The Benefits of Consolidating Data Silos

A Bentley White Paper

Robert Henderson
Transportation Solution Specialist
Asset Performance

Published:
June 2019

Bentley
Advancing Infrastructure

www.bentley.com

The Data Challenge

Rail and transit organizations are great at collecting various forms of data but are often not effective at analyzing these massive volumes of data. The rail and transit industry is not alone. On average, companies are effectively analyzing only 1% of their data. The other 99% of the data is called dark data, according to Gartner. Dark data is data that is collected and stored but not used to deliver any insight.

With the rail and transit industry going digital, dark data has become a significant problem for track maintenance. The data being collected comes from various hardware suppliers that provide their own inefficient software solution to analyze the data. This creates data silos within the organization. AssetWise Operational Analytics Linear Decision Support is a hardware-neutral solution that consolidates third-party data into a single system for further linear analytics for track maintenance.

Success requires organizations to quickly assess every dimension of the railway infrastructure and its condition over time while evaluating and learning from past maintenance activities. Data silos throughout the organization present a challenge for owner-operators making decisions about track and other maintenance-of-way assets. Determining what work needs to be done, where the work should be focused, and when the work should be performed requires an open, connected data environment and a single repository for the data. The requirement to reference and analyze more than 200 types of railway asset data simultaneously, including track layouts, mileposts, curves, planned work, work history, rail defects, gross tonnage, maintenance zones, track category, and predicted future asset conditions, could be successfully implemented with the consolidation of all these data silos.

Big Data for Track Maintenance

Data collection systems, such as track geometry vehicles, ground penetrating radar, laser scanners, video recorders, track walkers, and more, can produce many gigabytes of data in a day. Organizations need to efficiently store, provide links to, and correlate this data for rapid access and analysis, enabling them to make timely and accurate maintenance decisions. An informed decision for track maintenance requires the support of managing multiple datasets containing hundreds of thousands of surveys and can be terabytes in size.

One can argue with a purist on what "big data" is or what it represents. However, using the "six V's" against track maintenance objectives is a much simpler way of making big data relatable in rail and transit.

1. **Volume**: With big data comes volume. The substantial amount of data can make it difficult to drill into, to analyze, and to find the root cause of a problem. It is common for rail owners and operators to collect billions of geometry samples. Additionally, owner-operators can obtain millions of new samples per day.

2. **Velocity**: At what speed and how often is the data coming? If the track reliability team cannot keep up with the rate at which new data from inspections and sensors arrives, then the data has no value. The reliability team will fall behind in processing and determining issues.
3. **Veracity**: Many datasets used in rail are collected by measurement systems that are only understood by a small number of members at an organization. Veracity is the ability to go into a dataset or inspection and pick out the specific data as well as the meaning of the information. It is common for rail owners and operators to have structured and unstructured data, subjective from visual inspections and objective data from measurements, binary files, and data of variable quality.
4. **Variety**: Rail and transit organizations usually don't have one type of data coming into their systems; data comes from multiple sources and formats. An organization could potentially have 100 different classes of data that is being brought in and collected together to help understand how the linear assets are deteriorating more clearly.
5. **Visibility**: For rail and transit organizations to gain a complete view of the asset conditions, the information from different systems need to be combined. Accessibility to all factors stored in separate data silos can affect decisions. These factors must be visibly presented in a form that allows complex conditions to be easily understood and to identify the emerging requirements quickly. This type of visibility typically requires interoperability with third-party solutions such as, but not limited to, Enterprise Asset Management systems for work management and asset registry or geographical information systems (GIS) for an asset's physical location.
6. **Value**: The dataset must have value because there is an expense associated with the data collection. Therefore, it is important to determine the value the data will bring to the organization before collection. There is a substantial cost for maintaining and running a measurement fleet, and there are limitations to a track maintenance budget.

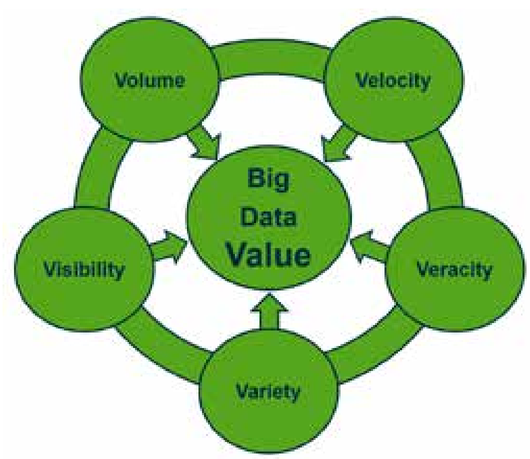

Figure 1: The six V's of Big Data

Rail and transit users currently have hundreds of terabytes of data with tens of thousands of miles of track, and they could have hundreds of repeat recordings against the entire network. This large volume of data can include ground penetrating radar, millions of daily measurements referencing the millions of assets on top of the linear network, and hundreds of users in the system.

The Data Silo Obstacle for Rail and Transit

Data comes from many sources and in many forms. The variety of this data often exceeds the understanding of a single person. Different teams use a range of isolated datasets, each team may use and understand different data types, but the system should allow for the sharing of datasets between the business units. Turning the data into a big data solution requires a rail and transit organization to address the four most substantial obstacles of the linear data silos.

1. **Datasets from Different Sources**: The track maintenance reliability team will receive data from track recording cars, either autonomous or human-operated vehicles. Data can also come from walking inspectors who identify defects. Datasets could also be images from ground penetrating radar scans or video and point cloud surveys. Furthermore, datasets can be work records that provide visibility into the performed work and the time it was completed. It is vital to understand if the maintenance was effective based on asset condition.
2. **Storing Datasets**: Isolated systems could be the most significant obstacle for a rail and transit organization. Typically, the industry works in a siloed environment and a change in methodology and culture is often required. Often datasets are segregated by product, region, business unit, or another grouping an organization considers valuable. There is usually very little collaboration between silos and this prevents coordination on the operating model.
3. **No Accessibility**: Beyond datasets being in isolation of each other, it is often the case that rail and transit organizations have no unified visibility across the isolated systems. Maintenance decisions are made based on several data streams from several silos and, in most cases, decisions are made without realizing the data was available. Issues are treated in isolation, ignoring history, affected connected assets, related information, and previous work history. A single asset may span across several silos and may be partly maintained because of different conditions, ages, or attributes, and they are not captured at one time by one system.
4. **Data is an Asset**: Data must be treated as an asset. Data maintenance across several systems brings additional obstacles because it may not be understood across data silos, or the information may be duplicated and inconsistent.

Once an organization has broken down these obstacles, it brings together linear asset data from any number of sources and correlates the data by location and time. This information can include many integration capabilities for importing files, linking to existing databases, and web services to communicate with other systems. Additionally, an organization can extend value to discrete or hierarchical asset data systems, like EAMs and other CMMS or capital planning systems, by combining linear with discrete data analysis, then feeding results back to the nonlinear systems.

Data Quality

Data has a significant impact on business decisions, including maximizing existing budgets and performing the right maintenance at the right location at the right time. Making accurate and informed decisions with poor data is impossible. Making decisions based on inadequate data will put your organization at risk, and the data quality can be grouped into two categories.

1. **Accuracy**: A nominated dataset (usually the most recent) is processed to align to the real world. This alignment can be done through several mechanisms, but the most common is to determine a curvature diagram for the route taken by the recording vehicle and align the curvature from the measurements against it. The result is a set of corrections that must be applied to align the recorded curvature with the surveyed curvature. Then, these corrections are used to create another channel of measurement data, normally a short wavelength dataset such as top or gauge, and the resulting data is saved as a location baseline. This baseline is the best alignment of data against the real world.
2. **Reproducibility**: With the creation of the baseline, all other recordings are aligned against it. This alignment is performed in the system by running comparisons along the entire length of the survey and determining how much the survey needs to be moved at each point to ensure it lines up with the baseline. It is important to note that this process is not merely a single shift of the survey because the location error between the two surveys is likely to drift over time. Experience in AssetWise Operational Analytics shows that corrections must be applied every 100 to 200 meters along the survey and that errors in location can often be +/- 100 meters.

For rail measurements, sometimes location accuracy is less important than location reproducibility. Data alignment is a critical step in data cleansing, as a track quality index can be very sensitive to location errors. When data is trended to predict track failure, it is essential to ensure that the observed trend is due to track deterioration and not location errors or instrumentation issues.

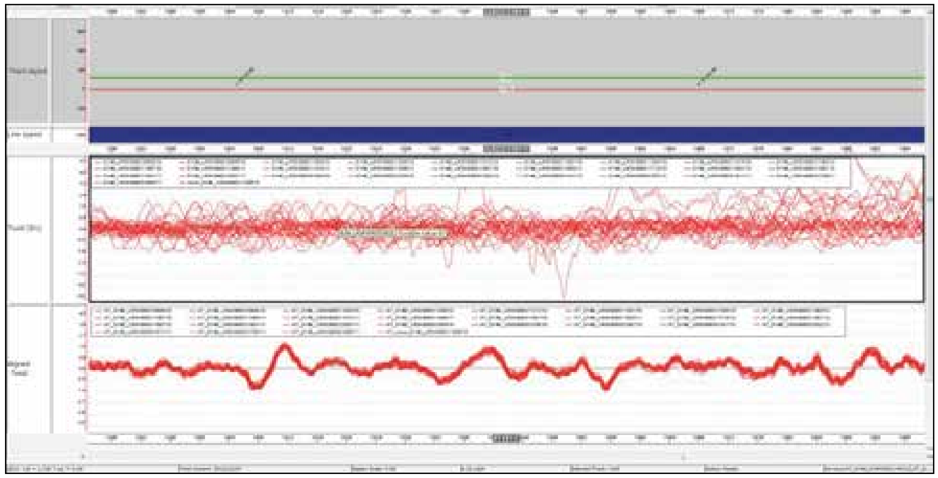

Figure 2: Data Alignment Example

Organizations spend an average of 20% of their work effort finding and validating data. The information is stored in several systems, and there is a lack of knowledge about which system has the needed information. As a result, individuals must search and verify every silo to make an informed decision. Furthermore, the search is often manual, requiring a lot of resources and wasted effort. Decisions made with incomplete or inaccurate data are expensive, with some causing track maintenance teams to rework areas of track based on mistakes when determining a specific location.

Another serious issue may result in an action creating a safety hazard. A fault indicated on the wrong track can not only create an environment where the maintenance team is working on the wrong track, but it could also result in shutting down the wrong line, while diverting traffic on the track that has the maintenance fault. This action could have severe and catastrophic consequences.

AssetWise Operational Analytics to Consolidate from Multiple Sources

The best decisions can only be made by generating a comprehensive overview of the situation. For example, a rail engineer analyzing an asset must see the current performance as well as the historical and future trends, while at the same time understanding nearby assets and the maintenance applied to the entire surrounding area. To gain a complete view of the asset conditions, the information from different systems needs to be combined and the factors stored in different datasets must be considered. Furthermore, the data must be represented in a form that allows complex conditions to be easily understood. Data visualization is critical in transforming vast quantities of complex linear data into actionable information that users can readily access, understand, and utilize. Straight line diagram visualization features provide instant access to an informative representation of any combination of configured data types at any location on the railroad network.

Consolidation of datasets is the core concept of Operational Analytics; the solution doesn't care who supplied the hardware, nor does it care what the data represents. This solution allows an organization to visualize the linear asset data from multiple sources concurrently and configure the information to tasks that are assigned to roles in an organization. This visualization configuration ensures that the right teams are viewing the information that is related to their decision-making process.

At the same time, there is an analysis engine that is configured to combine the organization's data so that the results of that analysis are visible. These could be quality indices, areas needing maintenance, predictions of deterioration, or effectiveness of past actions. The configuration of the analysis will ensure individuals can easily visualize and understand the linear data associated with their responsibilities.

The Operational Analytics engine has a library of more than 200 linear network-aware data processing rules and commands that can be applied to extract actionable information. These rules can find clusters of defects, trend track degradation to plan surfacing, filter spikes and flat lines from measurement data, and much more. Furthermore, these rules can be combined to perform sophisticated data analysis to extract more value from existing data.

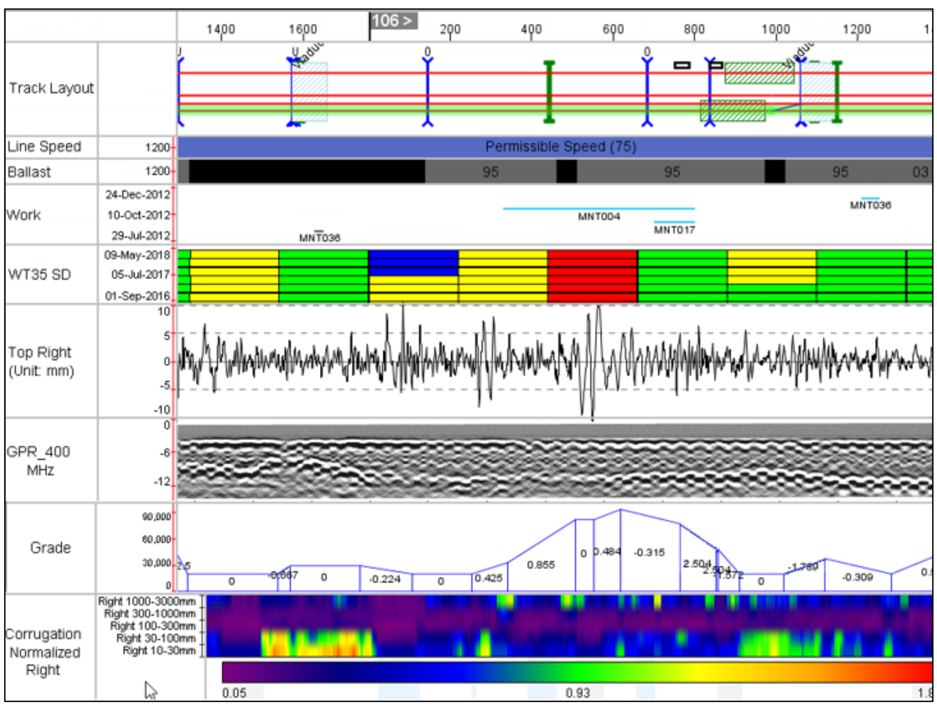

Figure 3: Holistic View Consolidating Different Data Silos

The Approach for Consolidating the Data

The concept of consolidating the data into a single source may seem simple at a high level. However, there is no international standard format for track-related linear data. Each source and vendor of hardware used to generate the data has its own independent standard form. The configuration of the dataset is important as it ensures future datasets from the same source can be automatically imported without any problems.

After data consolidation, the data must be aggregated, cleansed, normalized, coherent, current, and comprehensive. This data cleansing is completed in the Operational Analytics engine to provide a timely conversion of multiple datasets from multiple suppliers, which were previously held in siloed systems and imported into a single source filled with management intelligence.

The track maintenance teams can now analyze all factors of their linear assets within a single system that provides a single view. The single system approach will allow reliability teams to analyze different sources of data together and see historical and future trends against work management systems. Additionally, root-cause analysis can be applied by examining repeated problems, and previously attempted solutions. Once an organization has determined a fix for the root cause, it can perform a pattern analysis to discover the same issues on the entire rail network. This simple task has proven to be successful for rail and transit owners to move from a reactive maintenance culture to a proactive culture quickly.

The basic concept of operational analytics is to support decision making and applying for the right work, in the right place, at the right time. The methodology of consolidating data silos for track maintenance has five steps.

1. **Collection**: This is done through inspections, track recording surveys, maintenance records, manual walking inspections, GIS, ground penetrating radar, and ultrasonic imagery. Various business units or teams within a rail and transit organization will typically collect datasets that are of value to their business objectives.
2. **Datasets**: These datasets were previously data silos due to the various business units or teams collecting and managing the information in isolation of other teams. These datasets need to be collected in their original format and the original records must never be modified for auditing purposes.
3. **Aggregate**: This step requires cleaning and normalizing the data into a single system. All modifications of the original data are in isolation, and the cleansed data will be used for analysis.
4. **Visualize and Analyze**: The ability to configure views to suit the preferences required for an individual business unit and/or for individuals within the organization. This ensures that individuals do not receive information overload when analyzing data that has an impact on their decision-making process. Advanced statistical analysis will help determine the past, present, and future condition of the assets on the linear network. Trending the data to predict a functional failure ensures an organization can schedule maintenance before enforcing a speed restriction on the line.

5. **Make Informed Decisions**: Making the data actionable optimizes maintenance. Through optimization, rail and transit owners will improve the reliability of their networks but, most importantly, they will improve safety within the organization and for the public.

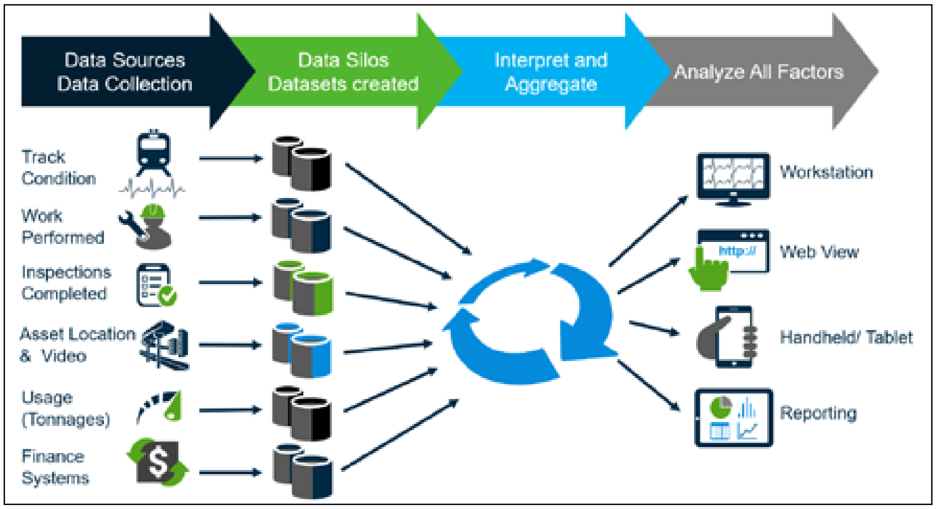

Figure 4: Collect, Consolidate, Aggregate, and Analyze to make Informed Decisions

Change in Culture

For decades, the rail and transit industry has successfully operated while continuing to work in siloed systems and often a siloed business structure. It is a significant challenge for an organization to go away from a mentality of "if it ain't broke, don't fix it" to a mindset of continuous improvement through the breakdown of existing silos and information sharing between business units. This challenge will result in a culture change, and this has been a significant challenge in many industries because of the fast-paced, changing world we live in today. Globalization and digitalization with the advancement in technology have contributed to this changing environment. Organizations that do not take advantage of these advancements have been left behind in the present time.

Addressing organizational change will impact three significant aspects of the current business process:

1. **The Technology**: Implementation of new technology can sometimes be challenging. A key component to success is to avoid technology for technology's sake; if the technology does not add value, do not implement the technology. Another critical element to understand during the implementation of new technology is that there must be support systems in place to support the users. This support requires quality and easily understandable training

programs. Organizations must understand that some investment in time is necessary before there is a return on the technology investment and that the investment in time varies between organizations. Furthermore, successful implementation of the technology is not the end. Continuous support and monitoring must be in place to ensure that the technology is bringing value to the organization. This understanding can be achieved using the technology itself. An organization can measure adoption and compliance using new systems with relative ease. Lastly, users who are adopting and complying with the new methods must be rewarded. Organizations need to celebrate their success through many communication channels to ensure success.

2. **The Process**: Implementing new technology without process change will result in guaranteed failure. No change in process will result in no difference in efficiency or safety because individuals will revert to their original ways of performing tasks. New systems and technology will require new processes to maximize the results of the new technology. For example, the rail and transit design engineers went from a world of manual drafting to computer-aided drafting (CAD) design and now to building information modeling (BIM). These significant changes require process change for the transition to be successful. The process change leads to a maximum return on investment (ROI) associated with the new technology. Additionally, the process change must evolve at a deliverable rate within the resource constraints of the organization. An organization will not suddenly go from a reactive, maintained state to a proactive and predictive state overnight. The process needs to evolve organically to ensure success, and the rate of growth is dependent on an organization. Leadership must understand that going from a reactive to predictive maintenance state could require several iterations of process change, each with a measurable ROI. Lastly, the most critical aspect of process change is to ensure that it is aligned with the vision, strategy, and standards of the organization and the overall business goals. This alignment will enable the required executive support for success.

3. **The People**: The most complex and often the most significant obstacle is the people affected by organizational change. Usually, this is referred to as “resistance” and management typically views individuals who are resisting change in a negative light. However, resistance is a natural part of the evolution of change, humans do not like endings, and change represents something lost. Individuals will have anxiety or fear with change, and leaders and management must not ignore these sentiments and understand that these feelings are natural through the transition process. Addressing individuals’ fears and anxiety will create trust with the people using the new systems. Furthermore, it is critical that the people are involved in the process at the start; this will ensure buy-in when the new system goes live. Success is user engagement, allowing individuals to use the right tool for the right job. This engagement gives them the support they need while giving them independence with a feeling of empowerment to do their jobs correctly and efficiently. Additionally, communicating a clear vision of the future through the transparency of the organization will ensure success moving forward. An organization must allocate time and resources for training programs, and additional resources should be assigned to individuals who are struggling with the implementation.

Figure 5: Digitalization Success

Digital Twin

At its core, the consolidation of data silos perfectly encapsulates the digital twin methodology. The digital twin concept goes far and beyond a 3D representation of a physical asset. Not only can the digital twin be a digital representation of the physical asset, process, or system, it is also the information that allows us to understand and model its performance. Typically, a digital twin can be continuously synchronized from multiple sources, including sensors and continuous surveying, to represent its near real-time status, working condition, or position. Digital twins are used to optimize processes, construction, and the operation and maintenance of physical assets. Furthermore, rail and transit owners have recognized the potential for leveraging digital twins in the application of analytics, artificial intelligence, and machine learning in simulations and decision support throughout the lifecycle of design, construction, and operations. Through a connected data environment, we will have the capability of visualizing and understanding the full lifecycle of an asset, from the original baseline created at the time of design, to its current condition on the linear network.

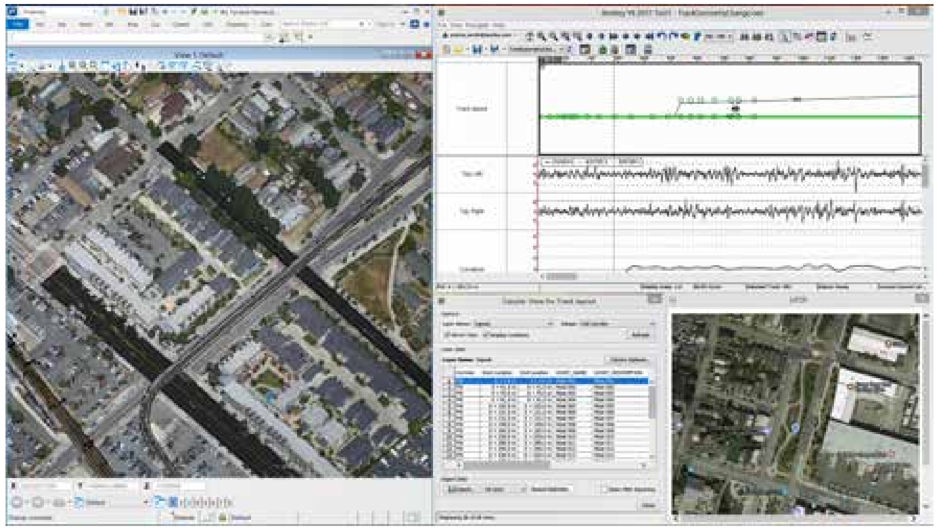

Figure 6: Synchronization of CAD (left), Linear Data (Top Right), EAM (Bottom Middle), and Maps (Bottom Right)

The Benefits with the Right Solution

Maximizing the efficiency of an organization's track maintenance team requires the right solution that will have the capabilities of joining the vast amount of data required for linear asset management. As the linear infrastructure ages, more problems will affect the assets' reliability. The infrastructure will continue to degrade, while the demand for that infrastructure will grow; these demands could be more operational hours, more vehicles, or longer vehicles. This additional demand will cause the linear assets to deteriorate faster. As a result, the network will have recurring faults and the costs associated with these faults will increase maintenance costs.

The need for rail and transit organizations to do more with fewer resources is an everyday reality. Solutions that will consolidate data will be essential moving forward. If an organization doesn't currently have a culture of continuous improvement, there will be a culture shock and this needs to be addressed at the start of implementation. There will be bumps and roadblocks moving forward but working together, keeping the vision at the target, success is certainly the future. This consolidation will allow track maintenance teams to perform analysis and forecast trends, make better-informed decisions about maintenance and renewals, and improve safety, security, and reliability of the linear network.