POLICY BRIEF

Leveraging Big Data in the Public Transportation Industry

Key Takeaways

1) Public transportation agencies generate large amounts of data as part of their daily operations
2) Public transit agencies and private sector partners are using these data sources to provide valuable insight and improve public transportation service and efficiency
3) Leveraging Big Data requires long-term investment in infrastructure and talent
4) Big Data is poised to have a larger role in public transportation funding decisions

According to most generalizations, Big Data is a term referring to large, continuous sets of data that must typically be structured before analyzing.[1] Big Data is also used as a label for new data sets that are utilized to inform decisions and solve problems. Organizations in other industries have turned their focus to applying significant quantities of data to reduce costs, increase efficiency, and make better-informed decisions.

APTA is undertaking an initiative on Big Data and the public transportation industry. As part of this initiative, selected APTA members engaged in a conversation on Big Data and its role in the industry. Multiple roundtable discussions with public transit agencies and private sector APTA members provided valuable insight on this topic. APTA also conducted a survey of transit agency members that revealed that 94 percent of agencies are using Big Data techniques and methods to improve their agency.

[1] https://www.sas.com/en_us/insights/big-data/what-is-big-data.html#dmimportance

How Public Transit Agencies Collect Data

Public transportation agencies generate large amounts of data as part of their daily operations. Automatic Vehicle Location systems track the positions of buses and trains and collect a constant stream of information. Passenger counting systems mounted on transit vehicles record how many people get on and off at each stop. Fare collection systems and smart cards record trips taken, transfers, and travel patterns among transit users. Other sources of data are not automatically created but entered by agency employees – items like time tracking, absenteeism data, safety incidents, and other employment data sources collected in the field. Public transportation agencies and private sector partners are working to use these data sources to provide valuable insight and improve public transportation service and efficiency. In the survey, agencies listed their AVL systems, fareboxes, and passenger counters as the most common sources of data, with more than 75 percent of agencies using each of those data sources.

New emerging technology and services are enabling this transformative data analysis. Multiple participants mentioned Amazon Web Services (AWS) products as crucial in providing several features that aid in breaking down Big Data at a lower cost. Cloud computing permits the sharing of data across an entire organization, allowing convenient access for different departments. Special software packages built into AWS allow for predictive modeling using large datasets. Thirty-nine percent of survey respondents said they were using cloud computing services as part of their data program.

Data Used for Maintenance and Operations

The central area where agencies are using Big Data principles is to improve and optimize operations and maintenance. Eighty percent of survey respondents indicated that they are using Big Data techniques and methods to improve their operations or maintenance. One member discussed how bringing disparate data sources together in one place has allowed the member to do more analysis on equipment and state of good repair issues. New systems like AWS allow for predictive modeling using information on asset condition and performance. In this case, the agency focused on bus breakdowns. By feeding performance information into a predictive model, the agency has been able to predict bus breakdowns and act to remove buses for maintenance before they break down in the field. This type of predictive action has the potential for significant cost savings.

Predictive modeling is being used in other capacities to improve efficiencies and reduce costs: for example, to monitor, manage, and react to potential operator absenteeism. By using historical personnel data along with information about weather and other events, the predictive model provides information on where operators are likely to be absent. This has allowed the agency to more efficiently station backup operators and limit service disruptions. Furthermore, additional Big Data sources (on-time performance, vehicle location) can be matched with schedules to show operators how they are performing compared to peers within their division and to themselves from week to week, leading to overall improvement.

Agencies are beginning to apply Big Data analysis in the safety and security fields; 62 percent of survey respondents indicated that they are using Big Data methods and techniques for safety and security purposes. Using Big Data tools for cybersecurity initiatives is of increasing interest to the public transit industry as hacking threats become more advanced. Machine learning tools analyze different dimensions of data for

supervision of the environment and on users of the network to determine if there are abnormalities from the normal environment. Another participating agency in the discussion group talked about using data from security cameras to detect abnormalities in the environment the camera is monitoring. While this information is typically of highest interest to the police department, there is potential for other agency departments to use it as well.

Data can also play a part in improving the safety of public transit operations for passengers and people in the outside environment. Data is collected by on-vehicle systems like collision avoidance sensors and on-board cameras that monitor the streets and intersections used by the transit vehicles. By combining this data with vehicle tracking information, agencies can pinpoint problematic streets and intersections where transit vehicles face the most conflicts. It also helps agencies understand what type of situations require additional training for operators.

Finally, one APTA member agency described how labor negotiations can benefit from Big Data analysis. By merging and cleaning a wide variety of datasets, the agency built a real-time negotiation model for use in working with labor partners. During negotiations, the team showed its labor counterparts the effects of different proposals on various categories. The real-time aspect of the model meant that when negotiators proposed a change to the contract, the team could show the resulting impacts at the negotiating table rather than having to spend time in the office re-running the model. This is still an emerging field of data analysis; only 22 percent of survey respondents said they were using Big Data in labor negotiations.

Data Used for Service and Planning

One agency described its efforts in using a new mobile fare app to generate data to help with service delivery and planning. The app collects information on trips and transfers and can even track trips among multiple regional transit agencies. This allows the regional agencies to make better service planning decisions; travelers no longer “disappear” when they leave one agency’s service for another. Additionally, a private organization detailed how cell phone data can be used to determine route patterns of public transit customers to help identify potential transfer and connection points between fixed routes. Communication with private partners will be important as they work on developing new applications for fare collection data.

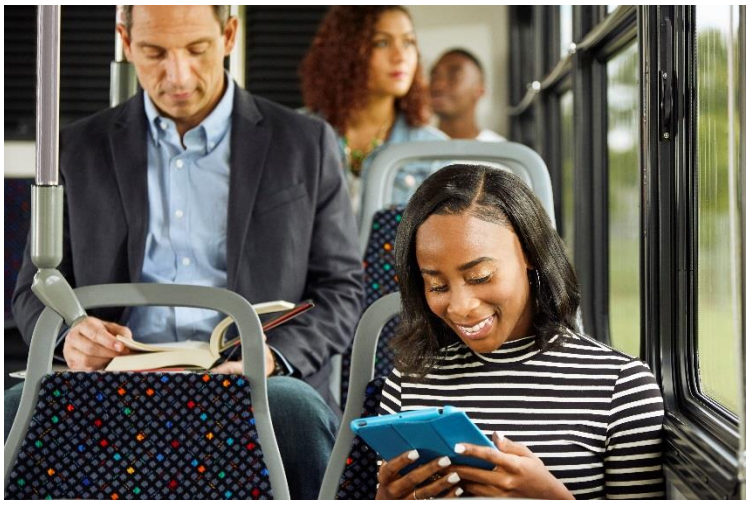
The addition of smart card systems around the country has provided information with the potential to alert individual transit riders about impacts to their trip. By integrating fare payment system technology and real-time operations datasets, a predictive model could alert transit users to potential delays to their usual trip and suggest alternative routes to avoid delays.

Public transit agencies are engaging in partnerships to conduct analysis using travel data generated by users and external data sources. For example, some agencies have been partnering with academic institutions to do travel behavior modeling. Another partnership, which involved Microsoft and Miami-Dade County, attempted to quantify how road traffic affected transit operations and how transit routes could be optimized based on existing commuting patterns. Continuing to integrate internal and external datasets presents a big opportunity for planners.

Data Used for External Purposes

Transit data applications have been focused, for the most part, on addressing internal factors. As new data sources come online and the analysis capability expands, however, there may now be a greater potential in using data for external purposes such as advertising, advocacy, and public affairs.

With more public transit agencies adopting new electronic fare systems and mobile payment options, there is not only greater convenience for transit customers, but also more travel data for agencies (as described in detail in the previous section). This trip data has the potential to be of monetary value to transit agencies, particularly through targeted advertising. When this data is shared with external parties, they have the capability to provide customized marketing messages and offers based on the travel location or travel patterns of individuals.[2] Being able to show travelers more relevant content means that agencies could bring in more revenue.



Aggregate transit data combined with other data sources can provide new insights on the value of transit networks. For example, Columbus, Ohio, received significant praise for the connection it made between public transit and infant mortality in its Smart City grant bid.[3] In bringing together various datasets (income, education, crime, and transportation accessibility), the city showed how the number of infant deaths in different Columbus neighborhoods correlated to external indices. To achieve its goal of reducing infant mortality by 40 percent by 2020, the city laid out a transportation plan to improve bus frequency, add transit amenities, and develop a specialized transportation pilot to connect expectant mothers to healthcare. Columbus received a $50 million U.S. Department of Transportation grant and additional private funding.

Other agencies are already noting the model that Columbus used and are looking at what additional connections can be made to show public transportation's value to assist their advocacy efforts. Linking public transit with job accessibility, health care, and education datasets would be further applications of this method. In an example of how transit can be connected to economic data, Mastercard transit payment data was integrated with retail data to analyze the effect of a car-free day in New York City.[4] Expanding on these types of connections could assist with measuring how transit infrastructure projects impact the economy in a holistic way.

Finally, for some agencies, new data sources and technologies provide a valuable opening for increased customer engagement. Current platforms already are able to modify web displays based on the viewer to provide them with more pertinent information. Social media platforms and search engines (Twitter and Google) may enable a monitoring of public opinion that is more efficient than conducting surveys and could

[2] https://home.kpmg.com/xx/en/home/insights/2017/06/turning-public-transit-ads-into-powerful-digital-tools.html

[3] https://www.transportation.gov/sites/dot.gov/files/docs/Smart%20City%20Challenge%20Overview.pdf

[4] https://newsroom.mastercard.com/press-releases/cubic-and-mastercard-study-shows-new-yorkers-shifted-to-public-transportation-and-sharing-services-during-councilman-ydanis-rodriguezs-car-free-day-nyc/

be utilized to respond to customer feedback more proactively. Furthermore, it may be a way to identify and communicate with non-riders who could be potentially serviced by transit.

Obstacles to Further Utilization of Transit Data / Recommendations

In detailing their experiences with data collections and analysis, public transit agencies have experienced several obstacles and have made remarks on what they would like to see happen in the future with Big Data.

To be able to interpret data, many transit agencies must first verify that the data being collected is of decent quality or if systems must first be upgraded. Some agencies, for example, are still working on expanding the number of automatic passenger counters on their vehicles and upgrading GPS devices. There is then the process of standardizing and developing commonality with the data. There must be a common reference point between the data sets to merge and derive comparative meaning from them. Applications from different vendors can be based on different technology, making accessing the data difficult. Fifty-nine percent of APTA members who responded to the survey indicated that data quality issues were an obstacle they have encountered, and 69 percent of respondents said a lack of data standardization was an obstacle.

One large agency has invested in a single location, a data warehouse, where information can be extracted from source systems. That ensures that all the agency's data can be accessed from one location, whether that be personnel scheduling, fare collection, or scheduling data. Centralizing more standardized, higher-quality, and more up-to-date data opens opportunities for new types of analysis to benefit the agency. With near real-time streams of data coming in, agencies can move from traditional "end-of-the-month"-type analyses toward predictive models that guide agency decision making in real time.

Additionally, there are some data sets that transit agencies would like to access but that may not be available. For example, ride-hailing data is only occasionally shared with agencies (in the case of mobility partnerships), preventing the agencies from taking advantage of even more information on customer travel patterns. External data sets (like various internal sets) may not be well equipped to be merged.

One reason that transportation network companies (TNCs) may not share data with outside parties is to protect the privacy of their customers. This privacy concern is also on the mind of public transit agencies when advancing their data applications. It has the potential to complicate some of the external data purposes listed in the prior section, including targeted advertising and precise route planning. Finding ways to further anonymize individual data will be critical if this data source is to be exploited by transit agencies and private parties.

Finally, continuing to promote a culture of data analysis should be a long-term commitment for transit agencies. At a leadership level, this involves the funding and prioritization of new technologies, data sources and data analysis. At a workforce level, this involves training and acquiring qualified employees who have the skills to analyze data. Sixty-three percent of respondents to the APTA survey indicated that a lack of internal staff expertise was an obstacle to further use of Big Data. Developing internal expertise, along with consultant input, will encourage this culture of data-based decision making.

Conclusion

As more industries become familiar with Big Data, public transportation agencies are following and are realizing the benefits of data-driven decision making. By leveraging Big Data, agencies can receive findings that are more accurate than past data collection methods (surveys), which can be used to inform decision making and increase awareness among the general public and business sponsors. The increasing affordability of data collection and analyzing tools have made this more accessible to agencies. It is now even common to see offices of innovation within transit agencies.

What is clear is that collecting and presenting quality transit data is becoming more important, especially from a funding perspective. In developing the Mobility on Demand (MOD) Sandbox program, the Federal Transit Administration required applicants to identify what data would be collected to inform decisions before, during, and after project implementation.[5] Applicants were also evaluated based on their plan to use this data to benefit the industry as a whole.

While public transit agencies are encouraged to take Big Data adoption head on, agencies with limited resources can still reach out for external help, insight, and counsel. Many private sector companies are stepping in to complement where transit agencies do not have adequate resources, for example with strategic approach and organizational design. With more next-gen contactless and mobile payment systems on the horizon for multiple agencies, data collection capabilities will only expand.

Working with all partners, public transit agencies can increase their ability to harness Big Data and increase operational efficiency. Learning lessons from other industries' experiences with analytics can help inform agencies about new platforms and technologies. APTA will continue to be a leading voice in monitoring developments in this area and helping public transit agencies connect with the resources they need to move forward with their data analysis.

Notes on the Survey

The survey was conducted in summer 2018. APTA received responses from 71 transit agency members. The results represent a sample of APTA's members and are not necessarily scientifically representative of the industry as a whole.

[5] https://www.transportation.gov/connections/how-internet-things-transforming-public-transit

Initiative Chair

William T. Thomsen, PE
President & CEO
Urban Engineers of New York, D.P.C.
2 Penn Plaza, Suite 1103
New York, NY 10121

The American Public Transportation Association (APTA)

APTA is a nonprofit international association of more than 1,500 public and private sector organizations which represents a $71 billion industry that directly employs 430,000 people and supports millions of private sector jobs. APTA members are engaged in the areas of bus, paratransit, light rail, commuter rail, subways, waterborne services, and intercity and high-speed passenger rail. This includes: transit systems; planning, design, construction, and finance firms; product and service providers; academic institutions; transit associations and state departments of transportation. APTA is the only association in North America that represents all modes of public transportation. APTA members serve the public interest by providing safe, efficient and economical transit services and products.

Authors

Matthew Dickens
Senior Policy Analyst
202.496.4817 | mdickens@apta.com

MacPherson Hughes-Cromwick
Policy Analyst
202.496.4812 | mhughes-cromwick@apta.com

For General Information

Policy Development and Research
Darnell Grisby, Director
202.496.4887 | dgrisby@apta.com
www.apta.com/resources/

APTA Vision Statement

APTA is the leading force in advancing public transportation.

FEBRUARY 2019