The Intelligent Transportation Society of America (ITS America)

The Use of Transit ITS Data for Planning and Management, and Its Challenges; a Discussion Paper

Brendon Hemily, Ph.D.

Final Report — Revised July 28, 2015

ITS America Research

Public Transportation ITS Discussion Paper #15-2

Produced by Intelligent Transportation Society of America (ITS America)
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology

Notice

Acknowledgements

This discussion paper is the result of discussions with many experts over the course of many meetings. The author would like to acknowledge the invaluable assistance of all those who participated in these discussions.

Technical Report Documentation Page

1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle The Use of Transit ITS Data for Planning and Management, and Its Challenges; a Discussion Paper		**5. Report Date** July 28, 2015 **6. Performing Organization Code**
7. Author(s) Brendon Hemily, Ph.D.		**8. Performing Organization Report No.** PT-ITS DP#2
9. Performing Organization Name And Address Intelligent Transportat ion Society of America (ITS America) 1100 New Jersey Ave., SE, Suite 850 Washington, DC 20003		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.**
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, S.E. Washington, DC 20590		**13. Type of Report and Period Covered** **14. Sponsoring Agency Code**
15. Supplementary Notes		

16. Abstract

From the earliest days of the development of Transit Intelligent Transportation Systems (ITS), there was a recognition among the most progressive transit systems that the data that would be created by Transit ITS (e.g. Automatic Vehicle Location, Automatic Passenger Counting, Advanced Fare Collection) would be an incredibly valuable resource, which could be used to create information to enhance planning and management, and support business processes and decision-making. This can greatly enhance the ability for managers to improve the effectiveness and efficiency of the services provided by transit systems.

However, transit agency staff often remain perplexed by the volume and complexity of the data and the challenges in using it. The challenge of how to use Transit ITS data remains one the most often heard concerns expressed by transit agency staff at Transit ITS meetings.

The objectives of this discussion paper are to:

1. Provide a high-level overview of the potential uses of Transit ITS data for planning and management purposes,
2. Identify the various challenges in using the data, and
3. Recommend research and other initiatives that would enable transit agencies to make more effective use of the data, and position the transit industry for a future of ubiquitous data and data-driven decsionmaking.

17. Key Words Public transportation, Intelligent Transportation Systems (ITS), Data	**18. Distribution Statement**

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**

orm DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Page left intentionally blank

Table of Contents

1. INTRODUCTION:

1.1. BACKGROUND

The Intelligent Transportation Society of America (ITS America) is the nation's largest organization dedicated to advancing the research, development and deployment of Intelligent Transportation Systems (ITS) to improve the nation's surface transportation system. Its Vision is to save lives, time and money and sustain the environment through the research, development and broad deployment of interoperable Intelligent Transportation Systems (ITS).

ITS America has been engaged in a long-term project to support the U.S. Department of Transportation (US DOT) in a range of activities related to research, development, and dissemination of information on ITS and its application to public transportation modes. It has been asked to develop strategic discussion papers on key topics that might help identify and understand challenges and barriers to ITS deployment, and to suggest recommendations for action, and thereby helping to achieve the full range of potential benefits that can be derived from the deployment of ITS by the public transportation industry.

These discussion papers build on the knowledge gained from a range of experts, including practitioners in the field, consultants, suppliers, and researchers, through workshops, listening sessions, and interviews, and supplemented by the review of pertinent literature.

1.2. THE ISSUE

From the earliest days of the development of Transit Intelligent Transportation Systems (ITS), there was a recognition among the most progressive transit systems that the data that would be created by Transit ITS (e.g. Automatic Vehicle Location, Automatic Passenger Counting, Advanced Fare Collection) would be an incredibly valuable resource, which could be used to create information to enhance planning and management, and support business processes and decision-making. This can greatly enhance the ability for managers to improve the effectiveness and efficiency of the services provided by transit systems. Today, one would refer to the notions of Business Intelligence and / or Data-Driven Decision-Making.

A small number of transit systems have carried this belief forward to continuously mine the data resources provided by Transit ITS, and these systems are consistently recognized as leaders in the industry. In addition, a small number of academic researchers have been exploring even more sophisticated uses of such data, to create origin-destination maps. However, the majority of transit systems have not opted, or been able, to pursue this avenue, and have rather focused most of their efforts in using Transit ITS to enhance real-time operations, and in particular incident management and security; the use of the data created by ITS remained in most cases an afterthought. In addition, there has been limited research or guidance of how to use this data.

However, this situation has been evolving in recent years. It has become increasingly popular for transit system boards and general managers to have available for internal use, and / or to provide to the public, “dashboards” that provide at a glance a visual snapshot of the transit agency’s system performance. And Transit ITS can in many cases create the aggregate data used to create these dashboards.

There has also been growing interest in the concept of “Business Intelligence” in management circles. In response to this interest, suppliers have started to offer data management tools that process the data created by the systems into a number of standardized “canned” reports.

Finally, there has been a growing public discussion of issues surrounding ownership of data, access to data, applications of open data, and the potential synergistic benefits to be derived from data fusion and data mining, popularly referred to as "*Big Data*". The transit industry has recently started to explore how it relates to these issues and developments, and is increasingly opening its data to third part applications developers.

However, transit agency staff often remain perplexed by the volume and complexity of the data and the challenges in using it (e.g. processes to clean the data, data warehousing cost, multiple sources of ridership data, relationship to NTD, etc.), and are often not trained into the processes of transforming data into information. The challenge of how to use Transit ITS data remains one the most often heard concerns expressed by transit agency staff at Transit ITS meetings.

But the Transit Industry is not fully using the data it collects, and is not yet positioned to expand its use in a future of ubiquitous data including Big Data, Smart Cities, and Connected Vehicles and Infrastructure.

The objectives of this discussion paper are to:

4. Provide a high-level overview of the potential uses of Transit ITS data for planning and management purposes,
5. Identify the various challenges in using the data, and
6. Recommend research and other initiatives that would enable transit agencies to make more effective use of the data, and position the transit industry for a future of ubiquitous data and data-driven decision-making.

2. WHAT DATA ARE WE TALKING ABOUT

ITS is a suite of different systems that are often inter-related. All systems produce data, including logs on events, faults, etc.

For the purpose of this discussion, the focus will be on a few primary systems that produce data of direct pertinence to transit planning and management. But there also exist secondary sources of data from systems that are being increasingly used in interesting ways. In addition, there is often need of context data that enables to correctly interpret information from primary sources.

2.1. Primary Sources of Transit ITS Data

There are three primary sources of data that are directly pertinent to transit planning and management.

- *Computer-Assisted Dispatch / Automatic Vehicle Location (CAD / AVL)*
  The CAD/AVL system is the heart of most Transit ITS deployments. It continuously tracks all transit vehicles in real-time, which enables efficient and effective operational control, incident management, security response, and service restoration.

  By comparing the real location of vehicles to their scheduled location, it enables continuous monitoring of schedule adherence. This can then be used to calculate Estimated Time of Arrival (ETA) of vehicles at all stops downstream and thus drive real-time information at displays at stops, on the internet, on mobile devices, etc.

  But more pertinent to this Discussion Paper, CAD/AVL provides a wealth of data from on-board devices (e.g. location, door opening sensors, odometer, etc.) that is geo-coded and / or time-stamped describing what the transit vehicles are doing. This in turn can be transformed into information on schedule adherence and On-Time Performance (OTP), running times, dwell times, delays, vehicle speeds, etc. It is important to recognize that it is not only just the GPS location that is important, but that the monitoring compares real-time outcomes to schedules.

  It should be noted that some systems provide a very detailed second-by-second log of events / locations of a bus between pull-out and pull-in. These files are extremely large with a complex file format, and it can be a challenge to extract the specific data required for a specific type of analysis.

  In addition, CAD/AVL systems are typically used to capture information from a number of other on-board sensors (e.g. passenger counters, wheelchair ramp, bicycle rack, etc.) that can also provide valuable information.

- *Automatic Passenger Counting (APC)*
  Automatic Passenger Counting systems can be developed as stand-alone systems, or integrated into AVL systems. A standalone APC system typically records passenger boardings and alightings with time and location coordinates. The collected data then is then off-loaded from the bus and matched to the stops and schedule database. After matching of the data to stops is completed, APC data can provide detailed profiles of customer activity by stop and time of day, as well as accurate estimates of passenger loads. This provides a wealth of information on customer demand.

  Some systems do calculate loads in real-time while the bus is in operation. However, imbalances between on counts and off counts can lead to escalating load estimates, and this can be a bit misleading if real-time loads are being broadcast in passenger info systems.

- *Advanced Fare Collection (AFC)*
  Data from legacy fare collection systems was typically of limited value because of its limitations. Data would be collected at turnstiles only on a periodic basis and provided aggregate total entries, etc. Electronic fareboxes on buses would only tally total information for a bus for an entire day, sometimes disaggregated by fare category. This was important for revenue control and ridership reporting, but of little use for planning or other uses.

  However, there has been dramatic enhancements to the data collected by more recent AFC systems, especially those using smart cards. Data is time-stamped, and increasingly geo-coded. The data can then be matched in post processing to stops. Alternatively, AFC systems are increasingly being specified to include an interface between the AVL and AFC systems so that each AFC event is automatically assigned to the current bus stop identified by the AVL system. As AVL uses an ordered list of stops scheduled to be observed by the bus, it allows "matching" of the AFC event to the correct stop as the AFC data is captured.

  In addition, the movement of individual smart cards can be tracked through the system, providing a wealth of information on customer behavior, including the possibility of building complete origin-destination matrices. Researchers are also using AFC data as a method for analyzing travel times and system performance.

2.2. OTHER SOURCES OF DATA

Other sources of data exist and are increasingly being considered as sources of information for planning and management purposes.

- *Transit Signal Priority (TSP)*
  TSP systems are designed to improve travel times and / or reliability at traffic intersections. First generation systems collected little information and did not permit matching data from the bus and the traffic controller. Newer generation systems should collect data on priority requests and responses, and this should enable planners to measure the effectiveness of priority strategies, as well as explore more aggressive priority strategies.

- *Vehicle Health (Mechanical Alarms)*
  CAD/AVL systems have always included mechanical alarms to alert control room staff about impending mechanical failures, in order to take remedial action and alert maintenance staff. While this data was not extensively used, with improved reliability of sensors, this should provide a potentially valuable source of information on vehicle health and enable more sophisticated fleet monitoring and maintenance planning and management.

- *General Transit Feed Specification (GTFS)*
  GTFS was originally developed as a simple but robust format to characterize transit routes, stops, and schedules that might be used to populate trip planners, such as Google Transit, but also by third party applications designed for mobile devices.

  More recently, many researchers and other experts have recognized that GTFS data also provides a remarkably simple way to build transit network models that can be combined with Geographic information Systems (GIS) or other forecasting models for a variety of purposes, including analysis of performance, accessibility, equity, etc.

- *GTFS-RT (Real-Time)*
  A real-time version of the GTFS format was developed to enable the provision of real-time information on travel times, incidents, etc. Where available, GTFS-RT is being used as a foundation for the third party trip planners on web sites and mobile devices. There are some efforts underway inn the Open Data community to explore how it might be used to measure performance.

2.3. CONTEXT DATA

In addition to the above, there is also need to consider data that provides context to the Transit ITS data for purposes of analysis. Examples include:

- Major seasonal periods (e.g. university summer break, etc.)
- Significant weather events
- Major events affecting ridership or traffic condition (e.g. festivals, sports events, natural disasters, etc.)

3. A TYPOLOGY OF USES OF ITS DATA FOR TRANSIT MANAGEMENT AND PLANNING

CAD/AVL and APC systems were first deployed in the 1980s in Europe and Canada, and in the 1990s in the U.S. From the earliest deployments, there was recognition among the most progressive transit systems that the data that would be created by Transit ITS would be an incredibly valuable resource for creating information to enhance planning and management.

Unfortunately, it has taken over two decades for this realization to filter throughout the industry, and in the interval, there have been very few efforts to capture the advanced reflections of the transit agencies that were on the forefront of using Transit ITS data for planning and management. These few efforts consisted of essentially three reports:

- Canadian Urban Transit Association, 1991: Report STRP #4 - *The Use of Automatic Vehicle Location for Planning and Management Information*
- Transportation Research Board (TRB), 2006: TCRP Report 113 - *Using Archived AVL-APC Data to Improve Transit Performance and Management*
- Transportation Research Board, 2008: TCRP Report 126 - *Leveraging ITS Data for Transit Market Research: A Practitioner's Guidebook*

Individual transit agencies have been struggling in isolation with the desire, and related challenges, to make effective use of the wealth of Transit ITS data to create valuable information for planning and management. However, these efforts have been scattered, and for the most part, hidden as well. Occasionally, one observes a pertinent presentation at the TRB Annual Meeting or at an APTA venue (primarily at the Multimodal Operations Workshop). Such presentations rapidly disappear, and thus provide no sustaining source of reference for transit agency staff looking for references or knowledge. The transit industry is particularly dispersed in regard to its efforts in this area, and individuals are very isolated. The consequence of this situation is that the collective knowledge, which exists in isolation, is not shared, and there has been no significant effort in the last eight years (nor much before that) to reflect and document the uses of Transit ITS data, and the related challenges, prior to the discussions leading to this Discussion Paper.

This is all the more unfortunate since the industry as a whole has been evolving:

- Many transit agency staff are becoming more interested in using Transit ITS data,
- ITS suppliers have made progress in the recent years in offering modules that enable the preparation of standardized reports based on the data created by their systems, and
- There is a growing discussion at many levels of the concepts of Business Intelligence, Open Data, Big Data, etc.

The time seemed opportune to reflect on the Use of Transit ITS Data and on the challenges related to its use.

The above-referenced CUTA report from 1991 provided a useful compendium of actual reports being produced and used at the time by Canadian transit systems with CAD/AVL and APC systems. Appendix 1 reproduces the lists of reports contained in the CUTA report. This is a valuable starting point for understanding what types of applications can be made of Transit ITS data.

The CUTA study classified the examples of reports into three categories, based on their primary purpose (and audience) within transit agencies. These were:

- Planning Information
- Line Management Information
- Executive Information

The TRB TCRP Report 113 identified a wide range of potential applications, and classified them according to the following categories of uses:

- Targeted Investigations (complaints, disputes, incidents)
- Running Time
- Schedule Adherence, Long-Headway Waiting, and Connection Protection
- Headway Regularity and Short-Headway Waiting
- Demand Analysis
- Mapping
- Miscellaneous Operations Analyses
- Higher Level Analyses

This discussion paper is based on numerous structured discussions with transit managers and staff in various venues, and a review of presentations at APTA, CUTA, and TRB as well as various research efforts in recent years. This information has been synthesized to develop a **Typology of Uses of ITS Data for Transit Planning and Management**, with the following categories:

- Corporate / Executive Monitoring and Reporting
- Service / Operations Planning
    - Service and Performance Monitoring
    - Strategic Analyses
    - Tactical Analyses
    - Policy Analyses
- Scheduling
- Line Management
- Real-Time Operational Control
- Exploratory / Data Mining

An overview of these categories of uses will be discussed in the following sections. A few examples have been provided for illustration purposes, but there are many more that have been developed by individual transit agencies; part of the challenge is to discover mechanisms to share this knowledge so that best practices can be adopted across the industry.

3.1. CORPORATE / EXECUTIVE MONITORING AND REPORTING

ITS data can be used to produce a substantial amount of the information required for executive monitoring and corporate reporting.

3.1.1. Executive Management Monitoring

ITS data can be used to monitor corporate performance for executive management on a daily, weekly, monthly, seasonal, annual basis.

Some agencies have in fact structured Executive Information Systems that perform a batch process overnight to assemble key management reports and provide the appropriate information to each level / area of corporate management. Such information can be structured to create color-coded alarms that highlight critical indicators (e.g. previous day's missed runs, etc.) so that enquiries can be made to the appropriate manager.

This is equivalent to internal dashboards.

3.1.2. Corporate Reporting

ITS data can be in many cases the source of data required for reporting to the National Transit Database, but it also can be used for calculating corporate Key Performance Indicators (KPIs). These in turn can be structured in an automated way to create reports or color-coded dashboards for Policy Boards, or for reporting to the general public. It should be noted that given the continuous issues in ensuring data quality, the creation of dashboards typically require a very strong data quality assurance process.

3.2. SERVICE / OPERATIONS PLANNING

ITS data can serve to enhance service and operations planning in several ways.

3.2.1. Service and Performance Monitoring

First and foremost, ITS data can provide a stream of data for continuous monitoring of service and performance. Before the deployment of ITS, data on schedule adherence, running times and maximum ridership load counts all had to be done manually, and as a result, was only collected on a periodic basis by traffic checkers or student interns in the field.

ITS enables the systematic collection of information on schedule adherence, on running times, and on ridership profiles at the stop level. The completeness of the ridership data depends on the percentage of the fleet that is equipped with APC equipment.

The service monitoring process typically produces standardized reports on important dimensions on a regular basis (daily, weekly, monthly, etc.). This information may be used for various important purposes such as the following:

- Measure On-Time Performance (OTP) and improve service reliability for customers,
- Analyze load vs. capacity, and overload / pass-ups
- Build ridership profiles by stop in order to identify key markets and transfer locations for route design, and to decide where shelters should be located,
- Identify recurring patterns of missed runs, overloads, early departures, etc.

Other examples are identified in Appendix 1.

More generally, the building of a systematic service monitoring system enables the transit agency to measure performance achieved against corporate service standards on an ongoing basis, often through the calculation of KPIs.

It should be noted parenthetically that the "Playback" feature on CAD/AVL systems has proved a great resource in customer communications to respond to customer complaints and disputes.

There are many issues related to ensuring the continuous quality of the data, as will be discussed in a later section. Given the need for quality data, the service monitoring system will include a number of special reports for diagnostics purposes. Discussions with transit agency managers often highlighted the lack, or poor quality, of diagnostic reports offered by the ITS technology suppliers. In such cases, transit agency may build their own standardized diagnostic reports, which are monitored in parallel to the above reports.

3.2.2. Strategic Analyses

Once a systematic monitoring process and resulting databases are established, this data can be used in various ways. It can be the foundation for strategic analyses, such as major networks redesigns, for example the creation of a Frequent Transit Network (FTN), or other significant campaigns (e.g. corridor priority treatment, transit network connectivity and resilience, etc.).

3.2.3. Tactical Analyses

ITS data can also be used for a variety of tactical analyses. Examples include:

- Appropriate locations for bus shelters.
- Transfer movements to improve schedule coordination and amenities
- Bus stop consolidation projects
- Delays by intersection in order to select where TSP would be of most benefit.

- Use of smart card data to calculate elasticity values

3.2.4. Policy Analyses

ITS and related data is starting to be used for policy analyses. Examples include

- Use of GTFS, or AFC data to measure accessibility. This in turn can be used to meet Title VI Equity requirements.
- Use of GTFS-RT data to benchmarking performance against peers.

Policy analyses such as these are for the most part being conducted by researchers outside of transit agencies, but the availability of Open Data will lead to an increase in such analyses and performance comparisons.

3.3. SCHEDULING

Beyond the use of ITS data for service and operations planning, it can be invaluable in improving the design of schedules. It provides valuable data on the location and time of maximum loads, and systematic data on the distribution of actual running times by month, day, time of day. It also includes deadheading running times. This disaggregated information can be sliced as finely as appropriate.

This can be fed into the scheduling system in order to adjust running times. The ITS provides an opportunity for schedulers and drivers to have informed discussions about running times, rather than ones based on perceptions of what is “real”. One benefit from better knowledge of the variability in running times is that it can serve to reduce recovery time, which saves operator hours and represents potential real cost savings. Unfortunately, there is a widespread perception in the discussions of experts that the transit agencies that actually use this wealth of running time data to tighten schedules are the exception, rather than the norm. This is surprising, and the cause is unclear.

3.4. LINE MANAGEMENT

Transit ITS data can also be used for line management purposes.

3.4.1. Operator Performance Analysis and Supervision

The discussions of experts led to some consensus that poor On-Time Performance is not just caused by external events and poor design of schedules, but is also highly dependent on the performance of individual operators. One large agency that had analyzed their OTP data suggested that more than half of the unreliability (within the control of the agency) was caused by individual operator performance.

As a result, there would be great value in using the ITS data to monitor operator performance.

The following operator-specific reports have been suggested:
- OTP with peer comparisons with other operators
- Speeding reports
- Early and leaving late reports
- Monthly chronic reports
- Bi-weekly location-specific reports
- Training reports
- Weekly probationary operator reports

This information can be used to structure field supervision work. The objective is to pinpoint operators who need assistance, supervision, or discipline.

3.4.2. Daily Service Reviews

ITS data can also be used to assist line management on a daily, or regular, basis. Reports from the previous day can be used to identify problems in order to suggest more active monitoring and remedial action.

Examples include:
- Missed runs
- Overloads / pass-ups
- Use of standby buses
- “Hot Spot” analysis: Allows road operations to concentrate on problematic locations, like areas with street repairs

Such reports can assist line management in daily operations. In New York Transit a System Data & Research Unit works closely with the Operations groups by providing the supportive information to decide on when and where to:
- Add extra bus service and deadhead trips,
- Structure field supervision work,
- Reposition standby buses, etc.

Visualization is increasingly important; it makes it easier to see performance indicators by location.

For higher accuracy AVL data is often integrated with other data sources such as:
- Schedules
- Timekeeping
- Road calls, etc.

3.5. REAL-TIME OPERATIONAL CONTROL

An emerging application of Transit ITS data is its use for real-time, or near real-time, operational control. These would qualify as decision support systems.

The Federally-sponsored Transit Operations Decision Support System (TODSS) is an example.

New York City Transit System Data & Research Unit has developed various such tools including:

- Operations Research Computational Analysis (in): Reporting and visualization platform:
    - Creates a Bus Bunching Dashboard
    - Pinpoints problem routes and locations
    - Allows Road Operations to investigate and take quick action
- Stringlines
    - A stringline chart represents each train trip by a “string” plotted with time on the horizontal axis and distance on the vertical axis
    - This tool is used for analysis of subway train movements
- Gap Table
    - Identifies the largest gap in service by line, direction, and console dispatcher’s territory
    - Allows immediate action for improving headway regularity

3.6. EXPLORATORY / DATA MINING

Finally, new tools and new approaches are emerging for data mining and data fusion. This is moving towards the concept of "Big Data"

Though still in its infancy, some examples are starting to appear:

- Safety Analyses: Portland TriMet mined a number of ITS and other data sources in an effort to bring to light patterns with respect to safety and collisions.
- Crowd-sourced data: Preliminary efforts are starting to emerge to mine twitter or facebook feeds. An example would be to use an "emotional quotient" classification of words used in tweets in an effort to assess emotional perception and identification.
- Transit data is being mined and merged with other sources of urban data (e.g. traffic, health, crime, etc.) as part of "Smart Cities" analytics, in an effort to reveal new relationships and provide a better basis for planning and policymaking.

4. CHALLENGES

Having discussed the uses of Transit ITS data, it is useful to discuss the various challenges related to the use of this data. Various structured discussions were organized and helped to provide a comprehensive overview of these challenges. The categories of challenges are as follows, and will be discussed in the following sections:

- Lack of Understanding by Senior Management
- Corporate Data Management Challenges
- Challenges Related to Ensuring Data Quality
- Challenges in Using ITS Data Once Cleaned

4.1. LACK OF UNDERSTANDING BY SENIOR MANAGEMENT

The challenges for those advocating for a more active use of Transit ITS data start at the top of the organizations. The following outlines some of the challenges that have been identified.

- Lack of interest from senior management and policy boards:
  Policy boards and senior managers of transit systems need to continuously focus on ensuring sufficient funding to operate and expand the transit system, and on building the stakeholder coalition to do so. Technology is for the most part a secondary concern, and they are often not very interested in ITS, even less so in the data that ITS creates. The transit industry is by-and-large characterized more by an operations-driven culture than by a data-driven decision-making culture.

- Design and Use of ITS is primarily operations-driven:
  Transit ITS projects are typically operations-driven, with the major focus on integrated voice and radio communications, dispatching, real-time operations, and safety / security incident management. Unfortunately, as a result, the use of ITS data is often an afterthought, after the Systems Engineering has been completed, and it is too late to structure the project.

- Confusion over benefits to be derived from ITS data:
  There is also confusion among senior managers with respect to the benefits to be derived from more extensive and intensive use of data. Benefits of the use of ITS data have been poorly articulated. Technical staff need to focus their messaging to senior staff on the benefits / processes that management understands, not the technology itself. There is also a very important need for the industry as a whole to better document how ITS data is being used, and the real benefits to be derived from its uses.

- Confusion over strategy with respect to data:
  Senior managers are often confused in how to effectively manage all the data being created by Transit ITS and other sources. In fact, one of the teams of the

Leadership APTA professional development program tackled this confusion head on by producing in 2014 a report entitled *Surfing the Data Tsunami; Agile Data / Information Strategies for Public Transportation*. This report provides much solid information and advice for senior managers in developing a data strategy. The link to the report can be found in Appendix 2, along with a list of the recommended strategies identified from best practices.

- Inter-Departmental Conflicts:
  Complex projects and systems, such as ITS, may lead to inter-departmental conflicts. There may be conflicting interests or perspectives with respect to the system and its operation, from different key internal stakeholder groups (e.g. operations, planning, scheduling, information technology, etc.). Senior managers rarely relish the prospect of addressing and resolving these internal conflicts, especially if they have relatively little interest in the system in the first place.

- "Single Source of Truth" Issue:
  To compound the challenge, different parallel systems may produce inconsistent data. The most common problem relates to ridership reports obtained from APC vs. AFC systems. The challenge of data credibility can aggravate any existing inter-departmental conflicts, and the discussions suggested that senior managers seek to have a " Single Source of Truth". Unfortunately, there is generally a lack of willingness to devote the resources to have an enterprise-wide data management strategy. It will be important for technical staff to resolve any issues related to data credibility.

The lack of understanding and focus on the importance of data from senior management has two important consequences.

The first concerns **the lack of ownership of the ITS data by the transit agency**. The lack of senior interest in Transit ITS data means that agencies often neglect the issue of ownership of data when developing the Request for Proposals for the procurement of ITS technologies. During the discussions there were all too numerous stories of transit agencies that had neglected to specify that all data created by the systems belongs in its entirety to the agency, only to find to their chagrin that the suppliers had standard wording inserted in the contracts asserting that the data and its use was proprietary. This enables the suppliers to charge license fees for its use, on top of development charges to build interfaces.

The second consequence affects the **lower priority given to resources and expertise**. The lack of senior interest and support often results in fewer resources being provided to build well-structured data warehouses, and to obtain the specialized expertise to effectively use the collected ITS data.

4.2. CORPORATE DATA MANAGEMENT CHALLENGES

There are various challenges related to the organization of data within the organization and its management. As previously mentioned, if the use of data is a lesser priority with senior management, the IT Department may be under-resourced, and data management will be a secondary priority compared to basic IT network hardware and software responsibilities.

Some of the typical data management challenges that have been identified include the following:

- Data storage: managing the volume of data, especially if there is a lack of an integrated data warehouse,
- Lack of (or unclear) data retention policies,
- Lack of systematic inventory of databases,
- Legacy special-purpose data spreadsheets or databases that are not integrated with the network, or that use non-standard definitions or formats making them difficult to use in conjunction with the ITS database,
- Lack of an integrated data warehouse and the resulting existence of multiple databases using different data formats, or with different coding of the same information (e.g. bus stop inventory),
- Use of proprietary, or just different, data formats and even definitions by the suppliers of ITS technologies, making interoperability and data integration challenging,
- Slow adoption rate of the Transit Communications Interface Profiles (TCIP) standard,
- Lack of clarity concerning policies and procedures with respect to the management and provision of Open Data, etc.

It should be noted that TCRP Report 84 Volume 9 *Transit Enterprise Architecture and Planning Framework*, may be of interest and is listed in the References. It presents multi-faceted methods, tools, and examples within a framework to help transit agencies successfully implement technologies. The report describes the connections between a transit agency’s business and the technology, assists with building the business case for specific investments, highlights different financing options, provides guidance on an enterprise-wide approach to create more efficient and effective system deployments, and provides a method to show the benefits of a technology investment.

4.3. CHALLENGES RELATED TO ENSURING DATA QUALITY

Beyond challenges related to the corporate management of data, there are also significant challenges that are specific to ensuring the quality of ITS-created data on an ongoing basis. Some of these challenges relate to external and internal factors, while others relate specifically to the nature of the data in that it needs to be cleaned and properly matched to schedules, runs and stops.

4.3.1. Factors Affecting Data Quality

- Lack of standardized methodology for measuring APC accuracy for acceptance testing:
  There are no standardized methodologies for measuring the inherent accuracy of APC systems. Individual agencies develop their own methodology, if at all, and the sophistication varies widely. As a result, the quality of the APC data varies considerably from agency to agency.

- Incomplete APC sampling:
  Theoretically, a fully representative base of ridership data can be obtained with only 20% of the fleet being equipped with APC sensors, provided a statistically sound sampling plan is developed and implemented. Some agencies make the fulfillment of the APC sampling plan a priority, and in at least one case make it an explicit objective of garage managers and dispatchers. Other agencies have great challenges in ensuring the assignment of APC buses according to plan because of competing operational priorities, layout and management of the garage lanes, etc. The most common solution in such cases is to strive toward 100% of the fleet being equipped, so the collection of data does not depend on the daily assignment process.

- Missing or corrupted data:
  These systems are complex and there can be a variety of reasons why ITS data ends up missing, or corrupted, including: failure of any of the on-board data collection sensors or storage unit, operator error in logging in (though modern systems have safeguards against this), problems in downloading the data in the garage, uploading the data to the back-end databases, or matching problems

- Lack of diagnostic tools provided by suppliers to determine cause:
  Unfortunately, experience by many participants in the discussions suggested that there is a lack of sophisticated, or even any, diagnostic tools being provided by ITS suppliers. The problem can in fact be compounded since bad data will be buried within the "canned reports" provided by suppliers.

4.3.2. Challenges in Matching and Cleaning Data

As mentioned, ITS data comes from various sources (AVL, APC, on-board sensors, etc.), and its value comes from being able to compare it to the agency's performance targets, as expressed in the theoretical schedule. ITS data therefore needs to be cleaned and properly matched to schedules, runs and stops. However, all agencies will experience significant challenges in this regard.

Some of the most commonly experienced problems are as follows:

- Imprecise geo-coding of bus stops
- Multiple bus stop databases, developed over time for different systems or purposes, with different bus stop IDs
- Data matching problems between the scheduling and the CAD/AVL systems. TCRP Report 113 is a good reference to many of these challenges. Common causes of matching problems relate to:
    - Layover locations: their selection, geo-coding, and operator behavior,
    - Operator movements on and off buses, at various, and unpredictable locations,
    - Loops on routes,
    - Branches on routes,
    - Negative loads, and algorithms for redistributing passenger numbers, etc.

Other sources of data quality problems may result from the following:

- Insufficiently frequent polling rates (every minute or more) of older systems,
- Multiple sources of GPS location, from different on-board systems (e.g. AVL vs. AFC),
- Algorithms to factor up passenger count samples when only a percentage of the fleet is equipped,

It should be noted that these challenges exist not only at the time of the deployment of the ITS technologies, but require an ongoing process to keep up with matching to ever changing schedules, and geo-coding to changing routes and even stop locations.

Since there are many challenges in ensuring the ongoing quality of the data, and few diagnostic tools, problems can get buried in canned reports provided by suppliers. It is therefore essential that transit agencies require not only ownership of the data, but also have access to the raw data.

4.4. CHALLENGES IN USING ITS DATA ONCE CLEANED

ITS data can be used to provide valuable information for planning and management. However, there are many challenges in making the most effective use of this data, even once it is cleaned and available. In many cases, the challenges listed below stem from, or are compounded by, the lack of understanding or even interest by senior managers with respect to ITS data and its uses. Articulating the benefits of using ITS data is a priority for technical managers and staff individually, but also for the industry as a whole.

- Supplier ownership of data:
  This situation should not exist and arises from a lack of attention at the time the requirements for the systems were developed, and the systems were procured. The Request for Proposals and subsequent contract should clearly state that all data produced by the systems is owned by the transit agency, not the supplier. TCRP Legal Digest Report 37 provides some guidance in this respect. In

addition, transit agencies should require access to all raw level data, for purposes of diagnostics, or analysis.

- Lack of internal resources and technical expertise:
  In most agencies there is a lack of resources and technical expertise for analysis using the ITS data. This requires expertise on the one hand on technical tools and processes for data mining and visualization, but on the other, on transit business processes. At the same time, there is generally a lack of resources for IT data management support.

- Lack of support within the transit industry itself:
  There are unfortunately few resources available within the industry to assist those who desire effectively using ITS data, beyond the handful of documents identified in the References section. Discussion participants identified several related challenges:
  - Lack, or poor quality, of supplier tools,
  - Lack of standardized methodologies for data mining, analysis, and visualization,
  - Lack of mechanisms to share reports, processes, visualization tools between transit agencies,
  - Lack of training specifically geared to effectively using Transit ITS Data,
  - Lack of research to explore the mapping of data to business processes, etc.

- "Bad Day" challenge:
  Significant unpredictable events, such as major weather storms, major road closures or security emergencies, natural catastrophes, can create significant disruptions to service and / or ridership patterns, as can other events (e.g. holidays, university breaks, etc.) Transit agency staff needs to understand how these events affect the data, and develop policies for how to handle "bad days", that may distinguish between analysis and reporting processes. Some suggestions were:
  - When looking at APC or AVL report, one should be able to check a box to exclude bad days (or perhaps checked by default so the analyst doesn't have to remember which days the university was closed.)
  - "Bad days" are still analyzed. You still need to know how to improve service on those bad days.
  - Identify the definition of a "bad day", and notify/indicate occurrences.
  - Develop different contingency scenarios and operating strategies.

- Other problems that require careful attention in the manipulation and analysis of ITS Data:
  - Resolution of “Single Source of Truth” issue, and
  - Standardized supplier reports that report Running Times with more detail than just Average Running Time; it should be reported by route segment (between timepoints or even stops)

5. RECENT NOTEWORTHY DEVELOPMENTS IN THE USE OF TRANSIT ITS DATA

The previous section identified a comprehensive list of challenges related to the use of Transit ITS Data. Despite these many challenges, there are some noteworthy efforts underway that are worth mentioning.

5.1. NEW TOOLS FOR ANALYSIS AND VISUALIZATION

First, there appears to be a growing suite of powerful tools that are emerging in various fields, that individual transit systems have been exploring in the application to ITS Data. These tools may assist in various respects for:

- Data storage and processing,
- Visualization and interactive reporting,
- Delivery of reports and tools to front line employees,
- Statistical methods for data mining and deeper analysis of trends and patterns, etc.

Metro Transit in Minneapolis for example, reported developing a suite of analytic tools that included:

- Reporting Tools Delivered with Systems
    - Playback
    - Fare data reporting
- Excel-Based Tools
    - Dialog window to generate queries
    - Pivot tables to aggregate / summarize data
    - Mix of on-the-fly and pre-generated reports
- Crystal Reports
- Interactive Web Tools
    - Web-based reporting environment
    - Input controls / filters create interactive tools
    - More robust visualizations
    - Combining reports/tools into dashboards
- Statistical Modeling / Predictive Analytics
    - Ridership forecasting
    - On-Time Performance modeling
- Real-Time Tools
    - Diagnostic reports
    - Decision support tools
    - GIS mapping of on-time performance, incidents, etc.

5.2. PARTNERING WITH UNIVERSITIES

There is a small, but growing, interest among certain academic researchers to pursue more applied research, and in particular through partnerships with transit agencies. In some cases, these are initiated by the academic researchers, and in other cases, by the transit agency. Irrespective, these can enable a combining of sophisticated knowledge on tools and methodologies, with "real-world" experience on business processes and needs. Examples include:

- University of Minneapolis and Metro Transit (see University of Minneapolis report in the References)
- MIT and Transport for London
- Polytechnique University (Montreal) and the transit systems in Gatineau and Montreal
- University of Waterloo and Grand River Transit

5.3. NEW APPROACHES TO ITS SYSTEM DEVELOPMENT AND DATA UTILIZATION

New York City Transit's System Data & Research Unit has been using a Fast, Agile Software Development process, using open source formats and software, and a highly dynamic and flexible collaboration between internal analytical experts and operations end-users, with great success. This has led to much more rapid and successful deployment of systems, but also to a whole generation of ways of using ITS data. An example is the use of Real-Time Stringlines and Gap Reports to support near-real-time operational control.

A key lesson learned is to improve data quality by integrating multiple sources. Not all data sources are available in real-time, but they try to use what is available for quick decision-making and short-term analysis (it is estimated to be 80-85% correct in representing actual service). Special processes and algorithms are then used to connect several data streams for a more complete picture, and long-term analysis planning studies. Their long-term goal is to have all data available in real-time for a comprehensive picture of service as it unfolds.

5.4. NEW DIRECTIONS IN USING SMART CARD DATA

Several academic researchers (e.g. MIT, Polytechnique, University of Kyoto, University of Queensland, University of Chile, etc.) have been focusing considerable effort to explore the many uses of smart card data, and the related challenges. An International Workshop was held in Gifu, Japan in 2014, and identified the variety of applications being made of AFC data. These included:

- Path choice / transit assignment (including impact of transfers, network choices, crowding, information, etc.)
- Transfer patterns
- Route / vehicle loading
- Service reliability
- Variations by time of day, day of week, etc.
- Inference of residence location and socio-demographic information
- Comparison with Travel Surveys to perform validity check
- Retention rates
- Impact of weather
- Shopping vs. mode access behavior

Smart card data provides patterns of customer behavior as a backdrop for analyzing:
- Service reliability as experienced by customers,
- Modifications to operational plans,
- Real-time intermodal coordination, and
- Incident management by assessing:
    - Traveler natural responses,
    - Contingency plans,
    - Emergency information provision,
    - Transfer management, etc.

The above applications represent applied research in nature, and will be valuable to the transit industry, though some may be of more interest to rail-based systems.

5.5. OTHER NEW DEVELOPMENTS

Other pertinent developments include:
- A new TRB *Subcommittee on Transformative Trends in Transit Data*, that is specifically focusing on the uses of ITS and other transit data,
- The use of GTFS for policy analysis and performance measurement,
- The development of new open source tools, including "Open Trip Planner Analyst", etc.

6. RECOMMENDATIONS

Finally, the Discussion Paper is making the following recommendations in order to address some of the challenges, and enable the transit industry, individually and as a whole, to make more effective use of Transit ITS data. These include:

6.1. GENERAL RECOMMENDATIONS

- Transit systems need to be reminded to require ownership of data in the design and procurement of ITS technologies, and this should include access not only to data aggregated into reports, but also raw data. (Note: TCRP Legal Results Digest 37 entitled **Legal Arrangements for Use and Control of Real-Time Data** provides some examples of wording related to ownership and access to raw data.)
- Suppliers should provide better diagnostic tools to identify missing and corrupt data problems, and determine the causes of the data problems when they occur (e.g. on-board sensors, downloading failures, data matching problems, etc.)

6.2. BENEFITS OF TRANSIT ITS DATA AND ENHANCED SHARING OF KNOWELDGE

There is a great need to identify and document benefits of using Transit ITS Data that would stimulate greater interest and understanding by senior management. There is also a more general need to expand the dissemination and sharing of knowledge. Suggestions include:

- Best Practices and Lessons Learned on Benefits of Using Transit ITS Data (key applications, Return on Investment, and visualization techniques)
- Find mechanisms to reach those that do not typically attend APTA conferences and workshops.
- Use the US DOT Professional Capacity Building Program to have a series of webinars in order to share information.
- Develop innovative mechanisms to share information on data and report formats, and visualization techniques. These might include:
    - Compendium
    - Web portal
    - Wiki page, etc.
- There is a need for more training on methodologies related to the management of data and methodologies for mining and analysis.

6.3. RESEARCH NEEDS

Finally various research needs have been identified. These include:

- Compendium of Examples of Financial Savings from Use of ITS Data
- APC Accuracy Measurement Methodology for Acceptance Testing
- Use of ITS Data for Real-Time Operational Control
- ITS Data and Performance-Based Management (including Asset and Safety Management)
- Best Practices: APC Data and NTD Reporting
- ITS Data as Input to Key Performance Indicators for Dashboards

KEY REFERENCES

American Public Transportation Association, 2014,: *Surfing the Data Tsunami; Agile Data / Information Strategies for Public Transportation*, Leadership APTA Report, Prepared by Mark Ellis, Roderick Diaz, Louis Cripps, Laura Minns, and David Huffaker http://www.apta.com/members/memberprogramsandservices/leadershipapta/Documents/Team 5 Surfing Data Tsunami paper.pdf

Canadian Urban Transit Association, 1991, *The Use of Automatic Vehicle Location for Planning and Management Information,* Report STRP #4

Transportation Research Board, 2003, *Uses of Archived AVL-APC Data to Improve Transit Performance and Management: Review and Potential*, TCRP Web-Only Document 23 http://onlinepubs.trb.org/Onlinepubs/tcrp/tcrp_webdoc_23.pdf

Transportation Research Board, 2006: *Using Archived AVL-APC Data to Improve Transit Performance and Management*, TCRP Report 113 http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_113.pdf

Transportation Research Board, 2008: - *Leveraging ITS Data for Transit Market Research: A Practitioner's Guidebook*, TCRP Report 126 http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_126.pdf

Transportation Research Board, 2011, *Legal Arrangements for Use and Control of Real-Time Data*, TCRP Legal Research Results Digest 37 http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_lrd_37.pdf

Transportation Research Board, 2011, *Transit Enterprise Architecture and Planning Framework,* TCRP Report 84 Volume 9 http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_84v9.pdf

University of Minnesota, 2011, *Data-Driven Support Tools for Transit Data Analysis, Scheduling and Planning, Final Report*, Report CTS 11-15, Prepared by Chen-Fu Liao http://www.cts.umn.edu/Publications/ResearchReports/reportdetail.html?id=2044

APPENDIX 1: *THE USE OF AUTOMATIC VEHICLE LOCATION FOR PLANNING AND MANAGEMENT INFORMATION-APPENDICES B, C, & D*
Report STRP #4: Canadian Urban Transit Association, 1991

The CUTA Report STRP #4 explored the uses of CAD/AVL and APC data by Canadian transit systems in 1991, and provided a compendium of examples of actual reports, organized in three sections on Planning (Appendix B), Line Management (Appendix C), and Executive Management (Appendix D). The following are the lists of reports contained in STRP #4.

Appendix B Examples of Planning Information

#		Type of Report	Format
B	1	Schedule Adherence Report	Table
B	2	Running Time Profiles	Graphic
B	3	Schedule Profile Deviation	Graphic
B	4	Actual verus Scheduled Headways	Graphic
B	5	Schedule Adherence Point Check	Table
B	6	Schedule Deviation Comparision	Graphic
B	7	Route Speed Profile	Graphic
B	8	Stop Time Analysis	Graphic
B	9	Route Passenger Load Profile	Graphic
B	10	Transfer Connections	Graphic
B	11	APC Route Evaluation	Table
B	12	Schedule Adherence 1 Week Summary	Table
B	13	Layover Report	Table
B	14	End of Line Schedule Adherence	Table
B	15	Distribution of Running Time	Table
B	16	Headway Report	Table
B	17	Route Operations Report	Table

Appendix C Examples of Line Management Information

#		Type of Report	Format
C	1	Summary of Incidents & Control Actions	Text
C	2	Trip Cancellations by Driver & Vehicle	Table
C	3	Utilization of Standby Buses	Table
C	4	Non-AVL Daily Service Summary	Text
C	5	Non-AVL Daily Pass-up Report	Table
C	6	Mechanical Failures by Vehicle	Table
C	7	Mechanical Failures by Driver & Vehicle	Table
C	8	Text Message Summary	Text
C	9	Driver Initiated Reports	Table
C	10	Short Turn/Lost Mileage Report	Table
C	11	Change-Off Report	Table
C	12	Mileage Report	Table
C	13	Average Hourly AVLC Calls	Graphic

Appendix D Examples of Executive Information

#		Type of Report	Format
D	1	Service Operated vrs Service Budgeted	Table
D	2	Service Operated-Service Budgeted by District	Table
D	3	Service Operated-Service Budgeted by Route	Table
D	4	Annual Schedule Adherence Summaries	Graphic
D	5	Daily Report on Schedule Disruptions	Graphic
D	6	Vehicle Cancellations	Graphic
D	7	Productivity and Efficiency Indicators	Graphic
D	8	Service Deviations	Graphic
D	9	Vehicle Change-Offs	Graphic
D	10	Monthly System Wide Schedule Adherence	Graphic
D	11	6 Month - System Wide Schedule Adherence	Graphic
D	12	One Day Schedule Adherence	Graphic

APPENDIX 2: LEADERSHIP APTA REPORT: SURFING THE DATA TSUNAMI; *AGILE DATA / INFORMATION STRATEGIES FOR PUBLIC TRANSPORTATION*
http://www.apta.com/members/memberprogramsandservices/leadershipapta/Documents/Team 5 Surfing Data Tsunami paper.pdf

The purpose of this paper is to outline strategies that, when employed, would help craft an agency-wide data / information management policy, a policy that is agile and can adapt as the needs of the agency changes.

Extracted from Table of Contents:

Best Practices and Strategies:

Assess Your Organization
Strategy #1: Define and Articulate your Vision
Strategy #2: Assess your Agency and Learn from Others
Strategy #3: Let Business Needs Drive Solutions
Strategy #4: Assess and Develop a Data Policy

Assign the Right People
Strategy #5: Engage CIO/IT Managers you can Trust
Strategy #6: Assign Data Managers in Functional Groups
Strategy #7: Encourage Collaboration

Engage Partners
Strategy #8: Decide how much Contracting is right for you
Strategy#9: Promote Positive Procurement Partnerships

Manage Data and Information
Strategy #10: Establish Data Management Protocols
Strategy#11: Establish an Enterprise Database

Maximize Data Value
Strategy #12: To Improve Performance, Turn Data into Information
Strategy #13: Look for Spin-off Opportunities
Strategy #14: Consider Revenue Opportunities
Strategy #15: Share Data

Recommendations for Data Management Policy Development
- Policy Goals
- Agile Internal Data Management Policy Elements
- Agile Data Dissemination Policy Elements
- Technology Collaboration Policy Elements