Surfing the Data Tsunami

Agile Data/Information Strategies for Public Transportation

Project Team 5

Leadership APTA

Class of 2014

Mark Ellis; *Maintenance Design Group, LLC - Katy, TX (Team Coordinator)*

Roderick Diaz; *Metrolink - Los Angeles, CA*

Louis Cripps; *Regional Transportation District - Denver, CO*

Laura Minns; *LYNX - Orlando, FL*

David Huffaker; *Sound Transit - Seattle, WA*

TEAM

5

Surfing the Data Tsunami
Agile Data/Information Strategies for Public Transportation

Table of Contents

Introduction

Previous Leadership APTA (American Public Transportation Association) project teams have focused on the utility of various technologies such as automatic vehicle location and real-time rider information. However, the analysis has mainly focused on the technology of the specific solutions rather than the organizational implications of investments in these technologies. The new technology generates a myriad of data that is taxing transit agencies in ways for which they have not been prepared. Project Team #5, Class of 2014 (the Team) focused on helping Public Transportation Agencies develop strategies to maximize the value of the data/information that is created from these technologies. Above all, the Team suggests that the creation of specific policies and practices addressing this data/information will yield the best solution for the agency, regardless of its size and service characteristics.

This paper aims to identify the current trends in data/information management and how transit agencies are addressing data storage, internal and external data distribution, and issues surrounding open and closed architecture of data generating systems. As the transit industry deals with this "tsunami" of data, they must also deal with the ever-growing appetite for more and more data by the transit riding public. Transit agencies' policy strategies must now be agile enough to adapt as the data requested by both internal and external users grows.

> **"You can get buried in technology"**
>
> **Gary Thomas, CEO of Dallas Area Rapid Transit**

The paper is directed at CEOs and decision-makers within the North American public transportation industry. The research collected for this paper and the resulting analysis reveals how transit agencies can be successful with their data/information management. Success comes from being agency. This success benefits the customers, the vendor community, and other users related to the industry.

Data/Information

Gary Thomas, CEO of Dallas Area Rapid Transit, said, *"You can get buried by technology."* A literal tsunami of data is being generated by your agency every minute of every day.

Is your agency ready for the Data Tsunami?

Data/Information management has become an integral component of every transit system. Some key factors in successful data management for the transit industry are:

- Storage
- Complexity

- Architecture
- Availability
- Value to the agency
- Impact on the agency
- Resulting policies

Data/Information management can be generally characterized in three terms - Volume, Velocity, and Variety.

Volume

The daily volume of data/information has increased more than tenfold since 2005. As an example, in 2004 the amount of internet traffic worldwide crossed 1 exabyte (EB)[1] per month. By 2010 that monthly traffic crossed 21 exabytes per month. The amount of data continues to grow exponentially.

Velocity

The speed of data generation is increasing as well. Think of how real-time information is refreshing automatically from the on-board technologies on your buses and trains. The analytics that digest this streaming data into manageable bytes has to interpret the data, and possibly take action, all in real-time.

---
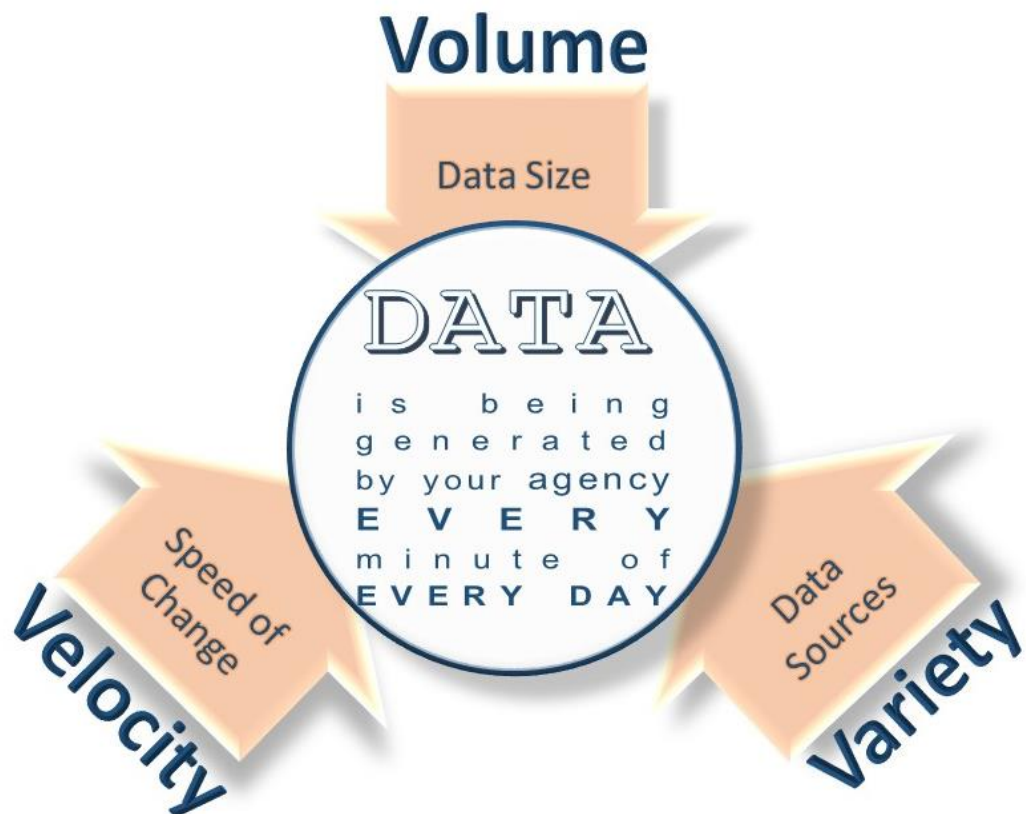
[1]The exabyte (EB) is a multiple of the unit byte for digital information. 1 EB = 1000^6 bytes = 10^{18} bytes = 1,000,000,000,000,000,000 bytes = 1000 petabytes = 1 million terabytes = 1 billion gigabytes

Variety

Perhaps the most challenging issue for transit agencies is the variety of the sources of data generated on a daily basis. Agencies have developed procedures to handle traditional sources of data, such as bus maintenance records. However, more recently, there are non-traditional sources of data coming from new media, for example? How many Facebook friends does your agency have? What are people tweeting about your service right now? The answers to these questions are pertinent to an agency's ability to provide the service its customers demand.

Methodology

Interviews

The methodology employed by the Team included personal interviews with transit agency CEOs, General Managers and technology vendors of products that either generate data for use by the agencies or utilize the data being generated by transit agencies.

Only **8%**

of Transit Agencies surveyed were <u>very satisfied</u> with their Policies and Practices

Survey

An on-line industry-wide survey was also conducted. The survey was designed to capture specific topic and quantitative data related to the subject matter of this paper.

This paper contributes to the existing literature by using revealed-preference information from the personal interviews and survey to discover the actual effects of data/information management decisions, and systems being utilized. The successes and failures of these systems were studied along with an analysis of micro level data to sort out causal trends and effects of strategies currently being employed by the industry.

Purpose

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.

Elements

Section One: Overview

This section provides a brief synopsis including introduction, data/information - supporting background, methodology, purpose, and elements of this paper.

Section Two: Research and Analysis

This section provides the results and analysis of the team's research, data collection, and analysis effort and provides insight into the state of the industry and trends related to data/information management strategies and policies.

Section Three: Best Practices and Strategies

This section provides a detailed description of the Best Practices and the specific supporting strategies related to effective data/information management.

Section Four: Case Studies

This section provides detailed information for three case studies developed by the team to illustrate how effective data management policies and practices can positively influence an organization.

Section Five: Recommendations for Data Management Policy Development

This section provides the recommendations by the Team on data management policy development strategies. In some cases, these recommendations may be immediately actionable and in other cases, they may lead to a more dynamic policy development philosophy related to data/information management.

> **"In God we trust; all others must bring data"**
>
> **W. Edwards Deming**
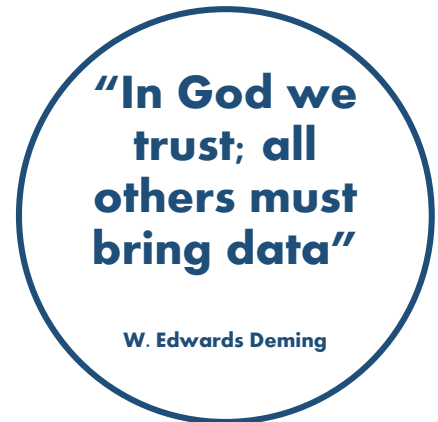
Appendix A

Transit Agency Interview Participants

Appendix B

Vendor Interview Participants

Appendix C

Interview Questions

Appendix D

On-line Cross Sectional Survey Data

	Research and Analysis	2

Transit Agency Interviews

The Team interviewed the General Manager's and CEOs from 21 transit systems of varying sizes and from varying regions of the country. This cross section of information allowed the team insight into the approach and policies being utilized by the transit industry.

The interviews covered a wide ranging discussion of technology and data management policies and practices at their agency. Also discussed were successful and unsuccessful technology initiatives, and what lessons learned can be gained from their examples, the role of the CEO, General Manager, and the Board in terms of technology and data management issues, suggestions of organizational structures and how they can support agile information management, and a discussion of best practices that these leaders use to improve their organization and their data management policies and practices.

The interviews with transit agency representatives included representation from a wide variety of sizes of agencies, a diverse geography, agencies with different combinations of modes, and agencies with different combinations of functions. It was important to the Team to include small and medium-sized agencies in order to glean insights that are applicable to a wider cross-section of the transit organizations. As with the survey results, the broad cross-section of interviews helped to generate findings and strategies that are universal in nature.

Private Sector/Vendor Partners Interviews

Solutions to manage transit generated data necessarily involves partnerships with private sector partners, including vendors, application (app) developers, and consultants. Interview research, therefore, consciously involved representatives of private sector partners. As with the transit agency representative interviews, the Vendor Partner interview topics explored a wide variety of issues, including the state of data/information management in the transit industry, engagement of private sector partners, and strategies to promote innovation. These interviews contributed directly to insights on implementation success factors, effective staff management and leadership approaches, procurement, data management and data standards, and positioning organizations to adapt to innovation.

On-line Cross-Sectional Survey Instrument

In order to obtain input from influential leaders in transit, the team developed and administered an On-line Cross-Sectional Survey[2]. This survey focused on transit

---

[2] The team used a cross-sectional "one-time" survey, or a single questionnaire or interview administered to each survey participant.

industry data/information management practices as of June 2014.

Goals and Hypothesized Relationships

Prior to the survey, the team hypothesized that attitudes and applications of data/information management would have correlations to property size, geographic region, and modes operated.

The goals of the survey were to provide clear, reliable, actionable insight on industry trends, and to guide the development of this paper[3]. In an effort to achieve these goals, the team addressed the following key demands during the development of the survey and related questions.

- Defining the survey goals
- Evaluation of available tools
- Deciding how using the data collected would deliver actionable results
- Providing a reliable and valid measurement of data/information management practices and trends

With these demands, the Team strove to design a survey to produce a high response rate, with high quality data. To accomplish this, the Team offered a return on investment to by the respondents by pledging to send them our findings.

The following survey development and deployment goals were accomplished:

- Personalized invitations
- A description of the project's nature
- Explaining the minimal amount of time required to participate
- Emphasized the importance of the respondents' participation in a successful project outcome
- All questions were distilled for simplicity
- Organized the instrument into a logical flow and specifically audience directed
- A consistent ratings scale was established and used throughout the survey

Survey test questions were examined for validity and adherence to generally accepted practices by an industry expert. The survey invitations and reminders were deployed at a time found to yield a large number of responses. The promise of feedback was found to be an effective incentive for eliciting survey participation, particularly among the executive respondent pool. After completion of the survey, the data files were evaluated for accuracy and internal consistency. The Team was prepared to adjust survey estimates to correct for identified errors, however, none were found.

To avoid selection bias, the complete population of CEOs from APTA's available resources was utilized by the Team. This provided representation of different property sizes, geographic regions, and modes operated of transit agencies and businesses. The survey obtained an excellent response rate, with 101 responses out of 230 invitations.

---

[3]Survey questions can be found in Appendix D of this paper.

This 44% response rate is particularly excellent considering the senior positions held by a majority of the respondents.

Using accepted industry standards, the results obtained are considered to adequately represent the views on data management among transit industry executives within North America. Consequently, for the purposes of this exercise, we were able to make several inferences from the results.

Figure A: Transit Agency Fleet Size Comparison

If you are a transit agency what fleet size best describes your organization ?

- Less that 100 Vehicles
- 100 - 249 Vehicles
- 250 - 499 Vehicles
- 500 - 999 Vehicles
- More than 1,000 Vehicles
- Not Applicable
- Not Sure
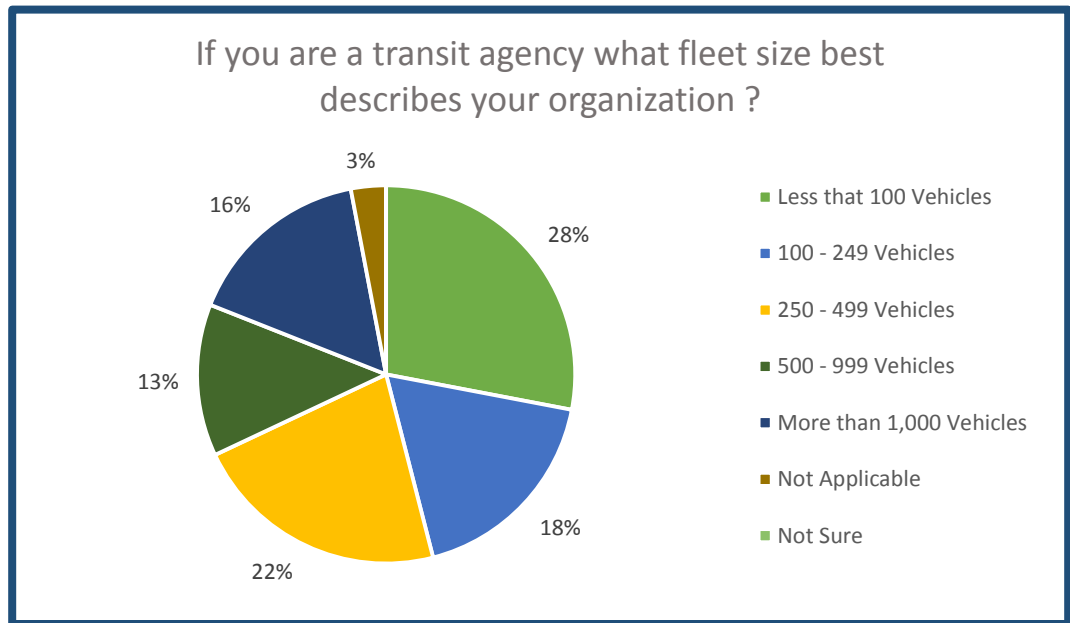
Table 1: Transit Agency Modes Operated

Answer Options	Response Percent	Response Count
Bus	92.1%	93
Light Rail	23.8%	24
Rail	30.7%	31
Bus Rapid Transit	28.7%	29
Paratransit	77.2%	78
Other (please specify)	26.7%	27
Answered Questions - 101	Skipped Questions - 0	

Overall, there was not a statistically significant difference in satisfaction with data/information management practices and policies when comparing property size, geographic region, and modes operated.   In comparison, the ratings for practices and policies were equal. Further, differences in satisfaction due to agency size were not conclusively demonstrated. While sample sizes appear small, the population size must also be considered. With very few agencies with greater than 1,000 vehicles, the number of participating executives from agencies of that size is substantial.

Figure B: Data Management Practices Satisfaction Ratings[4]

How satisfied are you with your organizations overall data management practices?

Rating	Value
1 = Very Dissatisfied	2%
2 = Somewhat Dissatisfied	29%
3 = Neither Satisfied nor Dissatisfied	11%
4 = Somewhat Satisfied	50%
5 = Very Satisfied	8%
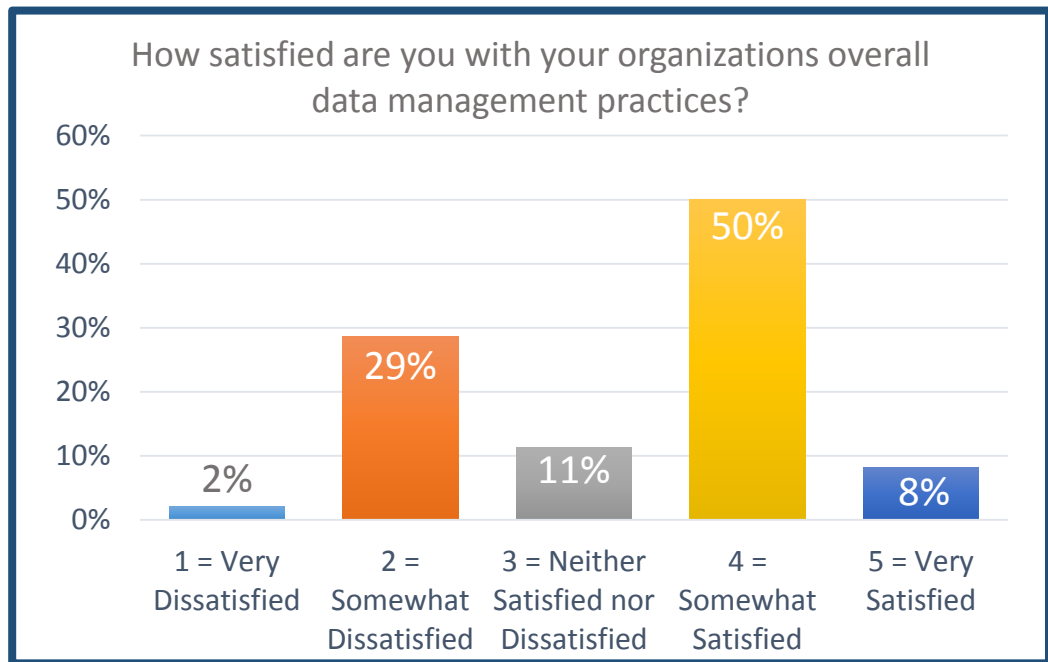
Figure C: Data Management Policies Satisfaction Ratings

How satisfied are you with your organization's data management policy/policies?

Rating	Value
1 = Very Dissatisfied	0%
2 = Somewhat Dissatisfied	25%
3 = Neither Satisfied nor Dissatisfied	16%
4 = Somewhat Satisfied	47%
5 = Very Satisfied	11%
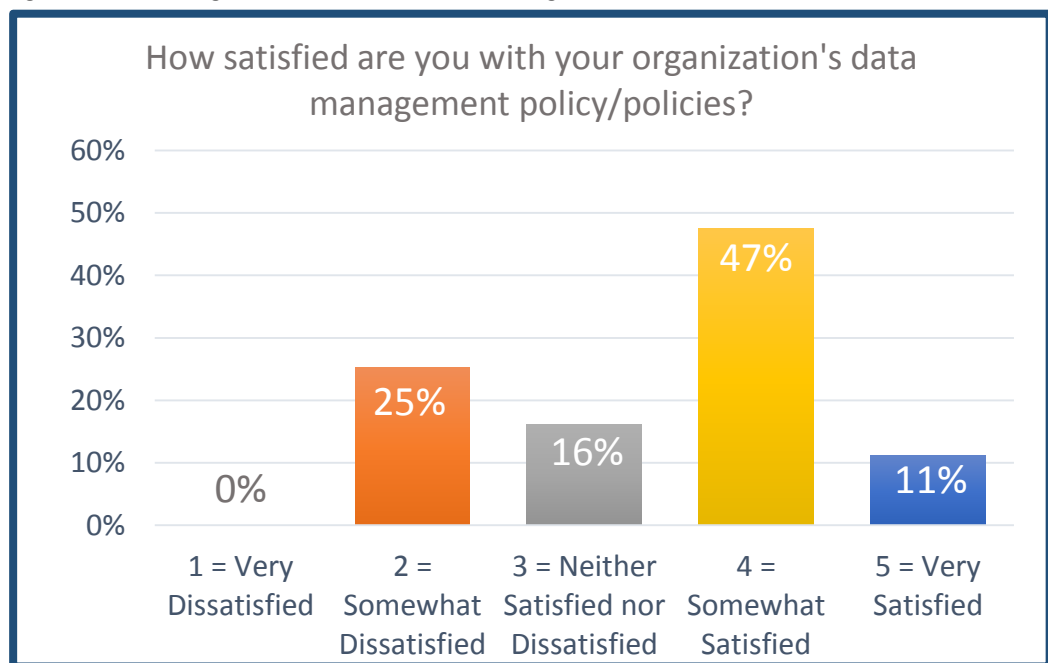
[4]Small sample sizes can affect the precision of a measurement.

Table 2: Data Management Practices and Policies Satisfaction Ratings

	How satisfied are you with your organizations overall data management practices?	How satisfied are you with your organization's data management policy/policies?	Number of Respondents
Less than 100 Vehicles	3.7	3.6	26
100 - 249 Vehicles	3.4	3.1	17
250 - 499 Vehicles	3.5	3.7	22
500 - 999 Vehicles	3.2	3.4	13
More than 1,000 Vehicles	2.8	3.4	16
Ratings Scale: 1 to 5			

Summary of Results

Transit agency leaders surveyed indicated the following:

- The industry is open to innovation, responsive to customers and data providers, and that data is well managed.
- The ability to use or maximize the value of data is evaluated less favorably indicating available data is not used to its maximum benefit within the industry. Ratings point out that, while data is well managed and responsive to stakeholder needs, use of the available data has not yet matured to the extent of maximum value being realized.
- Both data management practices and policies of individual agencies were similarly evaluated with Respondents either expressing satisfaction or dissatisfaction with both. While the majority of respondents indicated some level of satisfaction with current practices and policies, at least 25% of organizations denoted some dissatisfaction, and only 8% indicated that they were very satisfied with their practices and only 11% were very satisfied with their policies. We believe this signifies an opportunity to improve performance within the Industry.

Transit agency technology implementations have primarily focused on core operational needs. However, with the availability of new and innovative technologies, agencies are starting to explore and implement an increasing variety of technologies. As these technologies increase in sophistication, there is an ever increasing volume, velocity, and variety of data created.

Quotes from the Survey Sum up the Results

The following are quotes received from survey respondents and represent a qualifying summary of the survey results.

"The primary challenges with data management in the transit industry are the large amount of data generated, imprecise data definitions and a lack of agreement on what data is useful. It was very difficult to answer this survey as the extent and quality of data management varies greatly across transit agencies. Some do a good job but the industry as a whole has not been very effective in using data to tell its story or to make a point. The challenge is not to collect more data, instead it is how to interpret and summarize the data in an effective way. "

"We, as an industry, typically have done a poor job having data available such that it can be used nimbly and timely in service decisions"

"As a general rule, I find that as an industry we invest significantly in technology but rarely make use of the data generated, especially to support better service quality. We are also very afraid of the data or benchmarking of it with our peers…for all the wrong reasons. "

Best Practices and Strategies	3

Best Practices

Best Practices are defined as: *commercial or professional procedures that are accepted or prescribed as being correct or most effective*. The Team's research resulted in the coalescing of five best practices related to agile data/information management. These best practices are further outlined in this paper through specific supporting strategies. These strategies could be employed by transit agencies as they incorporate an agile data management approach.

Throughout the Team's many agency and vendor interviews and the comments received through the on-line cross sectional survey process it was clear that an industry-wide approach to data management is needed. It was also clear that although all agencies, regardless of size, must address the data tsunami. It was also clear that some agencies are already deploying policies and practices and that best practices are already developing. This paper organizes these developing best practices and outlines the specific strategies in a way that can easily be used to develop a consistent industry wide narrative and approach in dealing with data, the goal of which is to develop an approach that increases customer service abilities, enhances the customer experience, and positively influences operational effectiveness.

The five best practices identified by the Team are: Assess Your Organization; Assign the Right People; Engage Partners; Manage Data and Information; and Maximize Data Value. Each of these best practices has multiple specific supporting strategies. It is important to understand

Assess Your Organization

Assign the Right People

Engage Partners

Manage Data and Information

Maximize Data Value

that our research and by extension, this paper, only begins to define and address the real issue of increasing data requests by customers, agency data storage needs, the complexity of the data, and its true impact on an agency's ability to successfully deal with this growing issue. These best practices and supporting strategies will be useful to all agencies regardless of property size, geographic region, modes operated, or policies and practices already in place.

Assess Your Organization

Assessment is the cornerstone for understanding needs and preparing a useful gap analysis. For most organizations, the speed at which data and technology have been introduced and deployed far outpaces their data management policies and practices. Performing organizational and policy assessments results in benchmarking data that can ultimately be shared and used for decision-making. Four key strategies were identified by the team's research. These strategies are:

Define and Articulate your Vision

Strategy #1: Define and Articulate your Vision

A CEO or General Manager must set the tone within the organization for the importance of data in decision-making and ultimately the areas directly impacting customer satisfaction. Setting the tone is best accomplished with a well-crafted message and informed vision for data management, use, and distribution.

- Have an agile plan for data management and technology within your organization. Convey that plan and get a buy-in throughout the organization.
- Establish the need for data and information and how it will be integrated into your business model.

"Our biggest challenge was changing our culture. We had simply not been in the mode of using our data." (Elizabeth Presutti, General Manager for DART - Des Moines Area Regional Transit Authority)

Assess your Agency & Learn From Others

Strategy #2: Assess your Agency and Learn from Others

Assessing your agency is likely an ongoing endeavor and a goal for some other agencies. In either case the assessment of data management policies and practices is more than likely a new form of assessment for agencies, much like the new technologies they may be deploying. There are three components to this strategy:

1. **Conduct an Organizational Assessment.** Assess the data management related policies and practices. This assessment should look carefully at business solutions, operational benefits, positive customer impact, overall data volumes, storage methodologies, and the value data may or may not have on the organization. The assessment should yield a comprehensive understanding of where your organization is related to data management and how technology choices may have impacted current policies and practices. In some cases, technology choices may have dictated the content and agility of these policies and practices.
2. **Benchmark.** Develop a consistent method of measuring the use and effectiveness of your data/information management approach.
3. **Learn from your Peers.** Share your benchmarking results. Compare the results and learn from the analysis. Many transit agencies participate in forums such as the America Bus Benchmarking Group to share information and experience. There are innovative and agile data management strategies in use as a direct result of both failed and successful technology deployments. Agencies must adjust their

organizational data management strategies based on the successes and failures of their peers and the benchmarking comparative analysis.

Some agencies pride themselves on their level of innovation and technical expertise. *"If not on the bleeding edge, we're on the leading edge."* (Brian Lamb, Metro Transit). Other agencies acknowledge a more conservative approach, driven by resource constraints. *"Sometimes being on the cutting edge isn't where you want to be"* (William Steele, Manatee County Area Transit). Nonetheless, even in a resource-constrained environment, data management is critical due to the need *"To generate the required information in a more automated fashion"* and to *"Get the computers to do as much of the work as possible to eliminate the possibility for human errors"* (Jim Wilson, Blue Water Area Transit). This critical step of self-assessment will help agencies valuate and set their goals in data management.

Let Business Needs Drive Solutions

Strategy #3: Let Business Needs Drive Solutions

Public Transit is a business. It is a product consumed by a customer and as such every agency must focus on the business needs of the organization. Every aspect of the business of moving people is now touched by technology. In order for transit agencies to let business needs guide them in their technology solutions, they must ask these questions:

1. Does our data management approach, policies and practices support our business goals?
2. Does our data provide a true picture of our agency's performance and can these performance indicators be more accurate, timelier, and more useful in business and operational decision making?
3. Does our technology currently deployed support our long-term business, operational, and customer service goals? The evaluation of technology should always include analysis on the impact the technology will have on data management.

Assess and Develop a Data Policy

Strategy #4: Assess and Develop a Data Policy

Key to a successful approach to data/information management is a policy that responds to the organization. Data storage, dissemination, and use both internally and externally are all key considerations. Assessment of existing policy should focus not only on the legal requirements for collection and protection of data, but also on how data may be effectively disseminated to the customers. Development of any data management related policies must include elements that allow it to be responsive to the trend [in the transit industry] to incorporate more data and information into all aspects of the business.

Assign the Right People

The second major category of strategies involves internal management of the people involved with data management. Regardless of the size of the transit agency, it is important for transit agency leaders to adopt a proactive and agile approach to organizing a team to deal with data management.

Engage CIO/IT Managers you can Trust

Strategy #5: Engage CIO/IT Managers you can Trust

For effective data management, it is important for transit agency CEOs to designate a single internal lead and to set appropriate expectations for such a role.

The need for a single internal Data Management Lead (often called a Chief Information Officer (CIO) or IT Manager) is to eliminate any internal confusion on who holds ultimate responsibility for implementation and maintenance of various data management systems.   This single lead is also accountable on the data management strategy - keeping the organization up-to-date and adaptive to new developments in technology, information, and data management while ensuring that data management infrastructure effectively serves other functional groups within transit agencies.

The reporting relationship of the IT Manager to the CEO varies. Some IT Managers do report directly to CEOs, while others report through another departments such as Planning or Finance Department or Administration. In smaller agencies, a dedicated IT function may not exist and the responsibility may rest with departments such as Procurement (as is the case at a small agency such as Blue Water Transit), Planning, or Finance. Regardless of the official organizational structure, it is important for the CEO to have a working relationship with the Data Management Lead (IT Manager) that "*enables [transit agencies] to bring the appropriate level of technology to the organization*" (Steve Banta, Valley Metro) and to establish an environment where *"Senior staff is not afraid to bring ideas to you"* (Darrel Johnson, Orange County Transportation Authority (OCTA)).

Because requirements of data management can be complex and difficult to understand, Transit CEOs and vendors alike cited the need for the Data Management Lead to be ***an effective communicator***. The Data Management Lead must be able to convey complex concepts clearly, so that agency decision-makers can understand the purpose and the general parameters of a particular solution. One vendor cited an interesting test: *"If you can't draw it [the concept] on a piece of paper, then maybe you're the wrong person for the job."*

Strategy #6: Assign Data Managers in Functional Groups

Beyond the Data Management Lead, successful implementation of a data management solution requires a concerted effort throughout the transit agency. While the CEO needs to set a tone for the *"Importance of data for the agency"* (Mark Kroncke, Invoke Technologies), one of the primary ways of doing so is to assign the mission of effective data management to designated staff members, or "*Data Geeks*" (Mary Sue O'Melia, TransTrack), in various functional groups throughout the agency. This ensures that any integrated data solution incorporates data generated from various sources, but also responds to the various needs of each functional group. In resource-constrained agencies or in smaller agencies, these "Data Geeks" need not be exclusively assigned to the task of data management, but may have data management as one of a few key tasks assigned to them. In smaller agencies, this role may be filled by a general analyst or even a planning or operational manager.
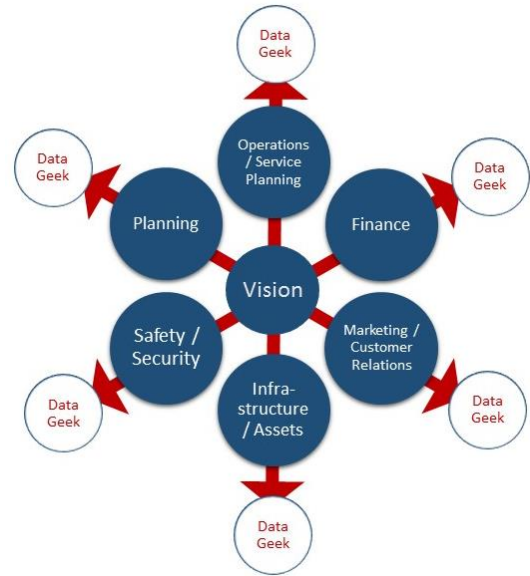
Strategy #7: Encourage Collaboration

It is important to empower all staff involved to collaborate and work together. It is also important that this collaboration be supported by lines of communication throughout the agency, but two levels are critical.

First, "Data Geeks" need to have a regular forum for interaction. This can take the form of a working group with regular meetings [at a minimum] and can be supplemented by any type of additional collaboration method or tool.

Second, there needs to be a forum for department heads to collaborate and collectively vet potential solutions, especially since technology and data management solutions inevitably involve and affect multiple functions and departments. For example, OCTA has a Technology Review Committee, comprised of various department heads. This committee reviews all new technology proposals and monitors the progress of implementation of new technology. This

committee structure allows the perspectives of each constituent group within the organization to be incorporated into the decision-making process. It also supports discussions about integration and relationships between systems and business processes. Metro Transit in Minneapolis has established a Strategic Initiatives Division to foster *"cross-functional innovation and coordination"*.

Engage Partners

In the realm of data/information management, there are many potential partners to assist a transit agency. It is very common that the team to support data management extends beyond the transit agency itself.

Decide how much Contracting is right for you

### Strategy #8: Decide how much Contracting is right for you

There is no universal model for an appropriate level of contracting and engagement of private sector partners in the pursuit of effective data management. Some agencies, especially larger multi-modal agencies like Tri-Met in Portland, are able to devote extensive internal staff resources to data management and perform many functions "in-house". These functions may include management and upkeep of a data warehouse, outreach to app developers, and data analytics. Other agencies, especially smaller agencies, opt to contract many of these functions to vendors and service providers.

The ability to contract is increasing due to several factors. First, new technologies, such as the development of "the cloud" provide an environment for data to be stored, managed, processed, analyzed, and secured off-site without significant capital investment in data servers and commitments to internal staff. Second, the market place for service providers in the realm of data management is growing and many functions that were once only performed directly by transit agency staff, such as data analytics and reporting, data integration, and data storage, now have products and services available in the commercial marketplace. The marketplace is quite dynamic with consolidations and mergers and new entrants to the market place emerging over time.

Given this environment, it is important to note that wherever a transit agency falls on the spectrum of contracting it is a choice that is specific to that transit agency. In addition, it could vary by the specific function or solution considered. There is no
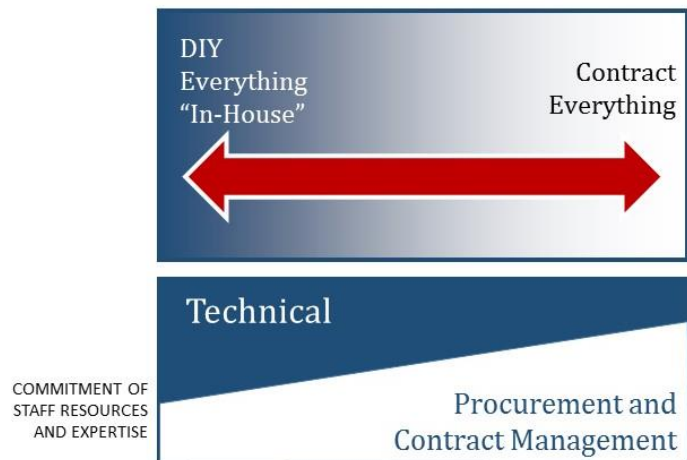
evidence that there is an "optimal" level of contracting. The decision on the extent of contracting is, however, somewhat contingent on an assessment of whether the transit agency can make a commitment to developing and retaining internal technical staff resources required to support a particular solution. Transit agencies that can make that commitment may contract less, agencies that do not have those technical resources available to them or that are resource-constrained, may contract more functions. The decision to contract more does require that the transit agency place a heavier emphasis on both procurement and contract management resources - both in budget and in expertise. It is often the case that the decision-maker in small agencies, such as at Blue Water Transit in Port Huron, MI, related to data management solutions is the Procurement Officer.

The point of engagement of vendors is the key point in time that highlights the importance of the earlier-described strategies of Defining and Articulating Your Vision (Strategy #1) and Letting the Business Needs Drive Solutions (Strategy #3). A key question when considering any solution is "*Does it meet the needs of the business?*" (Andy Byford, CEO of Toronto Transit Commission (TTC)) Having clarity on the particular business needs will help transit agencies navigate the complex world of data management service and product offerings. Furthermore, transit agencies need to have an *"Understanding between what you must have versus what's nice to have"* (Darrell Johnson, OCTA). This understanding will provide the focus necessary in entering the procurement process.

Strategy #9: Promote Positive Procurement Partnerships

Promote Positive Procurement Partnerships

For many transit agencies, data management is very closely linked to the technology that generates it. The data strategy is thus very strongly driven by the procurement and implementation of the technology solution. "*Successful data management thus depends on positive vendor relationships, which foster "symbiotic" benefits to transit agencies and vendors alike".* (Andy Byford, TTC - Toronto Transit Commission). Otherwise, transit agencies face the risk of "getting a bad name in the industry" and having vendors shy away from providing it with solutions. There are many ways to optimize the procurement and contracting relationship from beginning to end.

Industry Outreach

Even with a clear definition of a transit agency's vision in data management, it is important to understand how the marketplace is positioned to serve the needs of transit agency. Many transit agencies endorse industry outreach activities such as Requests for Information (RFIs) or pre-proposal meetings with vendors. For example, Metro Transit (Minneapolis, MN) used a sequential RFI-RFP process for its Real-Time Information Project that involved invitations to industry experts to give a vision of how the agency can move in a forward direction. These insights were incorporated into RFPs. Vendors also found pre-proposal information exchange useful for understanding the needs of transit agencies and the level of understanding that transit agencies had about particular solutions.

Some vendors suggested that innovation in industry outreach might even be useful. For example, if an agency were willing and able to operate a "test-bed" approach where multiple vendors or developers were engaged to develop solutions on a demonstration basis, it would allow for an agency to clarify its needs and see more concretely what new functionality is desired. For example, with some data solutions and technologies, *"You can't know exactly [what you want] until you see and touch it completely"* (anonymous vendor).

Clarity in Statement of Work

It is also important to have greater clarity in the vendor's statement of work. The type of clarity warranted is more on the performance of the system, rather than in the specific details of a technology. *"Focus on what you're trying to achieve as the end state."* (David Pickeral, IBM Smarter Cities) Emphasizing performance specifications allows multiple vendors to propose varied solutions to meet the needs of the contract. Furthermore, clarity in how the transit agency will assert its ownership of the data and will adhere to open standards and interfaces is also important to make sure that the data can be integrated with other agency data and also to provide assurances that agencies are *"Not trying to undermine the vendor"* (William Steele, Manatee County Area Transit)

Contract Structure

Different types of engagements can benefit from structuring the contract to meet the needs of the initiative.

For initiatives that are well-defined, agencies both large (such as OCTA) and small (such as Blue Water Area Transit) provided examples of how fixed-price contracts are used to maximize accountability of vendors. Given the technical complexity of data management, contracts can be difficult to manage if they are open-ended. For a contract related to maintenance of servers, Blue Water Area Transit used fixed price contracts, which *"Eliminated the need to understand Greek"* and *"Having to explain bizarre changes to [its] Commission."* OCTA implemented a performance reporting system using a firm fixed-price agreement with milestone payments. The contract emphasized delivery and budget in order to hold the contractor accountable. Interestingly, the contract focused less on schedule, which gave the contractor the incentive not to rush a solution, but to meet the needs of the contract.

Vendor/Contractor Engagement

Some technology/data management initiatives benefit from regular, frequent interchange and contact with the contract managers. Vendors find a high level of engagement useful *"To make sure we're meeting the spirit of the requirements"* with any technology project (Kirk Shore, Clever Devices).

Developer Solicitation with Release of Open Data

Some agencies are using open data standards to promote innovation, particularly among app developers. Agencies such as TriMet (Portland) and the Metro (Los Angeles County Metropolitan Transportation Authority) release certain types of data openly

and encourage developers to develop applications (apps) that can be made publicly available to a wide consumer audience.

Limitations

Admittedly, there are limitations to how far the public procurement process can be stretched. Business planning, budgeting, and public procurement all take a long time. Furthermore, many jurisdictions and the Federal Transit Administration have procurement rules which define contractor/vendor engagement. Many of these rules may not be specifically tailored to the pace of innovation within the data management sector. Specifically, it may be difficult to entertain unsolicited proposals or sole source engagements. Transit agencies and vendors alike expressed interest in exploring how the procurement process can be enhanced to support the specific needs of data management.

Manage Data and Information

With the discussion of managing data/information it is common for the term, Master Data Management (MDM) to come into the conversation. This is a system comprised of the processes, governance, policies, standards, and tools that consistently define and manage the critical data of an organization to provide a single point of reference. Also common with this discussion are tools that can be used to support data management by removing duplicates, standardizing data, and incorporating rules to eliminate incorrect data from entering the system in order to create a "single source of truth".

The tools include software packages but at the core are data networks, file systems, data warehouse, data marts, data mining, data analysis, and data visualization. At a base level the goal is to ensure that an organization does not use multiple and potentially inconsistent versions of the same data in different parts of its operations, which can occur in all sizes of organizations.

Before we delve into the creation of management processes, policies, standards and tools it is critical to have some strategies for managing data/information.

Establish Data Management Protocols

Strategy #10: Establish Data Management Protocols
Normalize the data

Data normalization is the process of reducing data to its recognized form. As an example, database normalization is the process of organizing fields and tables of a relational database to minimize redundancy and dependency. Data dependency is a situation in which a program statement refers to the data of a preceding statement. Normalization usually involves dividing large tables into smaller tables and defining relationships between them. The objective is to isolate data so that additions, deletions, and modifications of a field can be made in just one table and then propagated through the rest of the database using the defined relationships**.**

Validate and Clean Data

Data validation is the process of ensuring that a program operates on clean, correct, and useful data. A simple example of data validation is spell checking a document or checking to see that thelephone numbers follow a specified format.

Data validation refers to routines, often called "validation rules" or "check routines", that check for correctness, meaningfulness, and security of data that are input to systems. The rules may be implemented through the automated facilities of a data dictionary, or by the addition of explicit application logic.

Data validation is intended to provide certain well-defined guarantees for accuracy, and consistency. Data validation rules can be defined and designed using any of various methodologies, and be deployed in any of various contexts.

Data validation rules may be defined, designed, and deployed as part of the requirements-gathering phase in a software engineering specification.

Data validation and the related methodologies can become very detailed but the important factor is to have data validation rules, processes, and practices in place. Then the data should be routinely checked, tested, and updated.

Data Protection

Data protection encompasses an entire area that should be directed from a policy level. Protection should cover security, information privacy, data recovery, file encryption, retention, and decisions with contracting file storage. The protection of data was a special focus of many transit agency CEOs who noted recent breaches of consumer data in the retail industry. Two areas of data that were cited as especially important to protect were data related to the identity of customers – personal identity and financial account information.

Automate

Automation is the creation of automatic systems with an emphasis on minimizing or reducing human intervention. The major benefit of automation is that it saves labor and increases overall quality through improvements to accuracy. This is true for small and large agencies alike, especially when there are staffing and resource constraints.

Strategy #11: Establish an Enterprise Database

The process of working with data from disparate systems can make it difficult to build and maintain interfaces. This in turn makes it difficult to makes changes in software or hardware systems.

Why Enterprise Data Warehouse? An Enterprise Data Warehouse (EDW) is a system used for integrating data from one or more disparate sources to create a centralized repository of data for storage and use. Data warehouses store current and historical data and are used to create a "single source of truth" for an agency.

The creation of an EDW with standardized software tools and methodologies should be an early priority. Data will be copied from source tables to the EDW sometimes referred to as the target tables. This extract-transform-load (ETL) process uses staging, data integration, and access layers to house key functions. The staging layer, or staging database, stores raw data extracted from each of the disparate source data systems. The integration layer integrates the disparate data sets by transforming the data from the staging layer often storing this transformed data in an operational data store (ODS) database. The integrated data is then moved to yet another database, often called the data warehouse, where the data is arranged into hierarchical groups often called dimensions and into facts and or aggregate facts. The combination of facts and dimensions is often referred to as a star schema. In this description of the enterprise data warehouse we are primarily focused on data storage. The EDW is now the main source of the data, which is cleaned, transformed, cataloged and made available to selected users within the agency. Other benefits achieved with the creation of a single EDW are:

- Improved data quality, by providing consistent codes and descriptions
- Present the agency's information consistently
- Providing a common data model for all data of interest regardless of the data's source
- Restructure the data so that it makes sense to the business users
- Improvement in query performance
- Decision support
- Data security - data to be protected (personal identity and financial information) can be filtered and not brought into the EDW

Other Considerations

Enterprise databases require routine backup and maintenance with regular replication of environments as needed for tests and development processes. Managing such large databases, can add risk and cost to agencies and should be compared to other options
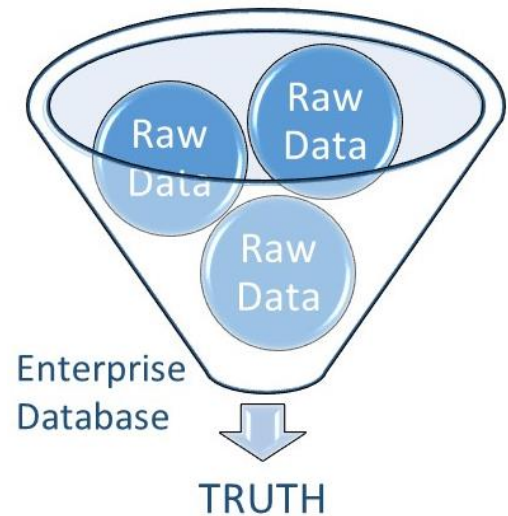
with an eye towards long-term sustainability. With an EDW, an Enterprise Data Management strategy will help reduce the burden on the agency to not only retain data but also maintain accessibility. Enterprise Database Management is a strategy to manage growth in any organization. Enterprise Database Management systems will require agencies to evaluate their data management strategies enabling them to manage data growth. Agencies can maintain access through the native enterprise application layer or other business intelligence tools to ensure accessibility.

As we have discussed, data growth is exponential in any organization, regardless of its size. Enterprise Data Warehouse and good Enterprise Data Management gives agencies the ability to stay ahead of data growth and achieve higher overall performance.

Maximize Data Value

The promise of maximizing data value is key to the success and acceptance of new technologies and often plays a significant role in the adoption of technology in an agency. The issue that this approach creates is that that evaluation and adoption often comes with no consideration of the agency's ability to capitalize on that promise or the impact data management policy or practices.

The strategies outlined for this Best Practice focus on the use of data and how the value of data can be maximized through agile policy and practices.

To Improve Performance Turn Data into Information

Strategy #12: To Improve Performance, Turn Data into Information

Data in its raw form is often difficult to digest or separate signals from noise.   To make decisions based on information that drive transit agencies they must understand the data and its context. Data consists of raw facts and figures. When that data is processed into sets according to context, it provides information. An example of this conversion would be recording the temperature continuously over a set period, this is data collection. From the data collected, information may be derived, such as the highest, lowest, and average temperatures over that period.

Andy Byford, TTC says, *"You are under achieving if you're not using your data for decision making."*

So the question then becomes, what value exists within our data and how can we maximize it? The primary way that transit agencies demonstrate turning data into information is regular reporting of system data and performance measurements. Often, this involves the use of performance dashboards. Transit agencies are also increasingly making these data available to their governing boards and to the public to enhance accountability and transparency.

Strategy #13: Look for Spin-off Opportunities

After an agency has decided on what data to collect, validated it, created methods and taxonomies for storage, and used it for making decisions; what other opportunities, both internal and external, may exist for its use? For example, San Diego Metropolitan Transit System (MTS) uses a non-formalized process were they regularly look for other opportunities to use data for other than originally intended uses. This practice has created a collaborative environment between areas of the company that improve overall agency performance.

Strategy #14: Consider Revenue Opportunities

The Team did not find many examples of monetizing the data collected and many transit agency CEOs acknowledged that seeking revenue from data is not one of their primary goals and that the trend appears to be in the opposite direction-making transit data available free of charge. However, this is an emerging area of interest across the industry. This is an area of interest shared by agencies, vendors, and consultants and warrants future investigation. One thing that everyone agrees on is that the data collected has value and revenue opportunities should be studied and regularly considered.

Strategy #15: Share Data

The data really does not have value if it is not being shared internally within the agency and within the transit industry. Data sharing is the practice of making data freely available to others. Although this "Open Source" approach seems new data sharing has a long successful history and consistently proves to be valuable to agencies. Data sharing, especially fostered by open and organic communities of professionals and data users, has demonstrated remarkable success in the transit industry through the widespread and worldwide use of the General Transit Feed Specification (GTFS) to share transit route, stop, and schedule information, both static and now, real-time.

A tradeoff in standardizing and openly releasing data should be considered. On one hand, it is the perspective of many that releasing all data that is not deemed to require protection is seen as promoting information innovation and opens up marketplaces to new products and service offerings. *"As an industry, we would never be where we are now if we had such a tight architecture."* (Brian Lamb, Metro Transit). On the other hand, for some types of data, a transit agency may face a significant level of effort or be charged a cost by a vendor to make the data releasable. In these cases, a transit agency must evaluate whether there is a strong business case (e.g., a large enough market of data users or a large enough consumer benefit) to release the data. Otherwise, it may *"spend a lot of money for no good reason."* (William Robert, Spy Pond Partners)

Case studies showed sharing data externally is the growing trend and is believed to be the first step in being able to monetize in the near future.

Case Studies

In an effort to develop an understanding of real world applications for data/information management and innovation the Team conducted a series of interviews with top-level managers at over a twenty transit agencies. A cross-section of agencies were chosen for interviews spanning across the US and interviewing large and medium sized agencies. The Team's goal was to gain an understanding of the challenges faced by transit agencies and to identify how agencies have leveraged data in innovative ways, to improve customer service, improve business functions, and stay abreast of technology. For this paper, three agencies will be profiled, that in our opinion, illustrate a variety of approaches to dealing with the data tsunami and best encapsulate the best practices and lessons learned. These case studies are by no means comprehensive, but illustrate a broad cross section of the industry and the various ways transit agencies may approach data management.

Additionally, while we used specific examples from Oregon, Texas, and Utah, this section will also discuss how other agencies have approached data management in similar ways.

TriMet - Portland, Oregon

Collaborative Approach

The Team's first case study is focused on TriMet in Portland Oregon. The agency provides bus, light rail, and commuter rail service to the Portland region in Oregon. General Manager, Keith McFarlane, a Deputy General Manager, and seven Executive Directors lead TriMet's executive leadership team. A key characteristic of TriMet's leadership is collaboration and accountability and this is illustrated in the executive leadership team's approach to managing data/information. The agency's vision statement is "To do our part in making our community the best place to live in the country" and the mission statement is "To provide valued transit service that is safe, dependable and easy to use".

Leadership is Comfortable Being an Innovator

Management teams generally fall within one of several camps when it came to leveraging and embracing the use of technology. Innovators are characterized as being on the 'bleeding edge' of technology. Some agencies may be an early adopter of technology and share similarities with innovators like TriMet. Other agencies, however, may be more comfortable adopting new technologies once they have been tested.

For TriMet, when it comes to managing data, the executive leadership team has no problem being an early adopter and a leader in the industry in developing new technologies that advance the agency's mission and vision. Examples of TriMet's innovative strategies are explained in more detail later in this section. However, these strategies align with the agency's mission by enhancing customer service and providing accountability to the public.

Collaborative Process

A second key characteristic of TriMet as an innovator in data management is their leadership's dedication to collaboration. Collaboration involves both internal stakeholders and active representation by all seven divisions, but also by reaching out to other industries to develop tools that can benefit the transit industry specifically. An example of this is TriMet's collaboration with Google© to develop the General Transit Feed Specification (GTFS) which developed as a way of importing transit schedules into google maps. This innovation was a game changer in the transit industry and has allowed many transit agencies across the world to include their schedules in google maps. TriMet also welcomes app developers and others in the information technology sector to develop tools that meet the demands of today's mobile technology and the next wave in technology that is not entirely known or defined today.

Collaborative Approach

TriMet also prides itself on being 'first' to do something. A recent tweet touts the agency's one year anniversary of the launch of its ticket app for IPhone© and Android© and goes on to thank passengers for making it such a success. This is yet another example of inclusion and collaboration that celebrates the development and use of technology across all divisions of the agency – from the developer side, to the operations/fare collection side to marketing and customer service.

General Transit Feed Specification Development

Bibiana McHugh, an IT Manager at TriMet, first conceived of the idea of incorporating transit schedules in trip planners like MapQuest© or Google Maps© to provide an easy tool to plan a trip by transit as easily as one by car. In 2005, McHugh sent inquiries to

MapQuest©, Yahoo!©, and Google©, asking if TriMet could partner with any of them in developing a transit component in their trip planning services.

A software engineer at Google©, Chris Harrelson, responded. Harrelson had been using his 20 percent time allotment for outside projects to include transit data with Google Maps© and the Google Transit Trip Planner© was born.

In developing the tool, TriMet worked with Google© to prepare TriMet's data set in a format that would work for Google Maps©. The challenge in developing the tool was the need to include a temporal and spatial element, which requires a relational database in order to manage all of the information. The resulting format is a collection of CSV files that model the agency's transit schedules. The resulting feed can be used to develop other useful applications such as service analysis and real time information.

Lessons Learned

Partnering with outside private industry can be a game changer.

GTFS has been a game changer within the industry and has met one of the most common demands from transit customers – an easy to use interactive trip planner for transit.

Another lesson learned is that providing open source data can be beneficial. This has been something many transit agencies have been reluctant to offer, due to concerns over accuracy of the data or that it might be used to misrepresent the agency in some way. TriMet leadership found the opposite to be true, that by offering the data free of charge, increased openness and even trust was developed between the agency and the community served. Other agencies questioned providing the data free of charge when they may have monetized their data. TriMet officials explained that the taxpayers paid for the data and the agency was somewhat obligated to provide it to them in an easily accessible and useful format. However, not all data generated by an agency should be free and open, as will be discussed in the case of Salt Lake City later in the paper.

DART - Dallas, Texas

Business Analytics

Dallas Area Rapid Transit (DART) serves the greater Dallas region covering approximately 700 square miles and 13 surrounding cities. Services provided include Bus, Light Rail, Commuter Rail, Van Pool, and Paratransit. President and Executive Director Gary Thomas who has been at the agency since 1998 and employs approximately 3700 employees leads the executive team. Other members of the executive team are comprised of a Deputy Executive Director and three Executive Vice Presidents overseeing Finance, Operations, and Growth and Regional Development. Under Thomas' leadership, DART has doubled its light rail system, becoming the longest in the United States at 85 miles.

Business Goals Drive Innovation

DART's approach to the use of data technology is focused on improving efficiencies and digesting massive amounts of data into easily understood formats. In interviews with Gary Thomas, he described his approach to the use and management of data as a business tool. Business goals drive the use and management of data, with an emphasis on customer satisfaction and understanding of the health/well-being of the agency from a business perspective.

It was apparent during the interview with Mr. Thomas, that what separated his approach from other executive leaders in the industry was his desire to treat DART as a business. He specifically mentioned taking a page from from Steve Jobs'(Apple©) playbook. Jobs received daily revenue reports and knew from one day to the next exactly how Apple© was performing, at least in one or two critical buisness metrics.

Business Analytics

Thomas pointed out that across much of the transit industry, ridership summaries, financial reports, and other dashboard type data is reported on a monthly basis with current reports lagging behind a month or more. In mimicking Jobs, Thomas created weekly reports of raw data that would show how many passes were sold for light rail, the number of boardings on its bus system, number of customer service complaints, and other data that could provide a snapshot of the health of the agency from week to week.

Thomas also uses similar reports in order to monitor the agency's supply chain and inventory. In this way he is able to monitor various metrics related to contractors, open contracts, late deliveries, and types of purchases. For example, an increased use of purchasing cards by staff may indicate an issue with procurement. Another key issue for any agency is the cost of inventory on hand. DART has nearly $30 million dollars in existing inventory, a significant investment for any agency.

DART also compares this data to the same period in the previous year. By having timely reports, the transit leadership is able to easily spot trends or issues needing to be addressed quickly and proactively rather than reacting months after the fact.

Leadership is an Early Adopter of Technology

DART can be considered an early adopter of innovation. Like TriMet, DART also developed its own ticketing app (GoPass) by contracting with an app developer. Because, DART paid for the development of the app, they are able to avoid some of the fees other agencies may pay for transaction fees and other costs of using a third party's technology. DART also receives a small percentage (1% to 2%) for each transaction made with the app.

The GoPass app is also shared among other transit agencies through agreements between DART and the participating agency. Agencies that wish to be a participating agency with Go App are charged a fee that is scaled to the size of the agency. DART maintains the app at this time, while the vendor/developer maintains the back of house for the operation. According to DART this is the only multi-agency electronic pass in the United States.

Lessons Learned
Use Data to Support Business Goals
Technology and data are used in various ways at DART and leveraged in order to meet business goals in terms of key business metrics such as improved customer service, use of analytics to drive business decisions, and to manage supply chain and improve inventory management.

Lessons Learned

UTA - Salt Lake City, Utah
Market Driven Technology
The Utah Transit Authority (UTA) is headquartered in Salt Lake City with a service area of 1,600 square miles of the Wasatch Front in Utah and includes the metropolitan areas of Ogden, Park City, Provo, Salt Lake City and Tooele. UTA operates fixed route buses, express buses, ski buses, three light rail lines (TRAX), a streetcar line (the S-Line), and a commuter rail train (the FrontRunner) from Ogden through Salt Lake City to Provo.

The authority is a local district political subdivision of the State of Utah. UTA is governed by a 16-member Board of Trustees primarily made up of appointees of elected officials and a five-member board made up of local representatives of the local jurisdictions included in the UTA service area.

The agency is led by Michael Allegra, General Manager who oversees approximately 2,000 employees. The mission statement of UTA is to "Strengthen and connect communities enabling individuals to pursue a fuller life with greater ease and convenience by leading through partnering, planning, and wise investment of physical, economic, and human resources."

Market Driven Technology

Leadership is Customer Focused
Allegra believes that his most important priorities as a leader are engagement, encouragement, and enthusiasm. This translates to a customer driven agency that sees technology and data management as a business tool focused on improving customer convenience. A key feature of UTAs approach to the use of data and technology is providing wide-open access to third party developers, in order to increase customer satisfaction and convenience. Additionally, data management

projects are focused on three distinct systems - those that increase customer value, those that improve business management and improve efficiencies, and those that improve safety/security.

Customer Demand Drives Innovation

In recent years, UTA has seen year over year ridership increases, with over 44 million trips in 2013. Realizing that over 70% of UTA's riders have access to a vehicle for transportation, management knows that in order to keep increasing ridership, understanding customer demand and changing demographics are key. Allegra's team has achieved this by introducing contactless fare cards, electronic pass cards that can be reloaded and managed online, and an active participant in joint developments around station areas to boost ridership.

Leveraged Choice Riders

UTA has also leveraged choice riders through customer driven efforts that make using transit as convenient as using other forms of transportation. Examples include a residential pass for City residents paid for annually through utility bills, prepaid cards with discounts for their 'unbanked' customers, and electronic far collection.

Lessons Learned

**Lessons
Learned**

Customer-driven Approach Increases Choice Ridership

UTA's approach to increasing ridership has been accomplished through embracing innovations that leverage technologies in response to customer demands. Management also embraces open data as a way of fostering the development of new tools and technology through third parties. Management also encourages third party developers by providing open access to UTA data and in turn publishing links to these applications and tools in their "App Center" which is placed prominently on their home page.

Conclusion

The above case studies are just a few examples of how many agencies approach data management. The above case studies were chosen because they illustrate three common approaches to data management and use of technology to improve efficiencies, meet customer demand, and collaborating with other industries to solve a common problem.

Recommendations for Data Management Policy Development	5

Recommendations

The following are the resulting recommendations for Public Transit Agencies related to Agile Data/Information Management Strategies and policy development. First, it is important to understand the following in order to expedite and simplify the process:

- You are not alone - use Peer Agencies as a resource
- Agencies must be "Deliberately Innovative" by focusing on the use of data to meet business goals and the agency's vision
- Apply Best Practices in Data/Information Management

We are generating more data than thought possible a generation ago.

Policy Goals

A policy should articulate general goals for data management to help inspire internal staff to incorporate positive data management techniques into their day-to-day work and to guide how data is managed over time, even as technology, new data sources, and uses of data evolve. The policy goals can address:

1. A statement about how the use of data shall serve the strategic objectives of the agency and be incorporated into its business practices
2. Assignment of responsibility for data management
3. Descriptions of which business units are responsible for which agency objectives and the relationship of IT Managers to support those business units
4. Goals for collaboration and transparency
5. Guiding principles for the sharing of data, both internal to the agency, external to other peer agencies, and to the public and vendors and developers

Agile Internal Data Management Policy Elements

The following are recommended policy elements for effective and agile data management policy development.

1. Increased convenience of data is correlated to increase in choice riders
2. Annual or regular audit of data to ensure accuracy
3. Auditing to ensuring process and practices are followed
4. Training of new employees
5. Recurrent employee training
6. Top down engagement

Agile Data Dissemination Policy Elements

All agencies receive requests for data, from other government agencies and the public. Some items to consider when developing policies for disseminating data to external customers are as follows:

1. Determine the agency's approach to data management in terms of how data will be shared. Identify which types of data should be freely available and which types of data should be protected (i.e. schedule data versus private customer data). Data which is high priority for protection (or separation from publicly available data sets) or data of a personal nature (related either to the personal identity or financial account information of customers).

2. Develop parameters for use of data by third parties, research, and understand the legal implications. Craft the necessary disclaimers the transit agency may wish to include with the data and Terms of Use that clarify what responsibilities any user of data conforms to and rights the transit agency reserves for itself.

3. Understand how the Freedom of Information Act (FOIA) will impact your data dissemination policies. Be prepared to provide data and other forms of agency information through the Freedom of Information Act (FOIA).

Agile Technology Collaboration Policy Elements

Collaborating with disparate industries has proven to be an effective way of developing leveraging emerging technologies and applications tailored to the unique needs of transit. Some policy elements to be considered include:

1. Use alternative procurement methods such as Letter of Interest, or Unsolicited Proposals.

2. Keep technology and software Request for Proposals (RFPs) and proposals requirements simple and open to allow for increased creativity in developing solutions by vendors. Focus on the problem to be solved rather than the specific technological requirement.

3. Use performance standards and measures in project solicitations to encourage variety in developing technological solutions.

4. Consider developing a test bed or proof of concept to increase competition and a greater variety of innovation in the use to data and information management

If your agency can employ these strategies, you will soon be Surfing the Data Tsunami!

Appendix A: Agency Interview Participants

A

Representative Name, Title	Transit Agency	City, State
1. Steve Banta, Chief Executive Officer	Valley Metro	Phoenix, AZ
2. Andy Byford, Chief Executive Officer	Toronto Transit Commission	Toronto, ON, Canada
3. Darrell Johnson, Chief Executive Officer	Orange County Transportation Authority	Orange, CA
4. Brian Lamb, Chief Executive Officer	Metro Transit	Minneapolis, MN
5. Bibiana McHugh, IT Manager of GIS Neil McFarlane, General Manager	TriMet	Portland, OR
6. William Steele, Transit Division Manager	Manatee County Area Transit	Bradenton, FL
7. Jim Wilson, General Manager David Frasier, Director of Procurement	Blue Water Area Transit	Port Huron, MI
8. Elizabeth Presutti, General Manager	Des Moines Area Regional Transit Authority (DART)	Des Moines, IA
9. Brad Miller, Chief Executive Officer	Pinellas Sun Coast Transit Authority (PSTA)	Tampa, FL
10. Gary Thomas, Chief Executive Officer	Dallas Area Rapid Transit (DART)	Dallas, TX
11. Michael Allegra, General Manager	Utah Transit Authority (UTA)	Salt Lake City, UT
12. Jason Weiss, Chief Information Officer	Sound Transit	Seattle, WA
13. Kenneth McDonald, President/CEO	Long Beach Transit	Long Beach, CA
14. Doran J. Barnes, Executive Director	Foothill Transit	Montebello, CA
15. Carolyn Flowers, Chief Executive Officer	Charlotte Area Transit System (CATS)	Charlotte, NC
16. John M. Lewis, Chief Executive Officer	Central Florida Regional Transportation Authority (LYNX)	Orlando, FL
17. Phillip Washington, General Manager	Regional Transit District (RTD)	Denver, CO
18. Kenneth Herrscher, Business Intelligence Architect	MetroLinkt	Los Angeles, CA
19. Nuria Fernandez, General Manager	Valley Transportation Authority	Santa Clara, CA
20. Mark Huffer, General Manager	Kansas City Area Transportation Authority	Kansas City, MO
21. Paul Jablonski, Chief Executive Officer	Metropolitan Transit System	San Diego, CA

Appendix B: Vendor Interview Participants

B

Interviewee	Organization	Line of Business
1. Mark Kroncke	Invoke Technologies	Payment Systems Integration and Consulting
2. Marshall Moyrene	CSiT	Communications, Security, and Integrated Data Systems
3. Mary Sue O'Melia	Transtrack	Transit Performance Reporting Vendor
4. David Pickeral	IBM Smarter Cities	Integrated Data Systems and Data Analytics
5. William Robert	Spy Pond Partners	Asset Management Consulting
6. Kirk Shore	Clever Devices	Operations and Maintenance Management, Passenger Information
7. Angela Miller	Cubic	Transportation Systems and Services
8. Paul Ernsdorff / Brett Koenig	Assetworks / Trapeze	Transit Software and related services Vendor
9. Daisy Wall	RouteMatch	Transit Scheduling and Software Vendor
10. Abe Padilla	Spectrum Data Analytics	Data Analytics Consulting

Appendix C: Interview Questions

C

Interview Questions, Leadership APTA Project Team 5

Team Members:
Leader: Mark Ellis, Central Region Manager, Maintenance Design Group, Katy, TX
Lou Cripps II, Asset Management System Administrator, RTD, Denver, CO
Roderick Diaz, Director of Planning and Development, Metrolink, Los Angeles, CA
David Huffaker, Director, Facilities and Asset Control, Sound Transit, Seattle, WA
Laura Minns, Sr. Project Manager, LYNX, Orlando, FL

Preamble

We are a group of public transportation professionals associated with a leadership program for the American Public Transportation Association (APTA). As part of this program, we are conducting research that will be directed at CEOs and decision-makers within the North American public transportation industry. Our research is aimed at understanding how the transit industry and transit agencies can position themselves and their data/information management approaches to be more agile, open to innovation, and able to maximize value for the organizations themselves, for their customers, and to the vendor community that serves organizations and users related to the industry.   We are hoping to glean insights from you and your experience for agencies of your size and service mix to be able to learn appropriate lessons for the industry.   This interview consists of a set of open-ended questions designed to find these insights.

Definition of "Data" for our project:

> Potential types of data that can be subject to this discussion:

> > Service Data: Service Performance/Schedules/Location-Based Data

> > Customer Data: Ridership / Travel Patterns / Demographics

> > Financial Data: Fare/Revenue

> > System/Infrastructure Asset Condition

> > Public Safety/Security Data

Information and Data Management – Overall Approach

When it comes to data and information management, how would you describe your overall approach?

Does your agency have an information and data management policy that defines and codifies this approach or aspects of this approach?

Roles and Responsibilities

What role do you see for top-level decision makers (CEOs and GMs) in enabling agile information management and removing barriers?

What role do you see for CIOs (Chief Information Officers) and IT Managers?

What is the role of your board, if at all, in data/information management and in promoting an agile approach to it?

Approaches/Strategies/Models

Leadership Approach

For CEOs:

How does your agency ensure that it is in alignment with current technology advances?

Given that IT and Information Management can be a technically complex discipline, what approaches do you take to provide direction to those in your organization involved in those realms? Do you ever find it difficult? How do you make sure you are given the right amount of information about technology, in order to make decisions?

Does your agency have a process in place for reviewing the applicability of advances in technology and how they might be applied to your customers or operations?

Do you have any strategies to support a culture of innovation in data and information management in your organization?

Are there some examples of successful projects related to data use or data management at your agency that you would like to highlight? What conditions would you say contributed to their success?

Do you have any examples of technology projects that failed to meet expectations? What lessons did your agency learn from those examples?

What qualities would you say characterize the more successful/agile transit agencies with respect to data management/use?

What are some common stumbling blocks that prevent your agency from becoming more agile in the realm of data and information management?

Agency Organizaiton/Models

To become more agile and responsive, how would you suggest transit agencies position themselves in each of the following aspects?

Internal organization and structure?

Decision-making processes and responsibilities?

Availability/openness of data and information?

Procurement and engagement of vendors?

Required infrastructure or information architecture?

Does your agency have a research & technology program? Which group in your agency manages this program?

Data Policies / Openness of Data

How does your agency model align with the growing openness of data, in general, and in the transit industry, specifically? How does your agency define the appropriate balance between openness of data and keeping information proprietary and protected?

Is there any case where openness of data is not desirable?

Has your organization done an enterprise wide assessment and created a policy to define all governance processes, rules and recommendations on data management etc.?

In what ways do you see that the public transportation industry is different from other industries that should cause it to approach data and information management in another way?

Benchmarking with respect to other industries

Does your agency participate in benchmarking with other transit agencies, specifically in technology and information management? What other agencies have you noted perform well in staying current with technology trends?

Does your agency compare to other industries?   What models would you point to in other industries for transit agencies to emulate?   What is so positive about these other models?

Involvement of the Private Sector / Third Parties

How do you see the transit industry encouraging or enabling private sector participation in overall data/information management and development of applications that maximize the use of data and information available?

How well do your private sector vendors support your agency in the realm of data management?

General:

Information for Customers

Do you see providing more information to customers as a way to maintain or increase ridership?

Which departments in your agency are involved in the deployment of information to customers?

       [IT, Operations, Training, Marketing, HR, Planning, Maintenance, Legal, Customer Service]

Are there other agencies that we should speak regarding "best practices" in customer information?

Operational information:

What tools do you deploy to gather information about the service performance of your system?

Do you have a data strategy for using that information?

How is service information used by your Operations and Maintenance departments?

How is service information used by your Executive leadership?

How is service information used by your Board?

Please describe any "lessons learned" regarding the gathering of service performance information.

<table>
<tr><td colspan="2">Appendix D: On-line Cross Sectional Survey</td><td>D</td></tr>
</table>

Question 1: What type of organization do you work for?

What type of organization do you work for?		
Answer Options	**Response Percent**	**Response Count**
Transit Agency (public Direct Operated)	84.2%	85
Purchased Transport (Private operated)	4.0%	4
Software Company	0.0%	0
Other (please specify)	11.9%	12
answered question		101
skipped question		0
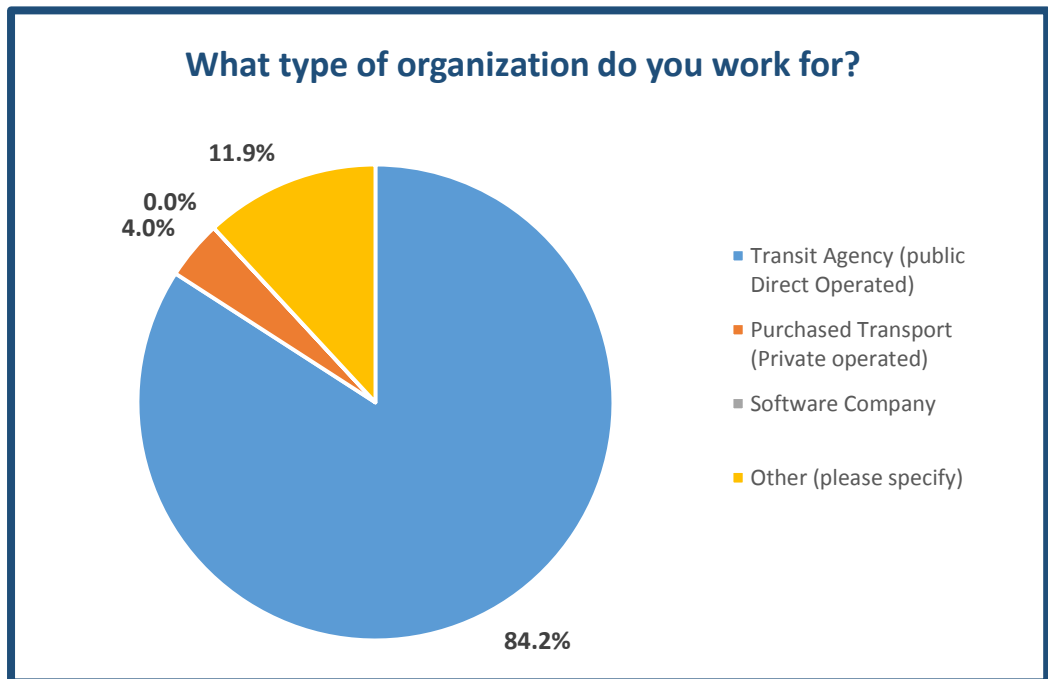
Question 2: If you are a transit agency what fleet size best describes your organization?

If you are a transit agency what fleet size best describes your organization?		
Answer Options	**Response Percent**	**Response Count**
Less than 100 vehicles	28.0%	28
100 - 249 vehicles	18.0%	18
250 - 499 vehicles	22.0%	22
500 - 999 vehicles	13.0%	13
More than 1000 vehicles	16.0%	16
Not Applicable	3.0%	3
Not Sure	0.0%	0
	answered question	100
	skipped question	1
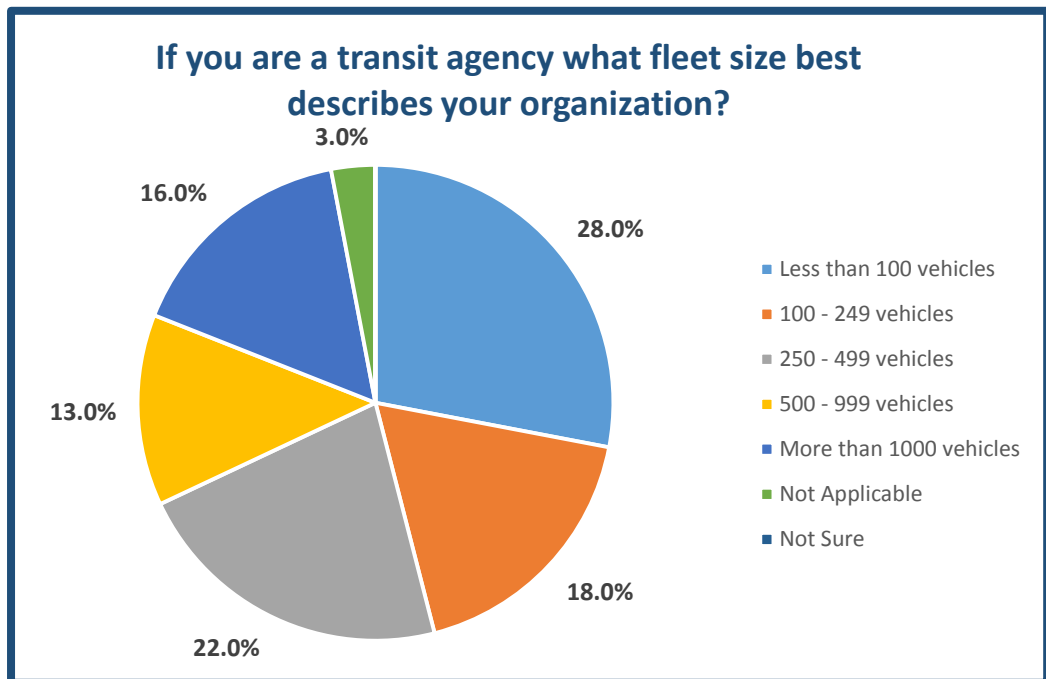
Question 3: What modes do you serve/provide?

What modes do you serve/provide (select all that apply)?		
Answer Options	**Response Percent**	**Response Count**
Bus	92.1%	93
Light Rail	23.8%	24
Rail	30.7%	31
Bus Rapid Transit	28.7%	29
Paratransit	77.2%	78
Other (please specify)	26.7%	27
	answered question	101
	skipped question	0

What modes do you serve/provide?

Question 4: Which of the following best describes you position within your company or organization?

Which of the following best describes your position within your company or organization?		
Answer Options	**Response Percent**	**Response Count**
CEO/ GM / President	75.2%	76
CIO	2.0%	2
Other C-Level Officer	12.9%	13
IT Professional	0.0%	0
Vendor / Consultant (Software)	0.0%	0
Other (Please Specify)	9.9%	10
answered question		101
skipped question		0
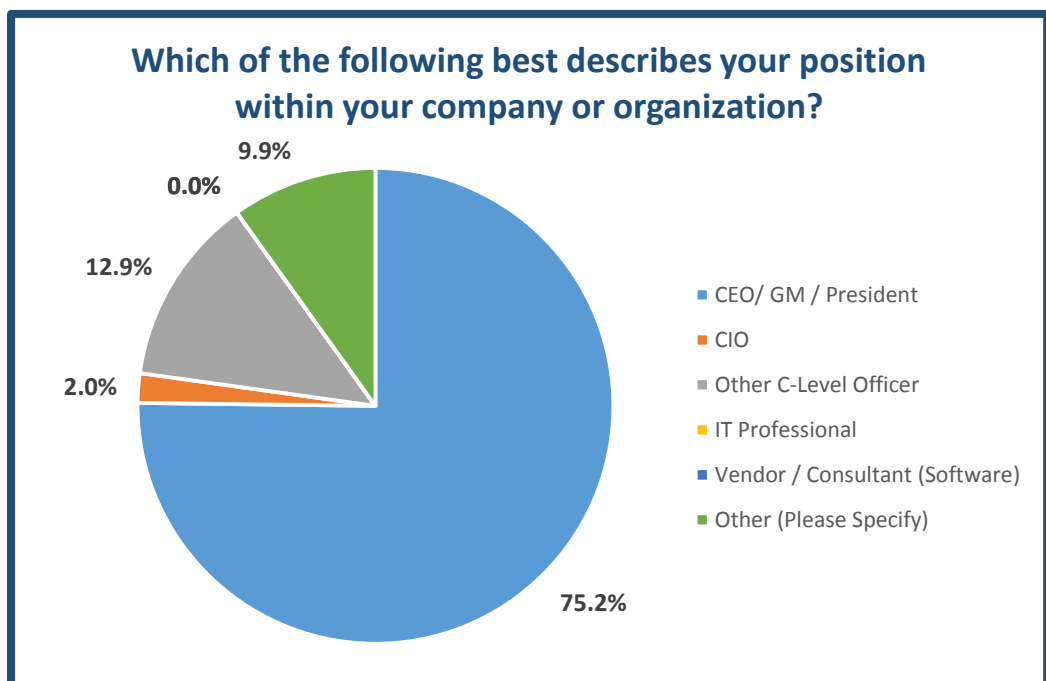
Question 5: Please provide your input on each of the areas of data management for the transit industry.

Please provide your input on each of the areas of data management for the transit industry.		1 = Strongly Disagree	2 = Somewhat Disagree	3 = Neither Agree nor Disagree	4 = Somewhat Agree	5 = Strongly Agree	Not sure/No Opinion	Response Count
Answer Options								
Data Management	The Transit Industry manages data effectively.	4	30	10	45	9	2	100
Information Management	Information Management - The transit industry manages and uses information gleaned from the various data sources effectively.	1	31	9	43	12	3	99
Organization of Data/Information	Organization of Data/Information - Data that the transit industry provides is well-organized and easy to use and process.	5	31	18	37	10	0	101
Organization of Information	Organization of Information - Information that the transit industry generates is well-organized and easy to use and understand.	4	35	13	35	14	0	101
Responsiveness as a Provider of Data	Responsiveness as a provider of data - The transit industry is responsive as a partner or provider of data to outside entities and private sector vendors.	1	28	19	35	15	3	101
Ability to Maximize Decision-making	Ability to maximize the value of data for agency decision-making - The transit industry and transit agencies use data to the fullest extent to support agency decision-making.	6	39	14	31	9	1	100
Ability to Maximize Improvement	Ability to maximize the value of data for improvement of agency operations - The transit industry and transit agencies use data to the fullest extent to improve agency operations.	8	31	22	33	4	1	99

Please provide your input on each of the areas of data management for the transit industry.								
Answer Options		**1 = Strongly Disagree**	**2 = Somewhat Disagree**	**3 = Neither Agree nor Disagree**	**4 = Somewhat Agree**	**5 = Strongly Agree**	**Not sure/No Opinion**	**Response Count**
Openness to Innovation	Openness to innovation - Transit agencies (including the culture within agencies) are open to innovation in the realm of data and information management.	3	25	16	33	23	1	101
Responsiveness to Customers	Responsiveness to Customers - Transit agencies manage data and information in a way that is responsive of customer's needs.	3	22	21	42	12	1	101
Maximizing Value for the Public	Maximizing value of available data for the public at large - Transit agencies manage data in a way that maximizes the value of data for the public at large.	2	37	24	33	3	2	101
Maximizing Value Stakeholders	Maximizing value of available data for other stakeholders - Transit agencies manage data in a way that other stakeholders and private developers can maximize value from it.	7	32	26	28	6	2	101
						answered question		101
						skipped question		0

Question 6: For each statement, describe you plans with existing technologies.

For each statement below, describe your plans with existing technologies.					
Answer Options	Currently use	Funded but not implemented	Evaluating/ Exploring	Don't intend to implement	Not sure/ Not applicable	Response Count
CAD/AVL	75	13	8	1	2	99
Electronic Fare Collection, Smart Card, Mobile Fare Systems	55	14	26	4	2	101
Automated Passenger Counting (APC)	71	9	11	7	3	101
Onboard Camera Security Systems	86	3	5	6	1	101
Station/Stops Camera Security Systems (CCTV)	67	1	12	12	6	98
Scheduling Systems	86	3	5	5	0	99
Materials Management Systems	71	6	9	5	10	101
Maintenance Management Systems	86	4	7	2	2	101
Facility Maintenance Management Systems	54	12	22	6	7	101
Enterprise Asset Management system	40	15	32	4	9	100
Business Intelligence / Business Analytics Systems	24	9	31	12	22	98
Positive Train Control / Automatic train Stop	11	18	9	10	50	98
Other Passenger Information Systems (Please Specify)	33	7	7	4	38	89
Other (Please Specify)						31
					answered question	101
					skipped	0

Question 7: How satisfied are you with your organizations overall data management practices?

How satisfied are you with your organizations overall data management practices?		
Answer Options	**Response Percent**	**Response Count**
1 = Very Dissatisfied	2.0%	2
2 = Somewhat Dissatisfied	28.3%	28
3 = Neither Satisfied nor Dissatisfied	11.1%	11
4 = Somewhat Satisfied	49.5%	49
5 = Very Satisfied	8.1%	8
Not Sure/No Opinion	1.0%	1
answered question		99
skipped question		2
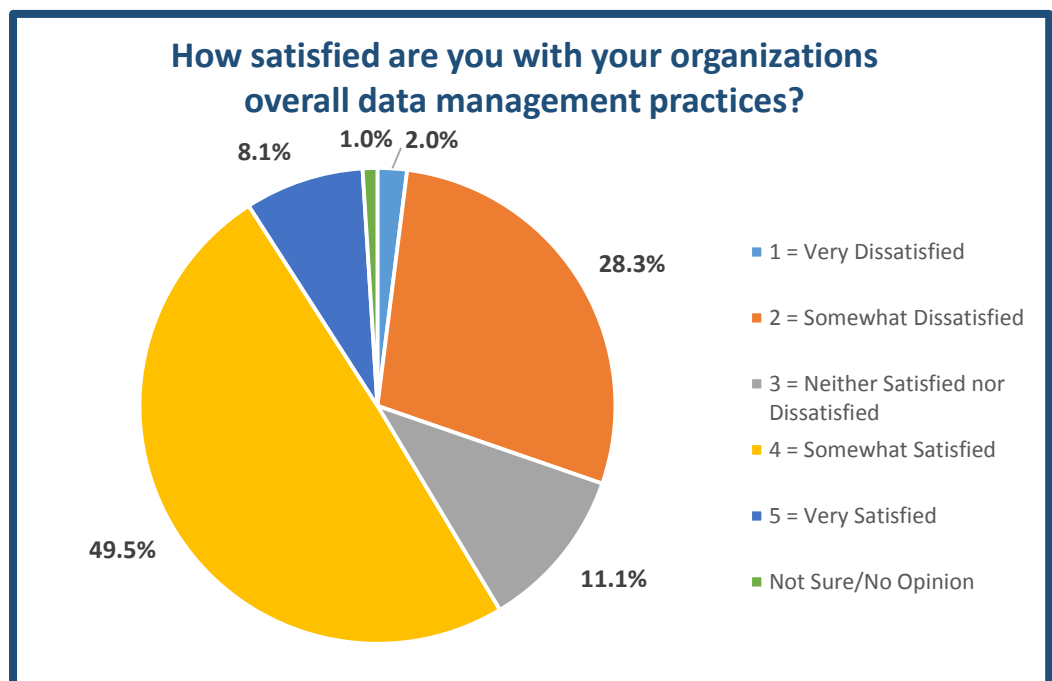
How satisfied are you with your organizations overall data management practices?

Question 8: How satisfied are you with your organization's data management policy/procedures?

How satisfied are you with your organization's data management policy/procedures?		
Answer Options	**Response Percent**	**Response Count**
1 = Very Dissatisfied	0.0%	0
2 = Somewhat Dissatisfied	24.8%	25
3 = Neither Satisfied nor Dissatisfied	15.8%	16
4 = Somewhat Satisfied	46.5%	47
5 = Very Satisfied	10.9%	11
Not Sure/No Opinion	2.0%	2
answered question		101
skipped question		0

How satisfied are you with your organization's data management policy/policies?