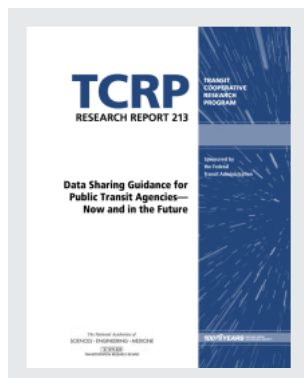


This PDF is available at <http://nap.edu/25696>

SHARE



Data Sharing Guidance for Public Transit Agencies Now and in the Future (2020)

DETAILS

0 pages | 8.5 x 11 | PAPERBACK

ISBN 978-0-309-67240-5 | DOI 10.17226/25696

CONTRIBUTORS

Glen Weisbrod Cecilia Viggiano, EBP, Shan Jiang, Emma Homstad, Tufts University, Kendril Melissa Chan, LLC, and Steptoe & Johnson Sarah Nural, LLP; Transit Cooperative Research Program; Transportation Research Board; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine 2020. *Data Sharing Guidance for Public Transit Agencies Now and in the Future*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25696>.

GET THIS BOOK

FIND RELATED TITLES

Visit the National Academies Press at NAP.edu and login or register to get:

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

Copyright © National Academy of Sciences. All rights reserved.

TRANSIT COOPERATIVE RESEARCH PROGRAM

TCRP RESEARCH REPORT 213

**Data Sharing Guidance for
Public Transit Agencies—
Now and in the Future**

**Cecilia Viggiano
Glen Weisbrod**
EBP (FORMERLY EDR GROUP)
Boston, MA

WITH

**Shan Jiang
Emma Homstad**
TUFTS UNIVERSITY
Medford, MA

Melissa Chan
KENDRIL LLC
Cambridge, MA

Sarah Nural
STEPTOE & JOHNSON LLP
Washington, DC

Subject Areas

Data and Information Technology • Passenger Transportation • Public Transportation

Research sponsored by the Federal Transit Administration in cooperation with the Transit Development Corporation



2020

TRANSIT COOPERATIVE RESEARCH PROGRAM

The nation's growth and the need to meet mobility, environmental, and energy objectives place demands on public transit systems. Current systems, some of which are old and in need of upgrading, must expand service area, increase service frequency, and improve efficiency to serve these demands. Research is necessary to solve operating problems, adapt appropriate new technologies from other industries, and introduce innovations into the transit industry. The Transit Cooperative Research Program (TCRP) serves as one of the principal means by which the transit industry can develop innovative near-term solutions to meet demands placed on it.

The need for TCRP was originally identified in *TRB Special Report 213—Research for Public Transit: New Directions*, published in 1987 and based on a study sponsored by the Urban Mass Transportation Administration—now the Federal Transit Administration (FTA). A report by the American Public Transportation Association (APTA), *Transportation 2000*, also recognized the need for local, problem-solving research. TCRP, modeled after the successful National Cooperative Highway Research Program (NCHRP), undertakes research and other technical activities in response to the needs of transit service providers. The scope of TCRP includes various transit research fields including planning, service configuration, equipment, facilities, operations, human resources, maintenance, policy, and administrative practices.

TCRP was established under FTA sponsorship in July 1992. Proposed by the U.S. Department of Transportation, TCRP was authorized as part of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA). On May 13, 1992, a memorandum agreement outlining TCRP operating procedures was executed by the three cooperating organizations: FTA; the National Academies of Sciences, Engineering, and Medicine, acting through the Transportation Research Board (TRB); and the Transit Development Corporation, Inc. (TDC), a nonprofit educational and research organization established by APTA. TDC is responsible for forming the independent governing board, designated as the TCRP Oversight and Project Selection (TOPS) Commission.

Research problem statements for TCRP are solicited periodically but may be submitted to TRB by anyone at any time. It is the responsibility of the TOPS Commission to formulate the research program by identifying the highest priority projects. As part of the evaluation, the TOPS Commission defines funding levels and expected products.

Once selected, each project is assigned to an expert panel appointed by TRB. The panels prepare project statements (requests for proposals), select contractors, and provide technical guidance and counsel throughout the life of the project. The process for developing research problem statements and selecting research agencies has been used by TRB in managing cooperative research programs since 1962. As in other TRB activities, TCRP project panels serve voluntarily without compensation.

Because research cannot have the desired effect if products fail to reach the intended audience, special emphasis is placed on disseminating TCRP results to the intended users of the research: transit agencies, service providers, and suppliers. TRB provides a series of research reports, syntheses of transit practice, and other supporting material developed by TCRP research. APTA will arrange for workshops, training aids, field visits, and other activities to ensure that results are implemented by urban and rural transit industry practitioners.

TCRP provides a forum where transit agencies can cooperatively address common operational problems. TCRP results support and complement other ongoing transit research and training programs.

TCRP RESEARCH REPORT 213

Project J-11/Task 31
ISSN 2572-3782
ISBN 978-0-309-48141-0

© 2020 National Academy of Sciences. All rights reserved.

COPYRIGHT INFORMATION

Authors herein are responsible for the authenticity of their materials and for obtaining written permissions from publishers or persons who own the copyright to any previously published or copyrighted material used herein.

Cooperative Research Programs (CRP) grants permission to reproduce material in this publication for classroom and not-for-profit purposes. Permission is given with the understanding that none of the material will be used to imply TRB, AASHTO, FAA, FHWA, FTA, GHSA, NHTSA, or TDC endorsement of a particular product, method, or practice. It is expected that those reproducing the material in this document for educational and not-for-profit uses will give appropriate acknowledgment of the source of any reprinted or reproduced material. For other uses of the material, request permission from CRP.

NOTICE

The research report was reviewed by the technical panel and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the National Academies of Sciences, Engineering, and Medicine.

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research and are not necessarily those of the Transportation Research Board; the National Academies of Sciences, Engineering, and Medicine; or the program sponsors.

The Transportation Research Board; the National Academies of Sciences, Engineering, and Medicine; and the sponsors of the Transit Cooperative Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of the report.

Published research reports of the

TRANSIT COOPERATIVE RESEARCH PROGRAM

are available from

Transportation Research Board
Business Office
500 Fifth Street, NW
Washington, DC 20001

and can be ordered through the Internet by going to

<http://www.national-academies.org>

and then searching for TRB

Printed in the United States of America

The National Academies of **SCIENCES • ENGINEERING • MEDICINE**

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, non-governmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at www.national-academies.org.

The **Transportation Research Board** is one of seven major programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to provide leadership in transportation improvements and innovation through trusted, timely, impartial, and evidence-based information exchange, research, and advice regarding all modes of transportation. The Board's varied activities annually engage about 8,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.

Learn more about the Transportation Research Board at www.TRB.org.

COOPERATIVE RESEARCH PROGRAMS

CRP STAFF FOR TCRP RESEARCH REPORT 213

Christopher J. Hedges, *Director, Cooperative Research Programs*
Lori L. Sundstrom, *Deputy Director, Cooperative Research Programs*
Gwen Chisholm Smith, *Manager, Transit Cooperative Research Program*
Dianne S. Schwager, *Senior Program Officer*
Jarrel McAfee, *Senior Program Assistant*
Eileen P. Delaney, *Director of Publications*
Natalie Barnes, *Associate Director of Publications*
Kami Cabral, *Editor*

TCRP PROJECT J-11/TASK 31 PANEL Field of Special Projects

Michael Aaron Berman, *Sound Transit, Seattle, WA (Chair)*
Matthew M. Barnes, *Oregon DOT, Salem, OR*
Amy Bernknopf, *Delaware Valley Regional Planning Commission (DVRPC), Philadelphia, PA*
Yvonne Carney, *Washington Suburban Sanitary Commission (WSSC), Laurel, MD*
Gigi Cooper, *David Evans and Associates, Inc., Portland, OR*
Mark Egge, *High Street Consulting Group, Bozeman, MT*
Gadi Glogowski, *Clever Devices Ltd., Woodbury, NY*
Ryan E. Harshbarger, *Avail Technologies, Inc., State College, PA*
Mohammad Jalayer, *Rowan University, Glassboro, NJ*
Brad Miller, *Pinellas Suncoast Transit Authority (PSTA), St. Petersburg, FL*
Matthew W. Pollack, *HNTB Corporation, Washington, DC*
Rachel Ann VerBoort, *City of Seattle Department of Transportation, Seattle, WA*
Michelle Daigle, *FTA Liaison*
Murat Omay, *FTA Liaison*
Darnell Grisby, *APTA Liaison*

AUTHOR ACKNOWLEDGMENTS

The authors wish to acknowledge the people who took the time to be interviewed for this study, including staff from transit agencies, private mobility companies, transit data companies, transit planning app companies, cities, state departments of transportation, electric utilities and electric utility data users, and academics.



FOREWORD

By Dianne S. Schwager

Staff Officer

Transportation Research Board

Transit agencies, as owners and users of data, seek to maximize the value of their own data and to access external data sets that can help them serve their communities and operate efficiently. *TCRP Research Report 213: Data Sharing Guidance for Public Transit Agencies—Now and in the Future* presents the results of a quick study that provides practical guidance for transit agencies regarding how to make decisions about sharing transit agency data and data from others, including how to evaluate benefits, costs, and risks.

EBP [formerly Economic Research Development Group (EDR Group)] conducted this project through a review of relevant academic, professional, and legal literature; a survey; and interviews. The research identified two types of models for sharing public transit agency data:

- **Public Data Sharing (Open Data).** Data is shared publicly in an online data repository or dashboard through an Application Programming Interface (API) or in a public-facing report. These sharing models promote transparency and can spur innovation, but they cede control over how the data is used.
- **Private Data Sharing.** In a private data sharing agreement, data is shared with a specific partner, often with a nondisclosure agreement. These types of sharing models can enable transit agencies to meet targeted goals. For example, many transit agencies have research partnerships in which they share data with researchers who help address transit agency planning and performance goals.

The report is action oriented and includes a how-to guide for transit agencies to prepare for and execute data sharing. It describes the key factors determining data sharing decisions, including benefits, costs, and risks, and addresses the legal context. The report presents models for sharing transit data as well as accessing external data sources. Finally, it documents the major challenges for data sharing, describes how transit data sharing is expected to evolve in the future, and notes topics for future research.

The report includes two appendices that (1) detail transit agency data types and external data sources with commentary on their sharing attributes and (2) present the interview guides used in this research. For reference, a list of acronym definitions is provided in Table 1.

CONTENTS

| | |
|----|--|
| 1 | Summary |
| 5 | Chapter 1 Introduction |
| 5 | 1.1 Need for Guidance on Transit Data Sharing |
| 8 | 1.2 Research Scope and Key Questions |
| 11 | Chapter 2 Guidance |
| 11 | 2.1 Ensuring Data-Focused Staff |
| 11 | 2.2 Establishing Goals and Objectives |
| 13 | 2.3 Defining Data and Analysis Needs |
| 13 | 2.4 Evaluating and Selecting Data Sharing Models |
| 17 | 2.5 Accessing External Data |
| 18 | 2.6 Responding to Public Records Requests |
| 20 | Chapter 3 Factors Impacting Transit Agency Decisions about Data Sharing |
| 20 | 3.1 Benefits |
| 25 | 3.2 Costs and Effort |
| 28 | 3.3 Risks |
| 33 | 3.4 Rules and Legal Issues |
| 36 | Chapter 4 Models for Sharing Public Transit Data |
| 36 | 4.1 Public Data Sharing |
| 40 | 4.2 Private Data Sharing |
| 41 | 4.3 Examples of Data Sharing Models for Customer Information |
| 45 | Chapter 5 Models for Accessing External Data Sources |
| 45 | 5.1 Purchasing Data or Analysis |
| 46 | 5.2 Accessing Data Through Mobility Service Partnerships |
| 47 | 5.3 Accessing Data Through Regulation |
| 49 | Chapter 6 Major Challenges |
| 49 | 6.1 Internal Data Management Structure and Protocols |
| 51 | 6.2 External Data Policies and Standards |
| 54 | Chapter 7 Conclusions and Next Steps |
| 54 | 7.1 Key Findings |
| 56 | 7.2 Future of Data Sharing for Transit Agencies |
| 57 | 7.3 Future Work |
| 58 | Bibliography and Other Resources |
| 61 | Appendix A Data Types and Sharing Attributes |
| 66 | Appendix B Interview Guides |



S U M M A R Y

Data Sharing Guidance for Public Transit Agencies—Now and in the Future

Data is becoming an increasingly valuable commodity. Transit agencies, as owners and users of data, seek to maximize the value of their own data and to access external data sets that can help them serve their communities and operate efficiently. Many transit agencies have realized benefits from sharing their internal data sets, ranging from improved customer information to innovative research findings that help the transit agency improve performance.

This report provides practical guidance for transit agencies regarding how to make decisions about sharing transit agency data, including how to evaluate benefits, costs, and risks. The report addresses the following:

- **Transit Agency Data.** How to make decisions about sharing transit agency data, including how to respond to public records requests.
- **Data from Others.** How to access external data sets and factors to consider when seeking access to external data sets.

Through interviews and information review, the research team identified two types of models for sharing public transit agency data:

- **Public Data Sharing (Open Data).** Data is shared publicly in an online data repository or dashboard through an Application Programming Interface (API) or in a public-facing report. These sharing models promote transparency and can spur innovation, they but cede control over how the data is used.
- **Private Data Sharing.** In a private data sharing agreement, data is shared with a specific partner, often with a nondisclosure agreement. These types of sharing models can enable transit agencies to meet targeted goals. For example, many transit agencies have research partnerships in which they share data with researchers who address transit agency planning and performance goals.

The research team identified numerous examples of benefits that transit agencies have realized through sharing their data. Sharing data can facilitate the following:

- Promote **transparency** and increase awareness of the transit agency and its engagement with transit customers.
- Spur **innovation** and support **research** that can help transit agencies plan better service and operate more efficiently.
- Enable **cost savings** for transit agencies by using outside resources for data processing and analysis.
- Generate **revenue** (e.g., through **advertising**).
- Support improved **customer information**.
- Support other community functions, such as informing **municipalities**, **real estate developers**, and even **law enforcement agencies**.

2 Data Sharing Guidance for Public Transit Agencies—Now and in the Future

- Facilitate **multiagency and multimodal mobility** solutions.
- Support **benchmarking** activities that help transit agencies track and improve their performance.

Through the interview execution and the information review the team also identified risks of data sharing that are perceived by transit agencies and documented in the literature.

- **Privacy risks** are present whenever data has personally identifiable information (PII). Sometimes, the potential for a data set to be combined with other data sets increases this risk. Transit agencies can take steps to protect privacy that include encryption of identifiers, aggregation, and addition of noise (random variation) to data to obfuscate individual patterns.
- **Security risks** can be present if data provides special insight into infrastructure and the locations of the people who use transit that could be used in a physical attack. Throughout the data management and sharing process, there is also risk of a cyberattack exposing private data.
- **Risks of data misuse** can be present whenever data is shared. Although transit agencies seek to mitigate this risk through data documentation, some users may intentionally or unintentionally misinterpret data, drawing conclusions that are incorrect.
- **Strategic risks** are defined as the risk that sharing data could compromise the transit agency's ability to serve its customers. This includes risks to the transit agency's reputation and the risk that the information will be used against the transit agency (e.g., by competitors).

Transit agency interviewees also noted the costs and effort required to share data. Effective data sharing is built on data collection and management. Many transit agency interviewees indicated these processes are challenging, they lack dedicated staff or a division responsible for data collection and management, and often data collection efforts are not designed with end uses and data sharing in mind.

In addition to maximizing the value they can accrue from sharing their own data, the transit agency interviewees indicated their agencies are interested in accessing external data sources as well. Transit agencies have successfully accessed external data sets through the following models:

- Purchasing data
- Accessing data through a mobility services partnership
- Accessing data through a third party
- Accessing data through legislation
- Accessing publicly available data

These models reflect the evolving nature of mobility, including technology-enabled mobility as a service (MaaS), which is changing the data that is collected and the data analytics needs of public transit agencies. In addition, legislation around data privacy is evolving, and transit agencies may choose to play a role in shaping it. Overall, a movement toward data standards, open data, and open data tools can help transit agencies generate more value from their own data and external data sets. Looking into the future, it is important for transit agencies to set goals that can be accomplished through data analysis and data sharing and develop staff capabilities and data sharing processes to work toward those goals.

The following are the key findings of this report.

Transit Agencies Share Data Frequently and See Many Benefits

- **Transit agencies collect data on the transit system, including route, schedule, and vehicle location data, which is commonly shared and contributes to customer information.** Private developers routinely use route, schedule, and vehicle location

data in customer-facing apps that help transit passengers plan their routes and find out when transit vehicles will arrive at stops and stations.

- **Transit agencies collect a wide variety of data on transit passengers. Sharing this data also generates value, including research that can improve system performance and increase advertising revenue for the transit agency.** Sharing passenger data can generate insights and innovation that are beneficial to the transit agency and may even generate revenue, particularly through advertising. However, these data types have the potential to be used to identify individuals, posing privacy risks.
- **Transit agencies share some data openly and share other data sets directly with partner institutions or individuals through private data sharing agreements.** Route, schedule, and vehicle location data were the most common type of open data shared, but transit agencies also share ridership, on-time performance, survey, and financial data publicly on their websites. In addition, all transit agency interviewees indicated that their agencies respond to public records requests for data. Several transit agencies have established data sharing relationships with research institutions and reported on beneficial insights gained through these relationships.
- **Information disclosure laws govern many aspects of data sharing by transit agencies.** These laws vary by state but may include exemptions for data pertaining to individuals or for specific data types.
- **Private companies in the MaaS industry, including private mobility providers and user information app developers, are interested in transit data.** Some expressed a willingness to further discuss the potential to purchase data from transit agencies; others questioned the notion of monetizing data collected by public transit agencies. They are especially interested in geospatial details of transit stations as well as data, such as passenger counts, that can help them plan their services.

Transit Agencies May Be Able to Increase the Value of Data Sharing in the Future with the Development of New Data Standards, Moving Toward Open Data and Tools, and Leveraging the Interests of the Private Sector

- **Data standards have the potential to increase the value of public transit data and make transit agency use of external data sets more efficient.** The majority of the transit agency interviewees were supportive of the idea of standards for public transit data types, noting that standards could promote the development of shared tools and other resources. Transit agencies are looking to external organizations for standards creation and adoption.
- **Open software tools could augment the value of public transit data and help transit agencies use external data sets.** A general movement toward open data and open tools can continue to benefit transit agencies.
- **Private companies in the MaaS industry, including private mobility providers and user information app developers, are interested in transit data.** Some expressed a willingness to further discuss the potential to purchase data from transit agencies; others questioned the notion of monetizing data collected by public transit agencies. They are especially interested in geospatial details of transit stations as well as data, such as passenger counts, that can help them plan their services.

Data Sharing Challenges are Part of Broader Data Management Needs

- **Often, transit agency data collection processes are byproducts of other functions of the transit agency (e.g., fare collection, operations, management).** More deliberate data collection efforts can ensure transit agencies can maximize the value of their data.

4 Data Sharing Guidance for Public Transit Agencies—Now and in the Future

- **Collecting, cleaning, processing, documenting, and cataloging data requires significant effort.** Those transit agencies that had developed procedures for processing and cataloging data found that this saved time responding to both public and internal data requests. Transit agencies may consider charging processing fees for public records requests that require significant effort (if allowed under state law).
- **Transit agencies identified internal organizational and technical needs to improve their processes for sharing data.** Data-focused staff can drive transit agencies' data sharing programs, developing goals, identifying needs, creating internal data management processes, including a data catalog, and evaluating data sharing opportunities.

Transit Agencies are Beginning to Harness the Value of External Data, But Challenges Remain

- **There is potential value in linking transit agency data sets to external data sets.** This can help transit agencies understand first- and last-mile trips and understand modal alternatives to transit.
- Transit agencies access external data sets, either by purchasing data or leveraging a mobility services partnership. Or in some cases, they may gain access to data through a third party. The transit agency interviewees acknowledged the challenges of negotiating data sharing agreements with private mobility providers, even when they have reached a service agreement.
- **Although private sector data, app, and mobility company representatives expressed interest in cooperating with transit agencies, they also cited privacy concerns as one reason their companies avoid sharing individual-level data with transit agencies.** In some states, transit agencies may need to work with state legislatures to ensure that data on individuals is exempted from state information disclosure legislation. Or transit agencies can work to access information through a third party.
- **Cities are beginning to exercise their regulatory power by demanding private mobility providers submit mobility metric data when applying for operational rights on city rights of way.** Transit agencies can work with cities to ensure that data requirements meet transit agency needs and that data is shared between the two public entities.



CHAPTER 1

Introduction

Data has become a critical component of many fields, and public transit is no exception. Public transit systems generate a substantial amount of data, including vehicle location records, records of passenger boardings and fare transactions, and information on routes, schedules, and real-time alerts. Private companies, including cellular network providers and smartphone app companies, also collect data on location and movement, which is routinely sold. Marketers use location information and mobility patterns to tailor ad campaigns—pushing mobile ads for nearby restaurants or targeting mobile users on their way home from work. Data from cell phone apps, Global Positioning System (GPS) probes in vehicles, and Bluetooth and Wi-Fi signals in mobile devices are collected, processed, and sold by companies, including HERE, INRIX, and Cuebiq.

“The transportation industry is a leader in creating the Internet of Everything, generating vast volumes of data each day through sensors in passenger counting and vehicle locator systems and ticketing and fare collection systems, to name just a few.”

(Rosado 2014)

Public transit customers in many cities are accustomed to receiving information on transit services in real time on their smartphones, using apps developed primarily by private companies and fed by public data.

The reliance of transit agencies and customers on data is only expected to increase, as predicted by the World Bank (Peralta Quiros 2018). New data sources, including location data from smartphones, and data from new mobility options, such as shared scooters and bicycles and transportation network companies (TNCs), can provide additional insight on how people move around their communities. Transit agencies are engaging in partnerships for service provision across modes and with private mobility providers (e.g., with bikeshare providers and TNCs like Uber and Lyft). These models can facilitate and often necessitate data sharing.

1.1 Need for Guidance on Transit Data Sharing

Just as there is value in having data, there can be value in sharing data. Sharing data has the potential to create benefits for transit agencies and their users. Sharing data can facilitate the following:

- Promote **transparency** and can increase awareness of the transit agency and its engagement with transit customers.
- Spur **innovation** and support **research** that can help transit agencies plan better service and operate more efficiently.
- Enable **cost savings** for transit agencies by using outside resources for data processing and analysis.
- Generate **revenue** (e.g., through **advertising**).

6 Data Sharing Guidance for Public Transit Agencies—Now and in the Future

- Support improved **customer information**.
- Support other community functions, such as informing **municipalities, real estate developers,** and even **law enforcement agencies**.
- Facilitate **multiagency and multimodal mobility** solutions.
- Support **benchmarking** activities that help transit agencies track and improve their performance.

A decision framework for data sharing must weigh these benefits against the cost of providing data and the risks of sharing it. Preparing data for sharing can be resource intensive and can be both technically and organizationally challenging for some transit agencies. Technically, data sharing requires knowledge of data cleaning, processing, and storing, including the application of appropriate cybersecurity and privacy protection measures. Data sharing often includes terms of use or a data sharing agreement, and transit agency staff must know which terms to attach in each case. Organizationally, transit agencies are working to find the most efficient staff structures to manage their growing volumes of data and the data sharing requests they receive.

Developing organizational structures and technical knowledge for data management can help transit agencies establish protocols for assessing risks when they make decisions about data sharing. The risks of sharing data include privacy risks, security risks, risks of data misuse, and strategic risks.

These risks are defined as follows:

- **Privacy risks** are present whenever data has PII. Sometimes, the potential for a data set to be combined with other data sets increases this risk. Transit agencies can take steps to protect privacy that include encryption of identifiers, aggregation, and adding noise (random variation) to data to obfuscate individual patterns.
- **Security risks** can be present if data provides special insight into infrastructure and the locations of the people who use transit that could be used in a physical attack. Throughout the data management and sharing process, there is also risk of a cyberattack exposing private data.
- **Risks of data misuse** can be present whenever data is shared. Although transit agencies seek to mitigate this risk through data documentation, some users may intentionally or unintentionally misinterpret data, drawing conclusions that are incorrect.
- **Strategic risks** are defined as the risk that sharing data could compromise the transit agency's ability to serve its customers. This includes risks to the transit agency's reputation and the risk that the information will be against the transit agency (e.g., by competitors).

Finally, data sharing decisions require an understanding of the laws that govern information sharing and data privacy. In many instances, as public agencies, transit agencies are required to share data if requested. However, these laws generally include provisions to protect individual privacy and other limitations. Including the correct contractual provisions in a data agreement requires technical understanding of the data as well as legal expertise.

Figure 1 summarizes the sharing characteristics of the major public transit data types, based on the literature review and transit agency interviews. Public transit data can be classified into two broad categories: (1) data pertaining to passengers and (2) data pertaining to the public transit system itself. One exception is incident data, because incidents may involve passengers as well as infrastructure, vehicles, and transit agency staff.

Data pertaining to passengers includes passenger count data and survey data as well as newer data types, such as fare or bank card transactions, video, Wi-Fi, Bluetooth, and app and webpage usage data. Most (but not all) types of passenger data contain records of individual passengers or records pertaining to a specific card or device that has the potential to identify an individual. This is a critical distinction for data sharing, because the sharing of individual records poses a privacy risk.

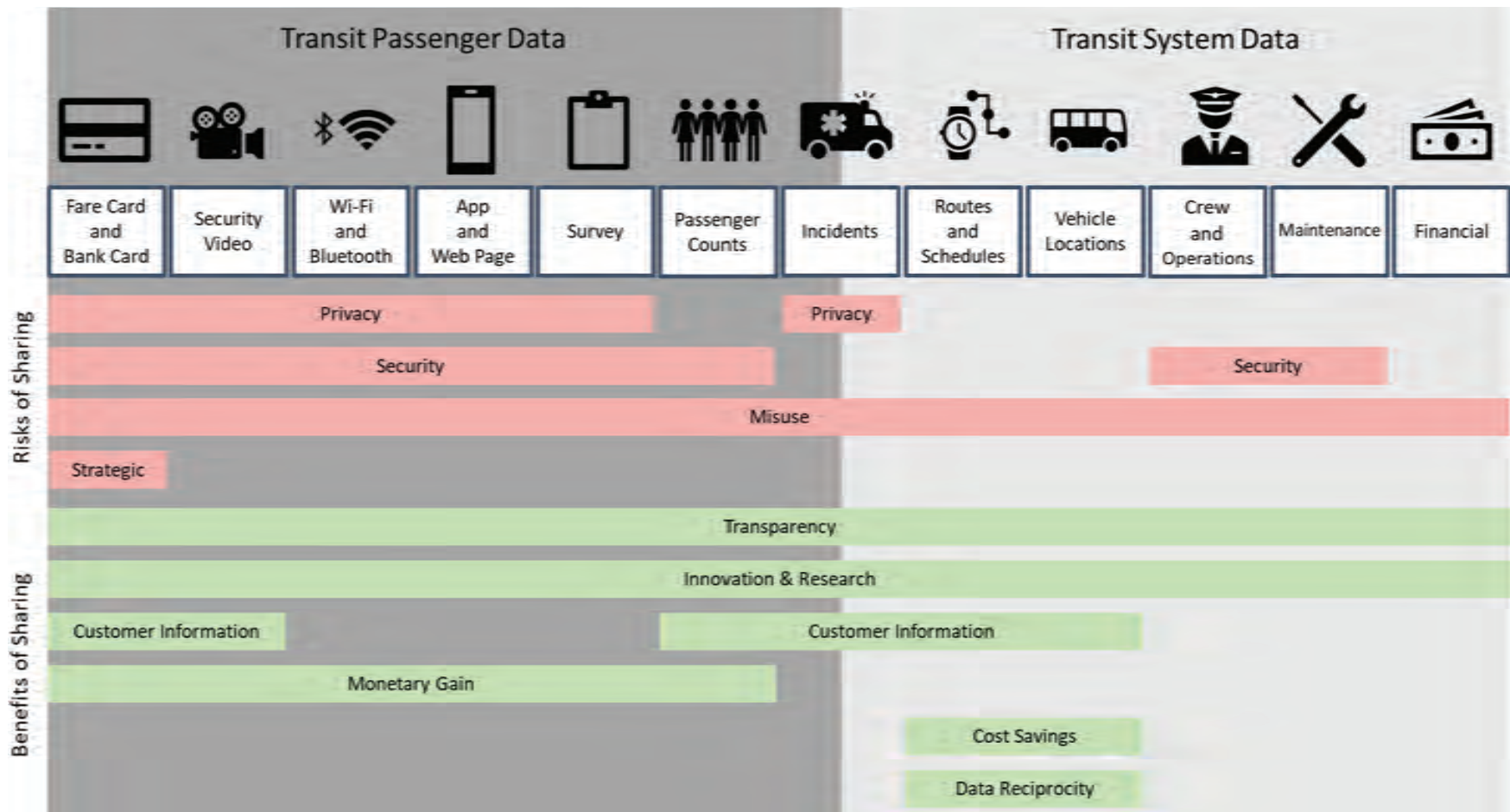


Figure 1. Transit data types and sharing characteristics.

8 Data Sharing Guidance for Public Transit Agencies—Now and in the Future

Data pertaining to the transit system includes route and schedule data, vehicle location data, maintenance, staff and operations data, and financial data. These data types typically do not contain privacy risks; however, there may be security risks associated with the sharing of some of these data types. Although security was not raised as an issue by the majority of the transit agency interviewees for this study, a few interviewees indicated that their agencies refrained from sharing some detailed transit system and transit system usage (passenger) data because of concerns that it could be used to stage an attack on public transit infrastructure and the people who use it. Additional information about transit data types and their data sharing attributes is included in Appendix A.

1.2 Research Scope and Key Questions

There are several models for data sharing. Broadly, this report considers the following types of data sharing:

- **Public Data Sharing (Open Data).** Data is shared publicly in an online data repository or dashboard through an API or in a public-facing report. These sharing models promote transparency and can spur innovation, but they cede control over how the data is used.
- **Private Data Sharing.** In a private data sharing agreement, data is shared with a specific partner, often with a nondisclosure agreement. These types of sharing models enable transit agencies to meet targeted goals. For example, many transit agencies have research partnerships in which they share data with researchers who address transit agency planning and performance goals.

Data Selling and Trading

The research team did not find documented instances of data **selling** by transit agencies, although there was one example of a transit agency charging for unique data analysis that was requested by a client. In general, transit agencies in the United States are subject to public records legislation (see Section 2.6). Although there are exemptions in some states for some data types, these laws generally require public transit agencies to share data when requested. In most states, public agencies may charge a fee for processing public records requests. In this report, charging fees for public records requests as a source of revenue and using transit agency data to increase transit agency advertising revenue are described.

Another potential mechanism to leverage value from public transit agency data is through a data **trade**, in which a transit agency provides data to a partner and receives data from the partner in return. This specific model appears to be rare for transit agencies. Instead, there are many examples of transit agencies providing data and receiving other benefits in return (such as data analysis, publicity, customer information platforms). There are also examples of transit agencies leveraging service agreements and their relationships with customers to access external data sources. These examples are discussed in this report in the broader context of data sharing agreements.

Research Methods

The analysis conducted for this study was based on two categories of sources: (1) a review of research and industry literature and (2) a survey and interview process.

The literature review covered a wide range of sources of information, including academic journals, online publications, news articles, and reports from organizations such as the International Association of Public Transport (UITP), FTA, Shared-Use Mobility Center, Transit-Center, and individual transit agencies. The research team also reviewed legislation and legal proceedings to summarize the legal context for public transit data sharing.

In addition, the research team conducted interviews to collect additional information about transit data sharing experiences from a variety of perspectives. The team used an online screening survey to identify transit agencies with diverse experiences with data sharing. The screening survey was distributed by the American Public Transport Association (APTA) to their members and was also distributed via several different TRB committees focused on public transit data and public transit planning, marketing, and policy. The team also reached out directly to transit agencies with valuable data sharing models that were identified through the information review process or based on their involvement in programs, such as FTA’s Mobility on Demand (MOD) Sandbox Program. The team interviewed representatives from 12 transit agencies of different sizes and across geographies.

The team also interviewed representatives from two cities, one state DOT, and one utility. This enabled the team to document how similar public or publicly regulated agencies address data sharing questions. The cities and state DOT were selected based on involvement in the MOD Sandbox Program. To gain the private sector perspective, the team interviewed representatives from three mobility and location data providers, three private mobility providers, two companies that provide transit data sharing and management services, and one transit planning app company. The team also interviewed representatives from an energy data start-up, an electric utility standards working group, and a nonprofit energy research group to gain additional perspectives. Finally, the team interviewed five experts, including academics and individuals, from organizations that support public transit and shared mobility.

The following sources were used in this report:

Interviews

- Transit agencies
- Other public sector staff (cities, state DOTs)
- Academic experts
- Private sector mobility providers
- Private mobility data owners
- Developers of mobility data platforms
- Representatives from analogous organizations in the utility sector

Review of Information

- Academic journal articles
- Publications from organizations, such as Shared-Use Mobility Center, International Association of Public Transport, and FTA
- News and web media
- Legislation and case law
- Transit agency websites and reports

Report Structure

This report tackles the broad concept of public transit data sharing from two perspectives. The first is the sharing of public transit agency data and the generation of value from this data. The second is transit agency access to external data sources.

Chapter 2 is an action-oriented, how-to guide for transit agencies, developed based on the information gathered and synthesized in this research.

Chapter 3 discusses the key factors determining data sharing decisions, including benefits, costs, and risks. It also includes a discussion of the legal context.

Chapter 4 summarizes models for the sharing of public transit data, and **Chapter 5** considers models for accessing external data sources.

Chapter 6 documents the major challenges for data sharing, and **Chapter 7** summarizes key findings and discusses how the transit data sharing context is expected to evolve in the future. It also notes topics for future research.

Appendix A provides more detail on transit agency data types and external data sources with commentary on their sharing attributes. **Appendix B** consists of the interview guides used in this research.

For reference throughout the report, a list of acronym definitions is provided in Table 1.

Table 1. List of acronyms.

| Acronym | Definition |
|---------|---|
| AFC | Automated Fare Collection |
| APC | Automated Passenger Counter |
| API | Application Programming Interface |
| APTA | American Public Transportation Association |
| AVL | Automated Vehicle Location |
| DOE | Department of Energy |
| DOT | Department of Transportation |
| FIPPs | Fair Information Practice Principles |
| FOIA | Freedom of Information Act |
| FTA | Federal Transit Administration |
| FTC | Federal Trade Commission |
| GDPR | General Data Protection Regulation |
| GIS | Geographic Information System |
| GPS | Global Positioning System |
| GTFS | General Transit Feed Specification |
| GTFS-RT | General Transit Feed Specification Realtime |
| HIPAA | Health Insurance Portability and Accountability Act |
| IT | Information Technology |
| ITS | Intelligent Transportation Systems |
| LBS | Location-based Service |
| MaaS | Mobility as a Service |
| MBTA | Massachusetts Bay Transportation Authority |
| MDS | Mobility Data Specification |
| MOD | Mobility on Demand |
| MPO | Metropolitan Planning Organization |
| NACTO | National Association of City Transportation Officials |
| NTD | National Transit Database |
| PII | Personally Identifiable Information |
| RFP | Request for Proposal |
| SORTA | Southwest Ohio Regional Transit Authority |
| SUMC | Shared-Use Mobility Center |
| TANK | Transit Authority of Northern Kentucky |
| TAZ | Transportation Analysis Zone |
| TfL | Transport for London |
| TIDES | Transit ITS Data Exchange Specification |
| TLS | Transport Layer Security |
| TNC | Transportation Network Company |
| USC | United States Code |
| UITP | International Association of Public Transport |
| WMATA | Washington Metropolitan Area Transit Authority |
| XML | Extensible Markup Language |



CHAPTER 2

Guidance

This chapter provides guidance to transit agencies for data sharing, based on insight gained from interviews and secondary sources, which is described and documented in Chapters 3 through 6. The guidance answers three sets of questions.

1. How can transit agencies maximize the value from sharing their own data? What data should be shared with whom? And what sharing model should be employed?
2. How can transit agencies access external data sources to meet their own transit agency goals?
3. What data must be shared? How should transit agencies respond to public records requests?

The first two sets of questions should be considered in the context of broader transit agency objectives. Organizationally, answering these questions requires the development of staff and processes around data management and data analytics. Establishing staff who are responsible for data sharing is the first of a four-step process shown in Figure 2. Instructions for how transit agencies can achieve each of these steps to answer the first set of questions are provided in the corresponding sections (2.1 through 2.4). Section 2.5 provides guidance on accessing external data (addressing the second set of questions), which is a distinct process from transit agencies sharing their data. The third set of questions involves institutional and legal issues. Instructions for this process are included in Section 2.6.

2.1 Ensuring Data-Focused Staff

The first step in improving data management processes that are critical for effective data sharing is to ensure your transit agency has staff with time and capabilities to make data sharing decisions. Transit agencies can use this Staffing checklist to determine if they have adequate staff to guide the data management and sharing process.

2.2 Establishing Goals and Objectives

Establishing goals and objectives that can be achieved through data sharing requires working across departments and with leaders of your organization to understand needs and goals and how these may be supported by existing or potential data.

Goals that depend on data and data sharing generally fall into the following categories:

- Public transit system performance, including innovation, planning, and prioritization
- Cost savings
- Revenue generation (e.g., from advertising)
- Customer information

Staffing Checklist

- ☐ Do you have a dedicated staff person or division focused on managing data?
If not, consider your transit agency's needs. Large transit agencies likely require a team of data-focused staff. For a small transit agency, a single staff person may be sufficient. For very small agencies, a staff member at a local government agency may play this role.

Data management staff should include individuals with the following skills:

- ☐ Database administration and maintenance, including understanding of security and permissions
- ☐ Data analytics, including the ability to use scripts to automate data analysis processes and work with larger data sets, and an understanding of how analysis of different data sets can answer key questions and achieve transit agency goals
- ☐ Knowledge of privacy risks and techniques that can be applied to preserve privacy of data pertaining to individuals, including personally identifiable information (PII)

- ☐ Do you have a lawyer to help interpret legislation that pertains to data storage and sharing?

Some transit agencies have a lawyer on staff. If not, there may be a lawyer at your state department of transportation (DOT) or local metropolitan planning organization (MPO) who can work with you.

With the help of a lawyer, transit agencies should ensure they understand the following:

- ☐ State-level data security laws
- ☐ State-level data breach notification laws
- ☐ State-level information disclosure (public records) laws and any exemptions that apply to transit agency data
- ☐ State-level tort laws that could be applied in instances of mishandling of private data

Our review of legislation indicated that, in general, federal legislation either does not apply to transit agencies or does not include specific legislation that applies to transit agency data. However, individual transit agencies should evaluate which laws apply to them and monitor changes to federal legislation.

- Transparency
- Facilitating multimodal travel and other community functions
- Benchmarking

Transit agency staff responsible for data management analysis should review their agency's goals across these categories and identify goals that internal or external data can help achieve. Furthermore, they can identify opportunities in which data sharing may help achieve those goals. These opportunities can then be evaluated using the framework outlined in Section 2.4.

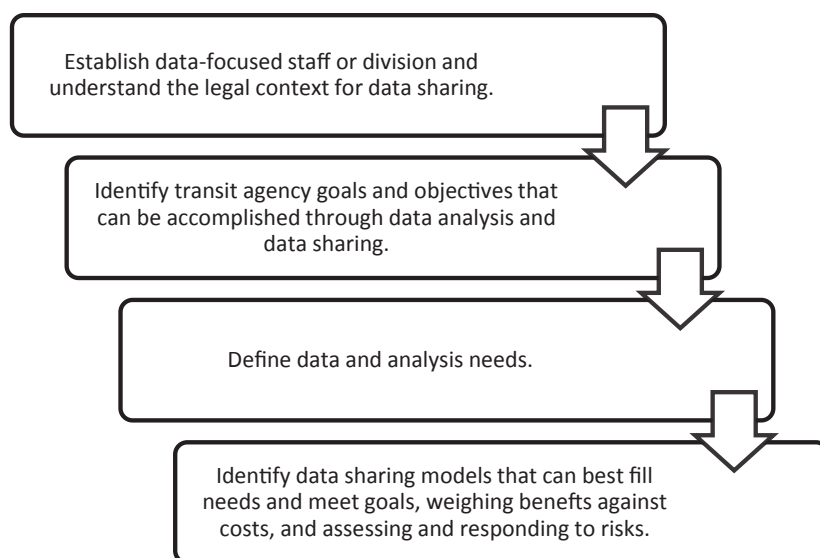


Figure 2. Organizational flow for data management and sharing.

Action Plan: Develop a plan that demonstrates how specific data sharing opportunities can achieve broader transit agency goals and objectives.

2.3 Defining Data and Analysis Needs

Effective data sharing depends on good internal data management. Achieving the goals identified in the previous step may require changes to data collection, data processing, and data documentation to ensure data is used effectively and appropriately when shared. Transit agencies can begin by inventorying the data management and sharing processes they have in place, using the Data and Analysis Needs checklist.

The Data Preparation Needs checklist can be used to evaluate data collection, processing, and documentation needs.

2.4 Evaluating and Selecting Data Sharing Models

Once objectives and needs are clearly established, use these as a basis to evaluate potential data sharing models. Data can be shared publicly, often called open data, or data can be shared with a particular partner, often under a private data sharing agreement.

To evaluate each potential model, consider the following:

- Benefits
- Costs
- Risks

Benefits

Evaluating benefits of sharing data is directly linked to the established transit agency goals and objectives. It is useful to consider the same categories when evaluating expected benefits:

- Public transit system performance, including innovation, planning, and prioritization
- Cost savings

14 Data Sharing Guidance for Public Transit Agencies—Now and in the Future**Data and Analysis Needs Checklist**

- ☐ Do you have a data catalog? Is it complete?

To check for completeness, consider the following:

- ☐ Have you checked in with points of contact across departments to ensure that all data is included in the data catalog?
- ☐ Are there other data types that are not collected but are needed to meet transit agency goals?

This can inform data collection, data purchases, and external data requests.

- ☐ Do you have data sharing protocols in place?

Specifically:

- ☐ Do you have a data sharing risk assessment methodology?
- ☐ Do you have data privacy protection protocols?
- ☐ Do you have a protocol for responding to information disclosure (public records) data requests?
- ☐ Do you have a method for making data sharing decisions and forming data sharing agreements, including designated decisionmakers?

Data Preparation Needs Checklist**Data Quality**

- ☐ Is the data sufficiently accurate and precise to meet objectives?
- ☐ Is it sufficiently clean (free of erroneous records)?

Data Coverage

- ☐ Does the data have sufficient coverage of transit users to draw insights?
- ☐ Are there biases in which people or vehicles are included?

Data Ownership

- ☐ Does your transit agency own the data? If data pertains to individuals, is there a mechanism to get their permission to use the data?

Data Processing

- ☐ Do data sets need to merge to answer questions and meet objectives?
- ☐ Is the data formatted to facilitate efficient analysis?
- ☐ Does the data need to be aggregated to protect individual privacy or to suit a particular audience?

Data Documentation

- ☐ Is data documented?
- ☐ Does each data set include a data dictionary?
- ☐ Are important caveats or assumptions included with each data set?

Data Sharing Models

Public Data Sharing (Open Data): Data is shared publicly in an online data repository or dashboard through an API or in a public-facing report. These sharing models promote transparency and can spur innovation, but they cede control over how the data is used.

Private Data Sharing: In a private data sharing agreement, data is shared with a specific partner, often with a nondisclosure agreement. These types of sharing models enable transit agencies to meet targeted goals. For example, many transit agencies have research partnerships in which they share data with researchers who address transit agency planning and performance goals.

Also consider: Is it more efficient or effective to perform the analysis and/or produce the data product internally?

- Revenue generation (e.g., from advertising)
- Customer information
- Transparency
- Facilitating multimodal travel and other community functions
- Benchmarking

Not all benefits can be quantified, but qualitative descriptions of expected benefits can guide data sharing decisions. For example, it may be useful to describe qualitatively what types of innovation may occur as the result of data sharing. How could public transit system performance improve? A key quantitative measure to include is how much internal effort can be saved.

Consider how benefits vary under different sharing models. Benefits of open data can be uncertain, because they depend on how the data is used. In a research partnership, the expected benefits are more likely to be clearly defined.

Costs

Data sharing requires internal effort by transit agency staff. It is important to consider how much effort would be required under different data sharing models and how this compares with the effort that would be required to meet data-driven objectives internally. Public data sharing (open data) can require significant data preparation to ensure the data is usable by diverse audiences. Private data sharing relationships may require ongoing management. In some cases, research partners are paid, requiring additional outlays.

Risks

In conducting risk assessments of data sharing, it is important not only to have a procedure to evaluate risks but also to have a protocol for what to do when risks are present.

Generally, there are four types of risks related to sharing of transit agency data:

- Privacy
- Security
- Misuse
- Strategic

According to transit agency interviewees, data privacy is a major concern. Each transit agency should develop a privacy risk assessment and protocol that it is comfortable with as an organization. Data privacy is an evolving topic. Legal definitions of data privacy vary and are expected to change over time. The Privacy Risk Assessment checklist can serve as a general guide for a privacy risk assessment.

Privacy Risk Assessment Checklist

- ☐ Does the data contain names, addresses, or other personal data, such as Social Security numbers?
- ☐ Does the data contain individual records or records pertaining to a small sample of individuals that could be used to identify an individual based on their travel patterns?
- ☐ Could the data be linked to other available data sets and used in combination with these other data sets to identify individuals?

A privacy protocol consists of a set of actions or rules to follow if data presents a privacy risk based on the assessment. Depending on which privacy risks are present, actions could include the following:

- Do not share data at all.
- Do not share data publicly.
- Process data before sharing. This could include censoring, aggregation, or adding noise (random variation). Specific processing protocols should be defined for different data types. Consider how processing impacts the potential to use the data to meet research and analysis goals (e.g., aggregation limits analysis of how individual users behave over time).
- Share data with partners contingent on requirements, including the following:
 - Ensure data recipients are trained in using private data.
 - Ensure data recipients have a secure method to store data.
 - Require data recipients to sign a nondisclosure agreement, ensuring that they will not share the data.

Security, misuse, and strategic risks should also be assessed. Each transit agency must determine how important these risks are for data sharing decisions. Risks are likely to vary by data sharing model. The Other Risk Assessment checklist provides an example risk assessment method that could be tailored.

Terms

Once a sharing model is selected, consider attaching terms. When sharing data that poses privacy concerns or other risks, it is important to attach terms that restrict additional sharing or publishing, and mandate cybersecurity measures.

Even with open data that does not pose privacy risks, terms can protect the data provider by including disclaimers about accuracy and use. At Transitland, an open data initiative

Other Risk Assessment Checklist

- ☐ Is the data likely to be misinterpreted by users and what would be the consequences of misinterpretation? Can this be avoided through data processing and documentation?

This risk is mostly limited to public data sharing, because private data sharing agreements can require public transit agency approval of products based on the data shared.

- ☐ Could sharing the data harm public perception of the transit agency?
Some transit agencies are hesitant to share data that reveals poor performance. However, this risk should be weighed against the benefit of transparency.

- ☐ Could sharing the data create an information asymmetry, in which a competitor to public transit has more information than the transit agency?
Given public records laws, this situation is often unavoidable. This risk must be weighed against the benefit of transparency and increased awareness of public transit through data sharing. Transit agencies can also try to access data from private mobility providers.

- ☐ Does the data pose a security risk? Could it be used in an attack?
Sharing any data on people's and vehicles' locations poses some security risk. Transit agencies must decide how much to weigh this risk against the benefits of data sharing.

focused on General Transit Feed Specification (GTFS) data, they suggest using an open data license and provide a model license to use as a starting point (<https://transit.land/an-open-project/#for-data-providers>).

2.5 Accessing External Data

Transit agencies may find that they need external data sources to achieve their established data-driven goals. Appendix A describes the types of external data sources that transit agency interviewees expressed interest in and their potential uses for transit agencies.

In general, transit agencies acquire external data through four possible models:

- As a direct purchase
- Through a service agreement with a private mobility partner
- Through a third party
- By accessing public data sources (e.g., census or the National Transit Database data)

As with decisions about sharing transit agency data, decisions about accessing external data should begin with definitions of objectives and needs. Transit agencies may need to spend significant effort engaging with external partners and finding a partner who will cooperate, particularly if transit agencies seek a cost-neutral solution. This effort should be considered when weighing the costs of acquiring external data.

Transit agencies should consider working with cities and state DOTs to coordinate efforts to access external data and leverage the power of these public sector collaborators.

When acquiring data, transit agencies should consider the following factors:

- Sample size
- Data collection biases and coverage across users and geographies
- Data cleaning and processing methods applied
- Data precision and aggregation
- Frequency at which data is provided
- Data format

Transit agencies should consider involvement in the development of data standards for external data sets. Data standards, such as the Mobility Data Specification, which specify required data components and formats, can make it more efficient for transit agencies to use data from multiple providers.

The Shared-Use Mobility Center released a white paper in 2019 that provides guidance to transit agencies seeking data sharing in a mobility partnership. The document includes a flowchart describing short-term actions and longer-term actions (Shared-Use Mobility Center 2019). Longer-term actions include working with lawmakers to modernize public records laws. This can be important in cases where external partners are hesitant to provide data on individuals that could be accessible to the public under existing information disclosure laws.

2.6 Responding to Public Records Requests

Understanding laws in your state is critical for responding to public records requests. Having a data-focused staff or division responsible for responding to public records requests pertaining to the transit agency's data is helpful, because this person or division can ensure they are well-versed and up to date on legislation and can track repeated requests and requests that require significant effort.

In general, information disclosure legislation contains exemptions to avoid releasing data that may pose privacy concerns or create other risks. However, legislation may not “keep up” with new data sets. Therefore it is important to consider risk assessment of public records requests. As shown in Figure 3, if public records requests pose risks, transit agencies may consider making internal changes about data storage or lobbying for changes to information disclosure legislation.

Transit agencies should also consider information disclosure legislation when requesting data from external partners. As described in Section 2.5, information disclosure laws can pose a barrier to accessing external data.

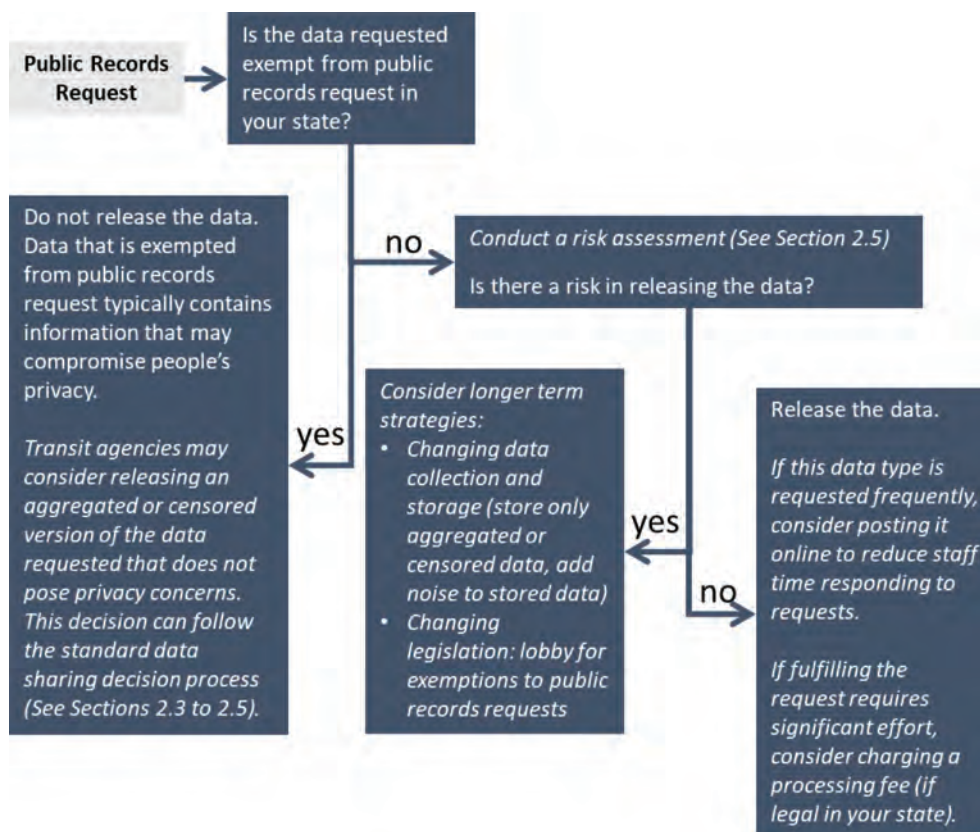


Figure 3. Flowchart for public records requests.



CHAPTER 3

Factors Impacting Transit Agency Decisions about Data Sharing

As described in Chapter 1, transit agencies share data with researchers, private companies, other public agencies, and the broader public for a variety of reasons. The interviews conducted and information gathered in this research effort revealed that, when transit agencies decide whether to share their data, who to share it with, and which model to use to share it, they consider several factors. Public transit agencies are motivated to share their data by diverse expected benefits, such as transparency and innovation, but they also evaluate risks and consider the costs of preparing data to be shared. Legislation around data sharing and data privacy underlies these decisions.

3.1 Benefits

The benefits associated with sharing data are wide-ranging and can be difficult to quantify. Transit agency interviewees frequently commented on the need for methods to assess the value of data and particularly the value of sharing data.

Transparency and Increased Awareness of Transit Services

According to the transit agency interviewees, the general public expects that public agencies publish data in free and open formats. Recent years have seen an emphasis on transparency in government and public agencies, precipitated from the federal level—a 2009 Office of Management and Budget memo encouraged transparency and prompted local governments to develop open data portals.

Publishing data helps transit agencies meet this transparency goal, which can positively impact public perception. Two of the transit agency interviewees identified transparency as a reason for sharing data.

In addition to transparency, data sharing can serve to publicize the transit agency, and may even encourage citizen engagement (Kassen 2013). Increased awareness of transit services was identified as one benefit of GTFS data sharing in a recent study (Schweiger 2015). One transit agency interviewee discussed how their agency’s open data spurred engaging online content and helped “build the agency brand.”

A Pilot to Increase Transit Visibility

In a recent pilot program, Denver Transit partnered with Uber to integrate transit service information for the City of Denver into the Uber app. Residents in Denver can use the ridesharing company’s app to plan trips on Uber and transit, and they can also buy transit tickets from within the app. The pilot program aimed to integrate multiple mobility service alternatives (including transit, bikes, scooters, and ridesharing) into one app to help reduce residents’ dependence on cars (Bosselman 2019). It also increases the visibility of the transit agency’s service to ridesharing riders.

Innovation and Research

In a report on the value of data, Abella et al. (2017) identify the ability of data to spur innovation as a key benefit of data sharing. The innovation impact of open data was pinpointed in the context

of public transit route, schedule, and vehicle location data in *TCRP Synthesis 115: Open Data: Challenges and Opportunities for Transit Agencies* (Schweiger 2015) and also in a report on the value of Transport for London's (TfL's) open data (Deloitte 2017).

Across the United States and abroad, private developers have responded to open streams of public transit route, schedule, and vehicle data by developing travel apps that provide trip planning and vehicle arrival information to customers. Transit agency interviewees commented on the ability of external partners to innovate in quick-changing contexts, such as app development. Open route, schedule, and vehicle location data have also led to the development of additional open-source resources, including products such as OpenTransit Indicators, which calculates performance indicators from this data, and TransitWand, a tool for collecting route and schedule data in the field (Lawson 2016).

Value Generated by Open Data

The gross value added to the economy from companies that develop apps using TfL's open data was estimated to be between £12 and £15 million (\$13 to \$18 million) and to directly support approximately 500 jobs (Deloitte 2017).

Standardization is a key factor that has increased the innovative impacts of route, schedule, and vehicle location. The GTFS has been widely adopted. (Data in this format is available for more than 1,350 public transportation providers as of August 2019.) The standardized format means that innovative tools and products that utilize GTFS can easily be applied across transit agencies. This increases the potential return on investment for innovators and enables sharing of innovations across transit agencies.

In addition to innovative products, data sharing supports public transit research. Nearly all the transit agency interviewees discussed the benefit of external research conducted using their agencies' data. Interviewees noted that within their transit agency there often is not time to focus on research-oriented questions. They named specific examples of research conducted by external partners that benefited their transit agency. These include the following:

- A bus turnaround dashboard
- An origin–destination inference algorithm
- A passenger segmentation model
- An electrification study
- Optimization of dispatcher assignment of work

Sharing data can also spur innovation through the combination of public transit data with external data sets. One transit agency interviewee discussed this potential, noting that transit data might be combined with other data sets, such as health care or census data, to create new insights. For transit agencies that seek to pursue innovative, multimodal collaborations, some level of data sharing is often a necessity.

Cost Savings

Outsourcing data analysis and processing work through data sharing can save transit agencies money. For example, although some transit agency interviewees noted their agencies have developed their own transit planning and real-time information apps, several interviewees noted that developing their own apps in house would be time-consuming and inefficient, compared with allowing external partners to develop them. One transit agency interviewee also noted that, by opening up data, external users of the data help the transit agency more quickly identify problems with the data sets. Increasing the data user pool saves the transit agency time spent looking for missing data and data anomalies. Clearly, this benefit must be weighed against the risk of releasing data that has not been fully vetted. However, for some data sets and partners, this may be a useful model.

Cost savings can also be accrued by releasing data publicly in batches, rather than repeatedly releasing data on a case-by-case basis through individual public records requests. Publishing

Fees for Data?

The electric utility sector is facing similar questions about whether fees should be charged for data products and data-driven services. The industry working group, Green Button Alliance (2018), believes that if data is to be sold, the pricing should be uniform for all potential customers. That is, data cannot be free for some users and for a fee to other users. However, in practice, some state public utility commissions allow utilities to charge a fee for data and others mandate the data be provided for free. Some organizations have developed revenue streams to support an internal team of information technology specialists, and still others see competitive advantage in developing a platform that can be licensed to other utilities that want to offer data products and data-driven services.

frequently requested, nonsensitive data online saves transit agency staff time in the long term.

Revenue Generation

Several transit agency interviewees expressed concerns over the risk of negative perception of data sales. They felt the public would not support the idea of the transit agency profiting off of data on the individuals who use the public transit system. In addition, several transit agency interviewees mentioned their agencies could not sell data, because they are required to provide it to anyone who requests it with a public records request. Under many state laws, transit agencies can charge public records requesters for the effort required to fulfill their requests; however, the transit agency interviewees did not report on any instances in which their agency had charged requesters.

Outside of public transit, there are examples of public entities generating significant revenue from the data they collect. Most state departments of motor vehicles charge for the release of vehicle registration data. Many states charge as much as \$5 per record. Although restrictions prevent data users from using the data to contact individuals directly, this data can be used in aggregate for market research and is frequently purchased both by vehicle manufacturers and data aggregators such as LexisNexis. The Florida Department of Highway Safety and Motor Vehicles was reported to have made \$63 million in 2010 through fees on registration records (Local 10 News 2011). There have been examples of individuals and organizations successfully challenging unreasonably high public records request fees (Grube 2013).

Is There a Market for Transit Agency Data?

Currently, the market for transit agency data is limited compared with the market for vehicle registration data (a significant revenue generator for state departments of motor vehicles). The market for transit ridership data is limited because transit riders are generally only a small share of a region's residents and alternate sources of mobility data (cellphone, GPS) cover a larger market. Route, schedule, and vehicle location data are widely used in private apps but have been made available for free. It is unclear whether or not these app developers would pay for this data. Other types of transit data are infrequently requested from transit agencies, based on the information received from the transit agency interviewees.

Representatives from Location-Based Services (LBS) companies that the research team interviewed for this study expressed interest in collaborating with transit agencies to find better use cases of their combined data sets. Examples from some geospatial mapping technology companies may also shed light on the possible use of transit data for the retail business. One spatial location mapping company interviewee indicated their company had developed visualization and mapping products from the open U.S. Census data. Their date-derived analytical results generate additional values for their business partners.

The private mobility provider interviewee indicated that they use transit data that is publicly available. Responses as to whether they would pay for this data varied. Several expressed particular interest in station geometry data, including station entrance locations and parking facility locations, because this information enables more detailed maps for multimodal connections.

Given the expectation of free and open data, public agencies have seen pushback when they attempt to sell data. Dutch agencies attempted to release data to some partners for free while selling data to others, which led to conflicts (Conradie and Choenni 2014). The tone of news stories about TfL's potential financial gain from Wi-Fi data similarly suggested that public agencies profiting off their data can provoke a negative response (Cheshire 2017).

Advertising

Although selling or charging fees for transit agency data is not a major consideration, several transit agency interviewees described the monetary benefits accrued by using their agency's data to increase advertising revenue. At least five of the transit agency interviewees indicated their agencies had already used data to generate advertising revenue or were considering doing so. One transit agency interviewee noted their agency is in the process of estimating the value of advertising in their transit system, including data-driven, targeted advertising. Two others described how their agencies used ridership data to price space in their transit system.

According to TfL's 2017–2018 Advertising Report, 20% of the UK's outdoor advertising by value is owned by TfL. In 2017, TfL provided customer segmentation data to advertisers. This depersonalized and grouped data from smart ticketing was overlaid with demographic market segmentation data from a private marketing company (Experian) to help prove to advertisers that they are reaching their target audiences. According to the report, advertising revenues in the fiscal year 2017/2018 were £152.1 million (\$185 million) (Transport for London 2018). With ongoing research on location-based advertising and its increased prevalence, there may be increasing latitude for transit agencies to generate revenue by leveraging their data.

Customized Data

One transit agency interviewee noted that their transit agency had occasionally sold bespoke analysis to clients. This consisted of specially requested analysis that would otherwise not be performed by the transit agency. This type of model avoids the privacy risks of sharing data directly and may alleviate the public perception risk of profiting off data that is perceived as a public good. Another potential revenue generator discussed in the transit agency interviews was the potential for transit agencies to sell the data infrastructure expertise they developed to share data, particularly expertise in the development of APIs that feed large volumes of real-time data to developers.

Customer Benefits

Perhaps the most significant benefit that transit agencies consider when sharing data is its potential to positively impact customers. Travel apps that help customers plan public transit trips and alert them to bus and train arrivals can save customers time. According to a study of open transit data, the primary reason transit agencies have cited for releasing route, schedule, and vehicle location data is to provide customers with more information (Schweiger 2015). In London, 42% of residents use mobile phone apps that use information from TfL's open data feed. These open data feeds provide customers with greater certainty about their journeys and potentially save passengers time. These benefits to customers from TfL data provided via apps were estimated at between £70 and £90 million (\$85 to \$109 million) per year in time savings (Deloitte 2017).

The innovative studies spurred by open data can impact customers as well. Research that helps transit agencies operate more efficiently or plan service better ultimately translates into benefits for public transit customers. The impacts of external research on customers may be

Getting More Value from Third-Party Apps

Transit agencies may be able to generate more customer benefits from third-party apps in return for the data they share. In Tampa, Florida, a pilot program embedded Open311, a service that allows users to report issues to the local government and transit agencies, in the open-source OneBusAway app (Barbeau 2018B).

Transit agency interviewees also described two additional information types that could be presented in customer-facing apps: **crowding information** and **fare information**. As data and processing methods improve, the possibility to provide reliable crowding information in real time is increasing. Although some transit agencies are concerned that reporting on crowded trains and buses may discourage public transit use, others see this as an important way to increase customer knowledge and improve their experience. Fare information is part of the GTFS standard but is not supplied by all transit agencies. Wang (2014) noted that the existing GTFS standard is insufficient in the way it describes fares and proposed an extension to GTFS to model complexities in fare structure, such as time of day variance, distance-based, and free transfers. At least one transit agency interviewee noted their agency was working to include fare information in its public information feeds and ultimately in transportation apps.

significant, particularly in data sharing models where transit agencies are able to influence external research to target their needs.

Facilitating Community Functions and Multimodal Mobility

Transit agencies also reported that data on passengers is requested by real estate developers, municipal planners, and law enforcement officers. Transit agencies attempt to support these community needs while also protecting private information on their customers. In a recent example in Boston, Massachusetts Bay Transportation Authority (MBTA) video surveillance data and fare card data were used to locate a kidnapped woman (Flanigan 2019).

Public transit is just one part of a multimodal transportation system. In some cases, public transit agencies partner directly with TNCs or micromobility providers. Data sharing is often critical to building a well-functioning multimodal transportation network. Some argue that the integration of public and private mobility options, which generally requires data sharing, makes cities more attractive to investors with private capital, increasing the number of skilled jobs available and widening the city's tax base (Hemerly 2013).

Benchmarking

Transit agencies share data for benchmarking, which helps them understand, track, and improve their performance. The National Transit Database (NTD) is a repository of transit agency information. Transit agencies that receive funding from FTA under the Urbanized Area Formula Program (§5307) or Other Than Urbanized Area (Rural) Formula Program (§5311) are required to submit data to the NTD. The data is frequently used by researchers to understand trends in public transit performance. The fact that it is standardized across agencies makes it easy to use for cross-agency studies.

Transit agencies can also pool data privately to benchmark performance. The American Bus Benchmarking Group is a consortium of bus agencies that share data and best practices. It aims to help its members understand their performance by making comparisons about practices and outcomes across agencies.

3.2 Costs and Effort

There are many steps required to prepare data for sharing. Figure 4 summarizes the common elements in preparing data for sharing. These steps require staff time and often also require contracting with external vendors. In many cases, these steps are required even for internal data use, a factor identified by several transit agency interviewees. Interviewees noted that good internal data management practices make data sharing easier. For example, a well-documented internal data repository helps transit agency staff make use of data and also reduces the additional steps required to distribute data. However, several transit agency interviewees noted that much of their agencies' data was not collected with analysis in mind, was not stored in a centralized location, and was not documented for external use. As such, preparing data in response to data requests often requires significant effort. There is additional effort required to conduct privacy and other risk assessments and to develop any licensing agreements necessary.

Data Cleaning

Although data cleaning is important, it can require significant effort. If a planning goal depends on high-level metrics that aggregate data across months or years, an imperfectly cleaned data set may be sufficient. In contrast, when data is shared to provide customer information, errors in the data can be problematic. If the route, schedule, and vehicle location information that transit agencies share are inaccurate, it may dissuade customers from using transit services, and could even have implications for customers' safety. Transit agencies interviewees emphasized the importance of vehicle arrival prediction quality and described issues such as “ghost buses”—in which bus arrivals are predicted but do not occur—that they are actively working to combat.

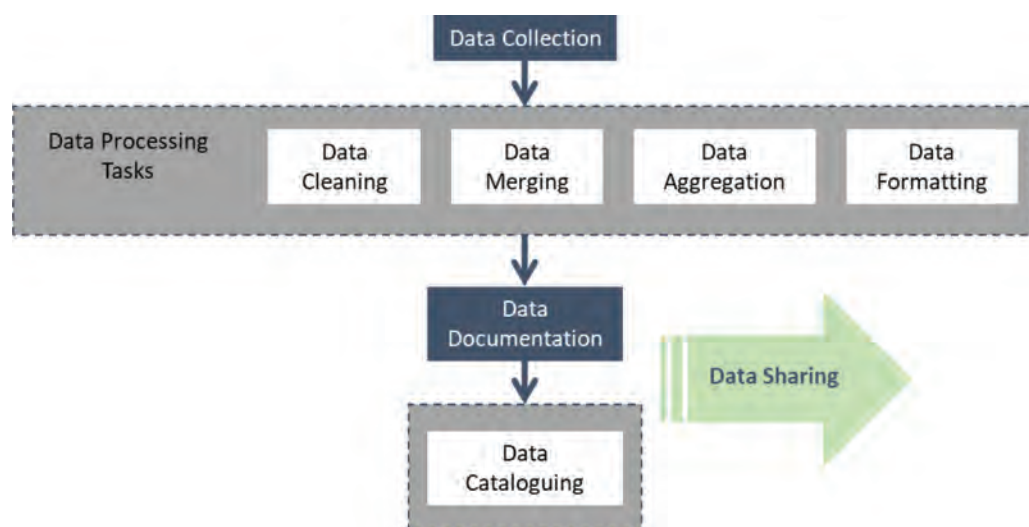


Figure 4. Process of preparing data for sharing.

This same level of data quality is not necessary for all data types. Two transit agency interviewees of agencies that release the most data publicly both noted that data does not have to be perfect to be released. They see benefits from releasing data even if it has minor flaws. As long as the issues and caveats are described in the data documentation, releasing data promotes transparency and can spur research and innovation. In some cases, external data users can actually help the transit agency identify and fix problems with the data.

Data Merging

Often transit agencies merge multiple data sources to produce more useful data products. Common examples include merging automated passenger counter (APC) and farebox data, connecting data to GTFS identifiers, and assigning data to trips or vehicles. These processes make data easier to use and increase its value both internally and externally.

Adding Value to Data

Sometimes internal analytical effort by transit agencies can pay off by making data more useful and desirable to other users. Looking to the energy sector, utilities in New York and California are creating “interconnection maps” to show third-party service providers exactly where they can provide distributed energy resources, and the price that the utility is willing to pay for load at those network nodes. Creating these maps is one step in streamlining the procurement of third-party services by publicly sharing localized electricity needs and creating a standard process for third-party service integration. In California, the utilities work independently, posting their interconnection maps on their websites (California Public Utilities Commission 2008). In New York, the utilities are collaborating with each other as well as the regional transmission operator to standardize data collection, management, and load forecasting methods (Joint Utilities 2016).

Transit agencies may create similar data products that make their data more accessible and valuable to third parties, including private developers and private mobility providers. Transit agencies can evaluate the potential benefits of creating these products, which could include revenue generated from processing fees and transit-supportive development against the effort required and the potential strategic risks of releasing these data products.

Other research questions require merging transit data with external sources, such as weather or census data. Transit agencies may opt to do this task internally, or they may share the data sets publicly or with a research partner that will complete the task.

Data Aggregation

Data aggregation refers to any process in which individual records are combined to produce summary data, for example, combining individual boarding or origin–destination data to provide estimates of average weekly ridership on a route. Transit agencies must make decisions about aggregation prior to sharing data. Transit agency interviewees reported that they aggregate data for a variety of reasons, including making data easier to use and understand (particularly for non-technical audiences), minimizing data storage needs, and protecting individuals’ privacy.

Aggregation is an important tool given the variety of audiences for transit data. One transit agency interviewee noted that different audiences are interested in different levels of aggregation. Although researchers typically prefer disaggregate data, journalists, advertisers, and real estate developers typically seek some level of aggregation so they can draw conclusions and make decisions about actions to be taken without having to perform a significant amount of analysis themselves. Providing aggregated statistics on things like ridership, on-time performance, and vehicle crowding can also prevent some types of data misuse, in which external users misunderstand aspects of the data and perform analysis that leads to incorrect conclusions. However, more detailed disaggregate data, when analyzed correctly, can spur research that generates new insights that can benefit the transit agency. Some transit agencies provide both disaggregate data for download and an interactive dashboard that allows the user to view aggregated information, with data grouped by time period and route.

The transit agency interviews also revealed many examples of data aggregation for privacy protection. This is described in Section 3.3.

Why Aggregate?

For internal use:

- Aggregation reduces data storage needs and protects against cyberattacks of individual records.

For external use:

- For some audiences, aggregation helps them understand the data and prevents misuse.
- Aggregation of individual records prior to sharing can protect individuals' privacy.

Data Formatting

Transit agencies may format data to make it easier to use or to conform to data standards. Standardizing data prior to sharing can produce additional value, for example, by encouraging standardized, open-source tools, as has been the case with GTFS. However, standardizing data also requires additional effort. A discussion of the advantages of data standards for data sharing and the challenges of developing and adopting data standards is included in Section 6.2.

Data Documentation

In general, some form of data documentation, typically including the development of a data dictionary, is required prior to sharing data. Although good, detailed documentation is critical when data is shared publicly, more basic documentation may be sufficient if data is shared with a partner under an ongoing collaborative relationship.

Data released without sufficient context and metadata (including information on assumptions inherent in the data and data dictionaries) is susceptible to misuse (Conradie and Choenni 2014). Data users need to know field definitions as well as any assumptions and caveats. Data field definitions and possible values are typically provided in a data dictionary. This process is especially important when data is shared externally. All the transit agency interviewees indicated their agencies provide some documentation with the data they share. Some expressed that this process can require significant effort and that it is sometimes a challenge to determine what level of detail of documentation is sufficient. Good documentation can help prevent misinterpretation and misuse of data but takes time to develop. The use of data standards can address this challenge, because transit agencies can rely on centralized documentation of data following the standard format.

Data Cataloging

Not all transit agencies have data catalogs, but they can be useful for data sharing. In fact, most of the transit agency interviewees noted that their agencies do not have a centralized data repository, and that data was stored in a variety of locations across the transit agency. The advantages and need for a centralized data catalog were explained in several transit agency

interviews. A centralized data catalog can serve both internal data analysis and data sharing. Parts of the data catalog may be made open to the public, with access to other parts granted to certain partners or limited to transit agency staff. The catalog can help internal staff find and use data collected across divisions and can also ease the data sharing process, saving the transit agency time responding to data requests.

One transit agency interviewee indicated their agency developed a public-facing dashboard where users can view and download many types of data. The interviewee noted that the dashboard saves time responding to internal data requests as well, because people from other divisions can “help themselves” to data. Having such a catalog requires staff effort to maintain. Many transit agency interviewees noted that the lack of a staff member or group dedicated to such an effort was the reason their agency did not have a catalog.

3.3 Risks

The primary risks that may impact public transit data sharing decisions are privacy, security, data misuse, and strategic risks. Section 2.4 includes checklists and guidance to assist transit agencies in identifying and addressing these risks. This section provides context and examples to illustrate these risks based on the interviews conducted and the review of literature and information.

Privacy

There are several sources of privacy concerns with public transit data, including the following:

- Personal data collected, such as registration information associated with fare cards (names, addresses, etc.).
- Anonymized individual data that risks re-identification when combined with other data sets.
- Anonymized individual data that risks re-identification even without additional data sets (PII).
As public transit agencies increasingly integrate their electronic fare systems with other modes and payment systems (such as credit cards) re-identification becomes increasingly possible.
- Facial recognition of video data.

Examples of re-identification of anonymized data occur across fields. For instance, in 2008, Netflix released data on movie ratings by individuals that they believed had been anonymized, but researchers at the University of Texas at Austin proved that they could identify individuals (National Academies of Sciences, Engineering, and Medicine 2018). Similarly, when the New York City Taxi and Limousine commission released data on taxi rides in 2014, a data scientist was able to identify individual trip origins and destinations and amount paid by combining the data set with medallion numbers visible in celebrity photographs (Lubarsky 2017). Transit agency interviewees expressed the need for guidance and protocols to follow to assess and reduce privacy risks.

How Important Is Privacy?

In assessing privacy risks, transit agencies may consider how important privacy is to their customers and their customers’ willingness to provide personal information to public agencies in return for benefits. Studies have shown that people are willing to trade privacy for benefits. According to a 2012 Pew study, almost three-quarters of smartphone owners get location-based information on their phones. However, people also appear to be selective in which sources they provide information to. The study found that more than half of app users surveyed had uninstalled or chosen not to install an app because of privacy concerns (Brakewood and Paaswell 2017).

In a focus group on transit agency apps, most users said they did not read app privacy policies, although 72% said they understood that their smartphone’s locations could be identified. In a survey on the same subject, most respondents said that transportation

apps should know their location (71%), and 60% said they were not concerned about this (Brakewood and Paaswell 2017).

However, people are concerned with their data being shared, especially if it is not for transportation planning purposes: 50% were “strongly concerned” about having data shared for marketing purposes, compared with only 13% “strongly concerned” about having it shared for transportation purposes. Brakewood and Paaswell (2017) also found that, although 35% of survey respondents were “strongly concerned” with data from transportation apps being shared with a private agency, only 18% were “strongly concerned” with this data being shared with a public agency.

Understanding these tradeoffs is important because there is a cost to maintaining data privacy. Erhardt (2016) argues that strong privacy restrictions, such as data obfuscation requirements, can limit the usefulness of smart card data. Lerner (2012) discusses data privacy regulations in the context of online advertising and suggests that they may inhibit innovation by posing obstacles to start ups and thus favoring large established companies. Transit agencies may be similarly burdened by privacy regulations relative to private mobility providers and private mobility data collectors.

Opt-in Models and Standards for Data Privacy

Researchers often want to access individual records, which some transit agencies are hesitant to share due to privacy concerns. There may be potential to address this challenge with opt-in models, in which individuals agree to share their data (International Association of Public Transport 2018). One transit agency interviewee discussed this option for accessing users’ Wi-Fi and app usage data. If transit agencies can show customers that they will use data to the customers’ benefit and establish trust with their customer bases, opt-in models can allow transit agencies to maintain sensitive data internally and use it for planning purposes. In some cases, transit customers may even opt in to sharing their data with external trusted partners, such as researchers and municipalities if they are made aware of potential benefits.

A possible model for opt-in data sharing comes from the electric utility industry Green Button Alliance. The Green Button Alliance’s DataGuard enables informed consent for customers to opt-in to data sharing and allows customers to decide how and when data is shared. The standard also includes secure maintenance and disposal of data, and self-enforcement or auditing to ensure security.

The Green Button Alliance has taken DataGuard a step further to develop two industry standards: (1) UtilityAPI (2) and Green Button. These standards require companies to educate customers of their data’s existence, and provide opt-in consent with options to set expiration dates on data sharing. Customers can either download their Extensible-Markup-Language-formatted (XML-formatted) data and send it to a third party, or they can use their utility website log-in credentials (as they would, for example, with a Facebook or Google account to log into multiple websites) to share their data with approved companies via Transport Layer Security (TLS) 1.2 encryption. The industry chose XML-format and TLS 1.2 encryption, because these are software standards for open sharing. Green Button data protects user privacy by splitting a user’s data into two parts: usage data and personal data. The usage data does not have personal identifying information, such as name, address, and geographic location. Personal data does not have any usage data.

How Much Aggregation Is Necessary to Protect Individual Data?

The majority of transit agency interviewees indicated that their agencies never shared individual records. Those that share individual records do so only with trusted partners who sign a nondisclosure agreement and undergo training in the handling of such data. Instead, transit agencies typically opt to aggregate individual records prior to sharing.

Transit agency interviewees revealed that the level of aggregation varies. One transit agency interviewee indicated that their agency never releases aggregated data containing fewer than 10 records within a given sample bin. For example, if a data requester asked for hourly boardings at a stop, and the stop had fewer than 10 boardings in 1-hour period, they would not release data for that hour. Another transit agency interviewee described a similar rule, but their agency set their minimum at five records. A third transit agency interviewee indicated that their agency only releases data aggregated to a census tract or Transportation Analysis Zone (TAZ) level. Yet another one specified that their agency only supplies average daily boarding information, typically aggregated to an entire year. The interviewee noted that this choice was not only for privacy reasons but also due to lack of consistent data and data quality concerns.

Multiple transit agency interviewees commented that their agencies' privacy policies felt arbitrary and that guidance on privacy protection would be appreciated.

Data Aggregation to Protect Privacy—Lessons from the Energy Sector

In the electric utility sector, publicly available data are also aggregated to protect individual users' privacy. State regulation of customer electricity usage generally covers the release of individual customer data, wherein a customer can release their electricity usage alone, or with their personal identifying information, to third parties. States also address how utilities make customer electricity usage data available for planning purposes, whether to state energy agencies and regulators or to third-party service providers.

Balancing the need for temporal and geographic granularity, which provides insights into consumer demand for electricity at a given time at a specific location, and the need for privacy is a challenge for state regulators. To achieve this balance, regulators define acceptable levels of geographic granularity.

- Vermont allowed utilities to release aggregated customer data, without personal identifying information, at the municipal level (i.e., aggregated across an entire town or city).
- Colorado developed the "15/15 rule," whereby utilities provide 2 months of customer usage data on a rolling basis (Colorado Public Utilities Commission 2015). It is aggregated across at least 15 customers of the same classification (e.g., large home, small home, small business) within the same ZIP+4 area, without PII. A single customer's load must not comprise more than 15% of the customer group. If there are fewer than 15 customers in a ZIP+4 area, or a single customer's load is more than 15% of the total data, the utilities expand the geographic area to ZIP+2. California (Lee and Zafar 2012) and Illinois (Illinois Commerce Commission 2014) also adopted the 15/15 rule.

Data Censoring

Data censoring may be required prior to sharing data that consists of written descriptions. For example, one transit agency interviewee noted that, although most of the information in their agency's incident reports was likely not sensitive and could be released for transparency purposes, some reports may occasionally contain descriptions of individuals that present a privacy risk. As a result, sharing this data would require significant effort to review the data and scrub any sensitive information.

Strategies and Lessons Learned About Data Privacy

Transit agencies have developed a variety of techniques to address privacy concerns when they collect and share data. One transit agency interviewee discussed the importance of transparency, noting that their agency used several methods to inform customers of Wi-Fi data collection in their stations. Notices described that data would be used to benefit customers through improved service planning. Being upfront about data collection helps mitigate the risk of privacy concerns being raised after the fact. Establishing the benefit to customers may create buy-in.

Another transit agency interviewee specified that their agency has a privacy officer who reviews data requests that have privacy concerns and who also conducted a privacy impact assessment. This organizational structure and proactive approach may also mitigate privacy risks.

Outside of transit, there are other frameworks for privacy that can guide transit agencies. One framework for assessing privacy risk categorizes data in three tiers: open data, restricted data, and highly restricted data that are collected under a pledge of confidentiality (National Academies of Sciences, Engineering, and Medicine 2018). Open data is data for which privacy concerns do not exist. Restricted data may have privacy concerns associated with it and should only be shared with appropriate provisions. Highly restricted data generally should not be shared, and individuals should be informed of its collection and uses.

The National Center for Health Statistics follows a “Five Safes” framework to guide decisions about data access (National Academies of Sciences, Engineering, and Medicine 2018). The Five Safes are as follows:

- Safe projects, in which they consider the specific use of data and determine whether it is “appropriate, lawful, ethical, and sensible”;
- Safe people, in which they evaluate the researchers who will be analyzing data;
- Safe data, in which they look at the information contained in the data and evaluate any potential confidentiality breach;
- Safe settings, in which they consider the security of the facilities where data is stored and accessed; and
- Safe outputs, which considers what types of findings will be released based on the data analysis and evaluates risks, particularly re-identification risks.

There are also technical approaches to privacy protection. One transit agency interviewee described their agency's process of encrypting data using a salt, which is an unknown character string that is added to a unique identifier prior to encryption. This serves as protection against decryption. The interviewee from this transit agency, which has a pilot to collect mobile phone data in collaboration with a private company, also noted that their agency had a process for automatically randomizing data relating to a sample of fewer than 10 individual devices.

Chen et al. (2012) describe the potential of the differential privacy framework, a statistical process for protecting user privacy in data sets consisting of individual user data by adding noise to the data sets. Their case study for the Montreal Transportation System demonstrated

that they could successfully apply the differential privacy framework to smart card data, producing a privacy-protected data set from which the transit agency could perform standard analysis tasks.

Security

In the context of public transit data sharing, physical security risks are defined as the risk of someone using transit data to inform an attack on transit infrastructure. As opposed to privacy issues, security concerns were not emphasized in the transit agency interviews; however, one transit agency interviewee noted that their agency was often prevented from releasing data (e.g., on stop-level boardings) that was deemed security sensitive. Another transit agency interviewee noted that their agency releases data if the requester can demonstrate a research or business need for the data. If not, the agency infers that the request may produce a security concern. Security was not mentioned in the other transit agency interviews with the exception of one interviewee who specifically noted that their transit agency does not perform a security risk assessment for data requests.

Cybersecurity is also a risk. Cyberattacks can compromise private data housed within a transit agency. When the transit agency shares data with an external partner, there is an additional risk that the partner is susceptible to a cyberattack. Cybersecurity risks were not raised in the transit agency interviews. However, this subject has been raised in forums on transit data sharing. For example, this was discussed at the Twin Cities Shared Mobility Data Workshop in July 2019. Additionally, the private company interviewees expressed concerns that transit agencies lack the capacity to guard or manage sensitive information that their companies share with the transit agency.

Misuse

Although security was discussed only occasionally in transit agency interviews, the risk of data misuse was raised in almost every interview. Misuse may be deliberate or accidental, with most transit agencies more concerned with accidental misuse, which they perceive as much more likely. One interviewee noted that data users often do not have the full picture. Because they see only part of the data, they may make incorrect conclusions. Another interviewee noted that they were concerned that users would select the wrong data source or use old, stale data to drive their analysis. One transit agency interviewee described an example in which a third-party app misrepresented the data the transit agency had published, leading to complaints to the transit agency from their customers.

Information about transit agencies that is relayed to customers through apps or published on websites and in newspapers can significantly impact the way customers view transit agencies. Although transit agencies cannot prevent misuse of published data, they can take steps to reduce instances of it. Transit agency interviewees noted the importance of checking data for errors before it is published and of fully documenting data that is published online or provided to partners. In terms of route, schedule, and vehicle arrival data shared with customers through third-party transit apps, several transit agencies are taking steps to actively manage what information is shared (see Section 4.3).

Strategic Risks

Strategic risks consist of any consequences of data sharing that impact the transit agency's ability to serve its function. For example, if data sharing can impact the way the transit agency is perceived by its customers or its ability to provide good service to its customers, there is a strategic risk.

Several transit agency interviewees described concerns about public perception. Particularly when asked about the possibility of selling data, they noted that this could cause their agencies to lose their customers' trust. There are also varying perspectives on the strategic risks of open data. One interviewee commented that some transit agencies are concerned with releasing data that shows things like poor on-time performance or overcrowding on their transit system. In contrast, the interviewee believed that releasing data promotes transparency and provides their customers with the best information available to navigate the transit system. In short, there may be strategic risks associated with releasing and not releasing data.

The UITP identifies a different set of strategic risks in their guidance document on data sharing. They discuss that there may be a strategic risk of sharing data of high commercial value for free (International Association of Public Transport 2018). Their report hypothesizes that, when certain data sets are shared, it may actually cause power to shift away from the transit agency. As an example, when transit agencies share GTFS and GTFS-Realtime (GTFS-RT) data, and these are used by third-party apps, the third-party apps collect information on customers that the transit agencies may not have access to. This information asymmetry may disadvantage transit agencies and hamper their ability to best serve their customers. This particular instance of information asymmetry was mentioned in several of the transit agency interviews and is discussed in more detail in Section 4.3.

3.4 Rules and Legal Issues

Laws around data privacy and data management can guide transit agencies in their data sharing practices. However, as Hemerly (2013) cautioned, technology has been developing more quickly than the legislation to keep up with it, which can lead to concerns and conflicts over what data is public and what data is private.

Legal Protection of Data Privacy

At the federal level, there is no general constitutional right to privacy of one's personal data. It is guaranteed only in two cases: "(1) where the release of personal information could lead to bodily harm . . . , and (2) where the information released was of a sexual, personal, and humiliating nature. . . ."¹

However, the Federal Trade Commission (FTC) has broad authority to protect consumers from unfair or deceptive practices that put consumers' personal data at unreasonable risk. For example, the FTC has pursued enforcement actions against companies for "failure to maintain reasonable and appropriate data security for consumers' sensitive personal information."² There is precedent, however, that transit agencies, as agents of the state, may be immune from FTC jurisdiction in many contexts.³ Similarly, a transit agency may be held liable under state tort law for mishandling a customer's personal data; however, some transit agencies may have sovereign immunity from these suits (Thomas 2017).

For some sectors, federal privacy laws regulate the collection and dissemination of certain types of information. For example, the Health Insurance Portability and Accountability Act (HIPAA) applies to sensitive health information. There are no transportation sector-specific federal laws that govern data sharing by transit agencies; however, FTA has published Open Data Policy Guidelines, which discuss best practices for public sharing of data by transit

¹ *Lambert v Hartman*, 517 F.3d 433

² *FTC v. Wyndham Worldwide Corp.*, 10 F. Supp. 3d 602, 607 (D.N.J. 2014), *aff'd*, 799 F.3d 236 (3d Cir. 2015)

³ *N.C. State Bd. of Dental Examiners v. FTC*, 135 S. Ct. 1101, 2015

agencies (Catalá 2016). These guidelines state that FTA should encourage transit agencies to embrace open data practices for data that does not contain private or personal information or that could create security or safety concerns. The guidelines do not address sharing this more sensitive data with specific parties.

State-Level Legislation

Most privacy law that applies to transit agencies in the United States is at the state level. Certain state constitutions protect an individual's right to privacy. In addition, some state courts have held that an individual's right to privacy must be balanced against a compelling state interest in disclosure (Thomas 2017). At least 28 states have enacted data security laws that apply to state, and sometimes local, government entities (National Conference of State Legislatures 2019). Although state laws vary widely in their scope and requirements, they generally require the development of guidelines and standards for data collection and retention. In addition, some of these laws require agencies to take specific measures to protect sensitive information from unauthorized access, destruction, use, modification, or disclosure. Some state laws also require public agencies to develop an information security plan based on standards and guidelines developed by the state's chief information security office.

State data security laws also vary in terms of which agencies are subject to their requirements. Certain state laws only apply to state-level agencies, not to local or other government entities [e.g., Cal. Civ. Code § 1798.14 (“Each agency shall maintain in its records only personal information which is relevant and necessary to accomplish a purpose of the agency required or authorized by the California Constitution or statute or mandated by the federal government.”); Fla. Stat. § 282.318 (State's Information Technology Security Act applies to each “state agency.”)]. Other state data security laws apply more broadly to other types of government entities [e.g., Ala. Code 1975 §§ 8-38-1 to 8-38-12 (This Alabama law, which requires covered entities to “implement and maintain reasonable security measures to protect sensitive personally identifying information against a breach of security,” applies to “government entities,” defined as “the state, a county, or a municipality or any instrumentality of the state, a county, or a municipality.”)]. The applicability of these state data security laws also depends on how the transit agency is formed, that is, whether the transit agency is a department within a city or state government, an independent authority, or a private operator that is publicly funded and overseen by a state entity.

State laws generally distinguish between publicly available information and personal information. Personal information is typically defined as a name or some sort of unique biometric or genetic print in combination with a Social Security number, driver's license number, or other identification number. Personal information is the subject of states' most stringent regulations. Aggregated, anonymous/de-identified, or publicly available data is often exempt from regulation (e.g., Md. State Govt. Code §§ 10-1301 to 10-1302).

One example of a state data security law is the Minnesota Government Data Practices Act, Minn. Stat. § 13. It regulates how government data is collected, created, maintained, used, and disseminated and applies to “political subdivisions,” including transit agencies that were formed pursuant on a local ordinance (this includes, for example, Metro Transit in Minneapolis–St. Paul). Under the Minnesota act, “Private or confidential data on an individual” may not be disseminated by a government entity for any purposes other than those stated in a warning provided to an individual at the time that individual is asked to supply private or confidential data (known as a “Tennessean warning”), unless the government entity receives the individual's informed consent or the dissemination has been authorized by statute. The warning must inform the individual asked to supply private or confidential data of the following:

- Purpose and intended use of the requested data within the collecting government entity;
- Whether the individual may refuse or is legally required to supply the requested data;

- Any known consequence arising from supplying or refusing to supply private or confidential data; and
- Identity of other persons or entities authorized by state or federal law to receive the data.

State data breach notification laws may also apply to transit agencies (e.g., the Alabama Data Breach Notification Act of 2018). A closer examination of these laws is required to determine whether any of them would impose a notification requirement on the transit agency if an entity with which it shared data experienced a breach.

Public Data Disclosure Laws

The federal Freedom of Information Act (FOIA) provides persons with the right to request access to federal agency records or information. Federal agencies are required to disclose the requested records unless an exemption applies. The federal FOIA does not apply to state or local government agencies, including most transit agencies. Instead, transit agencies are subject to state-level disclosure laws.

States vary in what types of records and which agencies are subject to their disclosure laws. In some states, disclosure laws do not apply at the local government or political subdivision level, meaning that they may not apply to a transit agency (depending on how that transit agency is structured). Some state disclosure laws include exemptions for disclosure of personal information. For example, Illinois exempts private information from disclosure requirements (unless otherwise required by statute to be disclosed). Private information is defined as “unique identifiers, including a person’s Social Security number, driver’s license number, employee identification number, biometric identifiers, personal financial information, passwords or other access codes, medical records, home or personal telephone numbers, [] personal email addresses[,] . . . home address[,] and personal license plates.”

In addition, in some states any information retained by an agency is subject to disclosure, even if the records originated outside the government. For example, New York’s Freedom of Information Law required disclosure of insurance company meeting minutes that were voluntarily and confidentially given to the New York Insurance Department.⁴ In some states there are special laws to govern data that private entities share with transit agencies. At least one state has issued regulations governing the sharing of PII by TNCs. Colorado Rules 6723(l) and 6710(e) require that these companies obtain consent from customers before disclosing any PII to a third party and that they keep records of any disclosure. However, the Colorado Public Utilities Commission has the authority to require these companies to disclose PII in specific situations without obtaining customers’ consent. A few states have laws specifically about sharing video data that require documentation of the “custodians” with whom data is shared (Thomas 2018).

⁴ *Washington Post v. Insurance Dep’t*, 463 N.E.2d 604, 607 (Ct. App. N.Y. 1984)



CHAPTER 4

Models for Sharing Public Transit Data

With the wide variety of data types that transit agencies collect, and the often resource and time-intensive processes required to prepare data for sharing, transit agencies are faced with decisions about which data to share with which audiences. This chapter describes different models for transit data sharing, identified both from the review of literature and web sources and from the transit agency interviews.

Data sharing often involves public–private partnerships, which have mixed assessments in the literature. Public–private partnerships have been used to decrease costs and increase efficiency (Lin and Mele 2012). Generally, they occur when the public entity has a capacity gap to fill. In the case of data sharing, either they have data they cannot interpret, or they want to use a private company’s data to enhance decisionmaking (Mackintosh 1992).

Broadly, models for sharing data can be classified as public or private. Public models, often referred to as open data, make data available to everyone, typically by publishing it online. In private models, data is shared with an individual, institution, or group of individuals or institutions. Much of the time, this data is shared under a data agreement, often including clauses about nondisclosure, preventing the data from being shared more widely. Within each of these classes of models there is considerable variation, as described in the sections that follow. Advantages and disadvantages of public and private data sharing are summarized in Figure 5 and elaborated on in the following sections.

4.1 Public Data Sharing

Public data sharing is prevalent among transit agencies. In a 2015 study that surveyed 67 transit agencies, 83% provided open data (Schweiger 2015). All the transit agency interviewees stated they publish data on their websites. Sharing data openly promotes transparency and can spur innovative use of their data. Additionally, public data sharing was touted by transit agency interviewees for its efficiency. Transit agencies receive large volumes of public records and other data requests, and publishing the data online allows transit agency staff to quickly point requesters to the online portal rather than have to fulfill requests individually. Data users may also download data directly from transit agency websites without interacting with the transit agency at all. Although there is an upfront cost to putting data online and providing necessary documentation, transit agencies believe putting frequently requested, nonconfidential data online saves them staff time in the long term.

According to the literature, open data programs increase the public’s perception of accountability. The assumption is, if the government is doing something wrong, it will show in the data (Brauneis and Goodman 2017). In addition, open data can lead to “citizen-sourcing” or utilizing the collective knowledge of residents to analyze and interpret data being released (Kassen 2013).

| Public Data Sharing | | Private Data Sharing | |
|--|---|--|--|
| ✓ Promotes transparency | ✗ Lack of control over how the data is used | ✓ Research partnerships can ensure data is analyzed to support transit agency needs | ✗ Significant effort and resources may be required to develop individual data agreements and respond to individual data requests |
| ✓ Can spur innovation | ✗ Public release of data that can be used to identify individuals violates customers' privacy | ✓ Training of trusted partners and non-disclosure agreements can enable sensitive data to be safely shared | ✗ Perceived lack of transparency and equity – data is only shared with certain partners |
| ✓ Saves transit agencies time responding to individual requests in the long run. | ✗ Depending on how data is shared, there may be significant effort required upfront | | |

Figure 5. Advantages and disadvantages of public and private data sharing.

By building data portals that respond to citizen needs and requests, cities can build cycles of trust and stewardship of their data (Dawes 2010; Homstad 2018). On the other hand, posting data publicly generally means that public transit agencies cede control over how the data is used. Lack of knowledge about the data available and lack of public capacity to utilize specific data types can also limit its impacts (Shelton et al. 2014). One innovative solution to this challenge was a data competition that one interviewed transit agency hosted. Not only did this bring attention to the transit agency's open data, but it encouraged participants to compete to answer questions posed by the transit agency. Several transit agencies are also considering different methods to ensure the route, schedule, and vehicle location data they publish are used to best serve their customers (see Section 4.3).

Transit Data Types That Are Shared Online

The data that transit agencies publish online includes route and schedule information, system alerts, and the real-time location of transit vehicles. In most cases, route and schedule information is published in the standard GTFS format, and vehicle location data follows the GTFS-RT standard. According to the 2015 study, the most common examples of open data among transit agencies are route and schedule information and vehicle location feeds (Schweiger 2015).

In addition to route, schedule, and vehicle location data, many transit agency interviewees indicated their agencies publish information on performance indicators, including route or line level ridership, passenger counts at bus stops and train stations, on-time performance, and reliability indicators. Transit agencies also provide summaries of survey data, including travel surveys and customer satisfaction surveys. Finally, at least one transit agency interviewee indicated that their agency publishes financial data.

Online Sharing Formats: Reports, Repositories, Dashboards, and Application Programming Interfaces

Transit agencies share data publicly in a variety of forms. All the transit agency interviewees indicated their agencies have data and reports that can be downloaded from their websites. In addition, two have interactive dashboards that allow users to interact with the data in a controlled way. Most of the transit agencies share route, schedule, and vehicle location data using an API, which is essentially a set of methods for retrieving data that makes it easy for developers to use the data.

These formats have advantages and disadvantages, as shown in Figure 6. Although static reports are easy for all audiences to understand, they do not allow researchers and innovators to manipulate the data, which can limit new insights that could be drawn from the underlying

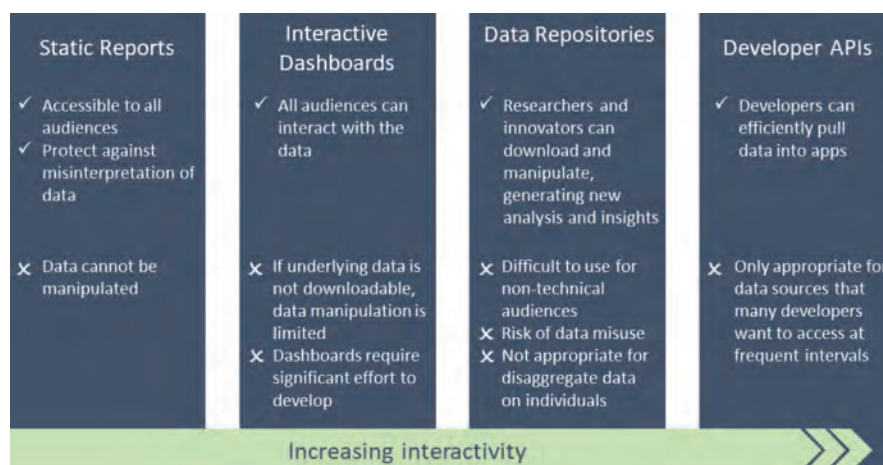


Figure 6. Methods for sharing data online.

data. On the other hand, static reports protect against data misuse, because the analysis is performed by transit agency staff.

Interactive dashboards typically also limit the chances of misinterpretation of data, because they allow for only controlled data manipulation. For example, one transit agency interviewee indicated their agency has a dashboard that allows users to look at service reliability and ridership for a specific route, date, and period of the day. Backend calculations are programmed by the transit agency, preventing incorrect analysis of the underlying data. Dashboards can be a convenient way for people with a wide range of technical abilities to interact with transit agency data. Of course, if the underlying data is not provided, dashboards do not promote new analysis. For example, although the dashboard described allows the user to retrieve reliability information, defined as the on-time percentage for a route (for low-frequency service), the underlying data would enable a researcher to answer more detailed questions about the extent and patterns of schedule deviation. In addition, developing an interactive dashboard requires significant effort on the part of the transit agency. In some cases, third parties have produced public dashboards based on open data. For example, the Bus Turnaround Coalition developed a dashboard that reports on the performance of New York City’s bus routes.

Most transit agency interviewees noted their agencies have a developer website designed for use by software developers. These sites house the transit agency’s API. Developers use these interfaces to access data that is then provided to customers in travel planning and real-time information apps. Some of the transit agencies require the users of this data to register to access an API key. At least two of the transit agency interviewees noted that this model enables their agency to cut off users who overburden the system with too many data requests. Travel planning and real-time information apps have become a key source of transit customer information. As such, transit agencies are reconsidering the best way to leverage their route, schedule, and vehicle location data to provide customers with information. Section 4.3 describes this debate in detail.

Many transit agencies use a combination of the mechanisms in Figure 6 to share data and information publicly. For example, the MBTA’s interactive performance dashboard allows users to select specific lines, dates, and periods when viewing reliability and ridership data. In addition, the underlying data is available for download. The transit agency also has a developer API and publishes reports that share transit agency insights and analysis of their data.

Developing a Dashboard for Multiple Audiences

Providing data that is relevant to citizens and providing it in forms that are usable and responsive is a challenge for public agencies (Abella et al. 2017).

It is possible to have a highly interactive dashboard that provides data in multiple formats. One such example comes from Pecan Street, a non-profit research and development organization that hosts electricity network data from public, academic, and commercial sources. Pecan Street developed Dataport as a platform to manage users, permissions, and data. Pecan Street's greatest challenge to date is incorporating new data resources without increasing the complexity of the platform, which supports online analyses. It began with energy data, and grew to include time-stamped electricity, water, gas, solar, weather, and transportation data. Pecan Street recognizes it has two types of users: "power" users and non-power users. Power users can directly query the database and join any data sets, thus using and manipulating any data set for insights. Non-power users can query and download data in Excel sheets.

Cross-Agency Data Sharing Platforms

Multiple public transit agencies share data with an organization that shares the data using a centralized platform. Such platforms typically require a standardized data format and may therefore require additional effort from transit agencies. Conversely, such models can allow transit agencies to share costs of processing, storing, and documenting data as well as addressing any legal implications of data sharing, potentially reducing transit agency effort.

FTA's NTD is an example of a cross-agency data repository. Recipients of FTA grants are required to submit transit system, ridership, and financial data to the database. The standardized format makes it highly accessible to researchers, who produce studies that can benefit transit agencies.

The National Association of City Transportation Officials (NACTO), which includes both cities and transit agencies, launched the SharedStreets initiative in 2018 (National Association of City Transportation Officials 2018). SharedStreets is an organization that provides open-source software tools and digital infrastructure that allows public entities and private companies to manage and share data about their physical infrastructure and vehicle activity. SharedStreets highlights four core functions of its collaborative platform: (1) to standardize data on physical infrastructure and vehicle activity, (2) to build open-source tools to use data, (3) to anonymize sensitive data on individuals, and (4) to establish a foundation of collaboration and trust (National Association of City Transportation Officials 2018). These functions illustrate the potential benefit of cross-agency platforms to transit agencies. Efforts to standardize and anonymize data can be shared across agencies, and open-source tools that operate on standardized data can provide value back to these transit agencies. These initiatives also represent a way for transit agencies to access private data. Uber, Lyft, and Ford Motor Company are all involved in the SharedStreets initiative. Chapter 5 describes transit agency access to external data sources in more detail.

Many state and local governments have created robust online repositories of their data. Some cities are automating and centralizing their data upload process. These repositories not only enable public access to data, but they facilitate the use of data across agencies (e.g., providing cities with access to transit agency data and vice versa).

Terms of Use for Public Data

The inclusion of terms of use or legal notices with open data provided online varies among transit agencies. For example, the performance data that powers the MBTA performance dashboard can be downloaded without agreeing to any terms of use. Houston METRO, however, includes a legal notice with its data downloads, indicating that METRO retains ownership of the data and that the data is provided “as is.” TransLink has similar terms of use for their Open API data (<https://developer.translink.ca/Home/TermsOfUse>). Section 2.4 provides guidance on the inclusion of terms of use, including model terms to protect transit agencies providing open data.

4.2 Private Data Sharing

Some data is more sensitive but still sharable under the right conditions. Other data types are not commonly requested and therefore have not been published on the transit agency’s website but can be shared when requested. To fulfill these needs, all the transit agency interviewees indicated their agencies share data directly with partners and data requesters. This category of data sharing includes data sharing with partners, such as research institutions, municipalities, and private sector contractors or real estate developers, and fulfillment of public records requests. When sharing sensitive data with partners, transit agencies may require nondisclosure agreements and training prior to releasing the data.

Research Partnerships

Most transit agency interviewees described the importance of sharing data with researchers. Five interviewees indicated their agencies have strong partnerships with a specific university or research institute. In at least two cases, the transit agency pays the university to complete research that aligns with the transit agency’s needs. In other cases, transit agencies do not pay the university, but nonetheless have a long-standing collaborative relationship that allows them to shape their partners’ research agendas. In some cases, transit agencies are willing to share sensitive data, including individual customer records, with researchers in these universities, after the researchers are trained and have signed nondisclosure agreements. Researchers are often able to spend time on questions that transit agencies are unable to, providing significant benefits to the transit agency. Long-standing partnerships allow the transit agency to have a standing data agreement and an established trust in the researchers.

Other Data Sharing Relationships

At least two of the transit agency interviewees indicated that their agencies shared data with local municipalities. In addition, several of the transit agencies provide data to real estate developers when requested. One transit agency interviewee specifically mentioned advertisers as a recipient of data.

Another private data sharing model is followed by the American Bus Benchmarking Group. Established in April 2011, the group consists of midsized bus organizations in the United States. Members share performance data that can be accessed confidentially by members of the group. The objective is to establish benchmarks that can help members understand their transit agency’s performance and identify best practices to improve performance (American Bus Benchmarking Group 2019).

Terms of Use for Private Data Sharing

As noted, sensitive data shared with a partner is typically accompanied by a nondisclosure agreement, according to the transit agency interviewees. In addition, at least two transit agencies

require any publications from their academic partners to be reviewed by the transit agency prior to publication.

On the other hand, it is common to provide nonsensitive data shared through public records requests without any provisions for use. At least two transit agency interviewees noted their agencies do not attach any provisions when they fulfill nonsensitive data requests, which may include aggregate statistics on ridership or boardings and transit system-level information.

4.3 Examples of Data Sharing Models for Customer Information

Route, schedule, and vehicle location data are among the most commonly shared types of transit data. According to information given in *TCRP Synthesis 115*, most transit agencies share this data and do so free of cost (Schweiger 2015). This data is typically shared in standardized GTFS and GTFS-RT formats, and this information has a clear value to customers planning transit trips and finding out when transit vehicles will arrive.

Across the United States and abroad, private app developers have created apps that use GTFS and GTFS-RT feeds to provide information to customers, and this is one of the most prominent examples of transit agencies sharing data. In London alone, there are 600 apps powered by public transit open data feeds, which are used by 42% of Londoners (Deloitte 2017). From *TCRP Synthesis 115*, approximately 40% of respondents to an APTA survey have developers using their open data. For large transit agencies, 68% reported that developers use their data (Schweiger 2015).

These apps have become a key component of how customers interact with transit systems. In many cases, customers are much more likely to receive information about transit services from private apps than directly from the transit agency. A review of different models for sharing data with app developers shows how some transit agencies are leveraging this data to exert more control over how apps serve their customers. In addition, transit agencies are developing models in which they receive data collected by transportation apps using a variety of mechanisms to facilitate this data transfer.

“Riders interact with these apps multiple times daily, making open data the most important customer communication channel agencies offer to the public.”
(TransitCenter 2018)

Figure 7 shows four different models for using route, schedule, and vehicle location data to provide customer information as reported in the transit agency interviews. Transit agencies can control the customer information received through an app in a variety of ways. Many transit agency interviewees indicated their agencies simply publish route, schedule, and vehicle location data and allow app developers to use it in their apps (dubbed an app-neutral approach). A relatively low-effort option that provides the transit agency with some control is to endorse an existing app. A more resource-intensive option, which one transit agency is in the process of completing, is to commission an app. Finally, for ultimate control, some transit agencies, including two of those interviewed, develop apps in house. This requires staff with specific technical skills to develop and maintain the app. It is important to note that, in the cases observed and documented, those transit agencies that endorse, commission, or develop an app continue to provide the data openly and allow other apps to use the data.

App-Neutral Approach

Many transit agencies publish GTFS and GTFS-RT feeds using a developer API, which in turn is used by third-party apps. Some transit agencies provide a list of vetted apps on their websites. Even with this safeguard, this hands-off data sharing model relies on developers to provide the



Figure 7. Mechanisms for customer information app development.

best information and user experience to customers and customers to find the best app through the app review process or word of mouth.

There are drawbacks to this approach, many of which were highlighted by transit agency interviews conducted. First, some apps may not provide accurate information. Even if all apps use the same data feeds, the algorithms they use to suggest routes and predict vehicle arrivals vary. One transit agency interviewee commented that one commonly used transportation app in their service area often provided inaccurate predictions, likely due to the algorithm used. This is a problem, because transit agencies want to ensure that customers access the most reliable information available.

Transit agencies are also concerned with the context under which information is presented. Third-party apps exclusively control the content and presentation of information and may rely on advertising revenue from other modes of transportation that are included in the app. Many apps, including Transit App and Google Maps, present information about the cost and travel time of Uber and Lyft on the same screen as transit information. Informed by customer research, at least one transit agency interviewee expressed concerns that this presentation of information may encourage users to choose TNCs over public transit. On the other hand, some transit agencies see TNCs as a potential complement and have worked to have transit information displayed within the Uber app. In Denver, not only are public transit options listed alongside Uber's offering in the Uber app, but customers can purchase tickets for these services in the app as well (Conger 2019).

A hands-off approach to sharing GTFS and GTFS-RT data also means that data on transit customers who use transportation apps is collected by the app developers rather than by the transit agency. The UITP argues that this can be strategically risky for transit agencies, because they are missing out on information about their customers. Instead, this information is accumulating to private developers. There is at least one example in which a private app developer, Citymapper, piloted bus service; however, in this example, the pilot was in cooperation with TfL, a public transit agency. In the “endorse an app” example that follows, the transit agency was able to negotiate access to a third party’s app data in return for endorsing that app.

Endorse an App in Return for Data

In Boston, Transit App, which the MBTA officially endorsed for real-time transit information and travel planning, provides the MBTA with data on app users' searches and locations (Enwemeka 2016). The MBTA made this agreement after releasing a nonfunded request for proposal (RFP) asking app developers to submit bids competing for endorsement by the transit agency. The transit agency evaluated bids and selected the best app to recommend to its users. The RFP also required that the selected app share data with the transit agency, enabling the transit agency to draw insights from app data.

Compared with an in-house or commissioned app, this model is much less expensive. However, as in the other cases, there is no guarantee that customers will use the endorsed app. In addition, in this model, the transit agency has less control over the presentation of information in the app. Rather than specifying these components explicitly, they select the best of the available options.

Commission App Development

Although most transportation apps are developed by private app developers without input or funding from transit agencies, transit agencies may opt to pay for some or all components of app development, giving them control over app specifications. Commissioned apps may be branded by the third-party developer or by the transit agency itself.

One transit agency interviewee indicated that their agency had released an RFP seeking a vendor to provide a trip planner to integrate into the agency's existing mobile payment app. Not only will the transit agency have control over the way information is presented, but they anticipate being able to provide more custom information such as detours, real-time alerts, and service change notifications. The RFP also specifies performance requirements, including a minimum level of prediction accuracy and a limit on the "ghost" bus and train incidence rate.

The transit agency will own all data produced by the app. The transit agency anticipates that the data will improve the transit agency's understanding of operations, predictions accuracy, customer preferences, and travel patterns. For example, this data could provide insights into customer origins and destinations, travel choices, and latent demand as well as customer responsiveness to changes in routes, frequency, service quality, and reliability. To protect user privacy, the vendor must abide by stringent General Data Protection Regulation (GDPR) standards now employed in the European Union.

A major drawback of this approach is the cost. In addition, the transit agency cannot guarantee that customers will use the app, because other apps will still be available. Also, one argument against a custom app, whether it is developed in house or commissioned, is that it works against the trend of universal transit planning tools. Apps like Google Maps and Transit App that are available across a large number of transit markets standardize the transit experience for visitors who ride transit in different cities. Standardization of apps may make transit easier to use and actually encourage transit use.

Develop an App in House

Two transit agency interviewees indicated their agencies have their own app, developed in house. This model gives the transit agency maximum control over the contents of the app and the data extracted from it. As long as the transit agency continues to put resources into maintaining and updating the app, they have flexibility to adapt the app over time, as data and customer needs change.

44 Data Sharing Guidance for Public Transit Agencies—Now and in the Future

Such a model is generally only feasible for large transit agencies with sufficient information technology (IT) and technical staff to devote to the effort. However, even some of the transit agency interviewees from larger agencies indicated their agencies opted out of developing their own app, because of the specialized and fast-changing nature of app development.

Transit agencies that either commission an app or develop it in house make decisions about the inclusion of other modes. On the one hand, some see a benefit to excluding potentially competing modes such as TNCs. On the other hand, several transit agency interviewees reported that they believe it is important to include other modes to facilitate multimodal journeys. In regions with multiple transit providers, integration of data from other agencies into a commissioned or in-house app is critical.



CHAPTER 5

Models for Accessing External Data Sources

In addition to getting value from sharing their own data, transit agencies see benefits from accessing external data sets. Mobility data from other sources can help transit agencies understand how people in their service areas get around, which can inform transit planning strategies. The transit agency interviewees were interested in external data sets including the following:

- Location data from cellphone connections, smartphone location-based services, and other GPS-containing devices.
- Data from transportation apps, including transit planning and fare payment apps, private mobility providers' apps, and MaaS apps.
- Other data collected by and pertaining to private mobility providers.

Motivations for accessing this data include the following:

- Enabling transit agencies to evaluate overall demand patterns in their service area and determine how to better meet peoples' needs.
- Enabling more detailed road speed data to inform bus operations and route alignment decisions, and to improve bus arrival predictions.
- Enabling analysis of how people behave when incidents or other disruptions prevent them from using the transit system.
- Enabling transit agencies to identify access and egress modes and distances.

In addition to private data sources, transit agencies use external, publicly available data sources, including census data, weather data, and Geographic Information System (GIS) data from cities, states, and regional agencies.

There are several methods that transit agencies are using to access private data sources. The following are examples that were identified either through the literature review or in interviews.

5.1 Purchasing Data or Analysis

One model to acquire private data is by purchasing the data or commissioning data analysis. TfL has documented its pilot project, producing transport model matrices derived from data from Telefonica UK, a cell service provider (Transport for London n.d.). The matrices were produced using depersonalized, aggregated data that will be used to better understand demand patterns for public transport users and drivers. They note that this data provides a better, more convenient alternative to roadside interview surveys. One transit agency interviewee indicated that their agency was in the process of acquiring cellphone

“Buying anonymous and aggregated mobile phone event data is a practical alternative to [roadside interview surveys]. It will provide data on trip patterns in a cost efficient, safe way without inconveniencing customers or compromising their privacy.”
(Transport for London n.d.)

data to better understand demand patterns in their transit system, and another indicated their agency had already purchased cellphone data.

Many city interviewees indicated their cities had also purchased data from technology companies, such as location-based services (LBS) companies, that generate traffic condition insights or travel demand flow information at the zonal aggregated or road segment levels. Some cities were able to share purchased data among departments in the city, while others were limited to using the data for specific projects, because of restrictions in their data usage agreements.

Although some public agencies purchase data that can be used indefinitely, another purchasing model is the service subscription-based software platform. In that case, the data may belong to the company that provides the service or may even belong to another company or organization.

Cities also purchase secondary data products from private companies or research institutions. Secondary data products include items such as “dashboards” [i.e., centralized webpages where residents can view multiple types of information about their city simultaneously (sometimes in interactive fashion) and predictive algorithms (i.e., programs that aggregate, model, and project data to assist in future decisionmaking)] (Brauneis and Goodman 2017).

5.2 Accessing Data Through Mobility Service Partnerships

Many transit agencies have partnerships with TNCs; a 2018 study cited 29 examples in the United States (Schweiterman et al. 2018). FTA’s MOD Sandbox Program provides funding and oversight for many of these partnerships and includes a data sharing requirement. Data sharing happens in a variety of ways. This section describes examples from three different transit agencies. In one case, the transit agency has taken an evolving approach to its partnerships, gaining access to more data over time. In another case, the transit agency negotiated a detailed agreement with their partner for specific data items and levels of aggregation. In the third case, the transit agency and its private partner provide data to a third party who conducts analysis combining the two data sets.

Evolving Access to Private Data

In this example, the transit agency interviewee indicated their agency has a partnership with a TNC to provide (1) first- and last-mile trips to transit stations and (2) door-to-door rides for members of its transportation disadvantaged program. This transit agency has been at the forefront of transit agency partnership with TNCs. Initially, the agency received only aggregated data from the TNC on the subsidized first- and last-mile rides, but with each contract amendment, they have negotiated to receive more data. Still, they do not receive trip-level data.

They use data from the TNC to track users of the program as well as program response time. They would like to use this data to better plan their transit services. They see the evolution in data transfer for both programs as a sign of their experience.

Custom Agreement

Another transit agency interviewee indicated their agency has taken steps to negotiate a detailed data agreement upfront. For their FTA MOD Sandbox-funded project, the transit agency used an informal procurement process to seek a mobility provider that was willing to share data. Once they selected a partner, the transit agency continued with a careful and

time-consuming process to iron out the details of their data sharing agreement. The agreement includes the following:

- **Variables.** A list of variables that the TNC will share and the level of granularity for each one.
- **Access.** Designation of which researchers would have access to the data, and where it would be housed.
- **Public data.** Designation of what data would be made public. The transit agency has committed to sharing aggregated data with municipal operators and city partners. In addition, there will be public reports on the pilot. Disaggregate data will not be made public.
- **Data ownership.** The transit agency will not own the disaggregated data but rather will have access to it for 5 years. They will be able to access aggregate data indefinitely.

Sharing Data Through a Third Party

A partnership between Uber and transit agencies in Cincinnati includes a “first-of-its-kind study” in partnership with transit agencies SORTA (Southwest Ohio Regional Transit Authority) and TANK (Transit Authority of Northern Kentucky), in which a transit consultant will combine Uber and transit agency data to draw insights that can inform strategic transit planning in the region (Salzberg 2018; Schweiterman et al. 2018).

In this case, Uber hired a consulting firm to analyze how people move in the city, including their use of rideshare and transit. The consultant worked directly with representatives from Uber, the city, and the transit agencies and has published a report on curb use and will publish a report on transit. The completed curb study has recommendations to the city for designing pick-up and drop-off areas and reallocating on-street parking (Teale 2019).

Using a third party generally means that the data provided by the private mobility provider is not subject to state public records laws. As such, mobility providers may be more willing to provide disaggregate data that may be proprietary or hold potential privacy concerns.

Other examples of third-party models include the SharedStreets initiative, supported by NACTO (described in Section 4.1), and University of Washington’s Transportation Data Collaborative, which was under development as of July 2019. The Collaborative provides policies, protocols, and platforms to enable data sharing and analysis of sensitive data (generated from public or private mobility services) with partnering agencies to create data-driven policy and support research uses. It creates an innovative model to address data ownership, access, and privacy and ethical issues in the interest of partner organizations. Not only will the data collected be exempt from public records requests, but the Collaborative aims to leverage the technical skills and storage and computing power of the university (Shared-Use Mobility Center 2019).

5.3 Accessing Data Through Regulation

Public transit agencies can work with cities and states to develop and push for regulation that can facilitate public agencies’ access to external data streams. Many cities have begun regulating and managing private mobility companies that operate on their public right of way. In response to the proliferation of micromobility services, various U.S. cities have restructured their management and regulation of transportation services. Although these policies vary in scope and detail, the core data sharing features of most fledgling micromobility policies are similar. Cities require micromobility companies to share data regarding trip and fleet availability, with many including specific expectations for the frequency at which data is shared (Migurski

48 Data Sharing Guidance for Public Transit Agencies—Now and in the Future

2018). A micromobility company interviewee indicated that, for them to enter this market, they regularly provide aggregated mobility data to cities and transit agencies

Transit agencies can also work with state legislatures to update public records laws that pose impediments to data sharing. The California Public Records Act has provisions that data that constitutes trade secrets will not be disclosed. Los Angeles Metro ensured that data it collected through an agreement with Via would be exempt from disclosure under those provisions. TriMet in Portland, Oregon, supported an update of Oregon law that exempts travel pattern data from public records requests (Shared-Use Mobility Center 2019).

CHAPTER 6

Major Challenges

In the quick-changing data management and sharing environment, the transit agency interviewees identified a variety of challenges, which are also reflected in the literature and shared across other sectors. Some challenges and needs are internal: protocols, organizational structures, and other changes that are required within the transit agency. At the same time, transit agencies are looking for external guidance and even regulation to govern these internal changes. Transit agency interviewees expressed frustration at the challenge of working out every detail of a data sharing agreement internally. They recognize the potential efficiency that could be gained through standardized protocols and policies. They also see regional or federal policies as a potential mechanism to encourage more cooperation from vendors and private mobility and data providers.

6.1 Internal Data Management Structure and Protocols

The majority of the transit agency interviewees identified a lack of coherent organizational structure for managing data internally as well as for data sharing. Interviewees noted that data was collected and stored across a variety of divisions or groups within the transit agency, and responsibilities for data sharing were therefore also spread across staff in different parts of the organization. For example, maintenance staff collect and manage maintenance data, operational staff have operations data, other data is housed by the transit agency's IT department, and planners access yet another set of ridership and route and schedule data. These data silos present challenges for internal use of data as well as for external sharing of data.

Many transit agencies described the fact that data requests may be sent to a variety of divisions within the transit agency. These are handled by different staff depending on the data type. Only one transit agency interviewee noted their agency had created an information management and governance group to handle all outside data requests. Due in part to these organizational challenges, transit agency interviewees noted that responding to individual data requests can be resource intensive.

Several transit agency interviewees described these challenges in terms of personnel and organizational needs (the need for a centralized data management staff person or group) and technical needs (the need for a centralized data repository and catalog). These needs are related in that a centralized data repository and catalog requires dedicated staff to develop and maintain. These data-focused staff could also take responsibility for other needs that were identified by the transit agency interviewees: the development of formal data sharing policies and protocols, including standard data licensing agreements and an established method for evaluating privacy risks.

Data Silo Problem

Transit agencies discussed challenges with data being stored and managed in silos across the United States. This problem occurs in organizations across sectors. Electric utility sector interviewees echoed the sentiment. Although data repositories and dashboards can help organizations make timely decisions, the dashboards are only as good as the data available to them. Without a complete data set, insights may go undiscovered.

As noted in the literature, developing these capabilities among transit agency staff likely requires staff training, particularly in small- and medium-sized agencies (Lawson 2016). As an additional challenge, staff turnover can make it difficult to ensure that progress in data management is sustainable. Establishing a staff member or team that is dedicated to data management is an important step in addressing these challenges.

Fulfilling these tasks can present technical challenges for some agencies that lack specialized resources (Brauneis and Goodman 2017; Lawson 2016). These challenges only increase with large-scale data that can require machine learning techniques for processing and scalable data storage and mining (Zaslavsky et al. 2013). In fact, this is often the reason that public agencies partner with private companies or universities who can help complete some of these data processing tasks both for the transit agency's internal use and for broader sharing. However, some argue that these partnerships take power away from the public agencies, particularly when external partners fail to transparently describe the methods they use to process data (Brauneis and Goodman 2017). The transit agency interviewees spoke positively about the technical assistance their agencies receive through data sharing.

In general, the transit agency interviewees were less concerned with having technical skills in house and more concerned with having the time to dedicate to data preparation tasks. Many expressed that the structure of their transit agency contributed to a lack of effort devoted to data management tasks. Most transit agency interviewees indicated their agencies do not have staff or divisions dedicated to data management, which means staff have other priorities. Those transit agencies that most actively analyze data internally tend to be most well-equipped, both technically and organizationally, to prepare data for sharing. Because of their internal capabilities, these transit agencies may be least in need of external research and innovation to use their data. The needs of small agencies to develop data sharing infrastructure require special attention.

One particular internal challenge facing transit agencies is in data collection. To maximize the value transit agencies can attain from sharing data, it is crucial that they collect valuable data.

Data Collection

Decisions about data collection determine the types of data and the data quality and coverage available to be shared. However, data collection is often more a byproduct of transit system design than a dedicated analysis effort. The data generated is often dependent on functional aspects of the transit system, such as operations and fare collection, rather than potential data analysis or the value that can be generated through sharing data. Some data collection processes predate modern conceptions of open data, and in some cases, considerable effort is required just to extract data from the transit system. Kitchin and Dodge (2014) note that automated data is generally collected as a result of an action, perhaps scanning a credit card or using a smart phone in which providing data is not the primary purpose. One exception is survey data, which is typically collected expressly for the purpose of analysis and evaluation.

Data collection issues can impact the value of data for sharing. Several of the transit agency interviewees identified data collection and harvesting data from existing systems as major challenges. Their agencies were hesitant to share data with gaps and inaccuracies.

Sometimes, data privacy concerns impact data collection processes, which can ultimately reduce the sharing value of data. For example, in almost all cases, smart card systems track individual card IDs to ensure that passes and discounts are applied appropriately. However, not all transit agencies store this information for analytical use. In some transit agencies, the data

stored for internal use by the transit agency has a new ID for each trip or for each day, preventing the tracking of smart cards across trips (in the former case) or across multiple days (in the latter). Although these measures protect individual privacy, they limit the potential for analysis. In many transit agencies, persistent encrypted IDs are stored for internal use, with precautions taken to preserve privacy when the data is shared externally. This shifts the decision about privacy protection to the data censoring phase.

Even when good data is collected, a final hurdle for transit agencies can be data ownership. This issue arises when transit agencies partner with private companies to provide services, in which case the private partner may not be required to turn over data to the transit agency. This is described in Chapter 5 on accessing external data. In other cases, vendors that install and maintain systems, such as automated fare collection (AFC), APC, or automated vehicle location (AVL), may retain ownership of the data generated. Transit agencies must be careful to consider the potential value of data sources and ensure that they have ownership of valuable data. Although there has been a shift in ownership of AVL and AFC systems to transit agencies, this issue may still persist for other data types. For example, if maintenance is outsourced, some maintenance data may be owned by the maintenance company rather than the transit agency.

The issue of data ownership is further complicated in the context of data on individuals. In the European Union, GDPR rules guarantee individuals' ownership of their own data (see Section 2.6). Although these rules do not apply in the United States, transit agencies should consider the possibility that laws around individual data will change, and specifically consider mechanisms in which individuals can give the transit agency permission to use their data as part of the data collection process.

Data Biases

The value of data can be limited by biases in terms of which data exists and which data is missing. In the mobility data field, this issue is discussed when app or GPS data from smartphones is used. This data excludes information on people who do not own or use smartphones, which may disproportionately include specific population groups such as low income and older people (Windmiller et al. 2014). Similar data biases can occur in transit agency smart card data. In most systems, not all passengers use smart cards, and it is important to identify which passenger demographics are more likely to use smart cards as well as the trip types more likely to be paid for with smart cards (Erhardt 2016).

6.2 External Data Policies and Standards

Transit agency interviewees noted that their internal development process would benefit from external guidance and policies. Although all transit agency interviewees identified internal improvements needed for data sharing, they also expressed interest in more external support. Transit agencies recognize that many data sharing challenges are shared across public agencies. Rather than devoting resources to solve these challenges individually, they seek external guidance around topics including the following:

- Handling sensitive or private data, including when small values need to be suppressed, and what precautions need to be taken to avoid re-identification risks
- Writing or selecting data licenses
- Documenting data

Transit agencies are also looking to external organizations for the development of data standards. This may require a regulatory push to encourage the adoption of new data standards, particularly to require private vendors to comply (Lawson 2016).

Developing Data Standards

Data standardization across transit agencies can enable external partners to repeat analyses for multiple transit agencies with limited additional effort. This can encourage private

companies and researchers to develop standard tools that can benefit transit agencies. Standardization is highlighted as a salient need in the transit data industry (Sánchez-Martínez and Munizaga 2016), but it is also a major challenge, with data formats varying significantly across organizations.

In public transit data, GTFS is the noted outlier, a standard format for route and schedule information that is widely used across transit agencies. The development of GTFS was initially pioneered by Google to integrate transit information into the Google Maps platform. Over time, it has become widely used outside of Google as well, particularly in apps that provide transit information to customers (Schweiger 2015). A newer standard, GTFS-RT, attempts to do the same for real-time vehicle location data.

The Vermont Agency of Transportation and Trillium Solutions developed the GTFS-flex specifications to support flexible demand-responsive transportation services, different from the original GTFS that only models fixed-route public transportation. It helps transit users get information about nonfixed-route transit services, which are common in less dense environments.

The NTD is another example of transit data that is both standardized and consolidated. To date, smart card data has not been standardized in the same way. Two organizations—the Integrated Transport Smartcard Association and the Secure Technology Alliance (formerly the Smart Card Alliance) in the United States—have developed standards for interoperability of smart cards, but these standards are focused on secure fare collection, not on data generation and formatting.

Transit Intelligent Transportation Systems (ITS) Data Exchange Specification (TIDES) and GTFS-ride are two projects developing standards for transit ridership data from passenger counters and fare collection. Still in their early stages, these standards look to support tools and applications for transit analysis.

The Los Angeles DOT has developed an emerging sharing data standard called Mobility Data Specification (MDS) which serves as a model for data sharing policy between cities and the private sector. This data is ideally shared through an API, which has the advantage of allowing cities to see a dynamic, continuous picture of fleet usage and placement. In addition, the standard can make data analysis more efficient.

How Do Data Standards Get Adopted?

1. Good standards require champions and resources to support a standards-making activity, including respected experts.
2. Well-designed standards have few optional fields and can evolve over time.
3. For adoption, there needs to be either a clear benefit or a mandate.

A major challenge with data standards is adoption. Transit agencies and the vendors they employ need to cooperate, and this requires effort from one or both parties to convert existing systems to meet new standards. Generally, standards are adopted either when there is a clear benefit (the proverbial carrot), or when their adoption is mandated (the stick). As an example, transit agencies quickly adopted GTFS because it allowed their information to be displayed in apps their customers were using. In contrast, many transit agencies submit standardized data to the NTD, because it is required to do so if they receive funding through §5307 or §5311 formula grants [Title 49 United States Code (USC) §5335(a)].

A standard needs champions as well as resources. Resources are key to supporting a standards-making activity. For traction, the activity should include experts who are respected within the industry. Well-developed standards minimize the number of optional fields, which limit the usefulness of the standards. The standards-making activity should include testing, and a certification system may need to be developed to evaluate compliance with the standard. Good standards can evolve over time. As an example, GTFS has limited ability to describe fares but has the potential to be extended to handle more complex fare policies (Wang 2014).

Data Standards in the Energy Sector

The electric utility sector is currently undergoing standards development. The Department of Energy (DOE) published high level guidance about privacy (SEE Action 2012) and data interoperability (ICF 2016). The DOE guidance points to federal initiatives that are relevant across industries—Fair Information Practice Principles (FIPPs), the Consumer Privacy Bill of Rights, FTC Codes of Conduct, non-binding industry standards, and emerging “privacy seal” initiatives. This top-down guidance notes that it is up to state regulators, utilities, and third-party service providers to define standards and implement them (U.S. Department of Energy 2015). Time will tell whether this patchwork approach is successful. Some state regulators are moving quickly, while others are not. The DOE’s report summarizes feedback from regulators and utilities about the challenges to develop interoperability: regulators lack the technological expertise and time to learn about interoperability needs; regulators lack access to industry publications and working group findings; and some utilities prefer proprietary systems rather than standardized services.

In summary, the process of data standardization depends not just on an individual transit agency’s technical and organizational ability to apply standards, but also on a strong coalition that has built effective, flexible, and respected standards, and on motivational carrots or sticks to promote the standard’s adoption. The majority of transit agency interviewees recognized the need for more data standards but felt that external organizations or regulators would be required to implement them.

Public Records Requests and Access to Data

Private sector interviewees cited that protecting user privacy was the most common concern about providing data to transit agencies. They are also concerned that, under state public records laws, the shared data from these private companies’ users could fall into the public domain, violating their customers’ privacy. For this reason, private companies often share aggregated data or provide access to an analytical platform rather than providing data directly. As described in Section 5.2, private companies also may share data with a third party rather than with a transit agency directly.

One transit agency interviewee pointed out that laws can appear arbitrary or out of date. A law in their state exempts smart card data from FOIA requests on the basis that it contains individual records. However, data from smartphone apps, which the transit agency is planning to collect, will not be protected from release under the same law. As described in Section 5.3, transit agencies are beginning to take an active role in shaping legislation.



CHAPTER 7

Conclusions and Next Steps

The research summarized in this report identified the critical factors that drive data sharing decisions and defined several models for data sharing by transit agencies, considering both sharing of transit agency data and access to external data by transit agencies. There were several key findings that represent common themes and critical issues and challenges observed across transit agencies and even across sectors. This chapter summarizes the important takeaways from this research. In addition, this is an evolving field. Mobility data is changing, as is the legal context and the conversation around data privacy and ownership. This chapter describes how transit agency data sharing may evolve in the future. Finally, this research identified several gaps where additional research is required. These are summarized in Section 7.3.

7.1 Key Findings

Based on the interviews and literature and information review, the following are the key findings about data sharing for transit agencies.

Transit Agencies Share Data Frequently and See Many Benefits

- **Transit agencies collect data on the transit system, including route, schedule, and vehicle location data, which are commonly shared and contributes to customer information.** Sharing these data types can promote transparency and generate insightful research. Additionally, route, schedule, and vehicle location data are frequently shared by transit agencies (10 of 11 transit agency interviewees stated that their agencies share this data publicly). Private developers routinely use route, schedule, and vehicle location data in customer-facing apps that help transit passengers plan their routes and find out when transit vehicles will arrive at stops and stations. Some transit agency interviewees reported receiving data from these travel planning apps.
- **Transit agencies collect a wide variety of data on transit passengers. Sharing this data also generates value, including from research that can improve system performance and increased advertising revenue for the transit agency.** Transit agencies collect data on passengers, including fare and bank card transaction data, Wi-Fi and Bluetooth data, video data, and passenger count data, all of which shows how the transit agency's network is being used. This data is often of interest to customers, journalists, real estate developers, and researchers. Sharing data can promote transparency and generate insights and innovation that are beneficial to the transit agency. It may even generate revenue, particularly through advertising. However, these data types have the potential to be used to identify individuals. Transit agencies need data privacy protocols to determine which data sets should be shared and what measures (aggregation, censoring, adding noise) should be taken to protect privacy.

- **Transit agencies share some data openly and share other data sets directly with partner institutions or individuals through private data sharing agreements.** The most common type of open data, according to the transit agency interviewees, was route, schedule, and vehicle location data, but transit agencies also share ridership, on-time performance, survey data, and financial data publicly on their websites. In addition, all transit agency interviewees indicated their agencies respond to public records requests for data. Several transit agencies have established data sharing relationships with research institutions and reported beneficial insights gained through these relationships.
- **Information disclosure laws govern many aspects of data sharing by transit agencies.** Information disclosure laws, which require public agencies to share information requested through a public records request, vary by state (most transit agencies are not subject to the federal FOIA). Transit agencies should ensure they understand the legislation in their state, including data exemptions and whether processing fees can be charged. If transit agencies find that data that poses privacy risks is not exempted from public records requests, they may consider working with state legislators to change legislation. Ensuring these exemptions are in place can also help transit agencies access external data sets.

Transit Agencies May Be Able to Increase the Value of Data Sharing in the Future with the Development of New Data Standards, Moving Toward Open Data and Tools, and Leveraging the Interests of the Private Sector

- **Data standards have the potential to increase the value of public transit data sharing and make transit agency use of external data sets more efficient.** The majority of transit agency interviewees were supportive of the idea of standards for public transit data types, noting that standards could promote the development of shared tools and other resources. Transit agencies are looking to external organizations for standards creation and adoption.
- **Open software tools could augment the value of public transit data and help transit agencies use external data sets.** A general movement toward open data and open tools can benefit transit agencies.
- **Private company interviewees in the MaaS industry, including private mobility providers and user information app developers, are interested in transit data.** Some expressed a willingness to further discuss the potential to purchase data from transit agencies; others questioned the notion of monetizing data collected by public transit agencies. They are especially interested in geospatial details of transit stations as well as data, such as passenger counts, that can help them plan their services.

Data Sharing Challenges are Part of Broader Data Management Needs

- **Often, transit agency data collection processes are byproducts of other functions of the transit agency (e.g., fare collection, operations, management).** More deliberate data collection efforts can ensure transit agencies maximize the value of their data.
- **Collecting, cleaning, processing, documenting, and cataloging data requires significant effort.** Several transit agency interviewees noted the significant technical needs and effort required to prepare data for sharing. Those transit agencies that had developed procedures for processing and cataloging data found that this saved time responding both to public and to internal data requests. Transit agencies may consider charging processing fees for public records requests that require significant effort (if allowed under state law).
- **Transit agency interviewees identified internal organizational and technical needs to improve their processes for sharing data.** The majority of transit agency interviewees

indicated their agencies do not have a centralized data repository, or a staff or group dedicated to data management. These interviewees noted that establishing dedicated staff could help make data management a priority. In addition, they expressed frustration that data is stored across divisions, making it difficult to find, use, and share. Data-focused staff can drive transit agencies' data sharing programs, developing goals, identifying needs, creating internal data management processes, including a data catalog, and evaluating data sharing opportunities.

Transit Agencies are Beginning to Harness the Value of External Data, but Challenges Remain

- **There is potential value in linking transit agency data sets to external data sets.** External data sets can help transit agencies understand first- and last-mile trips and modal alternatives to transit. Parking and curb-use information can also be of use to transit agencies.
- **Transit agencies access external data sets, either by purchasing data or leveraging a mobility services partnership. Or in some cases, they may gain access to data through a third party.** The transit agency interviewees acknowledged the challenges of negotiating data sharing agreements with private mobility providers, even when they have a service agreement. Despite these challenges, transit agencies see value in these data sources, and at least two of the transit agencies have negotiated access to data from TNCs.
- **Although private sector data, app, and mobility company interviewees expressed interest in cooperating with transit agencies, they also cite privacy concerns as one reason their companies avoid sharing individual-level data with transit agencies.** There are no transportation sector-specific privacy laws at the federal level that govern transit data sharing. In some states, transit agencies may need to work with state legislatures to ensure that data on individuals is exempted from state information disclosure legislation. Or transit agencies can work to access information through a third party.
- **Cities are beginning to exercise their regulatory power by demanding private mobility providers submit mobility metric data when applying for operational rights on city rights of way.** These requirements provide examples for public transit agencies establishing partnerships with private companies. In addition, transit agencies can work with cities to ensure that data requirements meet transit agency needs and that data is shared between the two public entities.

7.2 Future of Data Sharing for Transit Agencies

There are several factors that may change over time. First, transportation technology and data are evolving. Second, legislation around data management and data privacy may evolve. Third, there is a trend toward open data and open data tools.

Evolving Technology

As sensing and fare payment technology evolves, transit agencies will collect new and different data. In addition, new mobility options and MaaS platforms generate additional data. As transit agencies develop their internal data management staff and resources, it is important to build in flexibility to accommodate new data types and uses.

Evolving Regulations

States are increasingly regulating data collection, storage, and dissemination by public agencies. As data breaches and security of personal information become more of a concern, states may move to impose additional restrictions and disclosure requirements. States, such as California, have taken a leading role in enacting privacy data laws, and other states will likely

follow. Thomas (2017) hypothesized that a decision that impacts the way transit agencies manage data sets, including individual trips, may eventually be handed down by the U.S. Supreme Court.

In the European Union, the GDPR, approved in 2016 and enforced beginning in 2018, defines a comprehensive set of regulations around privacy. It requires that consent and terms be clear, transparent, and written in easily understood language. It specifies that individuals own their data and must be informed if their data is being transferred to another party or if there had been a data breach.

Data-as-a-Service companies that operate in Europe maintain that the owners of the raw data generated from smart devices are individual device users. These private companies leverage the usage of data to derive analytical data products and insights at various spatial-temporal levels. The technology companies claim ownership of the data derivatives generated from their algorithms, not the original data.

As described in Section 5.3, there has also been movement to update public records legislation to remove barriers to the sharing of private sector data with public sector agencies. These challenges are not limited to the transit sector and are likely endemic to the nascent nature of privacy protection laws. In the utility industry, some regulators have placed privacy restrictions on a third party's ability to share the customer's data with a partner provider. Although these restrictions are intended to protect consumers, some state regulators are reviewing and revising their rules to allow secondary release when it comes to enabling consumer convenience and guaranteeing customers reap the intended benefits of sharing data.

Changes to public records legislation could help transit agencies access more external data. Protecting data from public records requests could also enable transit agencies to monetize data or leverage it in data-for-data trades. However, the viability of these options will depend on how public perception and expectations of data ownership and data privacy evolve.

Open Data and Open Data Tools

There appears to be a general trend toward open data and open data tools. The success of GTFS is much touted in the public transit sphere, and there is a push for additional open data standards. Initiatives such as SharedStreets and the MDS seek to extend open data practices to private sector mobility data. The World Bank's Open Transport Partnership (2016) is another innovative data sharing model used mostly for developing countries, but may it be a potential model for developed countries as well.

7.3 Future Work

This effort identified several areas for additional study, including the following:

- A technical analysis of data privacy that identifies privacy risks by data type and provides methods to add noise or set aggregation and sample size requirements to protect privacy.
- Guidance on data cybersecurity for transit agencies to use internally and to share with partners who receive sensitive data.
- Analysis that quantifies the potential costs of data privacy and security risks.
- A 50-state survey of data privacy and information disclosure laws that apply to state and local government entities to provide more details on specific requirements that may be imposed on transit agencies with respect to data sharing.
- Collaborative standards-making activities to enable more effective sharing of both transit agency and external data sets and to promote open data and open data tools.
- A cross-agency study of the level of effort required for different data management tasks to help transit agencies better evaluate costs of data sharing and internal data analysis.



Bibliography and Other Resources

- Abella, A., M. Ortiz-de-Urbina-Criado, and C. De-Pablos-Heredero. (2017). A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. *Cities*, 64, 47–53.
- American Bus Benchmarking Group. (2019). https://americanbusbenchmarking.org/?page_id=11.
- Antrim, A., and S.J. Barbeau. (2013). The many uses of GTFS data—opening the door to transit and multimodal applications. Location-Aware Information Systems Laboratory, University of South Florida, Tampa.
- Barbeau, S.J. (2018A). Quality Control-Lessons Learned from the Deployment and Evaluation of GTFS-Realtime Feeds. No. 18-05585.
- Barbeau, S.J. (2018B). Closing the Loop: Improving Transit through Crowdsourced Information. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2672, pp. 224–234.
- Barry, J., R. Newhouser, A. Rahbee, and S. Sayeda. (2002). Origin and Destination Estimation in New York City with Automated Fare System Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1817, pp. 183–187.
- Bosselman, A. (2019). Uber Partners with Denver Transit—What Could Go Wrong? Retrieved from <https://denver.streetsblog.org/2019/01/31/use-uber-for-denver-transit-company-is-our-overlord-and-savior-as-it-crushes-the-epic-failure-that-is-rtds-mobile-app/>.
- Brakewood, C., N. Ghahramani, J. Peters, E. Kwak, and J. Sion. (2017). Real-Time Riders: A First Look at User Interaction Data from the Back End of a Transit and Shared Mobility Smartphone App. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2658, pp. 56–63. <http://dx.doi.org/10.3141/2658-07>.
- Brakewood, C., and R. Paaswell. (2017). Assessing NJ Transit's Mobile App for User's Receptiveness to Geotargetting. UTRC/RF Grant No: 49198-43-27. University Transportation Research Center—Region 2, The City College of New York, NY.
- Brauneis, R., and E. Goodman. (2017). Algorithmic Transparency for the Smart City. (SSRN Scholarly Paper No. ID 3012499). Retrieved from: <https://papers.ssrn.com/abstract=3012499>.
- California Public Utilities Commission. Decision Adopting the Renewable Auction Mechanism. D. 10-12-048. Filed 21 August 2008.
- Cambridge Systematics, Inc. (2018). *NCHRP Research Report 868: Cell Phone Location Data for Travel Behavior Analysis*. Transportation Research Board, Washington, DC. <https://doi.org/10.17226/25189>.
- Catalá, M. (April 2016). FTA Open Data Policy Guidelines. FTA Report No. 0095. U.S. Department of Transportation, Washington, DC.
- Chen, R., B. Fung, B.C. Desai, and N.M. Sossou. (2012). Differentially private transit data publication: a case study on the Montreal transportation system. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, Beijing, China, pp. 213–221.
- Cheshire, T. (2017). TfL plans to make £322m by collecting data from passengers' mobiles via Tube Wi-Fi. Sky News. <https://news.sky.com/story/tfl-may-make-322m-by-selling-on-data-from-passengers-mobiles-via-tube-wifi-11056118>.
- Chester, M., and A. Horvath. (2010). Life-cycle assessment of high-speed rail: the case of California. *Environmental Research Letters*, Vol. 5, No. 1: 014003.
- Colorado Public Utilities Commission. (2015). Recommended Decision of Administrative Law Judge G. Harris Adams Amending Rules. Proceeding No. 14R-0394EG. In the Matter of the Proposed Rules Relating to Data Access and Privacy for Electric Utilities, 4 Code of Colorado Regulations 723-3 and Data Access and Privacy Rules for Gas Utilities, 4 Code of Colorado Regulations 723-4.
- Conger, K. (Aug. 7, 2019). Uber Wants to Sell You Train Tickets. And Be Your Bus Service, Too. *The New York Times*.

- Conradie, P., and S. Choenni. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, 31(1), S10–S17.
- Corazza, M.V., U. Guida, A. Musso, E. Petracci, M. Tozzi, D. Vasari, and E. de Verdalle. (2018). A Predictive Maintenance System for Bus Fleets: Innovation and Research from the Case Study of Ravenna. Presented at 97th Annual Meeting of the Transportation Research Board, Washington, DC. No. 18-00906.
- Crunchbase. (2019). Location Based Services Companies. <https://www.crunchbase.com/hub/location-based-services-companies#section-investments>. Last retrieved July 15, 2019.
- Dawes, S.S. (2010). Stewardship and Usefulness: Policy Principles for Information-Based Transparency. *Government Information Quarterly*, 27(4), pp. 377–383. <https://doi.org/10.1016/j.giq.2010.07.001>.
- Deloitte LLP. (2017). Assessing the Value of TfL's open data and digital partnerships. <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>.
- Enwemeka, Z. (2016). The MBTA Says This Is the Best Transit App for Greater Boston Riders. WBUR Boston. <https://www.wbur.org/bostonmix/2016/09/06/mbta-best-transit-app>.
- Erhardt, G.D. (2016). How smart is your smart card? Evaluating transit smart card data with privacy restrictions and limited penetration rates. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2544, pp. 81–89.
- Flanigan, K. (2019). Video, Phone Data Led Cops to Missing Boston Woman. necn.com. <https://www.necn.com/news/new-england/Olivia-Ambrose-Kidnapping-Suspect-Victor-Pena-Court-Documents-504750322.html>.
- Green Button Alliance. DataGuard Energy Data Privacy Program. 8 Aug 2018. Webinar.
- Gordon, J., H.N. Koutsopoulos, N.H.M. Wilson, J.P. Attanucci. (2013). Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2343, pp. 17–24. <http://trrjournalonline.trb.org/doi/abs/10.3141/2343-03>.
- Goulet-Langlois, G., H.N. Koutsopoulos, and J. Zhao. (2016). Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies*, 64, pp. 1–16.
- Grube, N. (2013). Many States Charge Insane Fees for Access to Public Records. *Huffpost.com*.
- Hemerly, J. (2013). Public Policy Considerations for Data-Driven Innovation. *Computer*, 46(6), pp. 25–31.
- Homstad, E. (2018). *Data Driven Policy Making: Principles for the Smart Cities*. MS thesis. Tufts University, Medford, MA.
- ICF. (2016). *Standards and Interoperability in Electric Distribution Systems*. Submitted to the Office of Energy Policy and System Analysis, U.S. Department of Energy, Washington, DC.
- Illinois Commerce Commission. (2014). Investigation of Applicability of Sections 16-122 and 16-108.6 of the Public Utilities Act. Springfield.
- International Association of Public Transport (UITP). (2018). The Value of Data for the Public Transport Sector. Action Points of UITP. Brussels, Belgium.
- Joint Utilities. Central Hudson Gas & Electric Corporation, Consolidated Edison Company of New York, Inc., New York State Electric & Gas Corporation, Niagara Power Corporation d/b/a National Grid, Orange and Rockland Utilities, Inc., and Rochester Gas and Electric Corporation. Supplemental Distributed System Implementation Plan. Case 16-M-0411. In the Matter of Distributed System Implementation Plans. November 1, 2016. 217 pages.
- Kassen, M. (2013). A Promising Phenomenon of Open Data: A Case Study of the Chicago Open Data Project. *Government Information Quarterly*, 30(4), 508–513. <https://doi.org/10.1016/j.giq.2013.05.012>.
- Kitchin, R., and M. Dodge. (2014). *Code/Space—Software and Everyday Life*. The MIT Press, Cambridge, MA.
- Lawson, C.T. (2016). Transformative Trends in Bus Data: A Bright Future Ahead. *TR News*, No. 303 May–June 2016, pp. 28–33.
- Lee, A., and M. Zafar. (2012). Energy Data Center. Briefing Paper. California Public Utilities Commission, San Francisco.
- Lerner, J. (2012). The impact of privacy policy changes on venture capital investment in online advertising companies. Analysis Group, 1–2.
- Lin, J., and C. Mele. (2012). *The Urban Sociology Reader*. Routledge, Abington, UK.
- Local 10 News (July 22, 2011). Florida Makes \$63M Selling Drivers' Info.
- Lu, J., Z. Mao, M. Wang, and L. Hu. (2015). Goodbye maps, hello apps? Exploring the influential determinants of travel app adoption. *Current Issues in Tourism* 18, No. 11: 1059–1079.
- Lubarsky, B. (2017). Re-Identification of “Anonymized” Data. *Georgetown Law Technology Review*. Vol 1:1.
- Mackintosh, M. (1992). Partnership: Issues of Policy and Negotiation. *Local Economy: The Journal of the Local Economy Policy Unit*, 7(3), pp. 210–224. <https://doi.org/10.1080/02690949208726149>.
- Morency, C., M. Trépanier, and B. Agard. (2007). Measuring transit use variability with smart-card data. *Transport Policy*, 14.3: 193–203.
- Migurski, M. (2018). Micromobility Data Policies: A Survey of City Needs. Remix. <https://blog.remix.com/micromobility-data-policy-survey-7adda2c6024d>.

- National Academies of Sciences, Engineering, and Medicine. (2018). *Data Matters: Ethics, Data, and International Research Collaboration in a Changing World: Proceedings of a Workshop*. The National Academies Press, Washington, DC. doi: 10.17226/25214.
- National Association of City Transportation Officials (NACTO). (2018). Ford Motor Co., Uber, and Lyft Announce Agreement to Share Data Through New Platform That Gives Cities and Mobility Companies New Tools to Manage Congestion, Cut Greenhouse Gases, and Reduce Crashes. <https://nacto.org/2018/09/26/ford-uber-lyft-share-data-through-sharedstreets-platform/>.
- National Conference of State Legislatures. Data Security Laws. <http://www.ncsl.org/research/telecommunications-and-information-technology/data-security-laws-state-government.aspx>. Accessed June 14, 2019.
- Peralta Quiros, T. (2018). Data Analytics for Transport Planning: Five Lessons from the Field. The World Bank Group Transport for Development Blog. <http://blogs.worldbank.org/transport/data-analytics-transport-planning-five-lessons-field>.
- Rosado, W. Data Detour: Analytics Will Move Transportation Forward. (2014). *Wired*. <https://www.wired.com/insights/2014/07/data-detour-analytics-will-move-transportation-forward/>.
- Salzberg, A. (2018). A new way of partnering with cities. Uber Newsroom. <https://www.uber.com/newsroom/cities-as-partners/>.
- Sánchez-Martínez, G.E., and Munizaga, M. (2016). Workshop 5 report: Harnessing big data. *Research in Transportation Economics*, 59, pp. 236–241.
- Schweiger, C.L. (2015). *TCRP Synthesis 115: Open Data: Challenges and Opportunities for Transit Agencies*. Transportation Research Board, Washington, DC.
- Schweiteman, J.P., M. Livingston, and S. Van Der Slot. (2018). Partners in Transit: A Review of Partnerships between Transportation Network Companies and Public Agencies in the United States. Chaddick Institute for Metropolitan Development at DePaul University. Policy Series, Chicago, IL.
- SEE Action 2012. State & Local Energy Efficiency Action Network. A Regulator’s Privacy Guide to Third-Party Data Access for Energy Efficiency, December.
- Shared-Use Mobility Center. (2019). Objective-Driven Data Sharing for Transit Agencies in Mobility Partnerships. White Paper. SharedStreets. <https://sharedstreets.io/>.
- Shelton, T., A. Wiig, and M. Zook. (2014). The “Actually Existing Smart City.” *Cambridge Journal of Regions Economy and Society*, 8(1), 13–25. <http://dx.doi.org/10.1093/cjres/rsu026>.
- Teale, C. (2019) 1 Year in: Uber’s ‘rocket shot’ Cincinnati Partnership. Smart Cities Dive. February 11, 2019. <https://www.smartcitiesdive.com/news/uber-cincinnati-partnership-one-year-later/547964/>.
- Thomas, L.W. (2017). *TCRP Legal Research Digest, 48: Legal Issues Concerning Transit Agency Use of Electronic Customer Data*. Transportation Research Board, Washington, DC.
- Thomas, L.W. (2018). *TCRP Legal Research Digest 52: Legal Implications of Video Surveillance on Transit Systems*. Transportation Research Board, Washington, DC.
- TransitCenter (2018). *The Data Transit Riders Want: A Shared Agenda for Public Agencies and Transit Application Developers*. New York, NY.
- Transport for London. (n.d.). Event data from mobile devices. <https://tfl.gov.uk/corporate/privacy-and-cookies/event-data-from-mobile-devices>. Accessed March 8, 2019.
- Transport for London. (2017). Review of the TfL WiFi pilot: Our findings. <http://content.tfl.gov.uk/review-tfl-wifi-pilot.pdf>.
- Transport for London. (2018). Annual Advertising Report. <http://content.tfl.gov.uk/tfl-advertising-report-1718.pdf>.
- U.S. Department of Energy. (2015). United States Department of Energy (US DOE) data privacy and the smart grid: a voluntary code of conduct (VCC). Washington, DC.
- Wang, Y. (2014). *A Study of Open Payment Fare Systems: System Design, Fare Engine Algorithm and GTFS Extension*. MS thesis. Massachusetts Institute of Technology, Cambridge.
- Windmiller, S., T. Hennessy, and K. Watkins. (2014). Accessibility of Communication Technology and the Rider Experience: Case Study of Saint Louis, Missouri, Metro. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2415, pp. 118–126.
- World Bank. (2016). The World Bank Launches New Open Transport Partnership to Improve Transportation through Open Data. World Bank Group. <http://www.worldbank.org/en/news/press-release/2016/12/19/the-world-bank-launches-new-open-transport-partnership-to-improve-transportation-through-open-data>.
- Zaslavsky, A., C. Perera, and D. Georgakopoulos. (2013). Sensing as a Service and Big Data. ArXiv, Cornell University, Ithaca, NY. <http://doi.org/arXiv:1301.0159>.



APPENDIX A

Data Types and Sharing Attributes

Transit Data Types and Attributes for Data Sharing

Public transit data comes in many forms, with a variety of potential external uses if shared.

Fare and Bank Card Data

When passengers use smart cards or bank cards for boarding, as is becoming increasingly prevalent, it produces records of boardings or station entries and also introduces the possibility of tracking the use of a given card over time. Although this characteristic of the data raises privacy risks, it also enables transit agencies and researchers to understand more about how individual passengers use the network.

Some researchers and agencies, including MBTA, Washington Metropolitan Area Transit Agency (WMATA), and New York City Transit Authority have applied methods to infer destinations of passengers within their networks based on AFC data combined with AVL data (Barry et al. 2002; Gordon et al. 2013). This origin–destination level data is valuable to transit agencies and researchers because it typically provides a large, year-round sample of passenger demand patterns. Smart card data can also be used to understand how the travel behavior of users changes over time, including in response to fare or service changes, weather, or micro levels, such as a bus stop or zone (Morency et al. 2007).

Smart card data also facilitates the grouping of transit users by their distinct trip sequence structure. Using TfL data for a 4-week period, researchers clustered transit users based on their activity patterns. They found that 40% of frequent transit users did not follow a conventional trip activity sequence involving one trip to work in the morning and another trip home in the evening (Goulet-Langlois et al. 2016). Information on the passenger segments that use different stations or routes may be provided to advertisers to enable them to customize ads to different passenger types (commuters, visitors, etc.). As such, this data is not only valuable for research purposes, but it can also be used to generate revenue through advertising.

Wi-Fi and Bluetooth Data

Although smart card and bank card data can provide stop- or station-level origin–destination information, Wi-Fi and Bluetooth connection records enable even more detailed tracking of passenger movements within stations or within a gated transit system. For example, TfL engaged in a pilot study collecting Wi-Fi signals from passengers' phones, which they used to understand passengers' route choices within the subway network (Transport for London 2017). This aspect of Wi-Fi and Bluetooth data makes it especially valuable for research and also makes it an important source for informing advertising. Transit agencies can estimate not only the

number of passengers who pass through a station, but also the number of passengers who pass a specific location within a station. This has the potential to enable the transit agency to generate more advertising revenue by providing these detailed statistics to advertisers (Cheshire 2017). Like fare card transaction data, the fact that this data tracks individual devices means that there are some privacy risks associated with sharing the data in disaggregate form.

Video Data

Many transit vehicles and stations are equipped with video cameras. From a research and planning point of view, this data can provide insight on crowding and passengers left behind at stations. This data can also be valuable for police investigations, and there are examples of transit agencies sharing this data with law enforcement agencies. Video systems vary; some, but not all systems, enable “pan-tilt” or “zoom-in” features, which allow for facial recognition. Particularly in systems that permit facial recognition, there are significant privacy risks associated with storing and sharing this data (Thomas 2018).

Transportation App and Webpage Usage Data

Some transit agencies have developed or commissioned the development of a customer-facing transportation app for trip planning, ticketing, or both. When transit agencies develop or own their apps, they can harvest data from these apps. In trip planning apps, app users’ destination requests are saved. In bus or train arrival apps, the specific bus or route information requested is saved. In addition, many apps save data on users’ locations while they use the app, which can be used to infer origins and destinations for trips made (Lu et al. 2015).

In a research context, this data can be used to draw insights about mode choice and alternatives based on the behavior of app users. For example, users may search for transit directions in an app, but ultimately choose to use a different mode. This choice can be inferred from the user’s app usage and smartphone location data.

These apps and the data they generate also have the potential to be used in location-based advertising and other geotargeted information. According to a recent study, most transit agency ticketing apps are location-aware, but this feature is only used to locate nearby stops/routes. Although this may be an untapped revenue source for transit agencies, initial research suggests there may be pushback to this type of advertisement (Brakewood et al. 2017).

Any discussion of transportation app data merits noting that the majority of transit agencies rely on third-party developers to provide customer-facing transportation apps to their customers. Some transit agencies have access to the data from these apps, while others do not. Different models for app development and data access are described in Section 4.3.

Even transit agencies that do not have a proprietary app have a public website that often includes a route planning tool. Although usage of web planning tools may be more limited and does not provide location information, transit agencies may nonetheless be able to generate value from their web traffic analytics. Transit agencies can draw insight from the type of information their web visitors access, and this information may also be useful for local planners and developers.

Survey Data

Transit agencies regularly conduct surveys of their users. Although some of the information collected in surveys, such as origin–destination patterns, can now be inferred from other sources, surveys continue to be a valuable source of information on things like trip purpose, trip alternatives considered, and demographic characteristics of transit customers. Surveys are also

used to assess customer satisfaction and collect information about transit passengers' preferences and priorities. In short, surveys often provide information that cannot be gathered from other transit agency sources, nor can it be inferred from external data sources, such as cellphone and GPS data. As a result, this data may be valuable to researchers and others, particularly if it can be combined with other data sets. Raw survey data can pose privacy risks, because responses can contain identifying information, such as home address and demographic characteristics of the respondent. Instead of sharing survey data openly, transit agencies typically share aggregated reports on the surveys they conduct. Of the 11 transit agencies interviewed, one publishes survey responses to its customer satisfaction survey, aggregated by month. Six other transit agencies publish reports that summarize survey findings.

Passenger Count Data

Some systems for monitoring passenger movements provide aggregate information on passenger counts without tracking individuals. APC systems use sensors to estimate the number of passengers that board a vehicle, and “load weight” data (data on the weight of a train or vehicle and its occupants at different points along a route) allows analysts to estimate passenger loads. Some fare collection systems also produce anonymous count data, for example, estimates of the number of people who pass through a turnstile or interact with a farebox. Fareboxes often record all boardings regardless of fare payment type, and even unpaid boardings recorded by the driver. As a result, this data can support studies of fare evasion. More broadly, it is used to analyze crowding and productivity of routes and lines. Except in the case of very small samples (e.g., if just one person boards a bus at a stop during the period reported on), this data does not enable the identification of individuals and therefore does not elicit privacy concerns. Four transit agency interviewees reported their agencies share APC data with researchers (in three cases) or municipalities (in one case).

Incident Data

Transit agencies collect data on incidents, including details on the cause of incidents and the operational response. They also collect data on passenger injuries and claims. Sharing this data can support research that helps transit agencies improve incident response protocols or prevent incidents. In rare cases, incident data may pose privacy risks if individuals involved in the incidents are described in an identifiable way. One transit agency interviewee noted that their agency does not release incident data publicly because of the staff effort that would be required to read descriptive data fields to confirm that they could not be used to identify individual passengers. In general, the transit agency interviewees indicated that incident data is not released publicly (though some publish real-time alerts about incidents). However, one transit agency shares this data with a research partner who has analyzed incident responses and passenger disruption impacts.

Route and Schedule Data

Route and schedule data are commonly shared publicly by transit agencies. GTFS is a standardized format for this data, though some agencies use proprietary formats from scheduling software companies. Nearly all transit agencies responding to a 2015 survey provided this type of data free of charge (Schweiger 2015). This data is used in trip planning and real-time transit information applications (Antrim and Barbeau 2013; Schweiger 2015).

Transit agencies also have more detailed transit system data, such as station diagrams. This data can be useful for research on how passengers move through the network, but some transit agencies opt not to share it widely because of security concerns.

Automated Vehicle Location Data

AVL data tracks the location of vehicles over time. AVL data is often a critical input to analysis that infers passenger destinations (Gordon et al. 2013). In addition, AVL data can be used to track and display transit system performance, evaluating headway variability and schedule adherence.

Transit agencies use data from AVL systems to provide information to customers about the next train or bus arrival. Many transit agencies share AVL data streams publicly, and app developers use this data to fuel transit arrival apps (Schweiger 2015). In many cases, this is accompanied by real-time alert information. GTFS-RT is a standardized feed specification for this type of data, although not all transit agencies use this format for the published data (Barbeau 2018A).

Transit System and Vehicle Maintenance Data

This category of transit data may include records of failures and maintenance activities as well as maintenance facilities and maintenance costs. Transit agencies that report to the NTD report vehicle reliability statistics (defined as the average distance between major mechanical failures). Beyond this national reporting, sharing of transit vehicle maintenance data is not a prevalent topic in recent literature, and the interviewees for this study did not reveal any external sharing of public transit maintenance data. Researchers may use transit agency maintenance data for life cycle cost assessments (Chester and Horvath 2010). Some researchers have also considered the possibility of using sensor data to predict maintenance needs (Corazza et al. 2018). External research has the potential to support internal use of maintenance data for decision support. This data informs both maintenance strategies and capital investment decisions.

Staffing and Operations Data

Staffing and operations data includes crew and vehicle assignments, absenteeism data, and operational procedures. This data can support research on operational efficiency and scheduling, which may ultimately allow the transit agency to operate more efficiently. However, there was little discussion of these data types in the transit agency interviews on data sharing. In the area of crew scheduling, there was considerable research in the past, but there are now off-the-shelf solutions that transit agencies use. One transit agency interviewee indicated their agency provided operations data to a research partner who helped them pilot a new bus operations method.

Financial Data

Financial data includes transit agency spending and subsidies. The NTD collects information on transit agency spending. Sharing this data helps transit agencies maintain transparency and accountability. One transit agency interviewee indicated that their agency posts budgeted and actual expense and revenue data on a monthly basis.

Geospatial Data of Transit Facilities

Based on the interviews with several representatives from MaaS companies, such as ridesharing and micromobility companies, the need for open data on detailed (and accessible) transit station entry locations and parking facility locations is rising. For example, the digital information of specific transit station entrance locations is rarely provided by transit agencies to the public. Such information could help MaaS companies provide better and smoother integration with public transit services for first- and last-mile riders.

External Data Types and Transit Agency Uses

There are a wide variety of data sources that could have relevance to transit agencies, including financial data and social media data, such as Twitter. Although these data sets may be beneficial to transit agencies (spending patterns can reveal customers' movements or trip purposes; Twitter can be mined for tweets about public transit disruptions and other events), this section describes three classes of data that most directly measure travel patterns: trace data from cellphones and other GPS-enabled devices, data from transportation apps, and other data from private mobility providers.

Cellphone, Location-Based Services, and GPS Trace Data

Cellphone connection data is collected by cellular service companies, while smartphone apps that use users' locations collect LBS data. According to Crunchbase (2019), there were more than 3,300 organizations in the LBS sector in 2019. This includes fitness, navigation, social media, and dating apps, which collect data on people's whereabouts.

These data sources are aggregated by analytics companies who derive and sell speed and origin–destination insights (Cambridge Systematics, Inc., 2018). Transit agencies can use this data to understand characteristics of alternate modes, demand patterns on alternate modes, and transit access and egress behavior. Some companies and researchers use phone location and phone system data to infer a user's mode of travel. This mode-of-travel information can add value to this data for transit agencies.

Transportation Planning App Data

Transportation planning apps include navigation apps, such as Google Maps and Waze, and apps such as Transit App and NextBus, that provide information on transit vehicle arrivals and collect information including the following:

- Records for each session, including beginning and ending coordinates and time stamps
- Placemarks—stored home and work locations
- Carshare, bikeshare, and TNC bookings (if available through the app)
- Trip planning routes, stops searched, and favorite routes

Data from these apps provides an additional layer of insight about other location data from smartphone apps. Because an analyst can identify when and where a user looks at transit information for a particular location or route and then how they behave after (whether they take transit, book an alternate mode, or do not travel at all), this app data allows transit agencies to better understand their customers' decisionmaking processes.

Private Mobility and MaaS Data

Transit agencies are very interested in the travel alternatives that transit passengers have, as these are major determinants of transit demand. The transit agency interviewees were interested in TNC, scooter, carshare, and bikeshare data. One transit agency interviewee indicated their agency had already used bikeshare data to understand public transit's competitiveness with other modes. However, the small user base for the bikeshare system made it difficult to draw conclusions.

Some private mobility providers share some data publicly. For example, several bikeshare systems have released data on trip history (<https://www.capitalbikeshare.com/system-data> and <https://www.citibikenyc.com/system-data>) and Uber provides Uber Movement data, which shows zone-to-zone travel times based on Uber driver data. However, most TNC companies are hesitant to share demand data publicly.



APPENDIX B

Interview Guides

Transit Agency Interview Guide

1. Has your transit agency shared, traded, or sold any kind of transit agency data?

Sharing data includes open data initiatives and private data sharing agreements (for example, with a research institute). Trading refers to an in-kind arrangement (for example, providing data in exchange for services or advertising). Selling involves exchanging data for monetary compensation.

If yes:

- a. What data was shared, traded, or sold?
 - b. How was it shared, traded, or sold?
 - c. Who was it shared or traded with, or sold to?
 - d. Can you share the data agreement with us (or any related documentation)?
 - e. How did your transit agency identify and engage the partner who received the data?
 - f. Did your transit agency have to clean, process, or standardize the data?
 - g. What benefit(s) did your transit agency receive from sharing, trading, or selling data?
 - h. Can you quantify the benefit in monetary terms?
 - i. Were there negative consequences of sharing, trading, or selling data?
 - j. What (if any) barriers did your transit agency encounter in the process of sharing, trading, or selling data? For example, consider privacy concerns, technological barriers, or public perception challenges.
 - k. Is there any information or tool that you think would have made the process of sharing, trading, or selling data easier?
 - l. What advice would you give a transit agency considering a similar agreement?
2. (In addition to the experience just described), has your transit agency considered sharing, trading, or selling any kind of transit agency data?

If yes:

What barriers did your transit agency encounter that have prevented you from sharing, trading, or selling data at this time? For example, consider lack of partners, privacy concerns, technological barriers, security concerns, or public perception challenges.

3. For transit data types that your transit agency has not considered sharing, or trading, or selling, what barriers to sharing, trading, and selling exist? For example, consider lack of partners, privacy concerns, technological barriers, or public perception challenges.

Example Data Sources:

- Fare Collection Data
- AVL Data

- Route and Schedule Data
 - Vehicle Maintenance Data
 - Station Data
 - Survey data
 - Automatic passenger counter/load data
 - Safety data
 - Wi-Fi Data
 - App Data (for example, from a trip planner or fare purchasing app)
 - Other _____
4. For transit data types that your transit agency has not considered sharing, or trading, or selling, is there any information, tool, or structure that would allow your transit agency to overcome these barriers?
This could include information on partners or tools to connect with them, tools or structures for data standardization, organizations for data warehousing, etc.
 5. What changes do you anticipate in the future that could alter transit data sharing, trading, and selling?
 6. Are there external data sources that your transit agency would be interested in gaining access to? Explain the value these data sources would provide.
External data sources could include private mobility provider data, Wi-Fi data, GPS probe data, and app data.
 7. What value can transit data provide to external entities?
External entities could include public agencies, private companies (e.g., app developers, mobility providers, advertisers), and researchers.
 8. Are there questions you have for these external data providers regarding access to their data?
 9. What questions would you want to be answered in a guide on transit data sharing, trading, and selling?
 10. Is there anyone else that you know of who we should talk to for this project?
This could be someone else at your transit agency, at another transit agency, someone who consults with transit agencies, or some other subject matter expert.

Private Sector Interview Guide

Background Overview

1. Do you participate in any of the “Mobility on Demand Sandbox” or “Smart City Challenge” projects, collaborating with public partners (e.g., transit agencies or cities)?

Could you tell us more about it?

If Not:

2. Do you collaborate with public partners (e.g., transit agencies or cities) for data sharing?

Could you tell us more about it?

3. Do you collaborate with private partners for data sharing?

Could you tell us more about it?

If Yes or Relevant:

2. If you develop an app/or analytical platforms:
 - a. What type of data do you collect?
 - b. How do you facilitate data sharing among your partners?
 - c. Who manages the data? (e.g., data processing, documentation, catalog, etc.)
 - d. Who owns the data?

Data Sharing and Access

4. Do you share your data with public (private) partners?
 - a. What data is shared (traded)?
 - b. In what format is the data shared?

Interviewer will prompt the following categories:

- Raw data
- Data analytical results in aggregated forms
- Software with data analytical results
- Data visualization platforms
- Smartphone Application
- Other, please explain _____

5. Do you have access to (mobility) data from your public (private) partners?
 - a. What data is it?
 - b. In what format is the data shared?

Interviewer will prompt the following categories:

- Raw data
- Data analytical results in aggregated forms
- Software with data analytical results
- Data visualization platforms
- Smartphone Application
- Other, please explain _____

6. Could you please highlight the key takeaways of both parties in the process of negotiating the agreement?

Could you share the data agreement with us (or any related documentation)?

7. What (if any) barriers (risks) exist to sharing data with the public (private) partners?

Interest in Transit Data

8. Would your business be interested in sharing your data with transit agencies?
 - a. On a scale of 1 to 10, 1 being no interest at all, how would you rate your interest in sharing your data with transit agencies?
 - b. What type of data would you be interested in sharing?
9. Would your business be interested in getting data from transit agencies?
 - a. What data currently collected by transit agencies may have value to your business?

Interviewer will prompt with the following categories:

- AVL Data
- Route and Schedule Data
- Vehicle Maintenance Data
- Station Data
- Survey Data
- Anonymous Passenger Counter/Load Data
- Safety Data
- Wi-Fi Data
- App Data (e.g., from a trip planner or fare purchasing app)
- Other

- b. What types of data not currently collected by transit agencies may be useful to your business?
- c. On a scale of 1 to 10, 1 being no interest at all, how would you rate your interest in buying data from transit agencies?

- d. On a scale of 1 to 10, 10 being the most valuable, how would you rate the value of standardized transit data (such as GTFS) to your business?
- e. On a scale of 1 to 10, 10 being the most valuable, how would you rate the value of processed transit data (but not necessarily in a national or international standard), as opposed to raw data?
- f. Would data from transit agencies be more valuable if it were compiled from multiple agencies (as opposed to one agency) within one region (city, region, or state)?
- g. Would data from transit agencies be more valuable if it were compiled from multiple regions (as opposed to one city or region)?
10. What changes do you anticipate in the future that could alter the practice of data sharing and selling for your business?

City and State Interview Guide

Mobility on Demand Sandbox/Smart City Challenge Project Overview

1. What city departments or agencies are collaborating in the project?
2. What private entities have you partnered with to meet the project goals?

Data Sharing and Access

3. Do you share your data with private (public) partners?
 - a. What data is shared?
 - b. In what format is the data shared?

Interviewer will prompt the following categories:

- Raw data
 - Data analytical results in aggregated forms
 - Software with data analytical results
 - Data visualization platforms
 - Smartphone Application
 - Other, please explain _____
- c. If you have an app/platform:
 - i. What type of data do you collect?
 - ii. What type of data do your partners collect?
 - iii. How do you facilitate data sharing among your partners?
 - d. Who manages the data? (e.g., data processing, documentation, catalog, etc.)
 - e. Who owns the data?
 - f. What (if any) barriers (risks) exist to sharing data with the private (public) partners?
 4. Do you have access to (mobility) data from private (public) partners?
 - a. What data is it?
 - b. In what format is the data shared?

Interviewer will prompt the following categories:

- Raw data
 - Data analytical results in aggregated forms
 - Software with data analytical results
 - Data visualization platforms
 - Smartphone Application
 - Other, please explain _____
- c. What (if any) barriers (risks) exist to accessing data from the private (public) partners?

5. Could you please highlight the key takeaways of both parties in the process of negotiating the agreement?
 - a. Could you share the data agreement with us (or any related documentation)?
 - b. Was there a public process when deciding the partners?

Public Access Policies

6. Is the data you collect available to the public?
7. Is private sector data that you receive or purchase available to the public? In what form?
8. How does the public currently access and use data from your agency? How do you anticipate that changing in the future?

Utility Industry Interview Guide

1. Has your utility/organization shared, traded, or sold any kind of utility data?
Sharing data includes open data initiatives and private data sharing agreements (for example, with a research institute). Trading refers to an in-kind arrangement (for example, providing data in exchange for services or advertising). Selling exchanges data for monetary compensation.

If yes, interviewer will ask these follow up questions:

- a. What data was shared, traded, or sold? What criteria did your utility/organization use to define time intervals, time period, and frequency of providing this data?
- b. How was it shared, traded, or sold? What are the benefits to your method/platform?
- c. Who was it shared or traded with, or sold to? If smart meter data is shared—how is individual customer data used compared to aggregated/anonymized data used (are uses different, is the value of these types of data different)?
- d. How did your utility/organization identify and engage the data user/buyer?
- e. How does your organization process and standardize data? What do you recommend to transit agencies, to data sharing standards?
- f. What benefit(s) did your utility/organization receive from sharing, trading, or selling data?
- g. Can you quantify the benefit in monetary terms? If priced, how did your utility/organization determine pricing for the data?
- h. Were there negative consequences of sharing, trading, or selling data?
- i. What (if any) barriers did your utility encounter in the process of sharing, trading, or selling data? For example, consider privacy concerns, technological barriers, public perception challenges, expertise, and capital that needed to be developed.
- j. Is there any information or tool that you think would have made the process of sharing, trading, or selling data easier?
- k. What advice would you give a transit agency considering a similar agreement, considering they have many of the same challenges that an electric utility has in managing data?
2. Is there anyone else that you know of who we should talk to for this project?
This could be someone else at your utility/organization, at another utility/organization, a partner, or a subject expert.

Abbreviations and acronyms used without definitions in TRB publications:

| | |
|------------|--|
| A4A | Airlines for America |
| AAAAE | American Association of Airport Executives |
| AASHO | American Association of State Highway Officials |
| AASHTO | American Association of State Highway and Transportation Officials |
| ACI-NA | Airports Council International–North America |
| ACRP | Airport Cooperative Research Program |
| ADA | Americans with Disabilities Act |
| APTA | American Public Transportation Association |
| ASCE | American Society of Civil Engineers |
| ASME | American Society of Mechanical Engineers |
| ASTM | American Society for Testing and Materials |
| ATA | American Trucking Associations |
| CTAA | Community Transportation Association of America |
| CTBSSP | Commercial Truck and Bus Safety Synthesis Program |
| DHS | Department of Homeland Security |
| DOE | Department of Energy |
| EPA | Environmental Protection Agency |
| FAA | Federal Aviation Administration |
| FAST | Fixing America's Surface Transportation Act (2015) |
| FHWA | Federal Highway Administration |
| FMCSA | Federal Motor Carrier Safety Administration |
| FRA | Federal Railroad Administration |
| FTA | Federal Transit Administration |
| HMCRRP | Hazardous Materials Cooperative Research Program |
| IEEE | Institute of Electrical and Electronics Engineers |
| ISTEA | Intermodal Surface Transportation Efficiency Act of 1991 |
| ITE | Institute of Transportation Engineers |
| MAP-21 | Moving Ahead for Progress in the 21st Century Act (2012) |
| NASA | National Aeronautics and Space Administration |
| NASAO | National Association of State Aviation Officials |
| NCFRP | National Cooperative Freight Research Program |
| NCHRP | National Cooperative Highway Research Program |
| NHTSA | National Highway Traffic Safety Administration |
| NTSB | National Transportation Safety Board |
| PHMSA | Pipeline and Hazardous Materials Safety Administration |
| RITA | Research and Innovative Technology Administration |
| SAE | Society of Automotive Engineers |
| SAFETEA-LU | Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (2005) |
| TCRP | Transit Cooperative Research Program |
| TDC | Transit Development Corporation |
| TEA-21 | Transportation Equity Act for the 21st Century (1998) |
| TRB | Transportation Research Board |
| TSA | Transportation Security Administration |
| U.S. DOT | United States Department of Transportation |

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001

ADDRESS SERVICE REQUESTED

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.

www.national-academies.org

ISBN 978-0-309-48141-0



9 780309 481410