

Enterprise Data White Paper

What is Enterprise Data?

A data revolution is upon us as we now create approximately 2.5 exabytes - that's 2.5 billion gigabytes (GB) - of data every dayⁱ. So much data are now generated that as much as 90% of all of the data in the world today has been created in the last two years alone. This surge of new data has been accompanied by an increase in connectivity among the vehicles, sensors, people, and infrastructure in the transportation network.

Already we are seeing applications that share, use, and leverage datasets to improve current transportation operations or capabilities, such as Waze and Google Maps. USDOT's Enterprise Data program aims to create value from the data collected from intelligent transportation systems (ITS)-enabled technologies, including connected vehicles (automobiles, transit, and commercial vehicles), mobile devices, and infrastructure to make our transportation system safer, more accessible, efficient, and environmentally sustainable, while also protecting the privacy of its users.

Benefits of Enterprise Data

Enterprise data have several potential benefits including:

- **Providing new revenue opportunities** – New data sources open the door for innovators to develop applications and methods that can support economic vitality.
- **Monitoring performance and enabling more efficient responses** – Increased data from new sources provide a complete, more detailed view of the whole transportation system allowing decision makers to make informed choices on how best to increase system efficiency.
- **Increasing efficiency of information sharing** – Enterprise data will reduce the costs of data management and eliminate technical and institutional barriers to the capture, management, and sharing of data.
- **Improving the accuracy and timeliness of data** – More refined data collection methods will lead to higher quality data, which in turn will support faster data distribution.
- **Stimulating innovation of new research** – The increase in data will spark novel development of software and tools to use the data in new innovative ways.

Private Sector Uses of Enterprise Data

Various forms of enterprise data are used by the private sector in almost every aspect of our lives. From developing customer needs and relationships to better understanding our own health habits, enterprise data is playing a bigger and bigger role in our daily activities.

Here are just a few examples of enterprise data with potential for transportation applications:

- **Customer understanding** - Companies are now using customer data to better understand customers and their behaviors and preferences. Rich data sets from analytics and sensor data provide more insights into customer's habits, enabling service improvements tied to customer needs.

- **Real-time Traffic comprehension** - Traveler information providers analyze real-time data streams to intelligently identify current conditions, estimate how long it would take to travel from point to point within the city, and offer advice on travel alternatives, improving traffic in a metropolitan area.
- **Optimizing the supply chain** - The scale, scope and depth of the data generated by supply chains is providing abundant data sets to drive contextual intelligence about location and logistics. These data sets are revolutionizing how supplier networks form and grow into new markets and mature over time.
- **Personal health understanding** - Enterprise data is used by individuals for personal health monitoring. With the rise of wearable health devices individuals, now have personal data on their exercise and travel patterns which allow them to be more efficient in the ways they manage their own health.

Tools of Enterprise Data

To keep up with growing sources of data, new methods and tools to collect, transmit, sort, store, share, aggregate, fuse, and analyze and apply these new data sources will be needed for management and operations of future transportation systems.

New tools are emerging in the area of storage and retrieval of the data. Traditional database systems were not designed to handle the extensive amount of data that are now being produced, and their retrieval times are often too long to be useful in many situations. To solve this problem, data are now being stored in environments that support distributed storage and processing. Distributed computing is a process in which components of a software system are shared among multiple computers, typically in a cloud computing environment, to improve efficiency and performance of data access. By distributing the storage and processing of data in these environments, much larger data sets can be processed as more and more computing power is added.

To better comprehend these larger data sets, new data visualization techniques have been developed to evaluate the data so that whole patterns are discerned. Data visualization is the presentation of data in a pictorial or graphical format in order to gain additional insight. Data visualization techniques range from the very simple bar and pie charts to the more complex heat maps, word clouds, and interactive dashboards.

Enterprise Data Challenges

The increase in the amount of data also creates challenges when it comes to security, standardization and harmonization, and managing of high-quality data. Security is a top concern when looking to use new data sources that have been produced from the various transportation sources. Privacy and personally identifiable information (PII) in the form of names, vehicle information, financial information, login identification, and general location information must all be safeguarded to protect user privacy.

Among the challenges for broad adoption of enterprise data is ensuring the use of high-quality electronic data that can be effectively exchanged with other systems. This can only be done through the use of data standardization and harmonization across the sources. Standardization certifies that data from the same source will be in the same format and use the same units of measurement. On the other

hand, harmonization ensures data gathered in one location is also measured the same way in another location and promotes a single standard for the information.

Finally, for enterprise data to be successful, the data must be transmittable across multiple institutions, including the general public. This creates a challenge of how to provide a high quality data source that is easily accessible for end users. To address this challenge, data must: comply with ITS standards when applicable; be provided in non-proprietary formats; be disaggregated to the largest extent possible; be well-documented, including meta-data and context on why and how the data was collected; and be presented in a way that is easy to view, filter, and access.

Government Role in Enterprise Data

To promote the use of enterprise data, the United States Department of Transportation (USDOT) ITS Joint Program Office (JPO) has made enterprise data one of five strategic categories in the ITS strategic planⁱⁱ. This category focuses on USDOT efforts to increase operational data capture from stationary sensors, mobile devices and connected vehicles and expands into research activities involving the development of open data sources and mechanisms for large data analysis.

Smart City Challenge

The USDOT is promoting the use of enterprise data through its commitment to smart connected cities. A connected city leverages information and communications technology and urban analytics to create value from the data that is collected from connected vehicles, connected citizens, and sensors throughout a city or available from the Internet using information generated by private companies. Analytics that utilize data from across various systems in a city have tremendous potential to identify new insights and unique solutions for delivering services, thereby improving outcomes. In a connected city, all critical systems, including transportation, energy, public services, public safety smart payment, health and human services and telecommunications communicate with each other to coordinate and improve overall efficiency.

In December 2015, USDOT launched [The Smart City Challenge](#) to create a fully integrated, first-of-its-kind city that uses data, technology and creativity to shape how people and goods move in the future. The winning city will be awarded up to \$40 million from the U.S. DOT and connected vehicle technology is expected to be large part of this pilot

ⁱ <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>

ⁱⁱ <http://www.its.dot.gov/strategicplan/>