**Predictors of Adoption of the General Transit Feed Specification by US transit agencies**

**Carole Turley Voulgaris**

Assistant Professor of Urban Planning

Department of Urban Planning and Design

Harvard Graduate School of Design

Cambridge, MA 02138

Email: cvoulgaris@gsd.harvard.edu

**Charuvi Begwani**

Graduate Research Assistant

Department of Urban Planning and Design

Harvard Graduate School of Design

Cambridge, MA 02138

Email: charuvibegwani@gsd.harvard.edu

**Mengyao Li**

Graduate Research Assistant

Department of Urban Planning and Design

Harvard Graduate School of Design

Cambridge, MA 02138

Email: mengyao\_li@gsd.harvard.edu

Word Count: XX words + 1 table (250 words per table) = XX words

*Submitted [31st July 2022]*

**ABSTRACT**

The Abstract should be a stand-alone summary of the contents of the paper, equaling 250 words or less. It should present the primary objectives and scope of the study, techniques, methods or approaches briefly described and a concise summary of findings and/or conclusions reached.

**Keywords:** Format Example, Guide, Keyword, Keyword

**INTRODUCTION**

The development and widespread adoption of the general transit feed specification (GTFS) data format for transit route and schedule data has transformed the way travelers plan transit trips *(1)*. This data standard was introduced in December 2005 by TriMET, the public agency operating mass transit in Portland, Oregon in partnership with Google. Sixteen years later, most –but not all—transit agencies have begun publishing their route and schedule information in this format. What might explain a transit agency’s decision about whether and when to adopt this data format?

Little prior research has been done on the determinants of GTFS adoption. In one study focusing on California, Frick et al. evaluated the adoption of GTFS by transit agencies in California using eight independent variables - organization type, NTD reporting type, vehicles operated, service area size, population of the service area, population density, city level sales tax rates and mean income levels by zip code *(2).* They found that small transit agencies (reduced reporters) and rural transit agencies were less likely to have published GTFS feeds and independent public transit authorities are more likely to publish GTFS-r feeds than other types of agencies like departments within local governments.

The adoption of GTFS can be seen as an indicator of openness to innovation and technology transfer. Existing literature focuses more on factors that contribute to the diffusion of innovation amongst individuals *(3, 4)* with the applicability of those results to institutions being neither obvious nor well-established. Rogers’ work on diffusion of innovation highlights several characteristics of organizations that correlate with being early adopters of new technologies, including larger size, greater complexity – defined as high level of knowledge and expertise among its members, social interconnectedness, presence of an innovation champion, and organizational slack recognizing the availability of additional resources, especially for high-cost innovations *(4)*.

Literature on technology transfer and information sharing amongst transit agencies and more broadly, public institutions, is scarce. Studies on the adoption of other technologies by transit agencies may be informative in identifying agency characteristics that are generally associated with openness to innovation. Iseki et al. *(5)* have found that early adopters of smart cards for fare payment tended to be those with greater funding availability and those with established relationships with other transit agencies. A California-based study identifies several variables that explain the adoption and access to “Smart City” technologies – for instance such as electric vehicle charging station are associated with areas with high wealth concentration and larger proportions of older, Democratic residents; micro-mobility services are associated with middle- income areas, younger population and greater rates of public transit usage; and Transportation Network Company services are more likely to be found in cities with larger, higher-income and more Democratic populations in California *(6)*.

There is a need for empirical analysis to determine the factors that influence technology transfer and adoption. In this study we evaluate several factors that may have influenced the adoption of GTFS by 493 transit agencies in the United States that were providing scheduled transit service between 2005 and 2020 from the time of its inception. Based on the literature above, we hypothesize that a transit agency’s attitude towards the adoption of a new technology standard is related to its location, the population it serves, its neighboring agencies, the popularity of the new technology, its size, the household characteristics of the district, its expenditure distribution, and its financial status. Size can be represented by both agency size characteristics such as service area, passenger revenue miles, revenue, expenses and fare recovery rate, as well as city size and demographic characteristics such as urbanized area, population and density. Agencies of different regions may hold different opinions about technology upgrades and innovation. For instance, an institution serving a larger population, a denser district, or an area with a higher percentage of renter-occupied households, may face more pressure from customers to make its information available. Similarly, if a large number of fellow agencies nearby have adopted the new technique, or the national-level acceptance rate is high, the transit agency may be more eager to upgrade to catch up with local standards. Besides, institutions of bigger size, better financial status, or more spare money left after paying for general administration might be more favorable towards transitioning to new standards.

The results of this analysis can inform efforts by state- and national-level agencies seeking to encourage innovation by identifying agencies most likely to be open to adopting new technology. It can also help local agencies identify peers who are likely to have experience with innovation and experimentation. An FTA report on recognizes the use and adoption of GTFS as being a strong indicator of open data policy in different cities and states.

**DATA AND METHODS**

**Data**

For this research, we have used data from primarily three sources:

(i) Data about transit agencies taken directly from the Federal Transit Administration’s National Transit Database (NTD) which holds records of the financial, operating, and asset conditions of transit systems *(7)*. We identified 493 transit agencies in the United States that were providing scheduled transit services in 2005, when the GTFS data standard was initially published. Our sample includes transit agencies’ information each year from their establishment to their dissolution or adoption of GTFS standards between 2005 to 2020. The resulting sample consists of 5,528 observations of 493 transit agencies across the country.

(ii) Data about the adoption of GTFS drawn from the first publicly available feed available on three sources of archived GTFS feeds (OpenMobilityData, GTFS Data Exchange, and Transitland). We identified the earliest published GTFS feed available for each agency, if any and used the publication date to estimate the length of time it took for each agency to adopt the GTFS data standard.

(iii) Data about demographic characteristics of urbanized areas taken directly from the 2000 Decennial United States Census and 2010 Decennial United States Census (United States Census Bureau 2000, 2010).

The figure below shows the percentage of transit agencies that adopted GTFS between its introduction in 2006 to 2020 (**Figure 1**).

Chart

Description automatically generated

**Figure 1 Adoption of GTFS by transit agencies by year**

Diagram, map

Description automatically generated

**Methods**

We identified several variables to understand their potential relationship with GTFS adoption (**Table 1**). We categorize these 18 variables into 8 data categories:

(i) Locational characteristics of the agency: census region and division. This represents the geographic and regional location of the transit agency which also serves as a proxy for its distance from Portland where GTFS was first introduced.

(ii) Agency size: vehicles operated in annual maximum service, vehicle revenue miles, fare revenue, ridership, operating expenses, general administration salary

(iii) City demographic characteristics:

a) City size: population and population density

b) Household characteristics: percentage of renter-occupied households

(iv) Technology penetration: year, percentage, and the number of agencies that have adopted the GTFS standard. We look at the existing adoption rate for GTFS to understand its potential influence on other agencies to adopt it.

(v) Peer agencies in the same urbanized area: number of agencies in the same urbanized area, the percentage that the agency’s vehicle revenue miles account for in the whole urbanized area

(vi) Institutional characteristics of the agency: agency, organization, or institution type

(vii) Financial situation of the agency: fare recovery (fare revenue divided by operating expenses)

(viii) Organizational slack: overhead (the percentage that general administration salary takes in the whole operating expenses). We use expenditure distribution of the agencies, specifically wages and salaries as a way of understanding organizational slack and the availability of resources at hand for the agency to adopt new technologies.

**TABLE 1 Variables and sources of data**

|  |  |  |
| --- | --- | --- |
| **Variable Category** | **Name of Variable** | **Data Source** |
| Institutional characteristics | Agency Type | National Transit Database; Data Category: Agency Information (2005 – 2020)  2005 – 2007: agency\_info.xlsx  2008 – 2011: Agency\_Information.xlsx  2012 – 2013: Agency\_Information\_0.xlsx  2014: Agency-Information.xlsx  2015: Agency\_information\_1.xlsx  2016: Agency Information.xlsx  2017 – 2018: Agency Info\_1.xlsx  2019: Agency Info.xlsx  2020: Agency Information.xlsx |
| Organization Type |
| Institution Type |
| Agency Size | Service Area | National Transit Database; Data Category: Agency Information (2005 – 2020)  2005 – 2007: agency\_info.xlsx  2008 – 2011: Agency\_Information.xlsx  2012 – 2013: Agency\_Information\_0.xlsx  2014: Agency-Information.xlsx  2015: Agency\_information\_1.xlsx  2016: Agency Information.xlsx  2017 – 2018: Agency Info\_1.xlsx  2019: Agency Info.xlsx  2020: Agency Information.xlsx |
| Passenger Car Scheduled Revenue Miles | National Transit Database; Data Category: Annual Database Service  2005 – 2011, 2014, 2016 – 2020: Service.xlsx  2012 – 2013, 2015: Service\_0.xlsx |
| Vehicle Scheduled Miles |
| Unlinked Passenger Trips |
| Vehicle Revenue Miles |
| Vehicles Operated in Annual Maximum Service | National Transit Database; Data Categories:  Table 26: Fare per Passenger and Recovery Ratio (2005 – 2014)  Annual Database Service (2015 – 2020)  2005, 2007: Table\_26.xlsx  2006: Table 26 - Fare per Passenger and Recovery Ratio.xlsx  2008 – 2011: T26\_Pass\_Fare\_Recovery\_Ratio.xlsx  2012 – 2013: Table 26 Pass Fare Recovery Ratio.xlsx  2014: Table 26 Pass Fare Recovery Ratio\_2.xls  2015: Service\_0.xlsx  2016 – 2020: Service.xlsx |
| Fare Revenue | National Transit Database  Table 26: Fare per Passenger and Recovery Ratio (2005 – 2014)  Annual Database Fare Revenues (2015 – 2020)  2005, 2007: Table\_26.xlsx  2006: Table 26 - Fare per Passenger and Recovery Ratio.xlsx  2008 – 2011: T26\_Pass\_Fare\_Recovery\_Ratio.xlsx  2012 – 2013: Table 26 Pass Fare Recovery Ratio.xlsx  2014: Table 26 Pass Fare Recovery Ratio\_2.xls  2015: Fare\_Revenue.xlsx  2016 – 2020: Fare Revenue.xlsx |
| Operating Expenses | National Transit Database  Table 26: Fare per Passenger and Recovery Ratio (2005 – 2014)  Annual Database Operating Expense (2015 – 2020)  2005, 2007: Table\_26.xlsx  2006: Table 26 - Fare per Passenger and Recovery Ratio.xlsx  2008 – 2011: T26\_Pass\_Fare\_Recovery\_Ratio.xlsx  2012 – 2013: Table 26 Pass Fare Recovery Ratio.xlsx  2014: Table 26 Pass Fare Recovery Ratio\_2.xls  2015: Operating\_Expenses\_0.xlsx  2016 – 2019: Operating Expenses.xlsx  2020 Operating Expenses\_0.xlsx |
| Fare Recovery Rate | Fare revenue / Operating Expenses |
| City Size | Urbanized Area | National Transit Database; Data Categories:  Appendix D: 2000 U.S. Urbanized Areas (UZAs), Populations, Square Miles and Densities Reported by Transit Agencies (2005 – 2013)  Annual Database Agency Information UZAs (2014 – 2018)  Annual Database Agency Information (2019 – 2020)  2005: Appendix\_D.xlsx  2006: Appendix D – 2000 US Urbanized Areas, Populations, Square Miles and Densities Reported by Transit Agencies.xlsx  2007: AppendixD.xlsx  2008 – 2011: xD\_UZA\_Pop\_SqMiles\_Dens.xlsx  2012 – 2013: Appendix D.xlsx  2014: Agency UZAs.xlsx  2015: Agency\_uza.xlsx  2016 – 2018: Agency UZA.xlsx  2019: Agency Info.xlsx  2020: Agency Information.xlsx |
| Population |
| Area (Square miles) |
| Population Density |
| Total Number of Households | 2000 Decennial Census (Variable code: H004001)  2010 Decennial Census (Variable code: H004001) |
| Total Number of Renter-occupied Households | 2000 Decennial Census (Variable code: H004003)  2010 Decennial Census (Variable code: H004004) |
| Percentage of Renter-occupied Households | Total number of renter-occupied households / Total number of households |
| Locational characteristics | US Census Region | 4 US Regions (Northeast, Midwest, West, South)  9 US Divisions (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific) |
| Technology Penetration | Date of adopting GTFS data standard | Retrieved earliest published GTFS feed date from GTFS Data Exchange, OpenMobilityData, and Transitland |
| Number of Agencies that adopted GTFS Data Standard | Summary of agencies by year and GTFS status |
| Total Number of Agencies | Summary of agencies by year and GTFS status |
| Percentage of Agencies Adopting GTFS Data Standard | Number of agencies that has adopted GTFS data standard / Total number of agencies |
| Organizational slack | Expense Category | National Transit Database; Data Category: Annual Database Operating Expense  2005 – 2011: Operating\_Expenses.xlsx  2012, 2015, 2020: Operating\_Expenses\_0.xlsx  2013: operating expenses.xls  2014, 2016 – 2019: Operating Expenses.xlsx |
| Operator’s Salaries and Wages |
| Other Salaries and Wages |
| Fringe Benefits |
| Total Salary of General Administration | Summary of the total salary of general administration (sum of operator’s salaries and wages, other salaries and wages, and fringe benefits) by agencies’ ID |

**Model**

We use multiple logistic regression models to shed light on the relationships between the transit agency’s GTFS adoption status and different categories of variables. We started by selecting one variable from each category. For instance, Model 1 contains nine variables in total, each describing one influencing element: Division, Organization Type, population density, the share of the agency’s vehicle revenue miles in the whole urbanized area, the number of agencies that have adopted the GTFS standard, Vehicle Revenue Miles, the percentage of renter-occupied households, Overhead (the portion that general administration salary takes in the whole operating expenses), and the fare recovery rate.

Based on the best fit model from the first round of experiments, we followed up by testing combinations of variables inside each category. Apart from Division, Organization Type, Population Density, the share of the agency’s vehicle revenue miles in the whole urbanized area, the percentage of renter-occupied households, Overhead, and the fare recovery rate, Model II includes three variables describing the adoption rate (Year, the percentage of agencies that have adopted the GTFS standards, and the number of agencies that have adopted the GTFS standard) and three variables describing the size of the agency (Vehicles Operated in Annual Maximum Service, ridership, and Operating Expenses).

Model III serves as a control group, including very different variables from the list: Region, Organization Type, Population, the number of agencies that have adopted the GTFS standard, Vehicles Operated in Annual Maximum Service, and the fare recovery rate.

*Level 3 Header*

Fusce malesuada rhoncus nisl sit amet rutrum. Vivamus posuere elit dui, eu scelerisque urna interdum sed (*3*). Etiam quis libero id lacus bibendum interdum. Morbi et pretium lorem (*4*–*8*). Quisque mollis, enim vitae dapibus rhoncus, ex ipsum euismod metus, ultrices pharetra est augue et augue. Fusce et orci ut augue rutrum tempor. Quisque aliquet molestie nulla nec rhoncus. Mauris quis ante sem. Donec at finibus velit. Pellentesque commodo maximus nulla, in faucibus nisl vestibulum sed (*2,* *6–10*). Duis ut turpis maximus, semper sapien in, elementum metus, amet sem sit amet tortor blandit.

**RESULTS**

Nulla eu urna et erat bibendum venenatis sit amet sed sem. Nam quis quam tempor, placerat turpis a, imperdiet ante. Integer eleifend aliquam magna quis viverra. Nulla consequat sem vitae mi lacinia maximus. Fusce id ullamcorper nibh. Mauris sagittis, nunc et laoreet condimentum, lectus leo bibendum nisi, eu posuere nulla nisl non libero. Suspendisse libero sapien, fringilla a sagittis eu, venenatis a lacus. Proin id elementum metus. Cras dolor libero, dignissim in magna vitae, tincidunt porta tortor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Ut pretium dignissim nisi, in laoreet metus rhoncus quis. Curabitur pulvinar scelerisque ligula in malesuada.

**DISCUSSION**

Nunc eros felis, dapibus eu nisl convallis, pellentesque facilisis sapien. Duis vel magna ac ligula rutrum volutpat. Donec sed malesuada ipsum. Morbi sed metus sem. Etiam bibendum posuere elementum. In a tempor mauris. Mauris congue in diam ac malesuada. Praesent sed dignissim erat. Mauris sagittis lectus nec ornare condimentum. Etiam est ante, tristique vitae tellus id, venenatis tincidunt tellus. In hac habitasse platea dictumst. Interdum et malesuada fames ac ante ipsum primis in faucibus. Donec quis pellentesque augue, ac elementum turpis. Quisque nec tristique nunc, in porttitor magna. Donec eget laoreet tortor, euismod aliquam felis.

Donec libero velit, bibendum sit amet dignissim in, tempus vel lorem. Donec consectetur dui lorem, gravida eleifend metus blandit ut. Fusce metus lorem, volutpat sagittis orci id, lobortis semper lectus. Nunc eu erat porttitor, molestie elit nec, pharetra enim. Phasellus finibus risus at iaculis ornare. In hac habitasse platea dictumst. Suspendisse nec nulla in ante ultrices feugiat. Curabitur sed sodales ante. Donec nec gravida nibh, ac bibendum nulla.

Nunc porttitor magna sed ultrices condimentum. Suspendisse in nisl condimentum, posuere metus eu, ullamcorper mi. Fusce rutrum ipsum at felis pharetra porttitor. In facilisis ultricies enim, quis consequat ante sodales eu. Mauris in nunc vitae purus lobortis dapibus id sed erat. Vestibulum id vestibulum nulla, ut egestas lectus. Pellentesque ipsum ex, porttitor nec mauris quis, interdum aliquet turpis. Nullam vel purus id eros suscipit ultricies. Maecenas augue dolor, sollicitudin eget laoreet nec, tempus ut massa. Vestibulum ligula tortor, varius sed justo sed, mollis sollicitudin nulla. Nam vitae dapibus ante. Donec non neque hendrerit, luctus urna et, tristique lacus. Phasellus eu enim eu dolor dignissim euismod sed vitae tortor. Proin massa massa, interdum nec turpis in, auctor venenatis felis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque nec massa nulla **Equation 1**:

 (1)

Sed eu lacus sit amet risus sodales pretium a ac ex. Donec gravida auctor efficitur. Curabitur maximus pellentesque mollis. Nam eu ante sit amet ex feugiat dictum eget eu nisl. Nunc libero sem, molestie at justo malesuada, vehicula fermentum quam. Maecenas hendrerit eget sem ut blandit. Nam sit amet dolor volutpat, sollicitudin mi iaculis, bibendum felis **Equation 2**:

 (2)

Donec libero velit, bibendum sit amet dignissim in, tempus vel lorem. Donec consectetur dui lorem, gravida eleifend metus blandit ut. Fusce metus lorem, volutpat sagittis orci id, lobortis semper lectus. Nunc eu erat porttitor, molestie elit nec, pharetra enim. Phasellus finibus risus at iaculis ornare. In hac habitasse platea dictumst. Suspendisse nec nulla in ante ultrices feugiat. Curabitur sed sodales ante. Donec nec gravida nibh, ac bibendum nulla.

Nunc porttitor magna sed ultrices condimentum. Suspendisse in nisl condimentum, posuere metus eu, ullamcorper mi. Fusce rutrum ipsum at felis pharetra porttitor. In facilisis ultricies enim, quis consequat ante sodales eu. Mauris in nunc vitae purus lobortis dapibus id sed erat. Vestibulum id vestibulum nulla, ut egestas lectus. Pellentesque ipsum ex, porttitor nec mauris quis, interdum aliquet turpis. Nullam vel purus id eros suscipit ultricies. Maecenas augue dolor, sollicitudin eget laoreet nec, tempus ut massa. Vestibulum ligula tortor, varius sed justo sed, mollis sollicitudin nulla. Nam vitae dapibus ante. Donec non neque hendrerit, luctus urna et, tristique lacus. Phasellus eu enim eu dolor dignissim euismod sed vitae tortor. Proin massa massa, interdum nec turpis in, auctor venenatis felis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque nec massa nulla.

**CONCLUSIONS**

Nulla eu urna et erat bibendum venenatis sit amet sed sem. Nam quis quam tempor, placerat turpis a, imperdiet ante. Integer eleifend aliquam magna quis viverra. Nulla consequat sem vitae mi lacinia maximus. Fusce id ullamcorper nibh. Mauris sagittis, nunc et laoreet condimentum, lectus leo bibendum nisi, eu posuere nulla nisl non libero. Suspendisse libero sapien, fringilla a sagittis eu, venenatis a lacus. Proin id elementum metus. Cras dolor libero, dignissim in magna vitae, tincidunt porta tortor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Ut pretium dignissim nisi, in laoreet metus rhoncus quis. Curabitur pulvinar scelerisque ligula in malesuada.

**ACKNOWLEDGMENTS**

Nunc eros felis, dapibus eu nisl convallis, pellentesque facilisis sapien. Duis vel magna ac ligula rutrum volutpat. Donec sed malesuada ipsum. Morbi sed metus sem. Etiam bibendum posuere elementum. In a tempor mauris. Mauris congue in diam ac malesuada. Praesent sed dignissim erat. Mauris sagittis lectus nec ornare condimentum. Etiam est ante, tristique vitae tellus id, venenatis tincidunt tellus. In hac habitasse platea dictumst. Interdum et malesuada fames ac ante ipsum primis in faucibus. Donec quis pellentesque augue, ac elementum turpis. Quisque nec tristique nunc, in porttitor magna. Donec eget laoreet tortor, euismod aliquam felis.

**AUTHOR CONTRIBUTIONS**

The authors confirm contribution to the paper as follows: study conception and design: X. Author, Y. Author; data collection: Y. Author; analysis and interpretation of results: X. Author, Y. Author. Z. Author; draft manuscript preparation: Y. Author. Z. Author. All authors reviewed the results and approved the final version of the manuscript.

**REFERENCES**

1.McHugh, Bibiana. 2013. “Pioneering Open Data Standards: The GTFS Story.” In *Beyond Transparency: Open Data and the Future of Civic Innovation*, edited by Brett Goldstein and Lauren Dyson, 1st edition, 125–36. San Francisco: Code for America Press.

2. Frick, Karen Trapenberg, Tanu Kumar, and Alison Post. 2020. “Background Paper: The General Transit Feed Specification (GTFS) Makes Trip-Planning Easier — Especially During a Pandemic — Yet Its Use by California Agencies Is Uneven.” University of California Institute of Transportation Studies. 10.7922/G2ZG6QJZ.

3. Dedehayir, Ozgur, Roland J. Ortt, Carla Riverola, and Francesc Miralles. 2017. “Innovators and Early Adopters in the Diffusion of Innovations: A Literature Review.” *International Journal of Innovation Management* 21 (08): 1740010. https://doi.org/10.1142/S1363919617400102.

4. Rogers, Everett M. 2003. *Diffusion of Innovations*. 5th edition. New York: Free Press.

5. Iseki, Hiroyuki, Allison Yoh, and Brian Taylor. 2007. “Are Smart Cards the Smart Way to Go?: Examining Their Adoption by U.S. Transit Agencies.” *Transportation Research Record: Journal of the Transportation Research Board* 1992 (January): 45–53. https://doi.org/10.3141/1992-06.

7. Post, A., Ratan, I., and Hill, M. 2021. “Benchmarking Smart City Technology Adoption in California: An Innovative Web Platform for Exploring New Data and Tracking Adoption.” University of California – Berkeley, Institute of Transportation Studies. https://doi.org/10.7922/G26M355T

7. National Transit Database, Federal Transit Administration. Retrieved from: https://www.transit.dot.gov/ntd/transit-agency-profiles

8. Von Quintus, H. L., and A. L. Simpson. *Documentation of the Backcalculation of Layer Parameters for LTPP Test Sections.* Publication FHWA-RD-01-113. FHWA, U.S. Department of Transportation, 2002.

9. Shunk, G. A. Urban Transportation Systems. In *Transportation Planning Handbook* (J. D. Edwards, Jr., ed.), Prentice Hall, Englewood Cliffs, N.J., 1992, pp. 88–122.

10. Corbett, J. J. Toward Environmental Stewardship: Charting the Course for Marine Transportation. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2018.