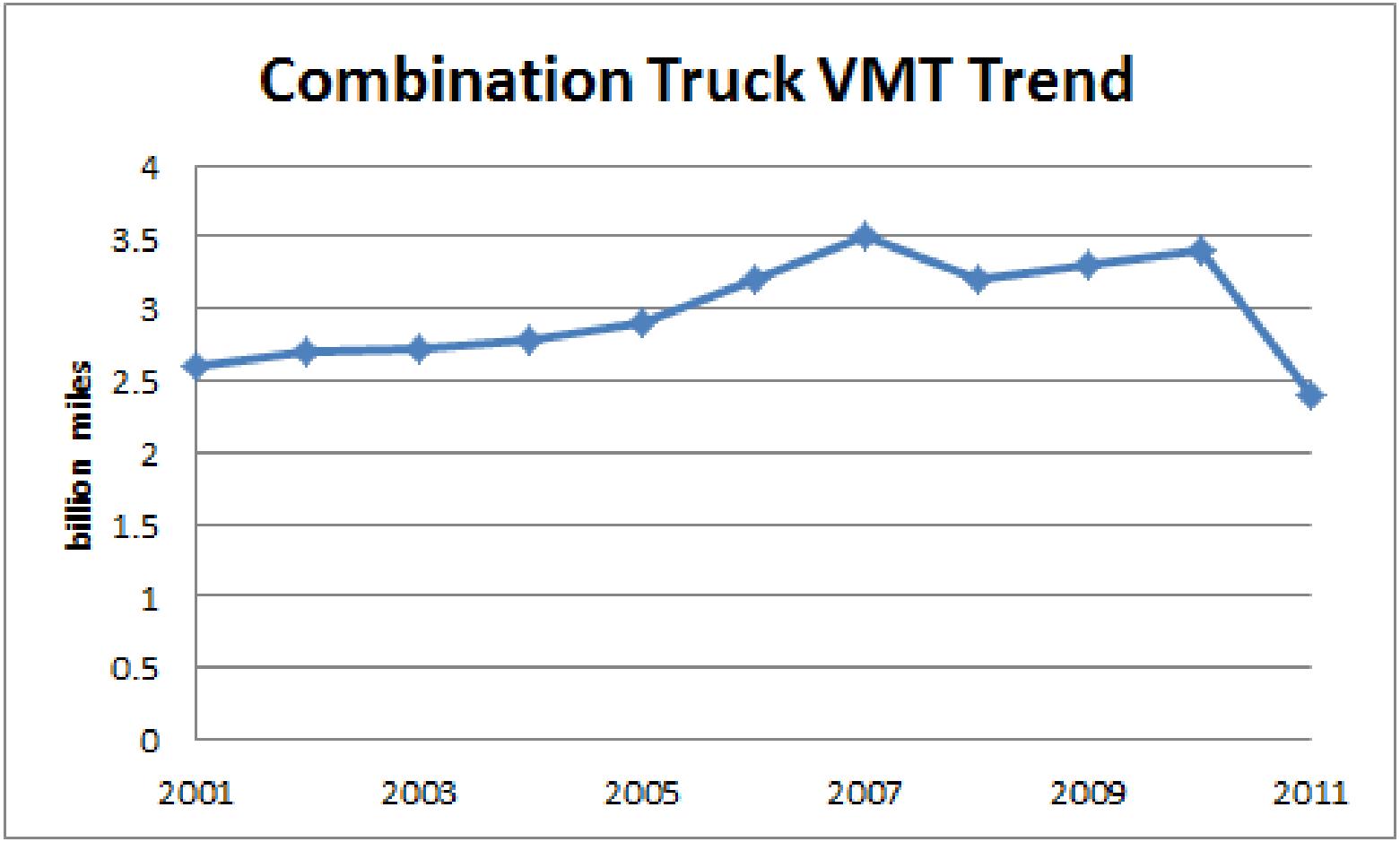
The degree to which the product meets the specific needs in terms of both content and coverage.

Data Quality Dimensions
2 - Believability

The capability of eliciting belief or trust of the program from both professional and the public covering nor only the data itself but also programs the data is tied to.

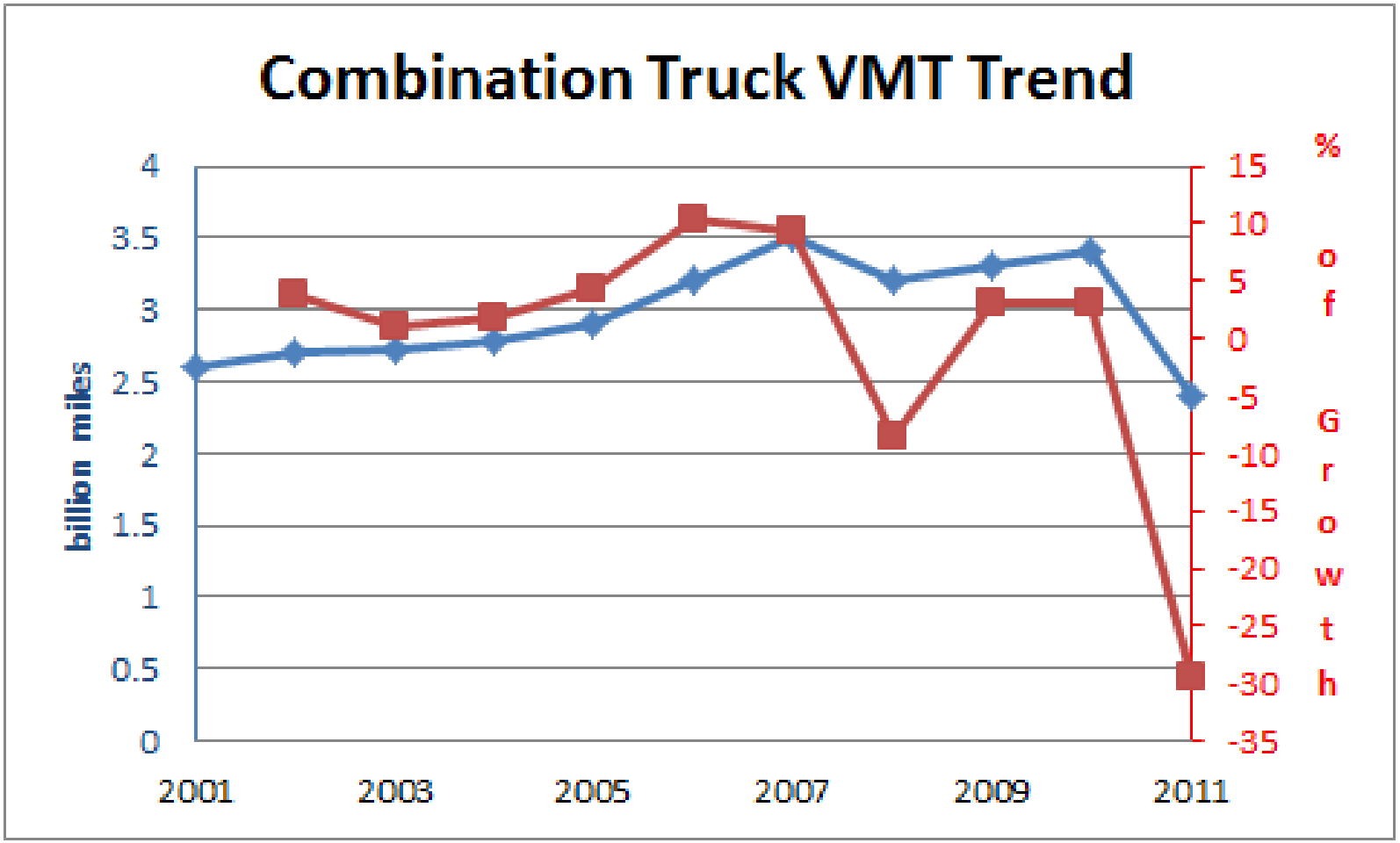
Data Quality Dimensions

Believability ?



Data Quality Dimensions

Believability ?



Data Quality Dimensions

Believability ?

Urban Freeway Daily Directional Factor: 0.65

Data Quality Dimensions

3 - Accuracy

The degree to which an estimate is the same as the true value.

Keep in mind that often than not, the truth is never known.

Data Quality Dimensions

Accuracy ?

Urban Freeway AM Peak Directional Factor: 0.5

Data Quality Dimensions

4 - Consistency/Coherence

The degree to which different sources or methods on the same phenomenon or subject are similar.

Often, a subject is characterized by several different parameters.

Data Quality Dimensions

Consistency/Coherence ?

1. **2012 Goods movement measured in tonnage increased 8% from last year**
2. **Manufacturing GDP grew 12% from last year**
3. **Tractor Trailer VMT decreased 9% from last year**

Data Quality Dimensions

Consistency/Coherence ?

1. **2012 Goods movement measured in tonnage increased 8% from last year**
2. **Manufacturing GDP grew 12% from last year**
3. **Tractor Trailer VMT decreased 9% from last year**
4. **Truck WIM data shows GVW stayed the same as last year.**

Data Quality Dimensions

5 - timeliness

The time between event or phenomena occurred and data are made available.

Decisions will be made regardless the data is available or not.

Data Quality Dimensions
timeliness

Do we believe our leaders will wait or you will wait until they or you have all the data in hand before making any decision?

Data Quality Dimensions

6 - Accessibility

The flexibility and easiness of information, data and documentation characterizing the data, procedures, processes and program from the responsible organization.

Data Quality Dimensions
Accessibility?

How can people get access to our data?

Web,

Email,

Phone,

FOIA,

Sunshine action

…

Data Quality Dimensions

7 - Interpretability

Availability of adequate information and support to allow users to properly use and interpret the data products.

To what degree all information and documentation will be released.

Data Quality Dimensions

Interpretability

While it is legitimate to concern data are misused, what have we have done to prevent such misuse or interpretation?

Have we provided adequate background information out of there for people to use?

Do we know why we collect the data and information we collect?

Data Quality Dimensions

8 - Traceability

The ability to verify the history, location, timing, sequencing, or application of an item by means of documented recorded identification

Data Quality Dimensions

Traceability

John, the genius left us. We do not know how the genius did it.

I think that I show we did it.

I do not remember exactly how we did it.

Summary - Data Quality Dimension

1. Relevance
2. Believability
3. Accuracy
4. Consistency/Coherence
5. Timeliness
6. Accessibility
7. Interpretability
8. Traceability

Data Accuracy Dimension

1. Coverage error
2. Sampling error
3. Non-response error
4. Measurement error
5. Processing error
6. Post-event adjustment error

Data Accuracy Dimension

Coverage Error

Error resulted from a mismatch between sampling frame and population.

e.g. sampling frame – “interstate highways”

population – “all public roads”

“Interstate highways” mismatches “all public roads”

Data Accuracy Dimension

Sampling error

Error associated with drawing samples as compared with census where the entire population are selected

Data Accuracy Dimension

Non-response error

Unit nonresponse error: a sampling unit does not respond to any part of the survey,

Item nonresponse error: a sample unit only partially completes the survey.

Data Accuracy Dimensions

Measurement error - Difference in the response value from the true value of the measurement.

Data Accuracy Dimensions

Processing error

Deviation between the value used in publication and the response value provided.

Data Accuracy Dimensions

Post-event adjustment error

The degree to which estimates are affected by adjustment procedures such as imputation.

Summary of Data Accuracy Dimension

1. Coverage error
2. Sampling error
3. Non-response error
4. Measurement error
5. Processing error
6. Post-event adjustment error

Quality Data Assurance Framework

1. Legal and Institutional Environment
2. Resources
3. Professionalism
4. Transparency
5. Ethical Standards
6. Methodology Soundness
7. Technology
8. Expertise
9. External Involvement

Quality Data Assurance Assessment Framework

1: Legal and Institutional Environment

The responsibility in collecting, processing, and disseminating of a given data program is clearly specified through various authority(policy, regulation, and legislation; the authority to collect data are explicitly provided, preferable with enforcement power; and the detailed “How” manual is maintained.

Quality Data Assurance Assessment Framework

2: Resources

Resources are commeasure with needs of the data program covering staffing, equipments. Mechanisms to ensure adequate resources are in place.

Quality Data Assurance Assessment Framework

3: Professionalism

Procedures and processes are guided by professional principles with no pre-perceived opinions and on a impartial basis.

Quality Data Assurance Assessment Framework

4: Transparency

Ensure all procedures and processes are transparent; and ensure what carried out is according to documented procedure or processes.

Quality Data Assurance Assessment Framework

5: Ethical Standards

Guidelines for staff behavior are established and well known to both staff and external entities

Quality Data Assurance Assessment Framework

6: Methodology Soundness

Adopt both industry and professional acceptable methods and approaches in methodology and ensure a policy to explore new methodology method is in place

Quality Data Assurance Assessment Framework

7: Technology and Expertise

Ensure technology is commensurate with methods adopted; and technical training in technology is in place.

Quality Data Assurance Assessment Framework

8: Expertise

Knowledge and skills people have possessed and the needed technical training is in place.

Quality Data Assurance Assessment Framework

8: External Involvement

Establish an external independent task force/committee to periodically review and recommend changes and improvements

Summary
Quality Data Assurance Framework

1. Legal and Institutional Environment
2. Resources
3. Professionalism
4. Transparency
5. Ethical Standards
6. Methodology Soundness
7. Technology
8. Expertise
9. External Involvement

Summary

1. Data Quality Dimensions – 8
2. Data Accuracy Dimensions – 6
3. Quality Data Assurance Framework - 9