Big Data Challenges in Railway Engineering

Nii Attoh-Okine
Professor
Department of Civil and Environmental Engineering
University of Delaware, Newark, DE, USA
Email: okine@udel.edu

IEEE Workshop on Large Data Analytics in Transportation Engineering
October 27, 2014
Railway Engineering

- Productivity
- Reliability
- Velocity
- Safety
Data Collection

- GPS
- AEI readers
- Electronic Data Exchange
- Video
- Hand Held Devices
5 V’s of Big Data

**Figure 1:** Five V’s of Big Data

**Variety**
- Time Series
- Spatial
- Images

**Velocity**
- Real Time
- Near Real Time
- Periodic
- Stationary
- Non-stationary

**Volume**
- Megabyte
- Gigabyte
- Petabyte
- Exabyte

**Veracity**
- Incomplete Data
- Subjective | Objective
- Inconsistent

**Value**
- $
5 V’s of Big Data

**Variety**

Structured and Semi-structured data.
- Sensor data
- Audio
- Video
- Combination of such data for efficient analysis.

**Velocity**

Increasing speed at which data is created, and the speed at which it can be processed, stored, and analyzed.

**E.g:** Trade data collection, Pavement data collection, Traffic streams.
## 5 V’s of Big Data

### Volume
It is defined as the amount of data generated in a period of time (e.g. Gigabytes, Terabytes) from different sources, locations to be stored and processed.

### Value
Mainly refers to the worth of the information being managed.

### Veracity
Refers to the trustworthiness of the data. It is an indication of the reliability and accuracy of information for making decisions.
Table 1: Types of Big Data

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro-data</td>
<td>The least aggregated level of data in very large data sets.</td>
</tr>
<tr>
<td>Meso-data</td>
<td>The mid-level of data aggregation in very large data sets, resulting from the collection of data that are a level-up from micro-data in terms of the kinds of information that is captured.</td>
</tr>
<tr>
<td>Macro-data</td>
<td>The most aggregated level of data in large data sets describes regional or geographic area.</td>
</tr>
</tbody>
</table>
Figure 2: Stages of Big Data Analysis
Figure 3: Example of Railway Big Data (Track Data)
Challenges

- Heterogeneity
- Inconsistency and Incompleteness
- Merging Data
- Time lines
- Privacy and Data Ownership
- Visualization