UC Office of the President
ITS reports

Title
Background Paper: The General Transit Feed Specification (GTFS) Makes Trip-Planning Easier — Especially During a Pandemic — Yet its Use by California Agencies is Uneven

Permalink
https://escholarship.org/uc/item/1f29b7dk

Authors
Frick, Karen Trapenberg, PhD
Kumar, Tanu, PhD
Post, Alison, PhD

Publication Date
2020-07-01

DOI
10.7922/G2ZG6QJZ

eScholarship.org Powered by the California Digital Library
University of California

BACKGROUND PAPER

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

The General Transit Feed Specification (GTFS) Makes Trip-Planning Easier—Especially During a Pandemic—Yet its Use by California Agencies is Uneven

Karen Trapenberg Frick, Ph.D., Associate Professor, City and Regional Planning, University of California, Berkeley
Tanu Kumar, Ph.D., Postdoctoral Fellow, College of William and Mary
Alison Post, Ph.D., Associate Professor, Political Science and Global Metropolitan Studies, University of California, Berkeley

July 2020

Berkeley Institute of Transportation Studies

Report No.: UC-ITS-2020-31 | DOI: 10.7922/G2ZG6QJZ

Technical Report Documentation Page

1. Report No.
UC-ITS-2020-31

2. Government Accession No.
N/A

3. Recipient's Catalog No.
N/A

4. Title and Subtitle
Background Paper: The General Transit Feed Specification (GTFS) Makes Trip-Planning Easier—Especially During a Pandemic—Yet its Use by California Agencies is Uneven

5. Report Date
July 2020

6. Performing Organization Code
ITS Berkeley

7. Author(s)
Karen Trapenberg Frick , Ph.D., orcid.org/0000-0001-8104-7254; Tanu Kumar, Ph.D., orcid.org/0000-0002-8707-9473; Alison Post, Ph.D., orcid.org/0000-0002-8962-6047

8. Performing Organization Report No.
N/A

9. Performing Organization Name and Address
Institute of Transportation Studies, Berkeley
109 McLaughlin Hall, MC1720
Berkeley, CA 94720-1720

10. Work Unit No.
N/A

11. Contract or Grant No.
UC-ITS-2020-31

12. Sponsoring Agency Name and Address
The University of California Institute of Transportation Studies
www.ucits.org

13. Type of Report and Period Covered
Background Report (July 2019 – June 2020)

14. Sponsoring Agency Code
UC ITS

15. Supplementary Notes
DOI: 10.7922/G2ZG6QJZ

16. Abstract
The General Transit Feed Specification (GTFS) is an open source data format public transportation agency use to share information about routes and vehicle arrival and departure times. A variety of trip-planning applications, including Google Maps, rely on GTFS feeds to incorporate public transit information. In April 2020, the California Integrated Travel Project conducted a Feasibility Study that called for the widespread adoption of GTFS-static (GTFS-s) and GTFS-realtime (GTFS-r) to make transit simpler for California residents; however, there is little research on patterns of information sharing across transit agencies. This background paper highlights findings from an analysis of GTFS use among agencies that report to the National Transit Database (NTD) in California. The prevalence of transit service modifications in response to the COVID-19 pandemic were also studied.

17. Key Words
GTFS, public transit, routes, schedules, data sharing, data communications, transportation planning, transit operating agencies, COVID-19

18. Distribution Statement
No restrictions.

19. Security Classification (of this report) Unclassified

20. Security Classification (of this page) Unclassified

21. No. of Pages
16

22. Price
N/A

Form DOT F 1700.7 (8-72)

Reproduction of completed page authorized.

About the UC Institute of Transportation Studies

The University of California Institute of Transportation Studies (UC ITS) is a network of faculty, research and administrative staff, and students dedicated to advancing the state of the art in transportation engineering, planning, and policy for the people of California. Established by the Legislature in 1947, ITS has branches at UC Berkeley, UC Davis, UC Irvine, and UCLA.

Acknowledgments

This study was made possible through funding received by the University of California Institute of Transportation Studies from the State of California via the Public Transportation Account and the Road Repair and Accountability Act of 2017 (Senate Bill 1). The authors would like to thank the State of California for its support of university-based research, and especially for the funding received for this project.

Disclaimer

The contents of this report reflect the views of the author(s), who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the State of California in the interest of information exchange. The State of California assumes no liability for the contents or use thereof. Nor does the content necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

Table of Contents

List of Figures

List of Tables

Introduction

The General Transit Feed Specification (GTFS) is an open source data format through which public transportation agencies share information about routes and vehicle arrival and departure times. Agencies can publish static transit schedules (GTFS-s), or even incorporate real-time information (GTFS-r) (Figure 1 shows examples of each type). A variety of trip-planning applications, including Google Maps and the Transit mobile app for iOS and Android, rely on GTFS feeds to incorporate public transit information. Through these widely-used applications, both types of GTFS feeds help can reduce the individual time costs of ridership and difficulties connecting between services operated by different transit agencies.

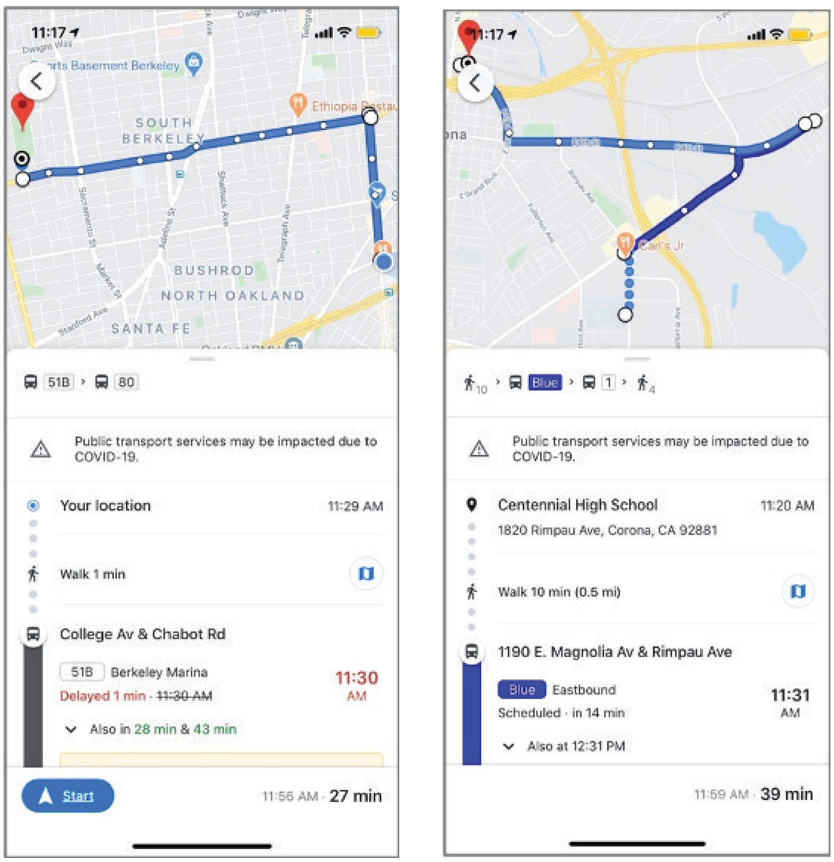

Figure 1 GTFS-realtime (left) and GTFS-static (right) to provide wait times in Google Maps.

In April 2020, the California Integrated Travel Project conducted a Feasibility Study that called for the widespread adoption of GTFS-s and GTFS-r to make transit simpler for California residents.[1] When centralized applications such as Google Maps or the Transit map integrate GTFS data, commuters can easily plan multi-modal trips with numerous connections. As we show at the conclusion of this paper, GTFS is also key to system resiliency as it can facilitate communication between agencies and riders during emergencies and major service modifications or outages, such as those experienced during the COVID-19 pandemic.

To date, there is little research on patterns of information sharing across transit agencies. This background paper examines the transit provider and ridership characteristics associated with the publication of GTFS feeds. In the analysis, GTFS publication is defined as the public sharing of data feeds that can be widely used by multiple actors, from public agencies to private sector app providers.[2] Learning about these patterns should help reveal barriers to publication, and potentially facilitate widespread adoption of this important data-sharing policy by transit agencies.

[1] The study can be found here: https://dot.ca.gov/cal-itp

[2] Several agencies might share their information with certain entities (such as the Transit mobile app) through private agreements.

Research Variables

Definitions and descriptive statistics for all variables included in this study are reported in Table 1.

Data

This study examined patterns of data publication across California transit agencies that reported to the National Transit Database (NTD) in 2018, the most recent period of publicly available data as of April 2020. The NTD contains all California agencies that receive funding from the Federal Transit Administration. The database excludes 16 other agencies that account for only one percent of ridership in the state.[3] The analysis in this study further excludes all agencies that provide demand-responsive service only (e.g. taxis, vanpools, and specialized services for seniors or disabled citizens). The final dataset has 172 agencies.

Outcome Variables

The study looked at whether or not agencies in the database published publicly available static (GTFS-s) or real-time (GTFS-r) feeds during April 2020. Data on publication of GTFS-s feeds was obtained from OpenMobility Data (www.transitfeeds.com). Data on the publication of GTFS-r feeds was coordinated with the California Integrated Travel Project. While several agencies might share their information with certain entities (such as the Transit mobile app) through private agreements, this study focuses on publicly available transit data feeds that are widely disseminated and integrated into third party-applications like Google maps. Future work may consider public agencies sharing with private entities as well.

Independent Variables

The 2018 NTD also provided data that may help explain why certain agencies are more likely to share data information, including each agency's organization type, NTD reporting type, vehicles operated, service area size, and the population of the service area.

The NTD classifies agencies based on the type of reporting required of them: full, reduced and rural. Agencies that are relatively small or based in rural areas do not have to make full reports to the NTD.[4] It also groups transit agencies by type: a) independent public agencies or authorities; b) city, county or local government units; c) other types, including universities, tribes, private corporations, and regional councils of government, of which the dataset only includes 9 total agencies.

The analysis also included data on city level sales tax rates and income levels. City level sales tax data was collected from the California Department of Tax and Fee Administration for 2020. Data on the mean income level as reported in the 2010 US Census by zip code comes from the Population Studies Center at the University of Michigan. These city and zip code level variables were joined with NTD data based on the zip code and city of the agency's main offices. Future work will merge tax and income level data with the NTD data by using transit agency service boundaries and a weighted average of the corresponding city and zip code level data.

[3] Find more information here: https://dot.ca.gov/-/media/dot-media/documents/cal-itp/calitp-feasibility-study-042420-a11y.pdf
[4] Agencies that either a) receive federal funding specifically designated for rural areas (section 5311 from the Federal Transit Administration) or b) operate 30 vehicles or less across all modes and types of service and do not operate fixed guideway (such as rail or Bus Rapid Transit) and/or high intensity busway service are classified as rural or reduced reporters to the NTD, respectively.

Table 1. Definitions and descriptive statistics for outcome and other variables

Variable	Definition	N for which available	Mean	Min	Max
Dependent Variables					
GTFS-s	Indicator for the availability of a public GTFS-s feed	172	0.540	0	1
GTFS-r	Indicator for the availability of a public GTFS-r feed	172	0.190	0	1
Independent Variables					
Service Area	Service area square mileage	133	812.000	0.000	1,736.000
Service Area Population	Population of service area	129	561,932.000	3,801.000	8,595,119.000
Density	Service area divided by population	133	5,012.000	0.000	6,999.000
VOMS	Number of Vehicles Operated to meet requirements on the data of Maximum Service	172	93.000	0	3,458
Tax rate	City-level sales tax	167	8.600	7.200	10.000
Mean income	Mean income for the zip code in which the agency headquarters are located	156	72,042.00	4,521.00	233,520.00
Type: City, County or Local Government Unit or DOT (Department of Transportation)	Indicator for agency type: city, county or local government units	172	0.610	0	1
Type: Independent Public Agency or Authority of Transit Service	1 if independent public agencies or authorities, else 0	172	0.37	0	1
Type: Other	1 if metropolitan planning organizations, councils of government, private firm, tribal agency, or university-based agency, else 0	172	0.952	0	1
Reporter Type: Reduced	1 if operating under 30 vehicles or less across all modes and types of service and not operating fixed guideway and/or high intensity busway service, else 0	172	0.34	0	1
Reporter Type: Rural	1 if receiving federal funding specifically designated to rural areas (section 5311 from the Federal Transit Administration), else 0	172	0.23	0	1
Reporter Type: Full	All other agencies reporting to the National Transit Database	172	0.44	0	1

Analysis and Results

Of the 172 agencies in our dataset, 93 (54 percent) had published GTFS-s feeds. Only 32 agencies in California (19 percent) have published GTFS-r feeds. Figure 2 breaks down the study agencies by reporter type and agency type and whether they have published either static or real-time information.

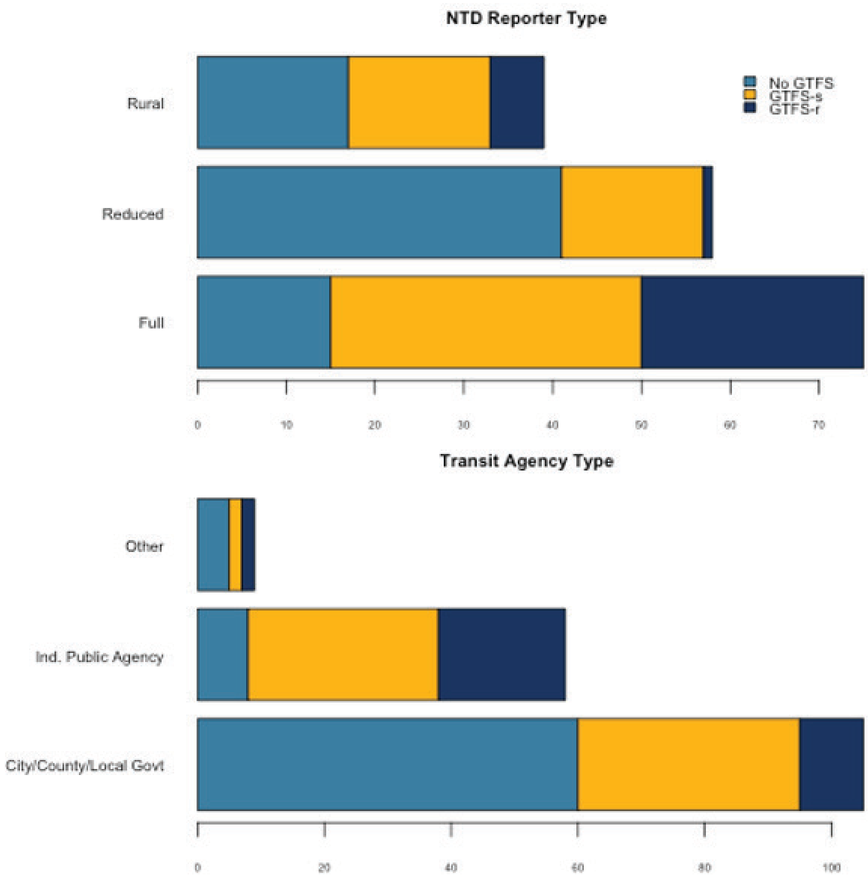

Figure 2. Publication of GTFS feeds by National Transit Database (NTD) reporter type and transit agency type. Note: The "Other" transit agency type includes tribal agencies, universities, private corporations, and regional councils of government.

Reduced and rural reporters are less likely than the larger full reporters to have published GTFS-s feeds. Table 2 demonstrates that these trends hold even when controlling for the other variables listed above using an Ordinary Least Squares (OLS) linear probability model. Model 1 includes all of the variables. Model 2 excludes variables that are collinear (or redundant) with the explanatory variable Reporter Type: Rural. "Other" type transportation agencies are also less likely than government run and independent public transit authorities to publish feeds, but this category of operator is small and has only 9 members, making inference from this pattern difficult.

These trends also hold for GTFS-r publication (Table 3). Furthermore, while the majority of transit agencies in California are government run, independent public transit authorities are more likely to publish GTFS-r feeds than other types of agencies, particularly those operated by a county, city, or local government unit. Across models 1 and 3, we see that independent transit authorities are more likely to publish GTFS-r feeds than local government units. Across models 1 and 2, we see that rural and reduced reporters are less likely to publish GTFS-r feeds than full reporters. Model 4 further shows that even when the sample is restricted to only agencies that have GTFS-s or GTFS-r feeds, independent agencies are far more likely to publish real time (as opposed to simply static) feeds than other types.

Table 2. Predictors of GTFS-s Adoption among CA Transit Agencies, 2020.

	Dependent variable:	
	GTFS-static feed publication[1]	
	(1)	(2)
Type: Independent Public Agency or Authority of Transit Service	0.140	
	(0.120)	
Type: Other	-0.700**	
	(0.320)	
Service Area Population	0.00000	
	(0.00000)	
Service Area	0.0001	
	(0.0001)	
VOMS		
Tax rate	-0.065	-0.051
	(0.075)	(0.045)
Reporter Type: Reduced[3]	-0.360***	-0.490***
	(0.110)	(0.089)
Reporter Type: Rural		-0.270***
		(0.100)
Mean	0.00000	-0.00000
	(0.00000)	(0.00000)
Density	-0.00000	
	(0.00004)	
Constant	1.100**	1.200***
	(0.560)	(0.410)
Observations	111	151
R^2	0.370	0.200
Adjusted R^2	0.320	0.180
Residual Std. Error	0.410 (df = 102)	0.450 (df = 146)

**p* < 0.1; **p < 0.05; *** p < 0.01. All models estimated using ordinary least squares.

[1] *Standard errors corrected to be heteroskedasticity-robust (HC2).*

[2] *Reference category is* Type: City, County or Local Government Unit or DOT.

[3] *Reference category is Reporter Type: Full.*

Table 3. Predictors of GTFS-r Adoption by CA Transit Agencies, 2020.

	Dependent variable:			
	GTFS-real time feed publication			
	(1)	(2)	(3)[4]	(4)[5]
Type: Independent Public Agency or Authority of Transit Service	0.260**		0.280**	0.360***
	(0.100)		(0.140)	(0.140)
Type: Other	0.280		0.280	
	(0.290)		(0.370)	
Service Area Population	-0.00000		-0.00000	-0.00000
	(0.00000)		(0.00000)	(0.00000)
Service Area	-0.0002*		-0.0002	-0.0002
	(0.0001)		(0.0001)	(0.0001)
VOMS			0.0001	0.0001
			(0.0002)	(0.0002)
Tax rate	-0.001	0.023	0.002	0.008
	(0.067)	(0.036)	(0.100)	(0.110)
Reporter Type: Reduced[3]	-0.230**	-0.300***		
	(0.095)	(0.072)		
Reporter Type: Rural		-0.160*		
		(0.081)		
Mean	-0.00000	0.00000	-0.00000	-0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Density	0.00002		0.00003	0.00005
	(0.00004)		(0.0001)	(0.0001)
Constant	0.190	0.078	0.130	-0.086
	(0.510)	(0.330)	(0.780)	(0.780)
Observations	111	151	67	60
R^2	0.210	0.110	0.110	0.160
Adjusted R^2	0.140	0.090	-0.012	0.051
Residual Std. Error	0.370 (df = 102)	0.370 (df = 146)	0.470 (df = 58)	0.450 (df = 52)

**p<0.1;* **p<0.05; *** p<0.01. All models estimated using ordinary least squares.

[1] *Standard errors corrected to be heteroskedasticity-robust (HC2).*

[2] *Reference category is* Type: City, County or Local Government Unit or DOT.

[3] *Reference category is Reporter Type: Full.*

[4] *Model includes full reporters only.*

[5] Model includes only agencies with either GTFS-static or GTFS-realtime.

Relevance of the GTFS During the COVID-19 Pandemic

As California's experience during the COVID-19 pandemic shows, the widespread adoption of GTFS-r is key facet of transit system resiliency. The introduction of shelter-in-place orders and social-distancing guidelines during the pandemic greatly reduced the use of public transit. Based on usage of the Maps app, Apple estimates that transit use in the United States decreased by 75% between January 13th and April 25th 2020.[5] Many public transit providers across the country cancelled or modified their services in response to this decline.

During the week of April 20th, we studied efforts by 30 transit agencies across California to communicate service modifications to the public through websites and GTFS feeds (Table 4). We examined randomly selected samples of 10 agencies that did not publish any type of public GTFS feed, 10 that shared static information using GTFS-s, and 10 that shared real-time information using GTFS-r. Data on GTFS-s feed publication were webscraped from a centralized online repository, and data on GTFS-r feed publication were provided by the California Integrated Travel Project.[6] While several agencies share their information with certain entities (such as the Transit mobile app) through private agreements, we defined GTFS publication as the public sharing of feeds that can be widely used by multiple actors, from public agencies to private sector app providers.

We found that service modifications were common during the pandemic. According to the websites of the 30 agencies studied, 23 had modified service schedules. We further found that use of GTFS-r was essential to the timely and accurate communication of service modifications and coordination of connecting trips between transit agencies. Of the 10 agencies using GTFS-r during normal operations, 8 continued to use GTFS-r during the pandemic and were thus able to directly communicate service changes to riders in an automated and efficient fashion. In contrast, while 8 of the 10 agencies using GTFS-s during normal operations made some type of service modification or cancellation, these changes were only reflected in the public GTFS-s displayed through Google Maps for 3 of the 8 agencies.

[5] Apple has made this data public beginning in January 13th. The data for this brief was collected on April 25, 2020. Find more information here: https://www.apple.com/covid19/mobility

[6] The online repository scraped can be found at https://transitfeeds.com. Information in this background paper is current as of April 1, 2020.

Table 4. Summary of service modifications and incorporation of changes into Google Maps by transit agency

	Agency Name	City	Service modification	Modification incorporated into Google Maps
Agencies using GTFS-realtime	Alameda-Contra Costa Transit District	Oakland	Yes	Yes
	City of Lakewood	Lakewood	Yes	Yes
	City of Santa Monica	Santa Monica	Yes	Yes
	Golden Gate Bridge, Highway and Transportation District	San Francisco	Yes	Yes
	Lake Transit Authority	Lower Lake	Yes	Yes
	Long Beach Transit	Long Beach	Yes	Yes
	Marin County Transit	San Rafael	Yes	Yes
	Napa Valley Transportation Authority	Napa	Yes	No
	San Francisco Bay Area Rapid Transit District	Oakland	Yes	Yes
	Western Contra Costa Transit Authority	Pinole	Yes	No
Using GTFS-static	City of Corona	Corona	Yes	Yes
	City of Fairfield	Fairfield	Yes	Yes
	City of Petaluma	Petaluma	No	N/A
	County of Siskiyou	Yreka	Yes	No
	Gold Coast Transit	Oxnard	Yes	No
	Lassen Transit Service Agency	Susanville	No	N/A
	Mountain Area Regional Transit Authority	Big Bear Lake	Yes	No
	Plumas County Transportation Commission	Quincy	Yes	No
	Southern California Regional Rail Authority	Los Angeles	Yes	No
	Ventura County Transportation Commission	Ventura	Yes	Yes
Not publishing any type of GTFS feed	Antelope Valley Transit Authority	Lancaster	Yes	No
	City of Arvin	Arvin	Yes	No
	City of Corcoran	Corcoran	No	N/A
	City of Glendora	Glendora	No	N/A
	City of Moorpark	Moorpark	Yes	No
	City of Whittier	Whittier	No	N/A
	Union City	Union City	No	N/A
	Los Angeles County Department of Public Works - South Whittier	Alhambra	Yes	No
	Pomona Valley Transportation Authority	La Verne	No	N/A
	City of Wasco	Wasco	Yes	No

Discussion

In summary, we have identified two predictors of GTFS feed publication that operate independently of one another, namely 1) whether or not an agency is reasonably large and operates in an urban area, and 2) whether or not it is an independent public agency, as opposed to a department within a local government.

An important area of future research is uncovering exactly why these patterns hold. Doing so will allow researchers to make policy recommendations to facilitate feed publication. It is possible, for example, that larger agencies have more resources to devote to publishing feeds or that independent agencies possess more capacity to develop innovative technological applications to attract customers. Potential barriers to publication at smaller agencies might be insufficient budget or personnel to implement and develop the feeds. As a result, state or federal level funding and coordinated planning at the state and regional levels may facilitate the timely publication and dissemination of this important information, thereby providing a stronger foundation for increased transit information for passengers and public agencies alike.

The COVID-19 pandemic shows that the widespread adoption of GTFS-r in particular is essential to the resiliency of communities to emergencies. In situations where services are continuously modified, it is costly for agencies to constantly update their GTFS-s feeds, and this information is less likely to reach users than GTFS-r feeds integrated into applications like Google Maps. This finding is critical from a social equity standpoint. A recent survey of over 20,0000 transit riders performed by the company behind the Transit mobile app found that current transit riders are more likely to be women, in lower-income brackets, and African American, Latino, or Native American as compared to transit riders prior to the pandemic.[7] These trends are in line with recent research finding that workers in sectors characterized by low rates of working from home or close physical contact with others are more economically vulnerable, and rely more on public transit (Mongey et al. 2020). The data from the COVID-19 pandemic shows that wider adoption of GTFS-r in particular would make it easier for vulnerable populations to learn of and cope with service modifications.

[7] More information about the survey is available here: https://transitapp.com/coronavirus

References

Analysis of Proposed Cal-ITP Initiatives: A Feasibility Study. *California Integrated Travel Project.* Sacramento, 24 April 2020

Mongey, Simon and Pilossoph, Laura and Weinberg, Alex, Which Workers Bear the Burden of Social Distancing Policies? (April 26, 2020). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-51. Available at SSRN: https://ssrn.com/abstract=3586077 or http://dx.doi.org/10.2139/ssrn.3586077