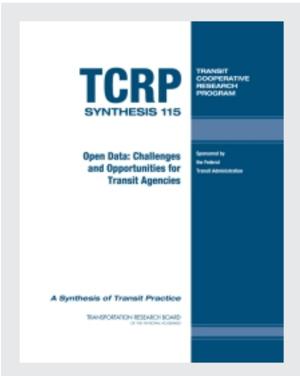


Open Data: Challenges and Opportunities for Transit Agencies

DETAILS

123 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-43272-6 | DOI 10.17226/22195

AUTHORS

Carol L. Schweiger; Transit Cooperative Research Program; Transportation Research Board; National Academies of Sciences, Engineering, and Medicine

TRANSIT COOPERATIVE RESEARCH PROGRAM

TCRP SYNTHESIS 115

Open Data: Challenges and Opportunities for Transit Agencies

A Synthesis of Transit Practice

CONSULTANT
Carol L. Schweiger
TranSystems Corporation
Boston, Massachusetts

SUBJECT AREAS
Planning and Forecasting • Public Transportation

Research Sponsored by the Federal Transit Administration in Cooperation with the Transit Development Corporation

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.
2015
www.TRB.org

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, to adapt appropriate new technologies from other industries, and to 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 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 longstanding and successful National Cooperative Highway Research Program, undertakes research and other technical activities in response to the needs of transit service providers. The scope of TCRP includes a variety of 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 Academy of Sciences, 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) Committee.

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 Committee to formulate the re-search program by identifying the highest priority projects. As part of the evaluation, the TOPS Committee 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 re-search programs since 1962. As in other TRB activities, TCRP project panels serve voluntarily without compensation.

Because research cannot have the desired impact if products fail to reach the intended audience, special emphasis is placed on disseminating TCRP results to the intended end 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.

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

TCRP SYNTHESIS 115

Project J-7, Topic SA-34
ISSN 1073-4880
ISBN 978-0-309-27171-4
Library of Congress Control Number 2014959838

NOTICE

The project that is the subject of this report was a part of the Transit Cooperative Research Program conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council. Such approval reflects the Governing Board's judgment that the project concerned is appropriate with respect to both the purposes and resources of the National Research Council.

The members of the technical advisory panel selected to monitor this project and to review this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. The opinions and conclusions expressed or implied are those of the research agency that performed the research, and while they have been accepted as appropriate by the technical panel, they are not necessarily those of the Transportation Research Board, the Transit Development Corporation, the National Research Council, or the Federal Transit Administration of the U.S. Department of Transportation.

Each report is reviewed and accepted for publication by the technical panel according to procedures established and monitored by the Transportation Research Board Executive Committee and the Governing Board of the National Research Council.

The Transportation Research Board of The National Academies, the Transit Development Corporation, the National Research Council, and the Federal Transit Administration (sponsor 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 clarity and completeness of the project reporting.

Published 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 at
http://www.national-academies.org/trb/bookstore

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Victor J. Dzau is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 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. **www.TRB.org**

www.national-academies.org

TOPIC PANEL SA-34
STEPHANIE CAMAY, *Parsons Brinckerhoff, New York, NY*
JOHN CROCKER, *Metropolitan Atlanta Rapid Transit Authority, Atlanta, GA*
TODD HEMINGSON, *Capital Metropolitan Transportation Authority, Austin, TX*
JAN-MOU "JAMES" LI, *Oak Ridge National Laboratory, Knoxville, TN*
BIBIANA McHUGH, *TriMet, Portland, OR*
RAJEEV ROY, *The Regional Municipality of York-Transportation and Community Planning, Richmond Hill, ON, Canada*
ERNEST TOLLERSON, *Metropolitan Transportation Authority of New York, New York, NY*
CHARLENE WILDER, *Federal Transit Administration* (Liaison)

SYNTHESIS STUDIES STAFF
STEPHEN R. GODWIN, *Director for Studies and Special Programs*
JON M. WILLIAMS, *Program Director, IDEA and Synthesis Studies*
JO ALLEN GAUSE, *Senior Program Officer*
GAIL R. STABA, *Senior Program Officer*
DONNA L. VLASAK, *Senior Program Officer*
TANYA M. ZWAHLEN, *Consultant*
DON TIPPMAN, *Senior Editor*
CHERYL KEITH, *Senior Program Assistant*
DEMISHA WILLIAMS, *Senior Program Assistant*
DEBBIE IRVIN, *Program Associate*

COOPERATIVE RESEARCH PROGRAMS STAFF
CHRISTOPHER W. JENKS, *Director, Cooperative Research Programs*
GWEN CHISHOLM SMITH, *Senior Program Officer*
JEFFREY L. OSER, *Senior Program Assistant*
EILEEN P. DELANEY, *Director of Publications*

TCRP COMMITTEE FOR PROJECT J-7

CHAIR
BRAD J. MILLER, *Pinellas Suncoast Transit Authority, St. Petersburg, FL*

MEMBERS
DONNA DeMARTINO, *San Joaquin Regional Transit District, Stockton, CA*
MICHAEL FORD, *Ann Arbor Transportation Authority, Ann Arbor, MI*
BOBBY J. GRIFFIN, *Griffin and Associates, Flower Mound, TX*
ROBERT H. IRWIN, *Consultant, Sooke, BC, Canada*
JEANNE KRIEG, *Eastern Contra Costa Transit Authority, Antioch, CA*
PAUL J. LARROUSSE, *Rutgers, The State University of New Jersey, New Brunswick*
DAVID A. LEE, *Connecticut Transit, Hartford*
ELIZABETH PRESUTTI, *Des Moines Area Regional Transit Authority–DART*
ROBERT H. PRINCE, JR., *AECOM Consulting Transportation Group, Inc., Boston, MA*

FTA LIAISON
JARRETT W. STOLTZFUS, *Federal Transit Administration*

TRB LIAISON
JENNIFER L. WEEKS, *Transportation Research Board*

FOREWORD

Transit administrators, engineers, and researchers often face problems for which information already exists, either in documented form or as undocumented experience and practice. This information may be fragmented, scattered, and unevaluated. As a consequence, full knowledge of what has been learned about a problem may not be brought to bear on its solution. Costly research findings may go unused, valuable experience may be overlooked, and due consideration may not be given to recommended practices for solving or alleviating the problem.

There is information on nearly every subject of concern to the transit industry. Much of it derives from research or from the work of practitioners faced with problems in their day-to-day work. To provide a systematic means for assembling and evaluating such useful information and to make it available to the entire transit community, the Transit Cooperative Research Program Oversight and Project Selection (TOPS) Committee authorized the Transportation Research Board to undertake a continuing study. This study, TCRP Project J-7, "Synthesis of Information Related to Transit Problems," searches out and synthesizes useful knowledge from all available sources and prepares concise, documented reports on specific topics. Reports from this endeavor constitute a TCRP report series, *Synthesis of Transit Practice*.

This synthesis series reports on current knowledge and practice, in a compact format, without the detailed directions usually found in handbooks or design manuals. Each report in the series provides a compendium of the best knowledge available on those measures found to be the most successful in resolving specific problems.

PREFACE

By Donna L. Vlasak
Senior Program Officer
Transportation Research Board

The report documents the current state of the practice in the use of open data for improved transit planning, service quality, and customer information; the implications of open data and open documentation policies; and the impact of open data on transit agencies, and the public and private sectors. It focuses on successful practices in open transit data policies, use, protocols, and licensing. This synthesis is intended for transit agencies, the public, and the private sector.

A literature review and detailed survey responses from 67 of 67 agencies surveyed around the world, including Canada and 14 European countries, a response rate of 100 percent, are provided. Also, four case examples offer more detailed information from agencies and organizations that have significant experience with providing open data.

Carol L. Schweiger, TranSystems Corporation, Boston, Massachusetts, collected and synthesized the information and wrote the report, under the guidance of a panel of experts in the subject area. The members of the topic panel are acknowledged on the preceding page. This synthesis is an immediately useful document that records the practices that were acceptable within the limitations of the knowledge available at the time of its preparation. As progress in research and practice continues, new knowledge will be added to that now at hand.

CONTENTS

OPEN DATA: CHALLENGES AND OPPORTUNITIES FOR TRANSIT AGENCIES

SUMMARY

In the past 5 years, more and more transit agencies have begun making schedule and real-time operational data available to the public. "Open data" provide opportunities for agencies to inform the public in a variety of ways about transit agency services.

The purpose of this synthesis is to document the current state of the practice and policies in the use of open data for improved transit planning, service quality, and customer information; the implications of open data and open documentation policies; and the impact of open data on transit agencies and the public and private sectors. The synthesis focuses on successful practices in open transit data policies, use, protocols, and licensing. A literature review and survey collected key information about open transit data. The survey was sent to 67 transit agencies around the world and had a 100% response rate. Of the 67 surveys received, three were from Canadian agencies and 14 from European agencies. U.S. responses represent agencies that carry a total of more than 5.4 billion passengers annually (annual unlinked trips), with U.S. agencies' annual ridership ranging from 1.8 million (a county transit system in Florida) to 2.6 billion (Metropolitan Transportation Authority in New York City).

The background of open transit data in the United States is as follows. Prior to 1998, data generated by technologies deployed by public transit agencies were not made available to the public. In 1998, Bay Area Rapid Transit released schedule data in the comma-separated values (.csv) format; this was the first known release of transit data to the public. Tri-County Metropolitan Transportation District of Oregon (TriMet) worked with Google in the creation of General Transit Feed Specification (GTFS, originally developed by Google and containing static schedule information for transit agencies, including stop location, route geometrics, and stop times) in 2005. Massachusetts Bay Transportation Authority (MBTA) opened the agency's data in 2009. As of April 2014, according to City-Go-Round, almost 29% of U.S. transit agencies provided open data. In 2003, the Digital Agenda for Europe, Public Service Information Directive was issued, requiring all European Union member states to release public sector information, including open public transport data. Many public transit agencies in the Asia-Pacific region are beginning to open their data as well, such as a recent initiative to combine and release the data from many public transit operators in Tokyo, Japan.

In addition, not only have public transit authorities benefited from providing open data, but the public, private, and independent sectors also have realized benefits. Transit authorities that have embraced transparency by providing open data have improved the perception and increased the visibility of transit. They also have been able to use the data they are releasing to the public to make internal improvements. The public now has access to many free applications that provide real-time and static transit information, which greatly facilitates travel using transit. Private businesses have been created or expanded to work with open transit data and have developed innovative applications that, in some cases, could not have been developed in a public agency. Finally, the independent sector, consisting of academic institutions and research and development organizations, has been instrumental in researching, analyzing, using, and promoting the creation and use of open transit data.

This synthesis examines and documents the state of the practice in open transit data using the following five elements:

- Characteristics of open transit data:
  - Reasons for choosing to provide open data
  - Standards and protocols for providing open data
  - Underlying technology used to generate the open data
- Legal and licensing issues and practices:
  - Legal and licensing issues
  - Public disclosure practices
- Uses of open data:
  - Applications
  - Decision-support tools
  - Visualizations
- Costs and benefits of providing open data
- Opportunities and challenges:
  - Techniques to engage users and reusers of data
  - Challenges associated with providing open data
  - Impacts on transit agencies and the public and private sectors.

SUMMARY OF PROJECT FINDINGS

Key statistics from the study follow.

- Fifty-seven or almost 83% of the survey respondents provided open data.
- The top four reasons for not providing open data, according to the survey, were
  - Too much effort to produce the data/we do not have the time or people to do the work required;
  - Too much effort to clean the data;
  - We cannot control what someone will do with our data; and
  - We do not know the accuracy of our data.

In addition to the four case examples conducted as part of this synthesis project and presented in chapter seven, two examples show the value and power of open transit data:

- Moscow's transit authority relied on open data to determine whether more investment in rail networks was necessary or if other services could better meet demand. Instead of building a new rail line at considerable expense, the authority restructured bus service, which allowed flexibility for future shifts in population. Not only did the authority avoid incurring more than $1 billion in infrastructure costs, but the restructured bus service also saved an average of 3 minutes per trip during the morning commute, amounting to 10 hours of travel time for each rider every year.
- In New Jersey, NJ Transit released data on passenger flows to the public in 2012. Third parties quickly analyzed ridership at different times of day and were able to pinpoint underutilized rail stops, which led to more express trains and a savings of 6 minutes in the average commuting time during peak hours.

In summary, based on the literature review, the responses to the questionnaire, and the case examples, the key findings of this synthesis project are as follows:

- The benefits to the agency strongly support open transit data. The availability of open transit data encourages innovation that could not be accomplished solely by agency staff.

  The top five overall benefits experienced by survey respondents were (1) increased awareness of our services; (2) empowered our customers; (3) encouraged innovation outside of the agency; (4) improved the perception of our agency (e.g., openness/transparency); and (5) provided opportunities for private businesses.

- Engaging application developers, other data users, and customers is an approach that can accomplish several critical tasks, including:
  - Obtaining feedback on data anomalies and data quality issues;
  - Ensuring that some portion of the applications developed by third parties meets the needs of customers; and
  - Finding out more about how people want to use/reuse the data.

  There are several ways to engage developers and customers. Results of the survey indicated that the most effective methods are conducting face-to-face events, conferences, and "meetups." Meetups are informal meetings to discuss particular topics, such as application development.

- The results of the literature review and survey indicate that standards and commonly used formats can be used to facilitate the generation and use of open data. Further, using standards makes it easier to transfer applications from one agency to another.
- Open transit data result in innovation that could not be accomplished within a transit agency. That is not to say that sufficient intellect does not exist in a transit agency; rather, it is an issue of having sufficient resources to develop applications and conduct analyses at the scale that can be done in an open market.

Specific findings based on the five elements include:

- Characteristics of open transit data:
  - The top three types of open data are routes, schedules, and station/stop locations;
  - The most prevalent underlying technologies that produce open data are scheduling software, geographic information system (GIS) software, computer-aided dispatch (CAD)/automatic vehicle location (AVL), and real-time arrival prediction software;
  - The overwhelming reasons for opening transit data are related to customer information—increasing access to this information and improving the information and customer service;
  - The primary factor that went into the decision about what data to open was the ease of releasing the data (more than half of the survey respondents indicated this);
  - A variety of standards and formats are being used, including GTFS (47 or 83.9% of respondents), Extensible Markup Language (XML) (26 or 46.4%), and .csv (18 or 32.1%), followed by GTFS-realtime (15 or 26.8%); and
  - Degree of openness
    - △ Thirty-two or 57.4% of the respondents reported that the data are completely open (everyone has access);
    - △ Forty-seven or 83.6% reported that the data are available in formats that are easily retrieved and processed;
    - △ Forty-nine or 87.3% reported that there is no cost for their open data; and
    - △ Forty-three or 79.2% reported that there are unlimited rights to use, reuse, and redistribute their data.
- Legal and licensing issues and practices:
  - Twenty-nine or 50.9% of the survey respondents reported that they require a license or agreement to use their agency's open data;
  - The top three elements that license agreements cover are the right to use the agency's data; nonguarantee of data availability, accuracy, or timeliness; and liability limitations for missing or incorrect data;
  - Only one respondent experienced any legal issues resulting from the release of open data to the public; and

  - The top three steps that respondents took to disclose their data publicly were converting transit data into formats suitable for public use; improving data quality to ensure accuracy and reliability; and adopting an open, nonproprietary data standard.
- Uses of open data:
  - The top five types of customer applications that have been developed as a result of providing open data are (in descending order of frequency) trip planning, mobile applications, real-time transit information (arrival/departure times, delays, detours), maps, and data visualization;
  - The top five decision-support tools that have been developed are data visualization, service planning and evaluation, route layout and design, performance analysis, and travel time and capacity analysis;
  - Almost two-thirds (33 or 63.5%) of respondents do not track usage of their open data;
  - The two most prevalent methods of tracking are to monitor data downloads and keep track of applications developed;
  - For mobile applications, an equal number of respondents reported Android and iOS applications; and
  - Sixteen respondents reported a total of almost 266 million Application Programming Interface (API) calls per month.
- Costs and benefits of providing open data:
  - The top five types of costs associated with providing open data are staff time to update, fix, and maintain data as needed; internal staff time to convert data to an open format; staff time needed to validate and monitor the data for accuracy; staff time to liaise with data users and developers; and web service for hosting data;
  - Almost 90% (43 or 89.4%) of respondents could not quantify how much time is spent on any of these activities;
  - There was limited information regarding the actual labor required from specific staff in the organization and the costs associated with open data; and
  - The top three benefits experienced by survey respondents are increased awareness of their services, empowerment of their customers, and encouragement of innovation outside of the agency.
- Opportunities and challenges:
  - Almost 70% (33 or 69.6%) of respondents engage or have a dialogue with existing and potential data users and reusers;
  - Twenty-five or 75.8% of respondents engage data users and reusers to obtain feedback on data anomalies and data quality issues. Twenty-four or 60% of the respondents use face-to-face events to engage these groups;
  - The organizational impacts on the agency resulting from opening the data ranged from increased transparency to better and more accurate internal data to lower costs to provide information. The majority of negative impacts were related to resources required to maintain an open data program;
  - Impacts on the customer were numerous, including better and more accessible information for customers; better perception, visibility and awareness of services, and improved customer satisfaction;
  - In terms of impacts on the public, creating and improving access to additional and higher quality public services was mentioned, along with improving public perception/image of transit, making transit more competitive, providing better regional coordination of services, encouraging innovation, and providing a better transit experience;
  - The impacts on the private sector are primarily providing business/commercial and development opportunities, including new and expanded companies (e.g., creating a new ecosystem of private entrepreneurs); enabling innovation and the creation of applications that may not have been created by the public sector; and adding value to existing public services; and
  - Challenges were noted by survey respondents in five areas: (1) resources and organizational issues; (2) data quality and timeliness issues; (3) standards and formatting

issues; (4) marketing issues relating to making the open data known and addressing branding issues; and (5) technical issues.

Several conclusions can be drawn from the results of the synthesis project, including:

- The benefits to the agency strongly support open transit data;
- The impacts of open transit data on customers and the general public are significant;
- The impacts on the private sector have been encouraging over the past several years. Applications and visualizations that could not necessarily have been conceived or developed by a transit agency have been created;
- The legal fears often thought to be barriers to opening transit data have not been realized;
- Standards greatly facilitate the use of open transit data, although this sometimes requires additional effort in producing the data;
- Engaging with data users and reusers has the potential to increase the value of the applications and visualizations;
- Five factors lead to a successful open data program: (1) obtaining and maintaining management-level support for such a program; (2) recognizing the need for the appropriate level of resources required to provide and maintain open data; (3) establishing ways to monitor data accuracy, timeliness, reliability, quality, usage, and maintenance; (4) creating and maintaining licensing or registration; and (5) having an ongoing dialogue with both developers and customers, a practice shown to increase the value of the data and products that are based on the data.

CHAPTER ONE

INTRODUCTION

PROJECT BACKGROUND AND OBJECTIVES

The primary focus of this synthesis is on the current state of the practice and policies in the use of open data for improved transit planning; service quality, customer information, and customer experience; implications of open data and open documentation policies; and their impact on transit agencies and public support. In addition, successful practices in open data policy, use, and protocols and licensing in the United States and abroad are documented.

Within the past 5 years, more and more transit agencies have made schedule and real-time operational data available to the public. "Open data" provide opportunities for agencies to inform the public in a variety of ways about a transit agency's services. For example, there is significant value to having web-based and mobile applications that are developed by people outside the transit agency—these applications allow riders to navigate public transit systems more easily. In this example, the agency does not bear the costs associated with the application development and encourages innovation in terms of how to present transit information to the public. Open data are being used to create enterprise-facing, decision-support tools that can help to optimize operations in real time, improve maintenance, and support capital planning and programs. However, in addition to opportunities, open data present challenges for agency operations and other business functions.

The use of open transit data was first reported by TRB in 2011's *TCRP Synthesis 91* (*1*). At the time that synthesis was prepared, there was a keen interest on the part of several U.S. transit agencies to provide open data, and legislation had just been passed that governed the required open data in parts of Europe. In addition, local, state, and federal requirements for open data (including the Open Government Directive issued by President Obama in 2009) were under development in the United States. Since that report was prepared, many more agencies have embraced and provided open data, and have realized benefits well beyond what was originally thought. This development occurs within a context of agencies requiring opening of their data. This synthesis documents the current state of the practice and policies in the use of open data for improved transit planning, service quality and customer information, and implications of open data and open documentation policies; and their impact on transit agencies and the public. It focuses on successful practices in open data policies, use, protocols and standards, and licensing.

This synthesis documents but is not limited to the following five major elements.

- Characteristics of open transit data
  - Reasons for choosing to provide open data
  - Standards and protocols for providing open data
  - Underlying technology used to generate open data
- Legal and licensing issues and practices
  - Legal and licensing issues
  - Public disclosure practices
- Uses of open data
  - Applications
  - Decision-support tools
  - Visualizations
- Costs and benefits of providing open data
- Opportunities and challenges
  - Techniques for engaging users and reusers of data
  - Challenges associated with providing open data
  - Impacts on transit agencies and the public and private sectors

A literature review, survey of selected transit agencies and other stakeholders, and detailed case examples or profiles were done to report on the state of the practice, including innovations, lessons learned, challenges, and gaps in information.

A review of the relevant literature in the field is combined with surveys of selected transit agencies and other appropriate stakeholders to report on the current state of the practice. Based on survey results, four case examples were developed to describe innovative and successful practices, as well as lessons learned and gaps in information.

TECHNICAL APPROACH TO PROJECT

This synthesis project was conducted in five major steps. First, a literature review was performed to identify the characteristics of open data; legal and licensing issues and practices; uses of open data; and the costs, benefits, challenges, and opportunities resulting from providing open data. See the References section for a list of sources.

Second, a survey was conducted to collect information on factors such as the reasons for choosing to provide open data; standards and protocols being used; public disclosure practices; customer applications and other data uses; techniques for engaging actual and potential users and reusers of the data;

and challenges associated with providing open data. In addition, information regarding the impacts on transit agencies and the public and private sectors was explored through the survey. The survey instrument is shown in Appendix A, and the list of agencies and staff titles responding to the survey is shown in Appendix B.

Third, the survey results were documented and summarized. Fourth, telephone interviews were conducted with key personnel at four agencies and organizations that have significant experience with providing open data; case examples from those four agencies are presented in chapter seven. Finally, the results and conclusions were prepared and documented.

REPORT ORGANIZATION

This report is organized as follows:

- Chapter one presents the goals and objectives of the synthesis and describes the technical approach used to conduct the project.
- Chapter two summarizes the literature review.
- Chapter three describes the characteristics of open data, including the reasons for choosing to provide open data; underlying technology used to generate the data provided to the public; and standards, protocols, and formats being used to provide open data.
- Chapter four describes the legal and licensing issues and practices, and public disclosure practices.
- Chapter five discusses the uses of open data, including customer applications, decision-support tools, other uses of open data (e.g., nontransit applications), and open data applications statistics.
- Chapter six presents information about the costs, benefits, challenges, and opportunities resulting from providing open data.
- Chapter seven presents case examples from selected agencies that have significant experience with open data.
- Chapter eight summarizes the results of the synthesis and presents conclusions.
- An Abbreviations and Acronyms section lists those elements.
- The References section contains the list of literature that was reviewed and referred to in this report.
- Appendix A contains the survey instrument.
- Appendix B shows the list of responding agencies and staff titles.
- Appendix C provides supplemental information regarding conferences, meetings, and agency events dedicated to open transit data.
- Appendix D shows the total annual ridership for each responding agency.
- Appendix E contains website addresses for agency license agreements and terms of use.
- Appendix F shows examples of customer information-related applications that are available through agency websites.
- Appendix G provides examples of applications noted through the Transport Innovation Deployment for Europe (TIDE) project.
- Appendix H has a list of Transport for London (TfL) open data available from the London Datastore.

CHAPTER TWO

LITERATURE REVIEW

The literature review revealed that a wide variety of reports, papers, articles, and press releases have been written about open data in transit. The literature review is divided into the following sections, including the five elements identified in chapter one:

- Characteristics of open data, including standard(s) used for open data
- Legal and licensing issues
- Uses of open data, including customer and other applications
- Costs and benefits
- Opportunities and challenges, including:
  - Engaging existing and potential data users and reusers
  - Impacts of open data

The first step of the literature review was to conduct an online Transport Research International Documentation (TRID) search. This TRID search yielded 30 documents, the most relevant of which were reviewed and used as input for this report. The second step was to obtain and review articles, press releases, and website information directly from agencies and open data organizations across the world. The third step was to review research reports from the FTA, FHWA, and TCRP. Finally, other papers and articles were obtained from a variety of sources, including the following:

- TRB Annual Meetings;
- APTA conferences and publications;
- Intelligent Transportation Society of America (ITSA) Annual Meetings;
- Intelligent Transportation Society World Congresses;
- Open Data Organizations;
- European Commission project documentation; and
- Internet searches.

The sources used in this synthesis are listed in the References section. Supplemental information regarding conferences, meetings, and agency events dedicated to open transit data can be found in Appendix C.

INTRODUCTION TO OPEN TRANSIT DATA

Before describing the types of data that are being released by transit agencies according to the literature, it is important to define the term "open data." Many definitions include the basic principle that the data are free to use, reuse, and redistribute. According to the Open Data Institute (*2*),

> Open data is information that is available for anyone to use, for any purpose, at no cost. Open data has to have a license that says it is open data. Without a license, the data can't be reused. The license might also say:
>
> - that people who use the data must credit whoever is publishing it (this is called attribution); and
> - that people who mix the data with other data have to also release the results as open data (this is called share-alike).

In addition, the Open Data Institute describes what constitutes "good" open data:

- Can be shared easily;
- Is available in a standard, structured format;
- Has guaranteed availability and consistency; and
- Is traceable, through processing, back to where it originates.

Finally, the conditions of open data are defined as follows (*3*):

- *Complete*—taking privacy into consideration;
- *Primary*—being as close as possible to the source;
- *Actual*—as automatic as possible in the exchange;
- *Accessible*—in digital format for as many users as possible and as many purposes as possible;
- *Readable* machine to machine;
- *Free*—mostly accessible for no cost and with no restrictions for use; and
- In an *open format* and to *follow a standard* [e.g., Extensible Markup Language (XML)].

Organizations that provide guidance about making a business case for open data; how to open the data; engaging with data consumers and reusers; licensing; and the effects of open data include:

- Open Knowledge Foundation
- Sunlight Foundation
- Open Data Institute
- The Open Data Foundation
- Project Open Data
- The Open Data Center Alliance
- Open Mobile Alliance
- Public Data Transit Community

On January 21, 2009, President Obama issued an Open Government Directive, which directed executive departments and agencies to take actions that support transparency, participation, and collaboration:

> Transparency promotes accountability by providing the public with information about what the Government is doing. Participation allows members of the public to contribute ideas and expertise so that their government can make policies with the benefit of information that is widely dispersed in society. Collaboration improves the effectiveness of Government by encouraging partnerships and cooperation within the Federal Government, across levels of government, and between the Government and private institutions (*4*).

In 2003, the Digital Agenda for Europe, Public Service Information Directive was issued, significantly affecting the release and "re-use of public sector information," including open public transport data (*5*). Through this directive, open data are "obligatory for all [European Union] member states." In February 2012, Claudia Schwegmann reported that developing countries are just now beginning to open their data, including public transport data (*6*).

The U.S. DOT supports and promotes open transportation data, as evidenced in its Open Government Plan for 2012 to 2014 (*7*). As shown in Figure 1, five initiatives in this plan include several open transit data-related activities.

The literature reviewed for this synthesis included several other concepts that are directly related to open transit data, including "open data movement" and "open transport." The open data movement and its relationship to public transit is defined by Eros et al. as coming

> from philosophical principles of open government, transparency, and accountability, and practical motivations related to increased returns on public investment, downstream wealth creation, more potential brainpower brought to examining complex problems,

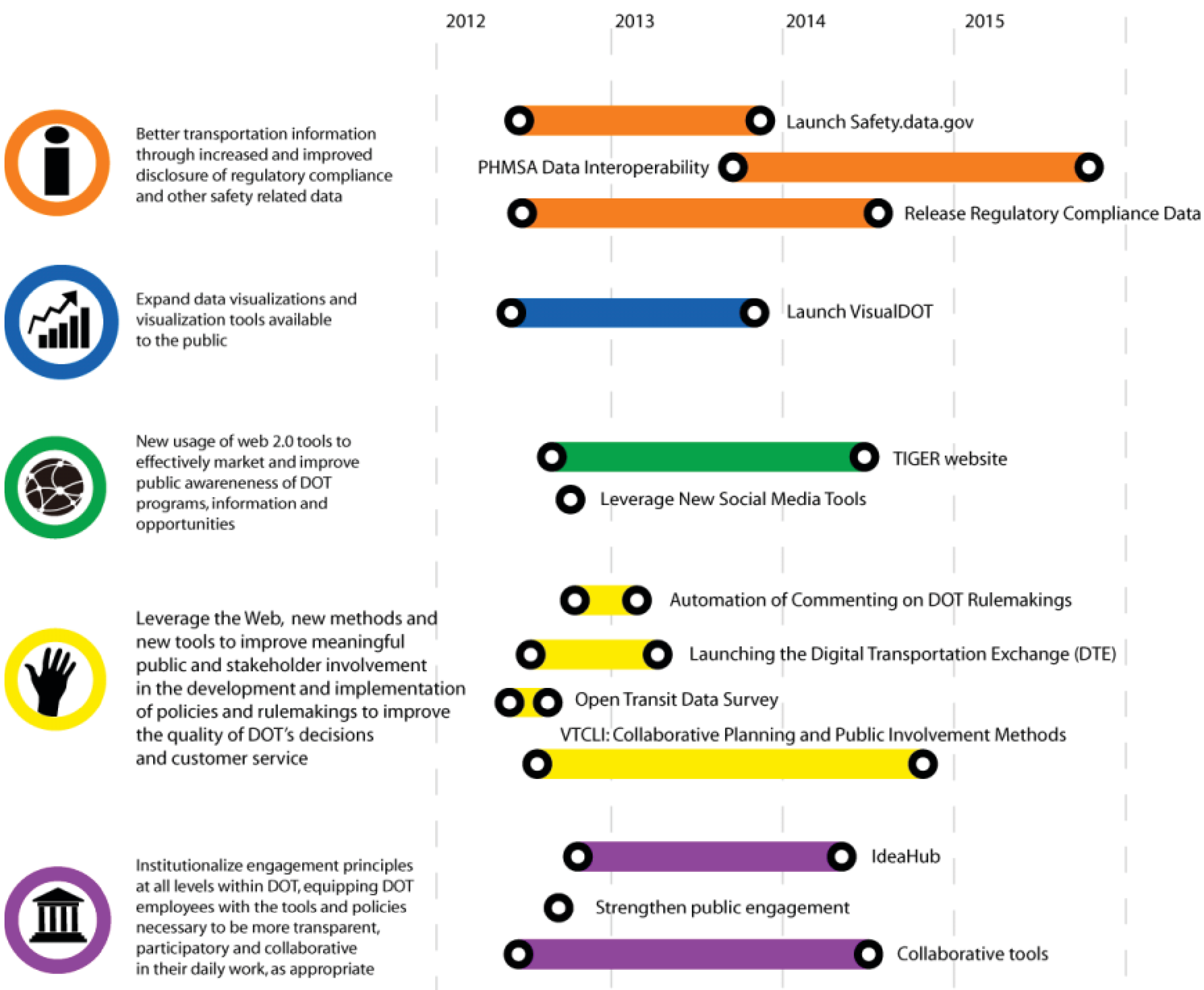


FIGURE 1 U.S. DOT Open Government initiatives.

> and enhanced public policy and service delivery. For transportation, the open data movement has fundamentally shifted how agencies communicate with users as an increasing number move from tightly controlling data and the products derived from them, towards generating and releasing data with minimal control over the end products. In transportation, the confluence of open data, GTFS [General Transit Feed Specification] (originally developed by Google and containing static schedule information for transit agencies, including stop locations, route geometrics, and stop times), and increasingly ubiquitous mobile computing, sensing, and communication technologies (epitomized by the 'smartphone'), has spurred numerous technical innovations from a range of actors. Tools include applications that assist with trip planning, ridesharing, timetable creation, data visualization, planning analysis, interactive voice response, and real-time information provision. Together, GTFS and GTFS-RealTime (containing real-time information related to vehicle positions, service alerts, and trip updates such as delays, cancellations, etc.) enable transit agencies and operators to engage the power of the software developer community and citizenry more generally to create new forms of information services about public transportation (*8*).

Open transport, according to the World Bank Open Transport Team,

> defines the next generation of tools and methodologies for managing and planning transport systems in resource-constrained environments. Open Transport is defined by three principles: Open data standards, open source software and open data. Open data standards are freely and publicly available, with no required use agreements. Open source software features universal access and redistribution rights via free licenses to a product's design, code, or blueprint (*9*).

From the open transit data perspective, "while not all transport data lends itself to be open, there are benefits to releasing some data, such as public transit service information, to achieve economies of scale in generation of third-party applications to support wider and more efficient use of a transit network" (*9*).

An overall history of the open data movement between 2006 and 2011 is shown in Figure 2, which shows "the evolution and role of [application] API in building influential and essential tools and applications for web users, plus the demographics of public data usage around the world" (*10*). This evolution had a significant impact on the introduction of open transit data.

A timeline of transit open data between 2005 and 2012, as detailed by Francisca Rojas, is shown in Figure 3.

> As agencies saw that disseminating transit information by electronic means was valuable to their customers, they began to create

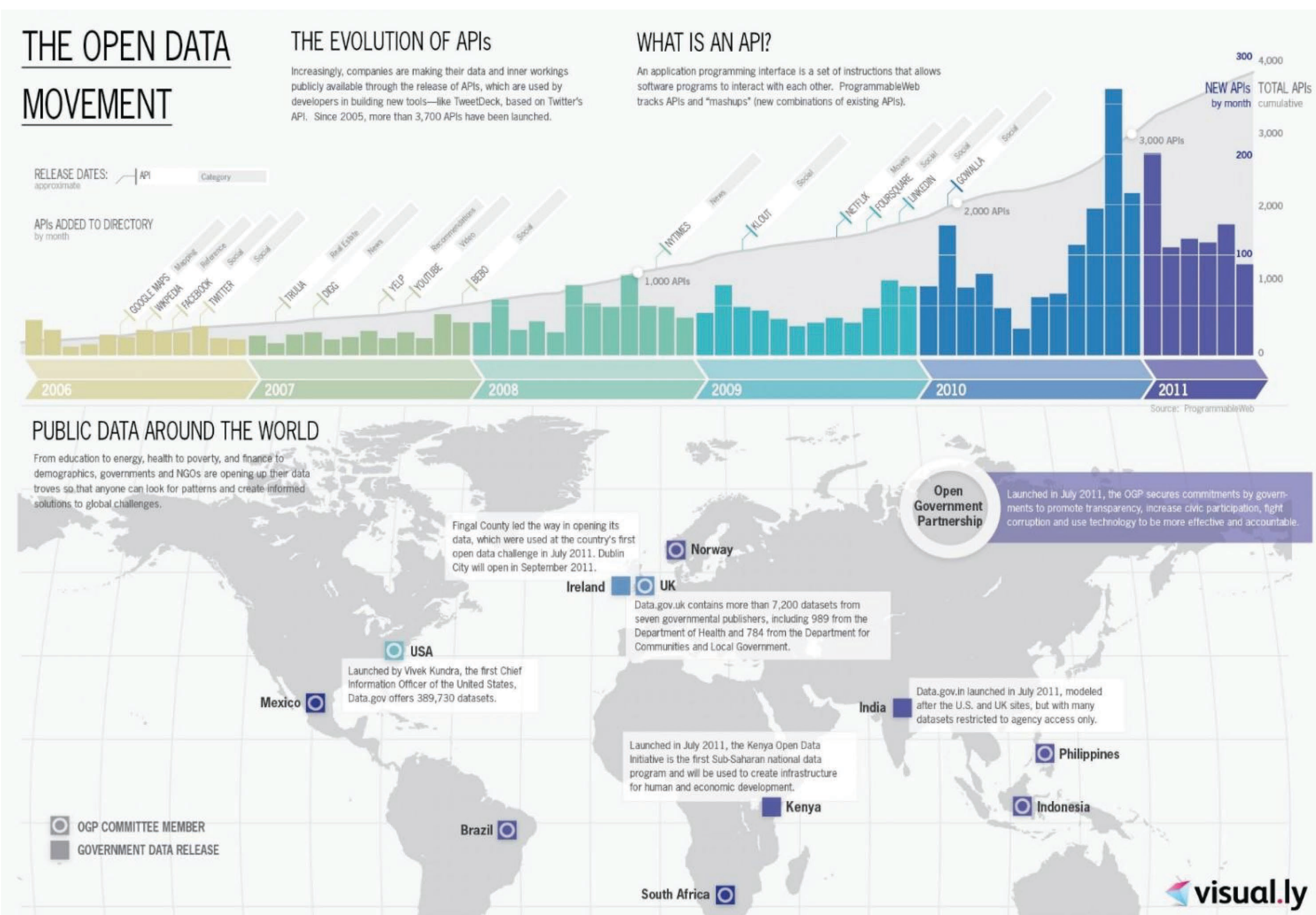


FIGURE 2 Open data movement 2006–2011 (*10*).

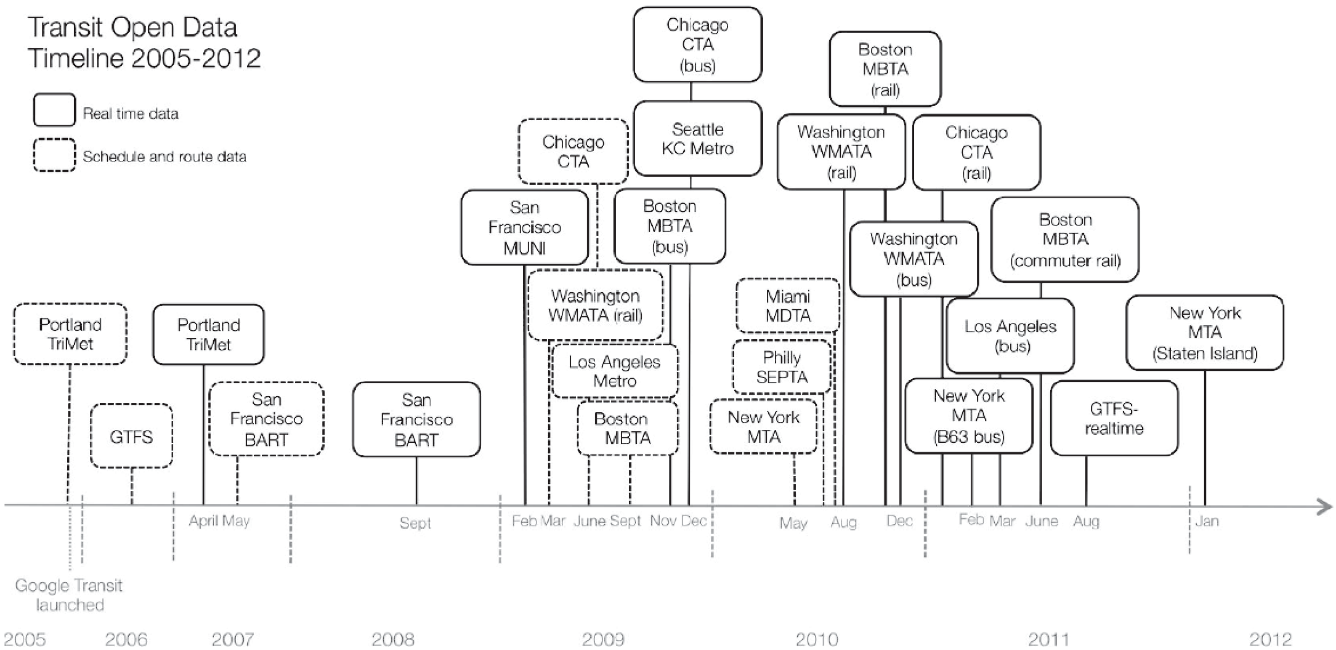


FIGURE 3 Transit open data timeline 2005–2012 (*11*).

> the points of interaction, or Application Programming Interfaces (APIs), needed for third-party software developers to access dynamic, real-time data feeds of bus and train location information. The solid-line boxes in Figure 3 indicate when agencies made available real-time data feeds to the public. This process gained momentum in 2009 and by the end of 2011 all major transit agencies in the United States had 1) posted their routes and schedules on Google Maps, 2) publicly released GTFS files of static transit information, and 3) created APIs for access to real-time information by third-parties (*11*).

Because this timeline ends in 2012, it does not display events in 2013 or 2014, such as Metropolitan Transportation Authority's (MTA's) completion of the Bus Time implementation and release of GTFS-realtime data.

Rojas suggests that TriMet, one of the U.S. transit agencies to pioneer the release and use of open data, started the spread of open transit data to some of the largest transit agencies because of the following factors (*11*):

- A de facto data standard offered by the GTFS format facilitated the process for agencies to integrate schedules and routes into Google Maps (now Google Transit), and for broader public disclosure of those same data sets;
- Demand from technologically savvy, networked transit riders for customized transit information because of the wide adoption of short message service (SMS)-enabled cell phones and location-aware smartphones, which enabled riders to view real-time information while traveling;
- Software developer communities that were eager to learn how to code mobile applications and sought available data sets to develop technical skills, and improve and expand access to transit information;
- Agencies willing to adapt intelligent transportation systems used for internal management of operations into data formats suitable for public disclosure; and
- Open data champions who built networks within agencies to share experiences and seek advice on technical and policy aspects of data disclosure.

The literature reviewed presented the questions that many transit agencies ask as they are determining whether or not to open their data (*12*):

1. What data should we provide?
2. How do we get the data?
3. Will the app developers use our data?
4. Will the developers provide a reliable service?
5. What are good examples of apps?
6. Do we need to produce an app ourselves?
7. What standards should we use?
8. How should we ensure data quality?
9. Should we preprocess our data?
10. Can we charge for our data?

A tool developed by the Center for Technology in Government at the University of Albany "provides a series of questions that take agencies through a review of their existing and proposed open government plans to quickly assess the public value of their open government initiatives" (*13*). This tool, called the Open Government Portfolio Public Value Assessment Tool (PVAT), evaluates each open government initiative using a "multistep question process, which includes making public value judgments, identifying stakeholders and uncovering the impacts that initiative will have on those stakeholder interests" (*13*).

Several pieces of literature covered the reasons transit agencies opened their data (*7*, *14–16*). These justifications include the following:

- Improved customer service
- Increased and equal information access
- Fostered innovative, diverse, and free apps
- Improved agency transparency
- Supported concept that public transit data are publicly owned
- Allowed information to proliferate, extending the reach of public transit information considerably
- Facilitated the development of technology enterprises, generating employment, a highly skilled workforce and wealth for the community, and prompting innovation
- Facilitated grassroots standards such as GTFS
- May result in increased ridership and increased revenue as the result of better passenger information
- Gave developers opportunities to make applications more easily with publicly accessible data, as well as realize a potential economic gain
- Supported better use of personal time for current and future riders because riders will have more information
- Contributed better transit data to help inform regional decision making

To provide some context into how many transit agencies provide open data, the GTFS Data Exchange reports that data are available for 726 transit agencies worldwide (*17*). City-Go-Round, supported by The Rockefeller Foundation, reports that of 866 U.S. transit agencies, 248 (28.64%) have open data in 2014 (*18*). Another point of reference is that, at the time of *TCRP Synthesis 91* (*1*, pp. 38, 52), on May 6, 2010, City-Go-Round reported 107 U.S. transit agencies of 780 (13.72%) were providing open data. This difference between 2010 and 2014 shows the rapid growth of open transit data. Yet another reference (see Figure 4) shows the number of unlinked passenger trips served by agencies with open transit data from 2009 to 2013 (*19*, p. 3).

The Open Knowledge Foundation developed an Open Data Index that reports on the following characteristics of open transit data (and many other types of open government data) by country and, in the United States, by major metropolitan area (*20*):

- Do the data exist?
- Are the data in digital form?
- Are the data publicly available?
- Are the data free of charge?
- Are the data online?
- Are the data machine-readable?
- Are the data available in bulk?

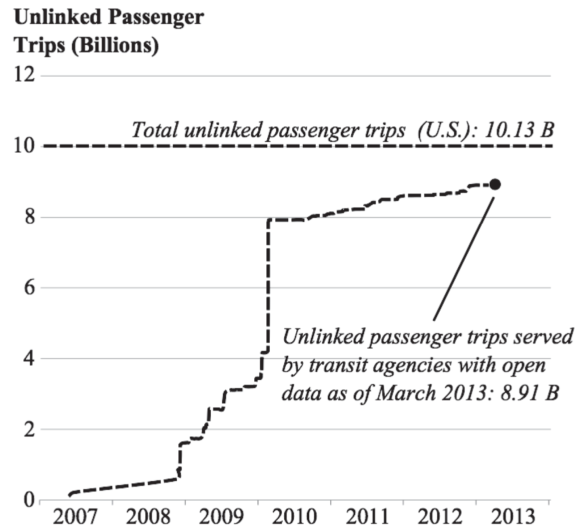


FIGURE 4 Growth of transit agencies with open data by passenger miles served (*19*).

- Are the data openly licensed?
- Are the data up to date?

Figure 5 shows a sample of this list by country. Figure 6 shows a sample by U.S. city. The transit portion of the index is shown on the far right. The transit index is based on answers to the nine questions listed.

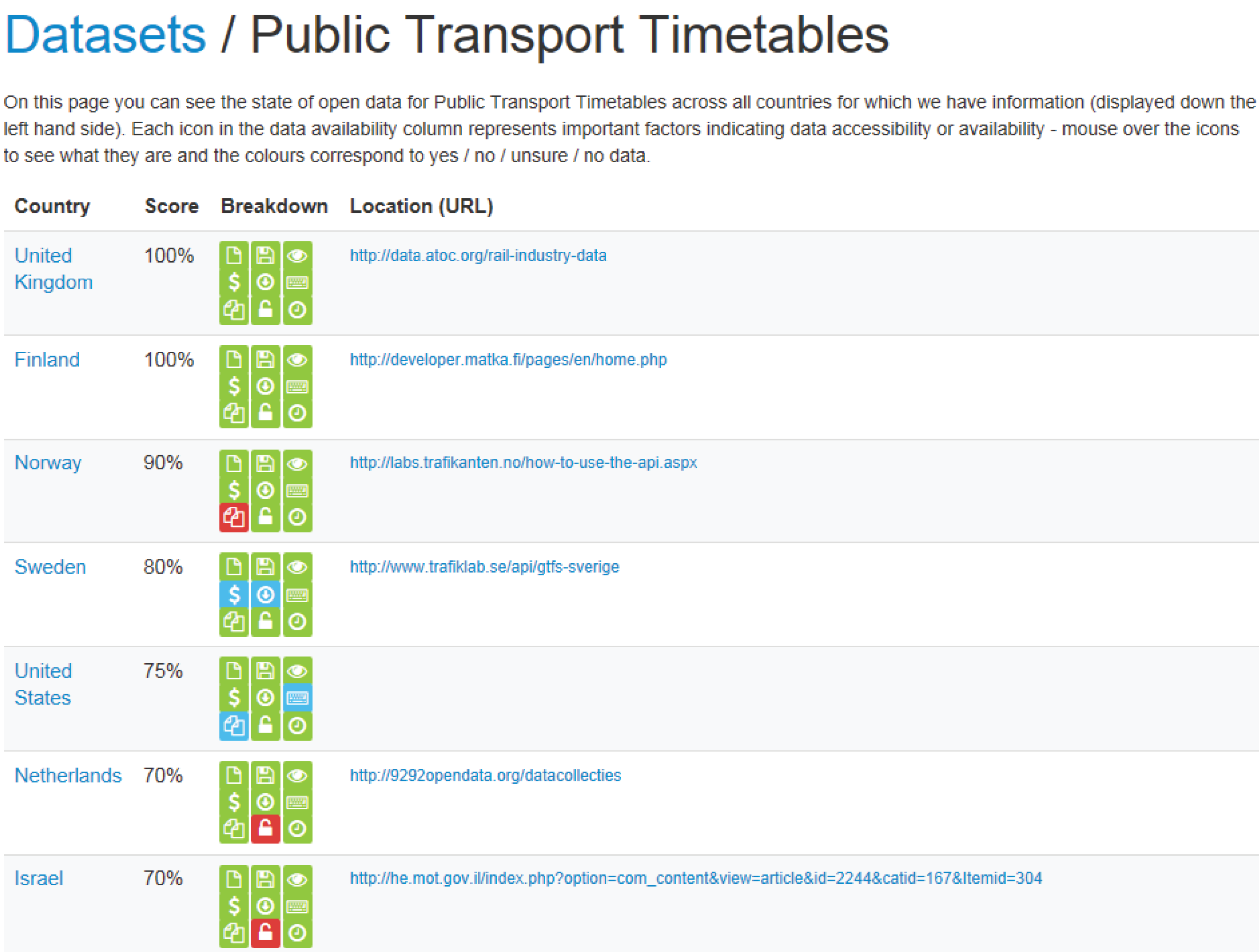
A variety of events focusing on open transit data are being held around the world. These events are described in Appendix C.

CHARACTERISTICS OF OPEN DATA

In guidance written in 2012 for transportation agencies interested in providing open data, the types of data that agencies should consider releasing include schedule data in GTFS format (see the following section for a definition of the de facto GTFS standard), route information in GTFS format, and "infrastructure locations, including stations, roadways and landmarks [from a geographic information system (GIS)]. For enhanced transparency, these data sets should be released [as well]: budgetary data, performance data and ridership data" (*23*), all in comma-separated values (CSV). Young-Jae Lee mentions two additional categories of data: real-time and origin-destination data (*24*).

From a developer's perspective, there are two types of open transit data: (1) static (e.g., transit schedules, routes, and stops), which changes only a few times a year, and (2) real-time (e.g., estimated arrival times, vehicle positions, and service alerts), which changes every few seconds (*25*). Further, there are "two magnitudes of open data," as follows:

- "Fire hose," which is a dump of the complete state of the transit system and is not directly suitable for mobile



Datasets / Public Transport Timetables

On this page you can see the state of open data for Public Transport Timetables across all countries for which we have information (displayed down the left hand side). Each icon in the data availability column represents important factors indicating data accessibility or availability - mouse over the icons to see what they are and the colours correspond to yes / no / unsure / no data.

Country	Score	Breakdown	Location (URL)
United Kingdom	100%		http://data.atoc.org/rail-industry-data
Finland	100%		http://developer.matka.fi/pages/en/home.php
Norway	90%		http://labs.trafikanten.no/how-to-use-the-api.aspx
Sweden	80%		http://www.trafiklab.se/api/gtfs-sverige
United States	75%		
Netherlands	70%		http://9292opendata.org/datacollecties
Israel	70%		http://he.mot.gov.il/index.php?option=com_content&view=article&id=2244&catid=167&Itemid=304

FIGURE 5 Open Knowledge Foundation's ratings of open transit timetables (*21*).

devices. In this magnitude, static data are all transit schedules/routes/stops, and real-time data are all estimated arrivals/vehicle positions/service alerts; and

- "Faucet," which is a precise subset of transit data and is suitable for mobile devices. In this magnitude, the data are specific. For example, static data could be "Stop ID 10 is served by Route 5," and real-time data could be "It is 2 minutes until Route 5 bus arrives at Stop ID 10."

The McKinsey Global Institute characterizes open data in terms of four characteristics (*26*):

- Accessibility: A wide range of users is permitted to access the data;
- Machine readability: The data can be processed automatically;
- Cost: Data can be accessed free or at negligible cost; and
- Rights: Limitations on the use, transformation, and distribution of data are minimal.

Figure 7 shows how data are open or closed based on these four characteristics. According to the McKinsey Institute,

> Open data sets also are defined in relation to other types of data, especially big data. 'Big data' refers to data sets that are voluminous, diverse, and timely. Open data are often big data, but 'small' data sets can also be open. We view open and big data as distinct concepts. 'Open' describes how liquid and transferable data are, and 'big' describes size and complexity of data sets. The degree to which big data is liquid indicates whether or not the data are open (*26*, p. 4).

The New Zealand government described a five-level model for open data.

> The World Wide Web Consortium (W3C) has developed a five star model to describe different characteristics of open data, and its usefulness for people wishing to reuse it. It is being used globally as a model for assessing data readiness for re-use. Applying this five star data model along with metadata standards will result in well understood and 'mashable' datasets (datasets easily joined together to create a new dataset). The three star level is considered the minimum standard for release of government's

Place	Score	Breakdown	Location (URL)	Information
Salt Lake City	60%			http://gis.utah.gov/data/sgid-transportation/transit/
San Francisco	70%			http://archives.sfmta.com/transitdata/google_transit.zip
Sacramento	60%			http://www.sacrt.com/gtfs.stm
New York City	70%			http://datamine.mta.info/
Louisville	100%			http://googletransit.ridetarc.org/
Asheville	100%			http://opendatacatalog.ashevillenc.gov/opendata/tag/5/?page=1&sort=name&dir=asc
Las Vegas	70%			http://rtcws.rtcsnv.com/g/google_transit.zip
Washington, DC	90%			http://developer.wmata.com/

FIGURE 6 U.S. City Open Data Census—State of Open Transit Data (*22*).

	Completely open	More liquid	Completely closed
Degree of access	Everyone has access		Access to data is to a subset of individuals or organizations
Machine readability	Available in formats that can be easily retrieved and processed by computers		Data in formats not easily retrieved and processed by computers
Cost	No cost to obtain		Offered only at a significant fee
Rights	Unlimited rights to reuse and redistribute data		Re-use, republishing, or distribution of data is forbidden

SOURCE: McKinsey Global Institute analysis

FIGURE 7 How data are open or closed (*26*, p. 3).

public data for re-use: non-proprietary, machine-readable, and accessible via the web, and licensed for reuse (*27*).

Those five levels are as:

1. Data are visible and licensed for reuse but require considerable effort to reuse: 1 star, on the web with an open license.
2. Data are visible, licensed, and easy to reuse but not necessarily by all: 2 star, machine-readable data.
3. Data are visible and easy to reuse by all (not restricted to using specific software): 3 star, nonproprietary formats.
4. Data are visible, easy to use, and described in a standard fashion: 4 star, Resource Description Framework (RDF) standards.
5. Data are visible, easy to use, and described in a standard fashion, and meaning is clarified by being linked to a common definition: 5 star, linked RDF.

In describing how Bay Area Rapid Transit (BART) creates value with open transportation data, Timothy Moore, web services manager, uses the graphic in Figure 8 to "demonstrate the flow of information in an open data ecosystem. Information flows in a continuous path clockwise from Customers to BART to Data to Developers" (and back to customers) (*28*). Rojas further defines these entities: the transit agency is the "discloser"; the developers are the "intermediaries"; and the customers are the "end users" (*11*, pp. 28, 32, and 33).

Hans Nobbe suggests the data can be provided at different levels of service, while still being open and free (*29*):

- Bronze: data is supplied with a limited bandwidth. There is no guarantee that data is supplied in time. The capacity of the system is maximized.

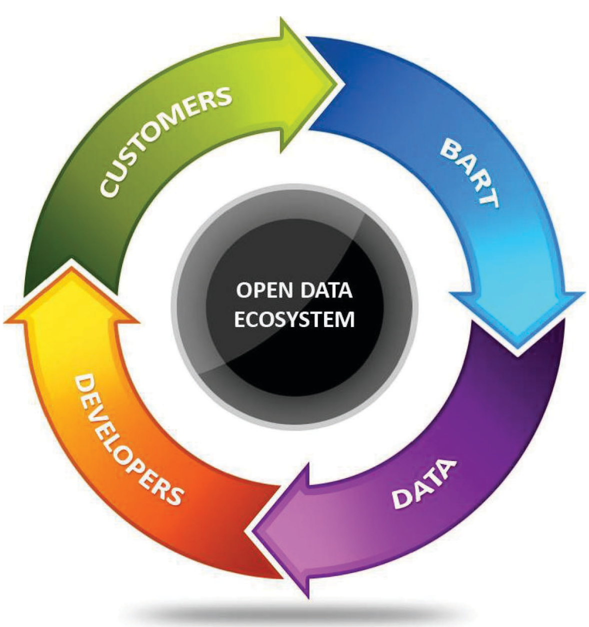


FIGURE 8 Open data ecosystem (*28*).

- Silver: data is supplied with a high bandwidth and guaranteed delivery. The extent of this level is determined unilaterally but in consultation with end users. For this service, a fixed fee will be charged.
- Gold: data is supplied by a mutual agreement. The fee is dependent on the contents of the mutual agreement. Which may be higher than in silver (for example, because 24 * 7 service requested) or lower cost (for instance as the end user guarantees transmission of safety related information).

A more multimodal perspective on open data types was discussed by Lee et al. (*30*). "Organizations of the government are recently supporting the common use of public information by preparing plans about information opening. Therefore, if public information is open and OPEN-API technology is introduced, then it would help promote the use of public information and improve the quality of information." The open data they included covered the following:

- Traffic flow information
- Traffic control information
- Incident information
- Closed-circuit television (CCTV) information
- Static and real-time transit information
- Bus arrival prediction information
- Bus transfer information
- Parking lot location operating information
- Parking lot guidance information

STANDARDS AND FORMATS USED FOR OPEN DATA

The use of standards in providing open transit data is critical and discussed in many pieces of literature. Kaufman identifies the basic standards and file formats for open transit data, as shown in Table 1 (*23*). Barbeau reports that successful open data formats are:

- *Organic:* Created and improved by the people actually producing and consuming the data;
- *Open:* Open process for evolution and data/documentation not hidden behind log-ins; and
- *Easy-to-use for app developers:* Is documentation simple to understand and are there existing open-source software tools? (*25*, p. 18)

Various sources define these and other standards used in open transit data are as follows:

- *GTFS* (https://developers.google.com/transit/gtfs/reference)—The General Transit Feed Specification, originally developed by Google, contains static schedule information for transit agencies, including stop locations, route geometries and stop times. "GTFS consists of a package of comma-delimited text files, each of which contains one aspect of the transit information and a set of rules on how to record it: six mandatory files (agency, stops, routes, trips, stops times, and calendar) and seven optional files (calendar dates, fare attributes, fare rules,

TABLE 1
COMMON DATA STANDARDS AND FILE FORMATS

	Champion	Where It's Used	Applicable Data Sets	Examples	More Information
			Data Standards		
GTFS	Google	Worldwide	Schedule data	Train line schedule	https://developers.google.com/transit/gtfs/
GTFS-realtime	Google	Select U.S. and European cities	Real-time data	"Train arriving in 3 min"	https://developers.google.com/transit/gtfs-realtime/
SIRI	European Committee for Standardization	European cities	Real-time data	"Train arriving in 3 min"	http://bustime.mta.info/wiki/Developers/SIRIIntro
TransXchange	U.K. Gov	U.K. Buses	Bus schedules and data	Bus route schedule	http://www.dft.gov.uk/transxchange/
DATEX 2	European Commission	European cities	Traffic data and management	Delays on Route 4	http://www.datex2.eu/content/datex-background
			File Formats		
CSV	Many	Worldwide	Data tables	Historic on-time data	http://www.ehow.com/how 5091077 use-csv-files.html
TXT	Many	Worldwide	Text	Textual information	http://en.wikipedia.org/wiki/Text_file
GIS	Many	Worldwide	Geographic mapping	Subway station entrances	http://en.wikipedia.org/wiki/GIS file formats
KML	Google	Worldwide	Google Maps and Earth	GIS road outlines	https://developers.google.com/kml/documentation/
XML	Many	Worldwide	Large data sets	Traffic numbers	http://www.w3schools.com/xml/xml what is.asp

Source: Kaufman (*23*, p. 4).

shapes, frequencies, transfers and feed info)" (*8*, p. 1). "The market success of GTFS has led to an unprecedented adoption rate by transit agencies as shown by total unlinked passenger trips for agencies with GTFS" (*32*, p. 1). For schedule data, GTFS adoption has substantially outpaced the Transit Communications Interface Protocols and Service Interface for Real Time standards in North America due to its relative ease of use for transit agencies to describe, implement and maintain data feeds (*34*, p. 3).

GTFS has evolved over the years to meet expanding requirements. The group that collaborates on these changes is the GTFS Changes Group (https://groups.google.com/forum/#!forum/gtfs-changes). The rules governing this group and how it is managed can be found at https://groups.google.com/forum/#!searchin/gtfs-changes/welcome/gtfs-changes/C5dgsKGkpDA/kyxN1DCS-dQJ.

- *GTFS-realtime* (https://developers.google.com/transit/gtfs-realtime/)—GTFS-realtime contains real-time information related to vehicle positions, service alerts, and trip updates (including delays and cancellations).
- *SIRI* (http://www.kizoom.com/standards/siri/)—The Service Interface for Real Time Information is a real-time data standard predominant in Europe and making significant inroads into the U.S. market, notably at the Metropolitan Transportation Authority (MTA) in New York. Recent change proposals to the SIRI standard include the definition of a structure for SIRI web services. The SIRI standard includes a component for schedule data (see later) but is designed for real time and thus is more complex than some other standards (*33*).
- *TCIP* (http://www.aptatcip.com/)—The Transit Communications Interface Protocols is an APTA standard with components that deal with passenger information and scheduling, as well as a host of other business divisions in transit. This standard was developed in the early stages of the transit information systems era; early studies encouraged its use for static and real-time data standardization. The standard's complexity results from its attempt to account for all the various operational procedures and service types offered by all transit agencies (*33*). "Some of the data elements needed for TCIP have since become part of the GTFS specification, including items such as agency name, block, route identifiers, trip identifiers and other similar information" (*33*, p. 3).
- *NextBus* (http://www.nextbus.com/xmlFeedDocs/NextBusXMLFeed.pdf)—A number of transit agencies use the NextBus XML API to deliver real-time arrival information.

Other formats reported in the literature that are being used for open transit data are as follows:

- Comma-separated values (CSV)—a file format used as a portable representation of a database. Each line is one entry or record, and the fields in a record are separated by commas (*34*). Agencies using GTFS have committed to producing and maintaining their schedule data in standardized CSV tables to display their system on

Google Transit's trip planner and, increasingly, opening these data to other third-party application developers (*32*, p. 1).

- Geo JavaScript Object Notation (GeoJSON) is a format for encoding a variety of geographic data structures. It is a geospatial data interchange format based on JavaScript Object Notation (JSON).
- Identification of Fixed Objects in Public Transport (IFOPT) defines a model and identification principles for the main fixed objects related to public access to public transport (e.g., stop points, stop areas, stations, connection links, entrances, etc.). IFOPT Standard builds on the TransModel Standard to define four related submodels (*35*).
- JavaScript Object Notation (JSON) is a data-interchange and text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages (http://json.org/).
- Network Exchange (NeTEx) is intended to be a general purpose format capable of exchanging timetables for rail, bus, coach, ferry, air, or any other mode of public transport. It includes full support for rail services, and can be used to exchange (*36*).
- Protocol Buffers—GTFS-realtime data exchange format based on Protocol Buffers. Protocol Buffers are a language- and platform-neutral mechanism for serializing structured data (think XML, but smaller, faster, and simpler). The data structure is defined in a GTFS-realtime.proto file, which then is used to generate source code to easily read and write your structured data from and to a variety of data streams, using a variety of languages (*37*).
- Resource Description Framework (RDF) is a standard model for data interchange on the web (*38*).
- Representational state transfer (REST) is a distributed system framework that uses web protocols and technologies (*39*).
- Really Simple Syndication or Rich Site Summary (RSS) is a format for delivering regularly changing web content (*40*).
- Simple Object Access Protocol (SOAP) is a method of transferring messages or small amounts of information over the Internet. SOAP messages are formatted in XML and are typically sent using HTTP (hypertext transfer protocol) (*41*).
- TransModel is the European Reference Data Model for Public Transport; it provides an abstract model of common public transport concepts and structures that can be used to build many different kinds of public transport information systems, including for timetabling, fares, operational management, real time data, and so forth (*42*).
- TransXChange (TxC) is the U.K. nationwide standard for exchanging bus schedules and related data (*43*). TxC provides a means to exchange bus routes and timetables between different computer systems, together with related operational data (*44*).
- Extensible Markup Language (XML) is more robust than GTFS in its abilities to represent large complex models, but the approach is more common in Europe and raises standardization challenges in the face of hyperflexibility (*33*).

Wong et al. said of GTFS:

> The GTFS, first introduced in 2005, is the result of a project between Google and TriMet in Portland to create a transit trip-planner using the Google Maps web application. Because of the collaborative approach to its development, the specification was designed specifically to be simple for agencies to create, easy for programmers to access and comprehensive enough to describe an intricate transit system. GTFS identifies a series of comma separated files which together describe the stops, trips, routes and fare information about an agency's service. Google opened the feed for general use in mid-2007 and it propagated widely as agencies translated their transit schedules into the format. The feed is the most used standard for static transit data exchange in the United States today (*33*, pp. 2–3).

As mentioned, according to data from the GTFS Data Exchange, as of March 11, 2014, data are available for 726 worldwide transit agencies (http://www.gtfs-data-exchange.com/agencies), and 239 agencies' feeds are available on http://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds, a guide to GTFS data that was written by Wong (*45*). A GTFS tool that can be used by small transit agencies was prepared by Williams and Sherrod (*46*).

Pioneering Open Data Standards: The GTFS Story describes TriMet's experience in helping develop GTFS and the effects that it has had on the transit industry (*47*, pp. 126–128). In summary, this story recounts the combined efforts of TriMet and Google in developing what would become a de facto standard for anyone to use to conduct transit trip planning anywhere in the world. The impact of GTFS was far reaching; 8 months after Google Transit was launched, five more transit agencies were added. "Within a year, Google Transit launched with fourteen more transit agencies in the United States and expanded internationally to Japan" (*47*, p. 128).

Those agencies that release their data through GTFS are required to provide the following data, at a minimum (*48*):

- Name or identification of the transit agency(ies) providing the data;
- Individual locations where vehicles pick up or drop off passengers;
- Routes, which are defined as groups of trips that are displayed to riders as single services;
- Trips for each route, which are sequences of two or more stops that occur at specific times;
- Times that a vehicle arrives at and departs from individual stops for each trip; and
- When service starts and ends, as well as days of the week when service is available.

Optional data types that can be provided through GTFS are as follows:

- Exceptions for when service starts and ends, and days of the week when service is available;
- Fare information for a transit organization's routes;
- Rules for applying fare information for a transit organization's routes;
- Rules for drawing lines on a map to represent a transit organization's routes;
- Headway (time between trips) for routes with variable frequency of service;
- Rules for making connections at transfer points between routes; and
- Additional information about the feed itself, including publisher, version, and expiration information.

Use of GTFS-realtime dictates that the following data, in addition to what is provided through GTFS (static information), are released (*49*):

- Real-time update on the progress of a vehicle along a trip (this is required information). This can specify a trip that proceeds along the schedule; a trip that proceeds along a route but has no fixed schedule; and a trip that has been added or removed with regard to schedule;
- Timing information for a single predicted event, either arrival or departure (this is optional). Timing consists of delay and/or estimated time, and uncertainty;
- Real-time update for arrival and/or departure events for a given stop on a trip (optional);
- Real-time positioning information for a given vehicle (optional);
- An alert, indicating some sort of incident in the public transit network (including cause and effect) (optional);
- Geographic position of a vehicle (required); and
- Identification information for the vehicle performing the trip (optional).

The SIRI standard (*50*) contains the following data elements (D.A. Laidig, Systems Engineering Manager, Metropolitan Transportation Authority, personal communication, June 18, 2014):

- SIRI-PT (Production Timetable): Exchanges planned timetables
- SIRI-ET (Estimated Timetable): Exchanges real-time updates to timetables
- SIRI-ST (Stop Timetable): Provides timetable information about stop departures and arrivals
- SIRI-SM (Stop Monitoring): Provides real-time information about stop departures and arrivals
- SIRI-VM (Vehicle Monitoring): Provides real-time information about vehicle movements
- SIRI-CT (Connection Timetable): Provides timetabled information about feeder and distributor arrivals and departures at a connection point
- SIRI-CM (Connection Monitoring): Provides real-time information about feeder and distributor arrivals and departures at a connection point. Can be used to support "connection protection"
- SIRI-GM (General Message): Exchanges general information messages between participants
- SIRI-FM (Facility Monitoring): Provides real-time information about facilities
- SIRI-SX (Situation Exchange): Provides real-time information about incidents.

MTA and the Utah Transit Authority (UTA) use open SIRI feeds. Of the list of SIRI messages, MTA Bus Time only supports VM, SM, and SX. UTA supports SM and VM (D.A. Laidig, Systems Engineering Manager, Metropolitan Transportation Authority, personal communication, June 18, 2014).

APTA conducted a survey of member agencies in 2013 to determine how transit agencies are providing static and real-time information to customers (*51*). The survey results show that GTFS was the most common format used by agencies, followed by proprietary formats from companies that provide scheduling software.

> Four out of ten agencies use a variety of other formats, and nearly two out of ten are using an internal agency format. Only two out of the 75 respondents said they did not use formats and standards. These tools and standards help agencies organize their routes and schedules internally, but they can also be used to create value for customers. The data organized by these standards can drive tools that customers can use to plan a trip, or they can create data streams that feed information to apps so customers can access this information on the go (*51*, p. 8).

Figure 9 shows the standards used by respondents to provide static information. The 25 respondent agencies that provided more than 25 million trips in fiscal year (FY) 2010 are classified as large agencies, and those providing fewer than 25 million trips are smaller agencies.

According to APTA,

> Looking at the split between large and smaller agencies, large agencies were more likely to use most of the listed formats, because those agencies were more likely to use multiple formats than the smaller agencies. A big majority—88%—of large agencies used multiple formats for static data. Only 58% of smaller agencies did so. Large agencies were much more likely to use tools from [technology companies] than smaller agencies. Smaller agencies were more likely to use a format in the 'other' category—these agencies used a variety of platforms provided by smaller software companies (*51*, p. 8).

The use of both GTFS and SIRI is exemplified in OneBusAway, an application whose "primary function is to share real-time public transit information with riders across a variety of interfaces" (*52*). Iley [in OneBusAway Application Suite (*53*)], describes OneBusAway as

> an open-source transit information software system, including several mobile apps, that was originally developed at the University of

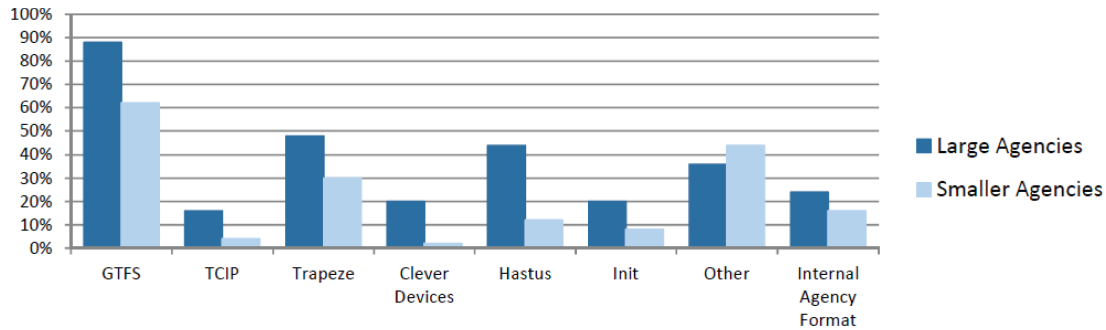


FIGURE 9 Results of APTA survey: Standards used to provide static data (*51*, p. 8).

> Washington and deployed in the Puget Sound area of Washington state. OneBusAway leverages the GTFS data format for schedule transit data, but the original Puget Sound deployment did not use a standardized interface for sharing real-time transit data with mobile apps. As part of their real-time Bus Time API pilot project in early 2011, Metropolitan Transportation Authority (MTA) in New York leveraged the OneBusAway software to build their own transit information system. MTA implemented a modified version of SIRI in their OneBusAway server to share their data with mobile app developers. In 2012, MTA [moved] on to the second step of deploying the same technology to other NYC boroughs.

In 2014, MTA completed its full five-borough rollout. Currently, OneBusAway provides real-time transit information in the Atlanta, New York City, Puget Sound, and Tampa regions.

Hazarika (*54*) describes the characteristics of two standards, SIRI and TransXChange (TxC), that are used in the United Kingdom to integrate various sources of public transit data. This is based on a survey conducted of "Local Authorities (LA's), Passenger Transport Executives (PTE's) and bus operators about their understanding, experiences and investment plans for these standards." This reference describes "the future growth/usage of the key features of TxC and SIRI against their customer's industry segments" (*54*, p. 4), and shows the key success factors (KSF), strengths, and weaknesses of SIRI and TxC (*54*, p. 33).

STANDARDS FOR OPEN PARATRANSIT DATA

One issue associated with using GTFS is how to portray paratransit services within this format because it was developed primarily for fixed-route transit service. Chambers, with Ride Connection in Portland, Oregon, addresses this issue in his presentation (*55*). He answers "yes" to his statement on page 7: "If a transit service can't be described in the GTFS, does it exist?"

Further, two other cases show how to use GTFS to describe paratransit or demand-response service. First, Klopp et al. (*56*) describe, in their Nairobi study,

> how paratransit networks can be mapped along with the collection of important transit data using off-the-shelf mobile phone technology. We also demonstrate the utility of the GTFS format when adapted for paratransit data collection. We found that GTFS allows for the incorporation of rich and useful metadata in a structured way. By using the GTFS data with Open Street Maps or Google maps, we created some of the first comprehensive visualizations of the Nairobi matatu paratransit system for the public and planners. By trying to fit aspects of the paratransit system into a GTFS format, however, it also became more clear where the fit is often hard to make because the standard was developed for planned formal transport system with fixed stops and schedules (even if they are not always strictly adhered to) and not the demand responsive and flexible paratransit system. Overall, it appears that modifications need to be made to GTFS to account for key differences between paratransit and more formal, planned systems.

Eros et al. (*8*, pp. 13 and 14) state that

> many cities across Latin America, Africa and Asia share this predicament; research indicates that flexible transport services constitute more than 90% of transit trips in cities such as Algiers. In Mexico a work-around was found by creating a variant to the GTFS feed based on defining fixed stops at regular intervals combined with the possibility for users to assess travel times and connections from any point between stations. Headway estimates, based on existing knowledge (including vehicle counts and speed data) substituted for schedules. Teams working in two cities, Manila and Dhaka, also had to deal with this challenge. Like Mexico City, Manila chose to avoid schedules, instead providing headway estimates for their jeepneys. In terms of stop locations, the Dhaka team included stop location based on where the bus stopped during the data collection ride. Manila's stops were interpolated every 500 meters along the route.

This discussion of standards for incorporating paratransit data directly relates to the use of GTFS for integrated trip planning. York Region Transit outside Toronto, Ontario, "is saving significant money having customers use fixed route for a portion of their paratransit trip" (information from Rajeev Roy, Manager, Transit Management Systems, The Regional Municipality of York-Transportation and Community Planning).

APPLICATION PROGRAMMING INTERFACES

Before leaving the standards discussion, it is important to draw the distinction between standards and an API. According to *Open Data Handbook Documentation*, an API is "A way computer programs talk to one another. [It c]an be understood in terms of how a programmer sends instructions between programs" (*57*).

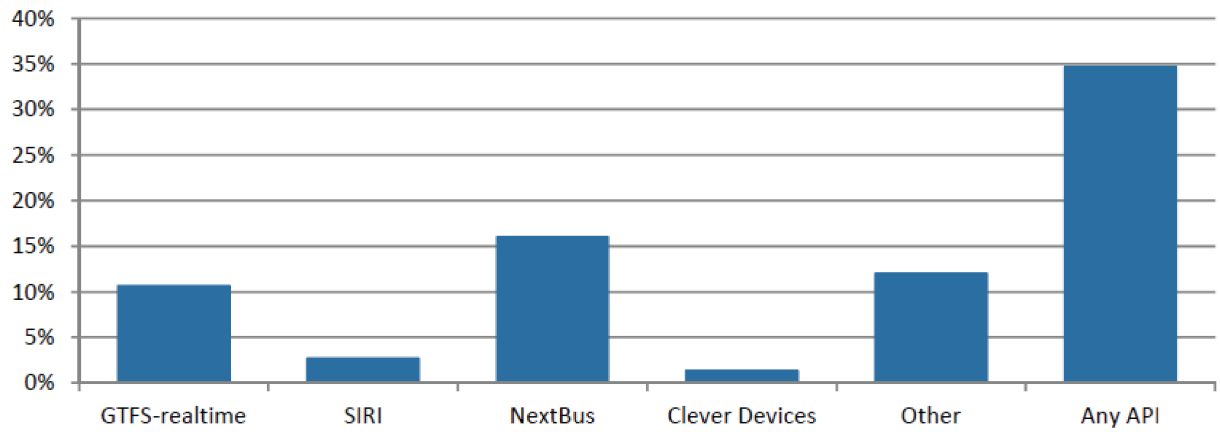


FIGURE 10 Results of APTA survey: APIs used to provide real-time information (*51*, p. 5).

As discussed, APTA's survey examined the use of APIs.

> Two-thirds of those agencies with real-time information provide an interface so app developers can utilize that information independent from the agency. Providing an interface for third-party applications allows developers to find innovative ways to provide this information to transit customers. Just more than one-third [37%] of all agencies provide an application programming interface (API) for third-party developers. APIs allow third-party applications to read and display data provided by transit agencies. The most common API used is NextBus, followed by GTFS-realtime (*51*, p. 1).

Real-time information is provided using the APIs, as shown in Figure 10.

LEGAL AND LICENSING ISSUES

In the Intelligent Transportation Systems (ITS) community, discussions regarding open data started 10 years ago at the 2004 ITS World Congress with "An Open Platform for Telematics" (*58*). Although the paper describes the technical aspects of telematics (defined as "the wireless provision of information and services to vehicles"), it raises issues about the functions and responsibilities among vehicle manufacturers, wireless carriers, information service providers, and call centers. These issues lead the reader to consider the legal and licensing ramifications of the development and implementation of the open platform.

When transit agencies began to consider providing their data openly, some became concerned about legal issues, including "misrepresentation of the agency, logo usage and brand identity" (*33*). In addition, agencies releasing data often are concerned about lawsuits and bad press (*59*). Other risks cited in the literature include "legal exposure due to the lack of accuracy of data, loss of advertising revenue on the agency homepage (if Internet traffic is directed to other sites, such as Google Transit, that provide transit services), and loss of control of dissemination of transit service information" (*60*, pp. 5–6).

For example, of the 50 reasons public agencies do not release data, there are several related to legal, privacy, and misuse issues (*61*), including the following.

- We can't legally do that.
- It will be misunderstood/misused.
- If we share our data/code, we'll be hacked.
- It might be presented in ways that result in people misunderstanding it. The media will misreport.
- People don't understand my data. It's complex/magical/for experts.
- The data source is a mess.
- The data might have errors or mistakes and could misinform the public.
- Privacy.

As shown in the discussion of the survey results, only one survey respondent (of 67) has experienced a legal issue arising from providing open transit data.

McCann addresses a few of these legal concerns (*61*). For example, the concern that "people might sue us" can be addressed by making "a policy that balances the public interest in accessing the data with the privacy concerns and stick[ing] to it." Another example is "we don't think it would be good PR to open this." This can be addressed by being "more proactive about good PR. Explain what the data means and why you are opening it."

Several open data organizations provide guidance regarding legal issues and licensing samples. For example, Open Data Commons presents sample legal tools and licenses in "Open Data Commons Public Domain Dedication and License (PDDL)" on http://opendatacommons.org/licenses/pddl/. The Open Knowledge Foundation provides numerous legal and licensing resources on http://opendefinition.org/licenses/.

In terms of open transit data, many agencies have developer license agreements and terms of use on their websites. Several examples are shown in Appendix E.

Timothy Moore, with BART, suggests that a simple license may be the most appropriate for open transit data (*62*). BART's license has the following characteristics:

- Short + sweet: 258 words
- We reserve the trademark
- Data provided 'as is' and 'as available'
- You don't have to sign anything (*62*).

Antrim and Barbeau (*60*), and The Finnish Transport Agency (*63*) report that open transit data agreements generally contain the following statements:

- The agency reserves the rights to its logo and all trademarks. These marks are an indicator used for official information from the agency only.
- The data are provided without warranties.
- No availability guarantees are expressed or implied.
- The agency retains full rights to the data (*60*).
- The license is free of charge and is made between the [agency] and a licensee.
- The licensee may freely copy and deliver; modify and use (e.g., for a commercial purposes); and combine and use as part of an application or service.
- It is not mandatory, but we highly recommend, that the name of the licensor (e.g., Finnish Transport Agency) is shown (*63*).

APPLICATIONS

The literature has numerous examples of open data applications. Examples of applications related to customer information that are driven by open data are rapidly evolving, and are available on agency websites or can be found on developer sites. Examples are provided in Appendix F.

An application that provides real-time information for MTA buses in New York City is called Bus Time. The application was piloted with route B63 on February 1, 2011, and was fully deployed citywide in 2014. Bus Time displays bus location and distance from stops, not time-based arrival predictions, on mobile platforms (*64, 65*). Bus Time provides open data between the MTA and software developers and customers, is based on open standards [between hardware components, between bus and server, and between server and other MTA/New York City Transit (NYCT) systems], and uses open source software (software code [OneBusAway] and APIs). Figure 11 shows how Bus Time works. Bus Time has had a positive impact on developers, with several apps developed: interactive phone application, arrival time predictions, and smartphone apps.

Samtrafiken, a nonprofit organization owned by 34 public transport operators and authorities in Sweden, has been a champion in open transit data. Their innovation manager, Elias Arnestrand, recognized that when developers started screen-scraping public transport data from various websites to create apps, agencies that "own" that data needed to consider releasing it. In addition, he realized that agencies could not keep up with the number of platforms that mobile devices were using. Further, apps that used screen-scraping technology could bring an agency's information systems to a halt because of the volume of data requests.

> [In] 2009 [we] created Trafiklab [http://www.trafiklab.se/]. It was formulated as an initiative to start to work with open data and open APIs. We wanted to make it simple to access this data

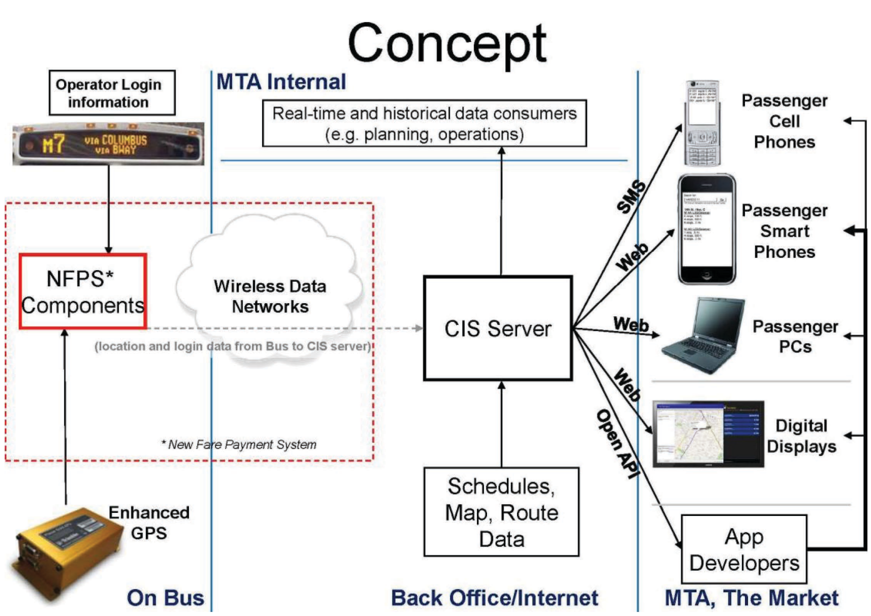


FIGURE 11 Bus Time concept (*65*, p. 4).

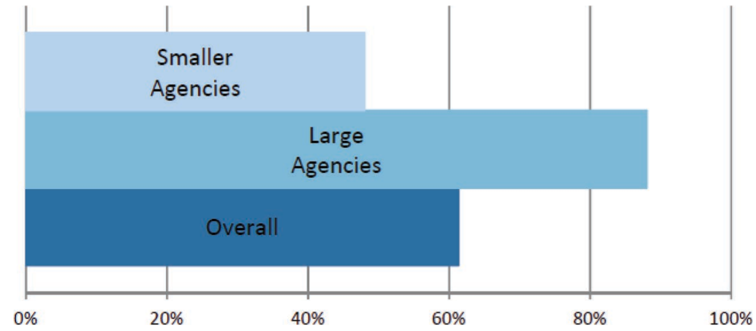


FIGURE 12 APTA survey results: Percent respondents that provide open static data (*51*, p. 10).

> and even make it fun for our industry and third party developers to discuss these issues. [T]his industry initiative [was] all together on one site instead of each public transport entity creating their own channel, data sources, set of agreements, and different types of APIs. This would have created a huge burden on third parties that wanted to access the complete set of public transportation data and services in Sweden. We looked at the external drivers that motivate developers [and recognized] that developers were driven by finding challenge, the satisfaction of getting their app to work, and the ability to showcase their work to the greater public. [W]e focused on [these drivers] to help get this initiative successfully off the ground.
>
> Today, our open APIs are a very important part of our strategy in providing customers with relevant public transport information and services. For example, for public transportation in Stockholm, more than 50% of the requests come from services created by third parties. APIs are a marketing and distribution channel for our public transportation information and services. We've realized that APIs are the cheapest and fastest way to build applications. And most importantly, APIs let third parties extend our products and services (*66*).

Another application that is based on open data is OpenTripPlanner (OTP), which was developed by OpenPlans. OTP "provides a robust multi-modal, multi-agency trip planner. This tool allows for multi-modal travel as one of the trip planning options for those looking to travel to transit via walking or bike" (*67*). This type of application greatly facilitates multimodal/multiagency coordination. It also uses crowdsourcing; OTP uses OpenStreetMap (OSM), a crowd-source open data set designed for routing. TriMet's customization of OTP "utilizes all open data including OpenStreetMap, GTFS, and the USGS National Elevation Dataset" (*68*, p. 10).

As mentioned, in 2012, APTA conducted a survey of its membership regarding customer information. A portion of this survey related to open transit data (*51*).

> Overall, a large percentage of agencies (80%) are providing static data such as schedules, routes, and fares to customers in some fashion. Around two-thirds participate in Google Transit, and a similar number make their data available to third-party developers. Just over six in ten agencies said that they make their static data available to third party apps. Large agencies were more likely to encourage third-party activities—88% of those agencies make their data available (*51*, pp. 10–11).

Figure 12 shows the percentage of APTA survey respondents that provide static data to third-party applications.

Overall, just more than 40% of APTA survey respondents had developers using their open static data; 68% of large agencies reported developers using their static data. "Around one-quarter of agencies surveyed indicated that app developers are using their real-time data. Forty-four percent of large agencies indicated that developers use their data and fourteen percent of smaller agencies indicated that this is the case [see Figure 13]" (*51*, p. 5).

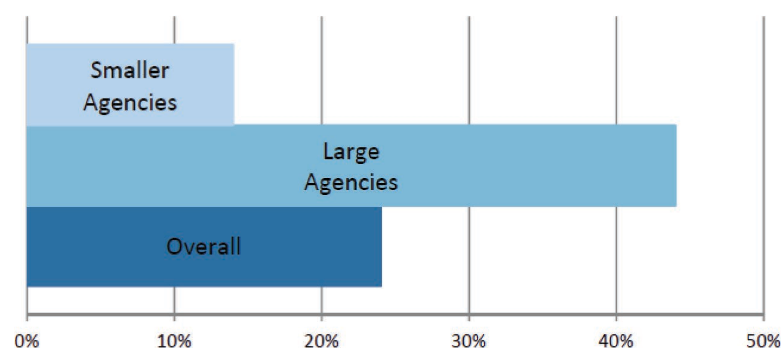


FIGURE 13 App developers using real-time data (*51*, p. 6).

Antrim and Barbeau describe some of the types of applications that use open transit data (*60*):

- *Trip planning and maps*—applications that assist a transit customer in planning a trip from one location to another using public transportation. Examples include Google Maps, OpenTripPlanner, Bing Maps, HopStop, MapQuest, and rome2rio.
- *Ridesharing*—applications that assist people in connecting with potential ridesharing matches. Examples include Parkio and Carma (formerly known as Avego).
- *Timetable creation*—create a printed list of the agency's schedule in a timetable format: TimeTable Publisher.
- *Mobile applications*—applications for mobile devices that provide transit information. Examples include Google Maps, Transit App for iOS 6 and beyond, Nokia Transport, RouteShout, and Tiramisu.
- *Data visualization*—applications that provide graphic visualizations of transit routes, stops, and schedule data. Examples include Walk Score, Apartment Search, and Mapnificent.
- *Accessibility*—applications that assist transit riders with disabilities in using public transportation. Examples include: Sendero Group BrailleNote GPS and Travel Assistant Device (TAD).
- *Planning analysis*—applications that assist transit professionals in assessing the current or planned transit network. Examples include:
  - OpenTripPlanner: Analyst Extension
  - Graphserver
  - Regional Public Transportation GIS Architecture and Data Model
  - Transit Boardings Estimation and Simulation Tool (TBEST)
  - TransCAD 6.0
  - GTFS-based Planning and Research
- *Interactive Voice Response* (*IVR*)—applications that provide transit information over the phone by means of an automated speech recognition system.
- *Real-time transit information*—applications that use GTFS data along with a real-time information source to provide estimated arrival information to transit riders. Examples include OneBusAway, NextBus, and TransLōc.

As mentioned in *TCRP Synthesis 104* (*69*), open data are being used to power electronic signs that display real-time information (see Figure 14 for an example). The digital signs originally developed by Mobility Labs have become commercially available. That synthesis reported,

> The widening world of open data availability is generating a deluge of mobile apps designed to deliver transit information to people on the move. But there is still an active—and growing—market for static displays, linked—among other factors in the United States—to the increasing importance of transit-accessible location in real estate markets. The system is for use by building owners—typically in lobbies—in any large metropolitan area providing open transit data. It draws together data streams from distinct agencies for presentation on large screens in multi-occupied residential and commercial properties.

In Europe, another vendor provides real-time transport displays (see Figure 15) in private homes and commercial properties (*70*). Another example of information displays driven by open transit data is shown in Figure 16 (*70*).

Wong described the use of GTFS to conduct several transit planning analyses (*18*). Table 2 "summarizes the fixed-route transit service measures from the [Transit Capacity and Quality of Service Manual] TCQSM and identifies those where GTFS feeds can be used as a data source. Two of the six measures can be calculated exclusively with GTFS feeds and three others can be calculated using GTFS feeds with supplemental data" (*18*, pp. 3–4).

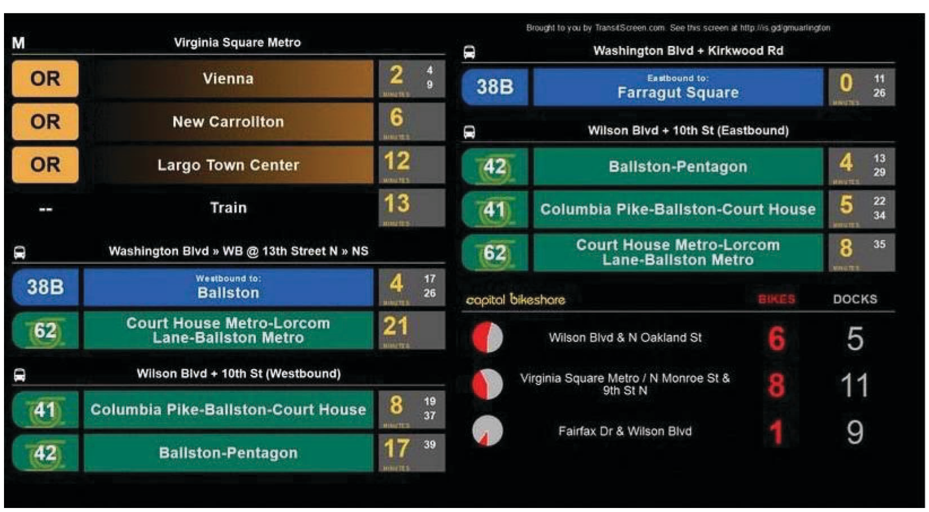


FIGURE 14 TransitScreen display in a commercial building in Washington, D.C. (*70*).

FIGURE 15 In-home portal (*70*).

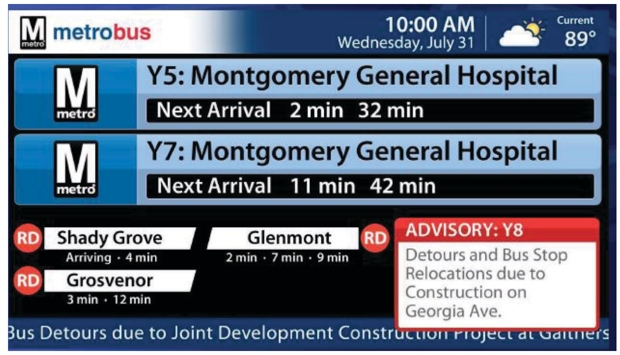


FIGURE 16 Sample digital sign in Silver Spring Metro Station (*71*, p. 6).

Figures 17 and 18 show the results of TCQSM analyses using GTFS data.

Antrim and Barbeau (*60*) provided other examples using GTFS applications, as follows:

- "The Delaware Valley Regional Planning Commission (DVRPC) used GTFS in developing their regional forecasting model;
- The Brookings Study of Transit and Jobs in America used GTFS data to determine how well transit connects people with their jobs; and
- The National Center for Transit Research identified opportunities to use GTFS data to support service planning and operational activity and developed a prototype application that integrated GTFS data with an automatic passenger counter (APC) for analysis and visualization."

Google Transit provides a real-time information application through Live Transit Updates for those agencies using GTFS-realtime and Google Maps (https://developers.google.com/transit/google-transit#LiveTransitUpdates). The cities covered by Google Maps can be found at https://www.google.com/landing/transit/cities/index.html.

The Transport Innovation Deployment for Europe (TIDE) project provides several examples of applications in Appendix G (*72*).

An application that calculates the average time it takes to travel between two stations of Capital Bikeshare is shown in Figure 19. This application examines trips between any two stations, comparing the average Capital Bikeshare rider's trip duration to estimated times for driving, transit, biking, and walking from Google Map's algorithms.

VISUALIZATIONS

The literature covers numerous visualizations using open data. The term *visualization* means "representing abstract business or scientific data as images that can aid in understanding the meaning of the data" (*73*). The literature also covers tools that assist in visualization, such as OpenTripPlanner (OTP) Analyst (*74*).

TABLE 2
DATA REQUIREMENTS IN TCQSM ANALYSES

Quality of Service Category	Resolution	Measure	GTFS Applicable	Additional Data Required
Availability	Transit stops	Average headway	Yes	None
Availability	Route segments/corridors	Hours of service	Yes	None
Availability	System	Percent transit-supportive areas covered	Yes	Employment, residential densities
Comfort/Convenience	Transit stops	Passenger load	No	Passenger counts
Comfort/Convenience	Route segments/corridors	On-time performance	Yes	Archived actual arrival times
Comfort/Convenience	System	Travel time difference	Yes	Traffic network

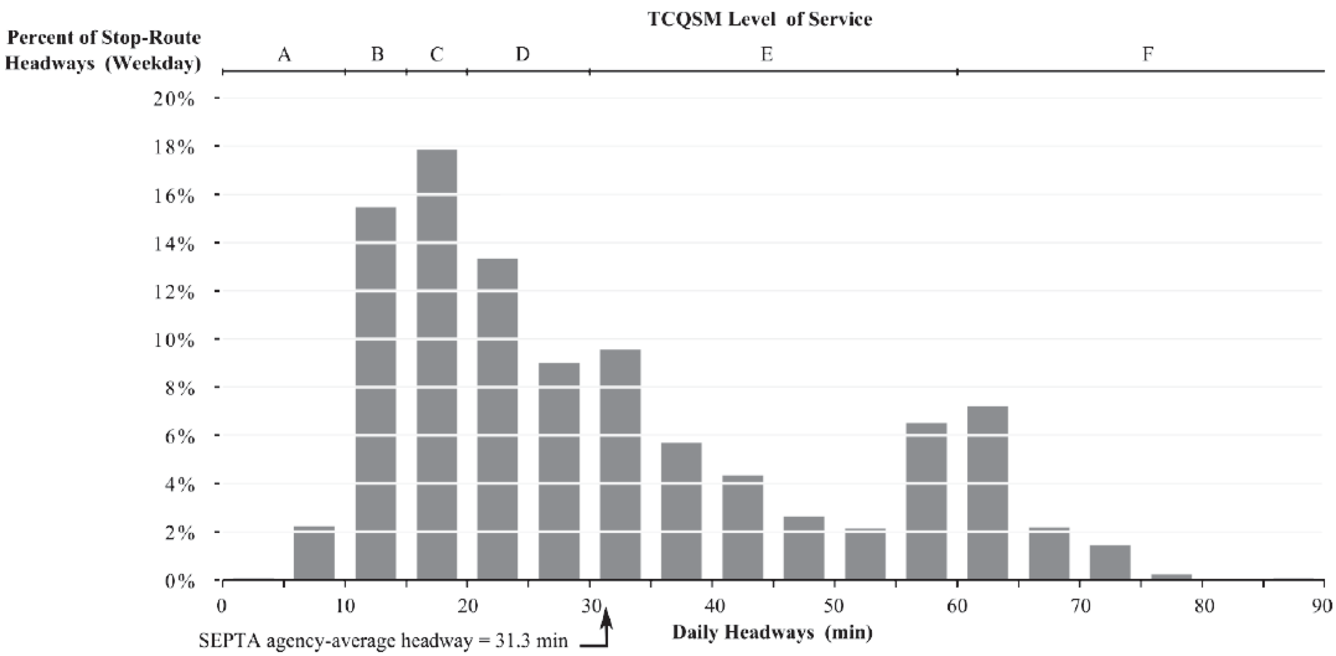


FIGURE 17 Distribution of stop-route level daily headways for the SEPTA bus system (*18*, p. 9).

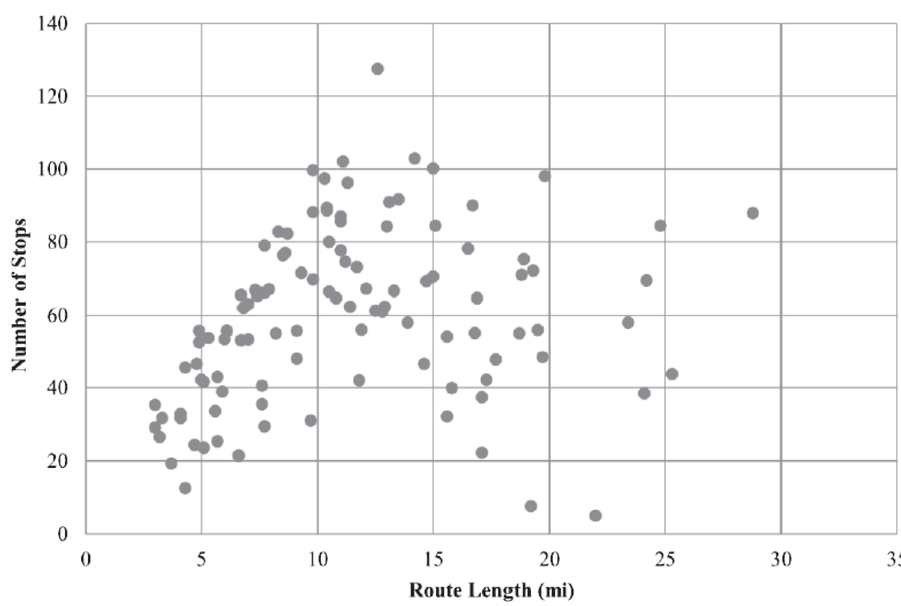


FIGURE 18 Length and number of stops for SEPTA bus routes (*18*, p. 9).

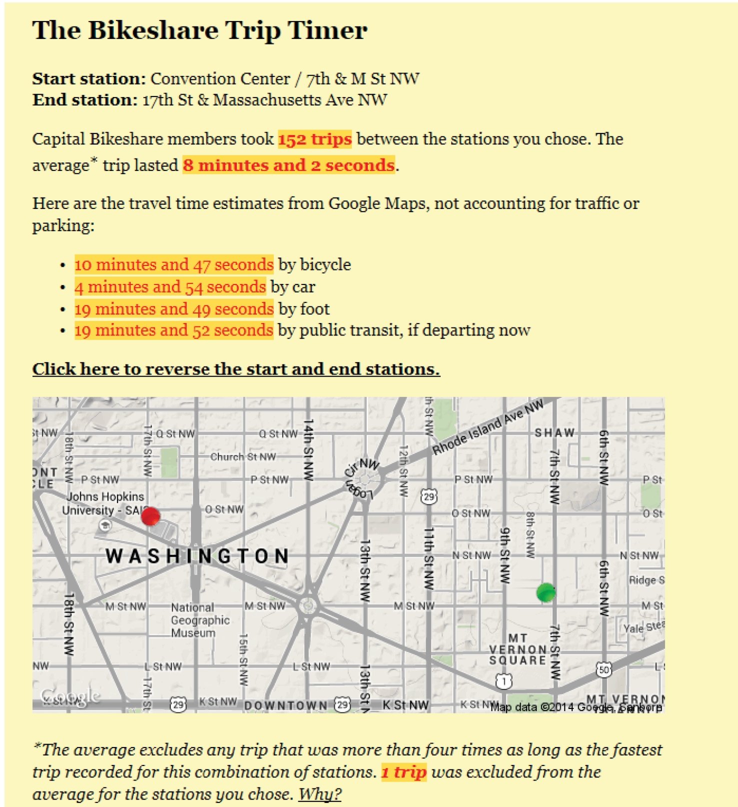


FIGURE 19 Bikeshare trip timer (http://rausnitz.com/bikesharetimer/results.php?stationstart=31223&stationend=31267).

> Mapping and data visualizations are effective tools for communicating the robust data and information within the GTFS by providing clarity to service levels which the data doesn't naturally produce. This can help articulate the impact of an agency's service changes and service cuts (*75*).

Open transit (and other types of) data allow many different types of visualizations that graphically display data to be easily interpreted. One example is showing the accessibility of Welsh schools by public transport using a program called Mapumental (*76*).

In 2013, there was an interest in showing the shortest time using public transit to get to any secondary school in Wales from any point in the country. Figure 20 shows this visualization.

> Time bands are in 15-minute increments, with red areas being those where schools are accessible within a 15-minute journey (the centres of the red dots therefore also represent the positions of the schools). Purple areas are those where journey time is between 1.75 and 2 hours, and the colours in between run in the order you see bottom right of the map. White areas (much of which are mountainous and sparsely populated) are outside the two-hour transit time (*77*).

There are several visualizations using open data from the Washington, D.C., area's Capital Bikeshare (*78*) shown in Figures 21 and 22.

Catalá et al. (*75*, pp. 32–41) provide several visualizations using GTFS data, including a Marey graph that shows the distribution of PATH trips and vehicles per hour at a particular stop (see Figure 23). The report also describes how GTFS data can be used to calculate service and performance evaluation metrics. Finally, the report describes opportunities to combine GTFS data with performance-related information (e.g., APCs).

There are many other articles and reports about using open data to conduct analysis and provide customer information (*79–94*), including an article about a visualization (see Figure 24) that shows access to jobs on public transit. "Specifically, it tells us how many jobs are accessible within 30 minutes—using the key at right—from each location by public transit, during the 7–9 a.m. peak morning window. The darker green areas have the greatest accessibility to jobs; the lighter green areas have the least. The red lines show transit routes" (*80*). Further, an initiative funded by Virginia Department of Rail and Public Transportation, is based on three factors: (1) open transport data standards such as GTFS and OpenStreetMap; (2) multimodal trip planning engines such as OpenTripPlanner; and (3) web-based visualization tools such as the D3 (Data-Driven Documents; e.g., https://github.com/mbostock/d3/wiki/Gallery) library (*82*).

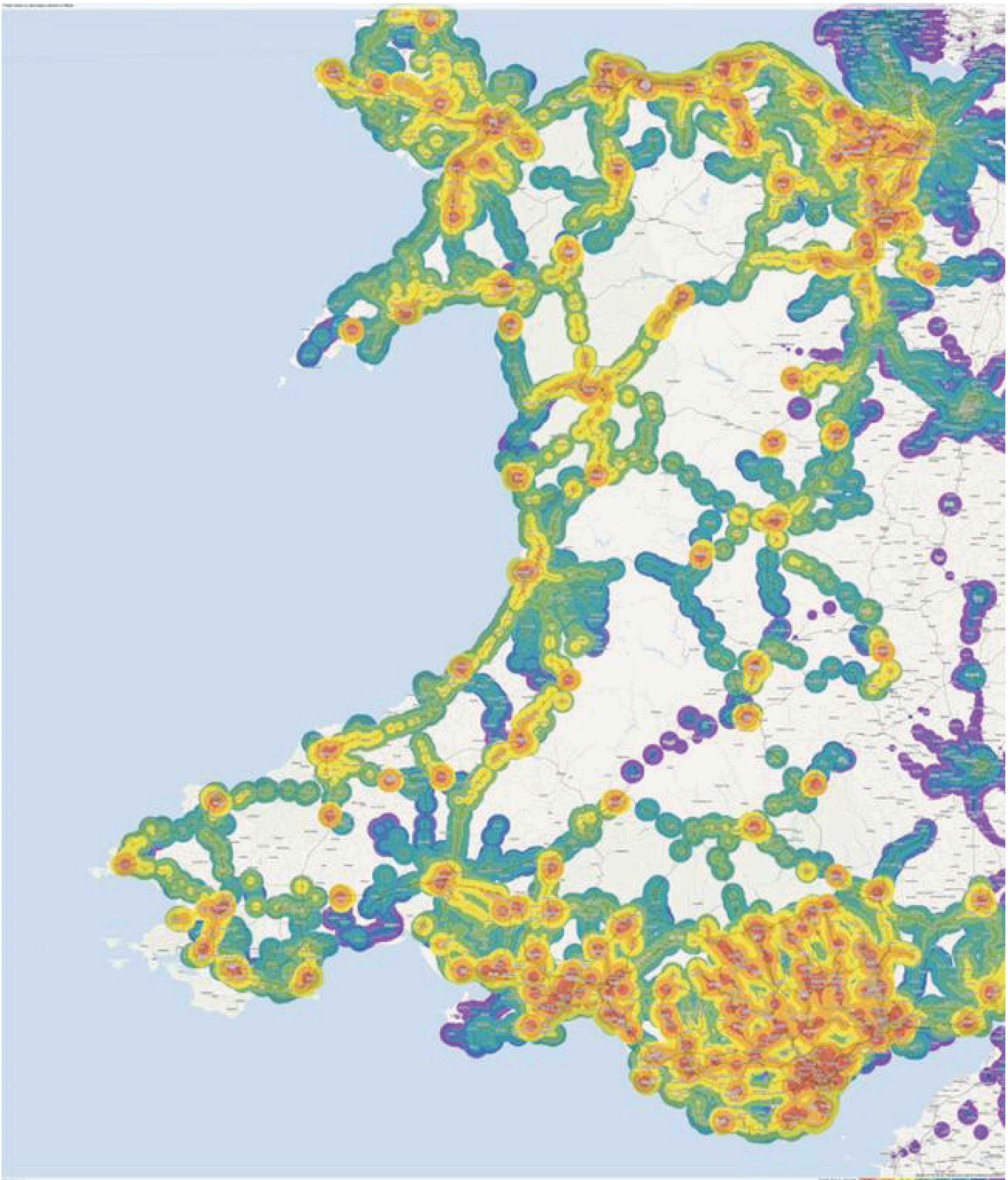
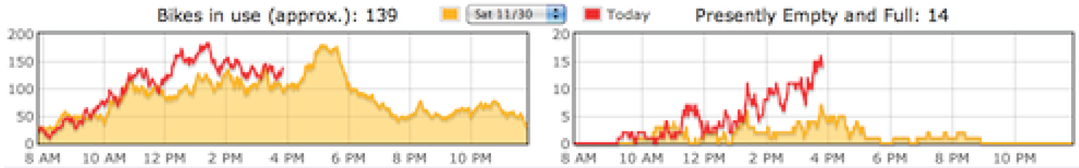


FIGURE 20 Transit times by public transport to secondary schools in Wales, with an arrival time of 9:00 a.m. (*77*).



Empty and Full Stations: 14				
Terminal	**Station Name**	**Status**	**Start**	**Duration**
31029	N Veitch & 20th St N	full	14:11:35	100
31614	11th & H St NE	empty	14:37:07	74
31017	Court House Metro / Wilson Blvd & N Uhle St	empty	15:17:48	34
31240	Ohio Dr & West Basin Dr SW / MLK & FDR Memorials	empty	15:19:50	32
31629	8th & East Capitol St NE	empty	15:20:51	31
31105	14th & Harvard St NW	empty	15:31:01	20
31602	Park Rd & Holmead Pl NW	empty	15:31:01	20
31216	McPherson Square / 14th & H St NW	full	15:32:02	19
31103	16th & Harvard St NW	empty	15:34:04	17

FIGURE 21 Capital Bikeshare DashBoard.

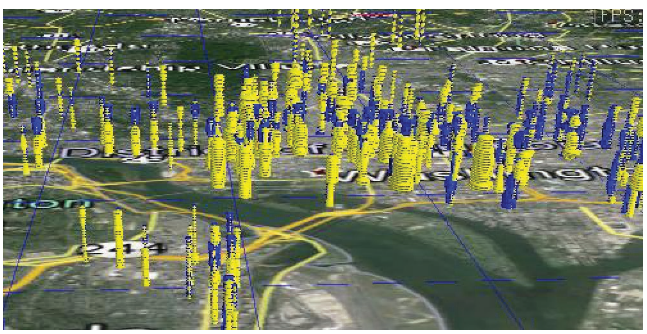


FIGURE 22 Three-dimensional visualization of trip history Capital Bikeshare data.

There are examples of nontransit visualizations using open data, such as the one shown in Figure 25, which displays the problem areas in the United States for flying, displaying airports with delays and flight cancellations.

COSTS AND BENEFITS

The costs associated with providing open transit data have been documented in a limited number of documents. Lee (*24*, p. 5) and Wong et al. (*33*) identified the costs as follows, based on the size and complexity of an agency's operations:

- Converting data to mainstream formats (e.g., GTFS), which may include an additional cost to purchase proprietary scheduling software (which has modules that

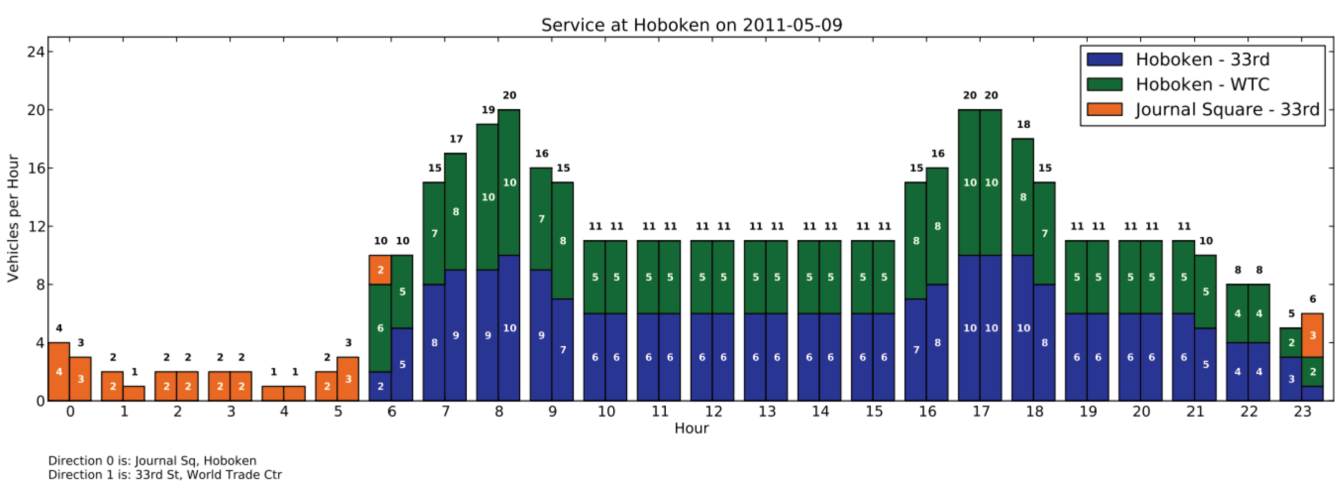


FIGURE 23 Port Authority Trans-Hudson (PATH) vehicles per hour at Hoboken.

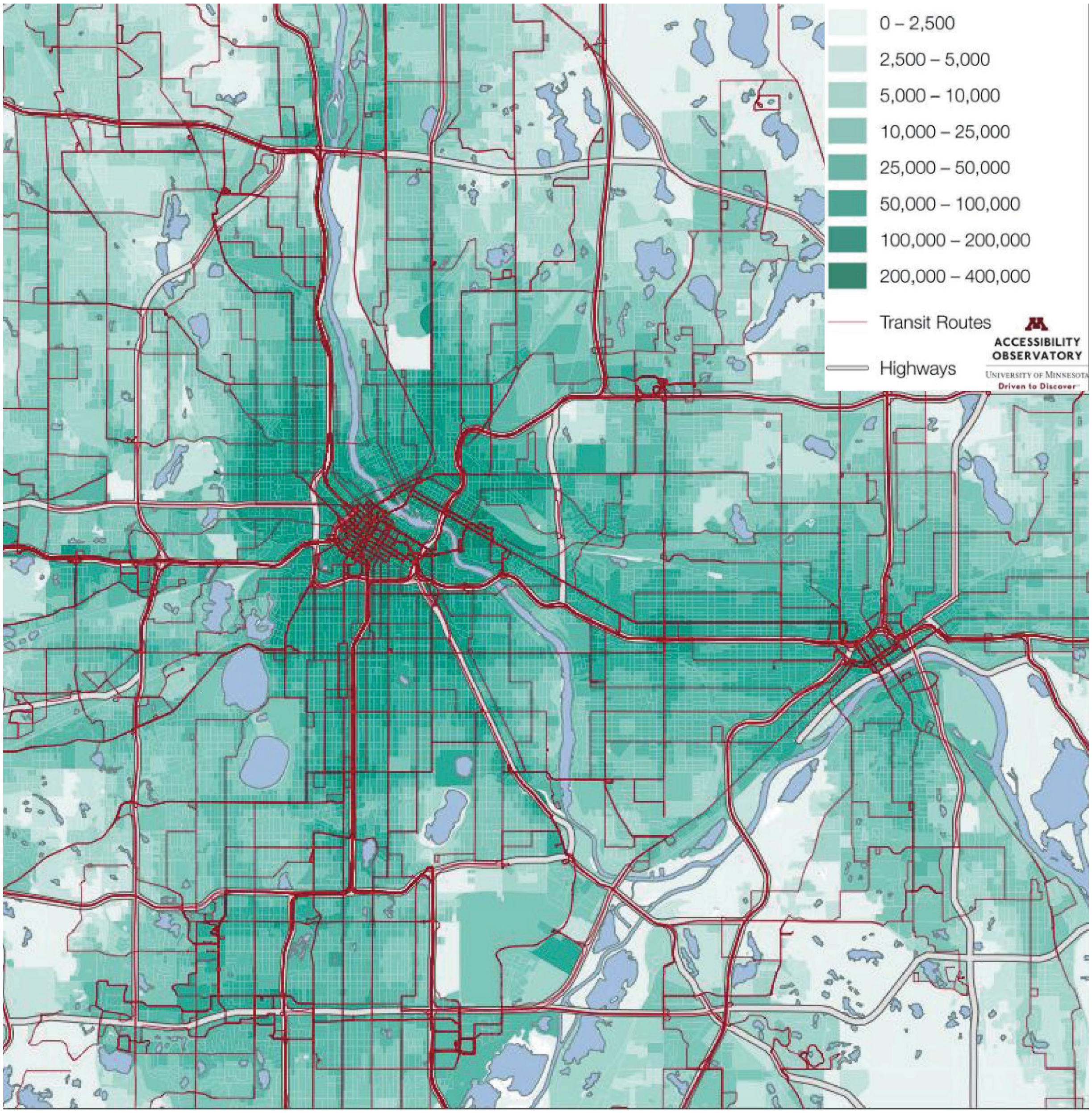


FIGURE 24 Job accessibility by transit (*80*).

automatically generate GTFS feeds) but allows for minimal human input which minimizes errors in data translation to GTFS;

- Web service for hosting data;
- Personnel time to update and maintain data as needed; and
- Personnel time to liaise with data users

One example of costs is provided in Wong et al. (*33*).

> Staff at BART discussed creating the original GTFS feed as an internal staff project, originally in less than one day. Since then, the agency reported spending less than $3,500 over the lifetime of a software product that was commissioned to specifically output GTFS from their existing scheduling database. This information is consistent with consultant estimates from the literature suggesting a cost for small agencies on the order of $3,000–5,000 based on simple networks with limited stops. Several free tools exist for agency use to generate and edit GTFS feeds including a project funded by the Transportation Research Board's IDEA program (*33*).

However, the benefits of open transit data are cited in several pieces of literature. Open transit data are often mentioned as having some of the most significant benefits because of the relationship of such data to riders. For example, Maltby (*95*) states that

> one of the most successful and prolific areas where open data has gone into mass public use has been through the multitude of transport information apps that allow citizens to better plan

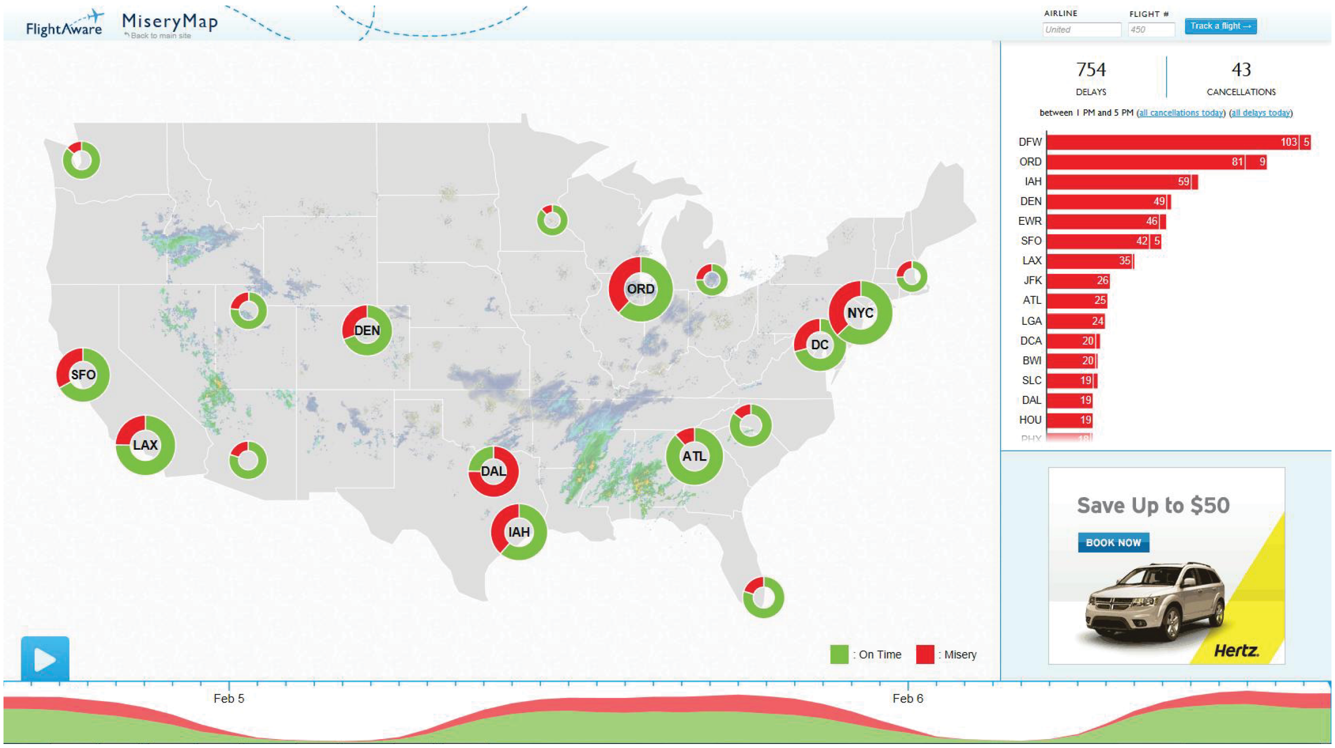


FIGURE 25 FlightAware's Misery Map (http://flightaware.com/miserymap/).

> their rail, tube or bus travel, find a parking space, or evade road works, or which through Google Now even help predict planning for journeys you are about to take. The Deloitte report for Stephan Shakespeare's independent review of Public Sector Information found there had been more than 4 million downloads of apps using transport data in London alone. Beyond the services delivered directly to the public, companies such as Placr are developing a thriving business aggregating transport data from a variety of sources and providing this as a service for app developers and organisations including Transport for London.

TCRP Legal Research Digest 37 (*96*), Traveline Information Ltd (TIL) (a partnership of transport operators and local authorities formed to provide impartial and comprehensive information about public transport in Scotland, England, and Wales) (*97*), and Lee (*23*, p. 14), identified the following benefits of open data:

- Perception that an agency is more "open" and transparent than other agencies that do not share their data;
- "Halo effect" of being involved in innovative third-party platforms and uses;
- Partnerships between agencies and their local developer communities;
- Higher quality (including accuracy) information for customers, resulting in improved customer service and experience, and potentially increased ridership;
- No customer confusion about the origin or location of "official" (agency) information;
- No additional customer services complaints;
- Third parties have developed applications that an agency:
  - may not have thought of;
  - did not know that customers wanted; and
  - could not afford to procure or develop;
- Time saved by agencies in developing customized applications;
- Crowdsourcing of data quality checking;
- Better understanding of the demand for data and customers' needs; and
- Better understanding of what services can be made available commercially, and those that cannot and need to be funded to ensure inclusion and accessibility.

The Polis Position Paper (*98*) reports that

> in addition to transparency, open data offers local authorities an opportunity to meet other local transport policies, notably to promote sustainable travel choices, by enabling them to:
>
> - Relook at their own business and *improve internal processes*, notably (i) seeking to understand which data a public body holds/gathers, (ii) thinking strategically about the value of data, (iii) improving the quality of data through feedback from the developer community.
> - Improve the *quality of service* to users by harnessing the creativity of the apps developer community to produce innovative services based on one or more data feeds.
> - Reduce the *cost of service provision*, by allowing the local authority to focus on data acquisition and management while the private sector takes on some of the burden of disseminating information to users.
> - Promote *economic development*, especially for local information services providers (see previous point).

TCRP Synthesis 91 discusses some of the benefits just being realized by open transit data (*1*, pp. 16–17).

> . . . Jay Walder, the New York Metropolitan Transportation Authority (MTA) Chief Executive Officer, stated that he hoped 'that the tools that might be developed using the agency's data would help transform the city's transit system into an even more useful resource for residents much faster and cheaper than it could do so itself.' Further, at the end of 2009, the Massachusetts Department of Transportation (MassDOT) launched the first phase of its open data initiative by releasing real-time information for five bus routes. The data released to software developers included real-time GPS locations of buses and arrival countdown information for every bus route. Within just one hour of releasing these data, a developer built an application showing real-time bus positions. Within two months, more than a dozen applications had been created including websites, smart phone applications, SMS text message services, and 617 phone numbers. All of these applications were created at no cost to MassDOT or the Massachusetts Bay Transportation Authority (MBTA).

One important set of benefits for the transit industry results from MTA's Bus Time project, as shown in Table 3. Transit agencies have traditionally procured proprietary solutions rather than open solutions, which is what Bus Time is based on.

McHugh (*47*, p. 130) notes that

> at TriMet, our process is automated, so there is very little overhead. TriMet has four major service changes a year, in addition to minor changes and adjustments in between. We may update and publish our GTFS data as frequently as twice a month. TriMet has not incurred any direct costs for this specific project, except resource time, which is a very small investment in comparison to the returns.
>
> Now that agencies have made GTFS freely available as open data, hundreds of applications have spawned worldwide. We found that by making our data easily and openly accessible, developers are getting very creative and expanding its use. This is not only beneficial because it expands the number of product offerings available, but it can also have emergent economic benefits for developers and the communities that they live in. In addition, because the standard allows for interoperability between cities, applications built to serve one city can be readily deployed to serve other cities for a much lower cost and effort than if the data wasn't standardized.

ENGAGING EXISTING AND POTENTIAL DATA USERS AND REUSERS

Lewis (*99*) describes a number of engagement strategies being used by transit agencies, as follows:

- MTA sponsors contests and hackathons, such as the MTA App Quest, which was their second challenge;
- SEPTA has sponsored three hackathons; and
- Madison Metro has developed "strong relationships with the local software community and universities."

> . . . Our advice would be to do your best to work with the developers ahead of time. Tell them your concerns and make sure to impress upon them the needs of your riders and the need for it to be accurate.
>
> After all, third-party developers aren't public transit employees or vendors, so system employees need excellent communication and a shared understanding of the goals and considerations in public transit if the project is to succeed. One of the students who developed a Madison Metro app continues to maintain it even though he's been hired by Google and moved to California, Rusch said. 'It's about fostering relationships with these people because they don't work for you directly and you're not purchasing a service from them,' he said.

As mentioned earlier in the report, APTA's survey regarding real-time information reported that app contests were held by 8% of the APTA survey respondents. Twenty percent of larger agencies used this engagement technique, as shown in Figure 26.

BART suggests that meeting with data users and reusers encourages discussing their needs and generating ideas. BART has used 10 engagement techniques (*62*):

- Meetups
- One-on-ones
- Google groups
- RSS feeds
- E-mail lists

TABLE 3
PROPRIETARY VERSUS OPEN SOLUTION

Traditional Single Vendor Solution	Our Method
Acquired knowledge largely resides in contractor	Knowledge is built within the MTA
Proprietary hardware	Open standards and off-the-shelf components allow easy replacement
	Allows concurrent use of hardware from multiple suppliers
Proprietary server	Open Source OneBusAway: anyone can download and use
Limited or proprietary inferfaces	Exposed data
$10k–$25k per bus (HW + on-bus SW)	$3,500 per bus

Source: "Bus Customer Information Systems" (*65*, p. 7).
Note: HW = hardware; SW = software.

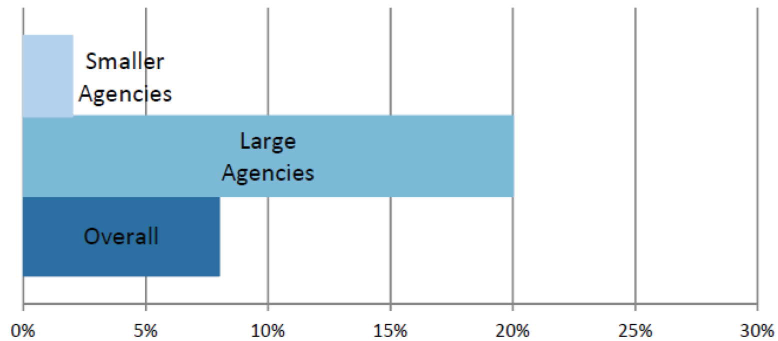


FIGURE 26 Number of agencies that held App contests (*51*, p. 6).

- Developer challenges
- Hack days
- Media events
- Transit camps
- "Find a politician willing to get in front of a camera!"

Barbeau (*100*) and Wong et al. (*33*) also discussed engagement approaches. First, the relationship with developers should be created using five steps: (1) open your GTFS data and share on GTFS Data Exchange; (2) share real-time data, too; (3) create a "developer page" with access to resources (e.g., GTFS license, data); (4) create a developer e-mail list/group for announcements, questions and answers, and collaboration; and (5) announce resources on "transit developers" group.

Wong et al. discussed various outreach strategies used in working with developers. "Agency interviewees reported that most developers seemed to have civic-minded motivations, aiming to provide travelers with better or more easily interpreted information. One staff member noted that agencies and developers were serving essentially the same people, and that an agency that supports developers with open data indirectly serves its own ridership and potential future customers" (*33*). Other examples of engaging developers included:

- Full description of GTFS and any other available data including limitations, service area and any non-standard data included;
- Participation in, preferably, agency-hosted e-mail groups or online forums to answer basic questions on or address issues in the data feed;
- Developer conference or "hackathon" where agencies describe major needs in traveler information and encourage developers to create software concepts that address those needs in a concentrated effort (sometimes held simultaneously with the release of agency data and/or in conjunction with other municipal datasets); and
- App 'showcase' where agency staff allows developers to present about their products to the staff to generate internal support for the projects (*33*).

There are a few disadvantages to using app contests to develop apps because conditions can change after the contest is held, and the winning apps may not be maintainable. For example, in 2012, the winning entry in an MTA app contest was later purchased by Apple. Thus, offering a reward to maintain an app for at least a certain period of time may be helpful.

IMPACTS OF OPEN DATA

An appropriate introduction to a discussion of the impacts of open transit data is to note a statement made in "Project Open Data: Open Data Policy—Managing Information as an Asset" (*101*): "When the U.S. Government released weather [in 1980] and [global positioning system] GPS data [in 1983] to the public, it fueled an industry that today is valued at tens of billions of dollars per year. Now, weather and mapping tools are ubiquitous and help everyday Americans navigate their lives."

Stephen Goldsmith discusses the fact that "amid nationwide public-sector budget cuts, open data are providing a road map for improving public transit and engaging an increasingly tech-savvy citizenry" (*102*). From a local perspective, Kurt Raschke sees open transit data as a way to meet the Washington Metropolitan Area Transit Authority's strategic goals set forth in the agency's latest strategic plan, called Momentum (http://www.wmata.com/Momentum/),

> . . . without the cost, delay, and dysfunction traditionally associated with government IT procurement. With open data, open source, and open architecture, we can deliver a cutting-edge product, while saving money and putting a better product in riders' hands faster. But to do this, we need the region's transit authorities to fully embrace open data and open system architectures. Breaking free from the cost, inflexibility, and data silos traditionally associated with transit passenger information systems means fully considering open-source products alongside their proprietary counterparts, and demanding that agencies work with and alongside—not against—open-source developer communities, both locally and across cyberspace. Leveraging open standards such as SIRI and GTFS-realtime, publishing high-quality open data, and collaborating with developers to enhance the quality and utility of data products all contribute to a better end product, and, ultimately, help advance the strategic goals which WMATA [Washington Metropolitan Area Transit Authority] has outlined in Momentum (*103*).

An example of open data creating a positive impact is the District of Columbia DC Circulator Dashboard (circulator-dashboard.dc.gov) (*104*), which was created to:

- Advance government transparency and accountability
- Facilitate data-driven decision-making
- Improve information availability to the community
- Engage the public on operations, routing, and safety
- Identify areas needing improvement
- Highlight success for replication.

The cost to develop the dashboard was minimal ($5,000 and two part-time staff over 6 months) and users need only a basic Internet browser and free plug-in to access it.

Other impacts are discussed by Eros et al., including the potential to

> . . . lower the barrier to innovation and enhance cross-fertilization of tools, approaches and ideas. In Mexico City, for example, almost immediately after the release of the GTFS feed, a number of apps made use of these data to provide value to users. The suite of apps is already growing; all have been created by American, Canadian, and Israeli developers as transfers of previously existing apps into the Mexico City environment. The nature of the GTFS format facilitates easy innovation transfers between different problems and contexts; as one city develops apps around a particular problem, others can benefit with relatively little additional investment. This is not limited to public-facing apps—it includes data collection as well as analysis and planning tools. However, more must be done, particularly in expanding the reach of this open-data culture to 'traditional' transport planning tools (*8*, p. 13).

Additional impacts wee noted by Suzanne Hoadley in her presentation at the 2013 Annual Polis Conference held in Brussels, Belgium, on December 3 and 4, 2013: "Opening Up Transport Data" (*105*). Polis is a network of European cities and regions working together to develop innovative technologies and policies for local transport. Since 1989, European local and regional authorities have been working together within Polis to promote sustainable mobility through the deployment of innovative transport solutions. See http://www.polisnetwork.eu/about/about-polis.) In some cities, the practice of opening up data has

> . . . helped create a relationship of trust with app developers; improved the quality of data itself; and genuinely harnessed the creativity of developer community (*105*, p. 3). . . . Further, the perceived benefits are improve[d] internal processes, including data inventory, data value and data quality; improve[d] quality of service by harness[ing the] creativity of developer community; reduce[d] cost of information service provision; and promot[ion of] local economic development (*105*, p. 4).

However, many challenges in opening data were mentioned and will be covered later.

The value of open data in the transportation area was described as follows by Barbeau (*25*).

> There are three major levers for unlocking value with open data in transportation:
>
> - *Improved infrastructure planning and management.* Open data on passenger flows and door-to-door travel times allows network operators (including municipal transit systems) to improve capacity and throughput. Open data already has been used to improve the design of transportation networks. For example, when Moscow's transit authority was modernizing its public transit system in 2012, it depended on open or shared data to determine where commuters lived and where they worked in the Russian capital. Officials used mobile phone location data along with government information on the ages, professions, and home neighborhoods of workers who commuted to specific business districts. Moscow then used this information to determine if greater investment was necessary in rail networks or if other services could do a better job of meeting demand. Based on the research, the city decided not to make a costly investment in a new rail line and instead met transportation needs by redrawing 100 bus routes. This limited Moscow's upfront investment costs and ensured that services could be flexible enough to meet the needs of a shifting population. In addition to avoiding more than $1 billion in infrastructure costs, the new bus routes reduced average morning commute times by three minutes per trip, saving ten hours of travel time for each rider every year. In New Jersey, NJ Transit released data on passenger flows to the public in 2012. Third parties quickly analyzed ridership at different times of day and were able to pinpoint underutilized rail stops, which led to more express trains and a saving of six minutes from the average commuting time during rush hour.
> - *Optimized fleet investment and management.* Real-time open data about vehicle location and condition and benchmarking of vehicle cost and maintenance information can help operators purchase, deploy, and maintain fleets more efficiently.
> - *Better-informed customer decision making.* Detailed open data about costs, reliability, environmental impact, and other factors can allow customers to make better decisions about which mode of travel to use and when (*25*, p. 32). . . .
>
> Based on our analysis, we estimate the global potential economic value that could be unlocked through these open data levers in transportation to be $720 billion to $920 billion per year. Optimized fleet operations (fuel savings, more effective maintenance, higher utilization) could enable as much as $370 billion a year in value. Improved infrastructure planning and management and improved consumer decision making can each lead to value of as much as $280 billion per year (*25*, p. 32).

The challenges associated with opening public transit data are covered extensively in the literature. In "Open Data Presents Opportunity, Challenge for Public Transit Systems" (*99*), many challenges were noted.

> The innovation potential of open source data brings with it associated challenges, not the least of which are cost and developing a sound process for releasing data and maintaining oversight of its use. Any public transit system interested in jumping on the bandwagon of open data—or expanding existing open data programs—could learn valuable lessons from the experiences of other agencies making the transition. 'You want to be in a position where you're giving information, you're supporting ideas, and you're encouraging creativity,' said Ron Hopkins, assistant general manager of operations for Philadelphia's Southeastern Pennsylvania Transportation Authority (SEPTA), which collects 1.5 million data points each month to measure on-time performance. 'The concern was, how do we manage all that?' (*99*).

In terms of challenges, James Wong, Landon Reed, Kari Watkins, and Regan Hammond discuss data integrity and maintenance as critical issues (*33*). In terms of data integrity, because open data can be used as input to traveler information tools, such as trip planners, "any inaccuracies in the GTFS cascade down to inaccuracies among [these] tools. Data maintenance relies on not only the agency maintaining a public file with up-to-date information, but also software developers who commit to update the information on their own projects when the data is updated" (*33*, pp. 6–7).

Antrim and Barbeau (*60*) discuss the challenges in terms of resource requirements as follows.

> Transit agencies must make the decision whether to format and maintain a GTFS dataset using their own personnel, or if they are going to outsource this task. It is important to consider that a new GTFS dataset will need to be produced every time there is a change to the schedule to keep the transit services based on GTFS data up-to-date. Major schedule changes can occur 3 to 4 times a year for large agencies, although, depending on the impact on the transit rider, the agency may want to update their GTFS data more frequently to reflect smaller changes in service on a weekly or monthly basis. Therefore, when identifying a GTFS creation process, the maintenance and sustainability of the process must be considered (*60*, p. 4).

Additional challenges are discussed in the Polis Position Paper (*98*), by the Finnish Transport Agency (*106*), by Marples et al. (*107*), by Traveline Information Limited (TIL) (*108*), in the McKinsey report (*25*), by Beasley of the Reading Borough Council (*12*), and by Watkins (*14*, pp. 27–30, 40–42):

- Opposition from information service providers, due in some cases to the fear of the threat of competition. Deployment of open data principals for data exchange between private sector firms and service providers could alleviate this challenge.
- Data control and ownership, for instance where data are owned by a cross-agency institution (e.g., passenger transport authority) or data are provided by a contractor. Some concerns cited by contractors include "competitor or commercially sensitive," "fear of use for measuring operational performance," and "extra burden on operations." Usage agreements could be helpful in expressing agency concerns.
- Organizational in the sense that there may not be a clear process/practical framework to guide transport authorities in opening their data. In addition, there may be a lack of understanding of the value of open data to improve performance and a lack of capability/expertise to implement an open data program. Also, having staff-level champions and strong leadership often leads to successful deployments.
- Architectural in that some systems are not designed for publishing open data. Further, they may have been developed ad hoc for a single operator, making it challenging to integrate the data among multiple agencies. Integrating data from multiple sources requires consideration of system capacity, load and response time, frequency of updates, nonstandard feed formats, different interpretations of standard protocols, nonstandard referencing, and data complexity. An example is where location references between one mode and another may differ. In addition, a gap in data standards may challenge release and use of open transit data.
- Data coverage, quality, privacy/confidentiality, accuracy, and timeliness: data may need significant amounts of "cleansing" or anonymizing before publication. One way to overcome this challenge is for agencies to take no responsibility for the data or information provided. However, having processes in place to report problems, see progress, and achieve fixes helps address this issue. Definitions for minimum quality of service requirements also would help.
- Unrealistic expectations or dependency from the public around the authority's capacity to provide consistent, convenient, and reliable data all the time (e.g., data latency following the detection of an incident). This challenge includes managing public reactions and expectations about changes in the transportation system that arise from the use of open data. Further, the data may not be what is desired by the public, which highlights the fact that the determination of which data to open should be based on the data users' needs. Sometimes "the private sector is better placed to provide the end user services and can help advise on what data [an agency] should focus on" (*12*).
- Cost of opening up data, which is pertinent in the current climate of public sector cuts and in view of the fact that most authorities do not have a dedicated budget for their open data activity. This cost does not just relate to building and providing the open data facility but also relates to the ongoing costs of maintaining open data (ensuring that authorities have the resources to update/refresh the data once it is published) as well as the support that must be provided to the developer community (*98*).
- Developer relationships need to be at different levels of engagement and promote support for mutual customers.
- Working with app developers should be sustainable and holistic, and include open communication lines.
- Performance measures are to be used to track success, including number of app downloads, number of apps developed, an app accessibility inventory, and market research surveys.
- Consider accessibility and equity.

CHAPTER THREE

SURVEY RESULTS: CHARACTERISTICS OF OPEN TRANSIT DATA

The synthesis survey covered several key characteristics of open transit data, including the justifications and reasons for choosing to provide or not provide open data; the underlying technology being used to generate the data; and the standards, protocols, and formats used in providing the data. Table 4 and Appendix B list the 67 responding agencies. Before examining these characteristics, the study team noted the overall annual ridership and modes operated by each respondent. U.S. responses represent agencies that carry a total of just more than 5.4 billion passengers annually (annual unlinked trips), with U.S. agencies' annual ridership ranging from 1.8 million (a county transit system in Florida) to 2.6 billion (MTA). The total annual ridership for each agency is shown in Appendix D.

JUSTIFICATIONS AND REASONS FOR PROVIDING OPEN DATA

The first question in the survey was "Has your agency provided open data?" Of the 69 (two agencies provided two responses each from different departments of the agency) responding agencies, 57 (82.6%) provide open data and 12 (17.4%) do not. Half of the agencies began providing open data in the 2010 to 2012 time frame. One agency started providing open data in 1981, two in the mid-1990s, 21 in the 2000 to 2009 time frame, and three in 2013.

Fifty-one agencies (almost 90% of those agencies that provide open data) provide it to increase information access to transit riders. The next most prevalent reason (49 responses, almost 86%) is to improve upon existing customer information and customer service or create new customer information services. The next most prevalent reason for providing open data is to foster a more positive perception of transit or encourage more people to try public transit. Table 5 shows all of the reasons agencies provide open data.

The most prevalent reason transit agencies are not providing open data is that it is too much effort to produce the data or the agency does not have the time or people to do the work required. The next most prevalent reason is that it takes too much effort to clean the data. All of the reasons are shown in Table 6.

The survey results show that almost all agencies providing open data see it as a way to maintain or increase ridership. Fifty-four or more than 96% of responding agencies that provide open data did not use any evaluation measures to assist them in deciding to open their data.

The major factor in agencies deciding which data to open is based on the ease of releasing the data. The next major decision is based on observing what other transit agencies have done regarding open data, and the third most frequent decision was done internally without asking any groups outside their agencies. All the survey responses are shown in Table 7 and Figure 27.

UNDERLYING TECHNOLOGY

In terms of the underlying technology that is generating the open data, the survey results indicate that scheduling software is the primary system being used. The next most used is GIS. The third most used is computer-aided dispatch (CAD)/automatic vehicle location (AVL). All survey responses about underlying technologies used to generate open transit data are shown in Figure 28. The "other" category includes the following responses:

- AVL/CAD vendor-supplied
- GTFS generated in Microsoft (MS) Excel
- Manually entered data
- Schedules. Scheduled information only
- Trapeze
- Trillium Transit
- We collect and host data from transit operators in the region
- All auto-generated from our enterprise relational database management system (RDBMS)
- Open source editing tool (https://github.com/conveyal/gtfs-editor)
- Ride checks
- Clever Devices (CD) BusTime Developer's API, database script to convert scheduling information in Hastus and CD Bustools to GTFS
- Ontologies (protégé) SPARQL.

TABLE 4
RESPONDING AGENCIES

Agency Name	City	State/Province/ Country
Alameda–Contra Costa Transit District (AC Transit)	Oakland	CA
Ann Arbor Area Transportation Authority (AAATA)	Ann Arbor	MI
Arlington Transit (ART)	Arlington	VA
AtB AS	Trondheim	Norway
Atlanta Regional Commission	Atlanta	GA
Bangor Area Comprehensive Transportation System	Brewer	ME
Bay Area Rapid Transit (BART)	Oakland	CA
Bilbao City Council	Bilbao	Bizkaia, Spain
Blacksburg Transit	Blacksburg	VA
Chittenden County Transportation Authority	Burlington	VT
Capital Metropolitan Transportation Authority	Austin	TX
Central Florida Regional Transportation Authority	Orlando	FL
Central New York Regional Transportation Authority	Syracuse	NY
Champaign–Urbana Mass Transit District (CUMTD)	Urbana	IL
Charlotte Area Transit System (CATS)	Charlotte	NC
Chicago Transit Authority (CTA)	Chicago	IL
Consorcio Regional de Transportes de Madrid	Madrid	Spain
Des Moines Area Regional Transit Authority (DART)	Des Moines	IA
Delaware Transit Corporation (DTC)	Wilmington	DE
Empresa Municipal de Transportes de Madrid, S.A.	Madrid	Spain
Fairfax County DOT/Fairfax Connector	Fairfax	VA
Grand River Transit (Region of Waterloo)	Kitchener	Ontario, Canada
Greater Bridgeport Transit	Bridgeport	CT
Greater Cleveland Regional Transit Authority (GCRTA)	Cleveland	OH
Helsinki Regional Transport Authority	Helsinki	Finland
Interurban Transit Partnership (ITP)	Grand Rapids	MI
Kansas City Area Transportation Authority (KCATA)	Kansas City	MO
King County Metro	Seattle	WA
Los Angeles County Metropolitan Transportation Authority	Los Angeles	CA
Manatee County Area Transit (MCAT)	Bradenton	FL
Massachusetts Bay Transportation Authority (MBTA)	Boston	MA
Metrolinx	Toronto	ON
Metropolitan Atlanta Rapid Transit Authority (MARTA)	Atlanta	GA
Metropolitan Transportation Authority (MTA)	New York	NY
Metropolitan Transportation Commission (MTC)	Oakland	CA
Monterey–Salinas Transit District	Monterey	CA
NJ Transit	Newark	NJ
Norwegian Public Roads Administration (NPRA)	Trondheim	Sør-Trøndelag, Norway
New Hampshire DOT	Concord	NH
North County Transit District (NCTD)	Oceanside	CA
Norwalk Transit District	Norwalk	CT
Orange County Transportation Authority (OCTA)	Orange	CA
Oregon DOT Rail + Public Transit Division	Salem	OR
Pennsylvania Public Transportation Association	Harrisburg	PA
Pace Suburban Bus	Arlington Heights	IL

(*continued on next page*)

TABLE 4
(*continued*)

Agency Name	City	State/Province/ Country
Pinellas Suncoast Transit Authority	St. Petersburg	FL
Roaring Fork Transportation Authority	Aspen	CO
Regional Transportation Commission of Washoe County	Reno	NV
Regional Transportation District (RTD)	Denver	CO
Stark Area Regional Transit Authority	Canton	OH
Suburban Mobility Authority for Regional Transportation	Detroit	MI
Samenwerkingsverband Regio Eindhoven (SRE)	Eindhoven	Brabant, Netherlands
Syndicat des transports d'Île-de-France (STIF)	Paris	France
San Mateo County Transit District	San Carlos	CA
Société de transport de Laval	Laval	Quebec, Canada
Sound Transit	Seattle	WA
Tampere City Public Transport	Tampere	Pirkanmaa, Finland
Transport for London (TfL)	London	United Kingdom
Transport for Greater Manchester	Manchester	United Kingdom
Tri-County Metropolitan Transportation District of Oregon (TriMet)	Portland	OR
Organization of Urban Transportation of Thessaloniki	Thessaloniki	Greece
Urban Transport Administration	Gothenburg	Vastra Gotaland, Sweden
Utah Transit Authority	Salt Lake City	UT
Votran	South Daytona	FL
Wiener Linien	Vienna	Austria
Worcester Regional Transit Authority (WRTA)	Worcester	MA
York Region Transit	Richmond Hill	Ontario, Canada

TABLE 5
REASONS FOR PROVIDING OPEN DATA

Reason for Providing Open Data	Number of Respondents	Percent of Agencies Providing Open Data
Increase information access to transit riders	51	89.5
Improve upon existing customer information and customer service or create new customer information services	49	86.0
Foster a more positive perception of transit/encourage more people to try public transit	44	77.2
Foster/encourage innovation around the agency's data or help third parties develop skills and services (e.g., with which the agency can contract)	36	63.2
Facilitate information sharing within the agency and with partners and customers	34	59.6
Agency transparency	33	57.9
Availability of data standard(s) for transit information (e.g., GTFS)	33	57.9
Improve effectiveness of the agency and its services	32	56.1
Increase customization for customer information	31	54.4
There was demand for us to open our data/we were requested to provide open data	29	50.9

TABLE 5
(*continued*)

Reason for Providing Open Data	Number of Respondents	Percent of Agencies Providing Open Data
Help achieve other agency goals (e.g., by providing a wider audience for published information)	27	47.4
Provide ways to better understand and use transit information within our agency	26	45.6
Participate in the latest trend in the transit industry	26	45.6
Improve or provide new private products and services	25	43.9
An information gap existed that could be bridged by better public data	20	35.1
Cut costs to our agency	12	21.1
Provide incentives for others to help maintain data sets, reducing the maintenance cost for the agency	12	21.1
Other (only one respondent specified that their "other" response meant "part of agency culture, esp. information technology."	6	10.5
Measure the impact of transit on the community(ies) that are served	6	10.5

From survey responses.

TABLE 6
REASONS FOR NOT PROVIDING OPEN DATA

Reasons for Not Providing Open Data	Number of Respondents	Percent of Agencies Not Providing Open Data
Too much effort to produce the data/we do not have the time or people to do the work required	5	38.5
Too much effort to clean the data	4	30.8
We cannot control what someone will do with our data	3	23.1
We do not know the accuracy of our data	3	23.1
Our data could be misused or misinterpreted	2	15.4
It will put a strain on our systems	2	15.4
Proprietary vendor contracts preclude us from sharing data with third parties	2	15.4
Our agency is too small	2	15.4
Our agency will be liable if erroneous data are provided to the public	2	15.4
Our agency does not know how to open our data	2	15.4
There is a lack of interest internally	1	7.7
There is a risk-averse culture within our agency	1	7.7

Other (please specify)

- We are in the planning phases of opening our data, but the above have been our roadblocks.
- We plan on it, need to verify accuracy.
- Working with vendor to make data available through GTFS–realtime interface.
- Our software doesn't have a secure location for customers to access the data. The regional transit authority uses another software module to provide real-time information to the public.
- We are in process of having an API created by a third party since we are restricted by our vendor from dealing with an agency that has experience with creating an API for Transit.
- Due to limited staff resources, information has not been shared. This will change with AVL/GPS project in process.
- Governor's office has requested datasets for a statewide effort. We will provide, but it is taking time due to staff constraints.
- Have been requested to provide select open data sets by the Governor's Office. We will do so but it is taking a long time due to staff constraints.

From survey responses.

TABLE 7
FACTORS IN DECIDING WHICH DATA TO OPEN

Decision Factor	Number of Respondents	Percent
Based on the ease of releasing the data	33	58.9
Based on observing what other transit agencies have done regarding open data	21	37.5
Decided internally without asking any groups outside our agency	17	30.4
Asked potential users of the data	11	19.6
Based on the cost associated with producing or cleaning the data	11	19.6
Asked the community in which your agency operates service	8	14.3
Asked riders	1	1.8

Other:

- Approached by Google.
- Approached by transit enthusiast.
- Based upon what our AVL/CAD vendor provided.
- I don't know.
- Open Government Data (OGD) Vienna.
- Requests to access data.
- Some in the developer community encouraged us to release items.
- User demands.
- Based on demand.
- Defaults to GTFS and availability of Clever Devices Bustime API.
- Worked with developers.
- We were already using web services for internal purposes, we merely exposed it with documentation for the third party developers; a no brainer.
- Supported University of Washington graduate study project to provide scheduled data to the public via third-party application (One Bus Away).
- Based on requests from third party service providers.
- Asked experts in the University field.
- Decided both internally, and from developer community.

From survey responses.

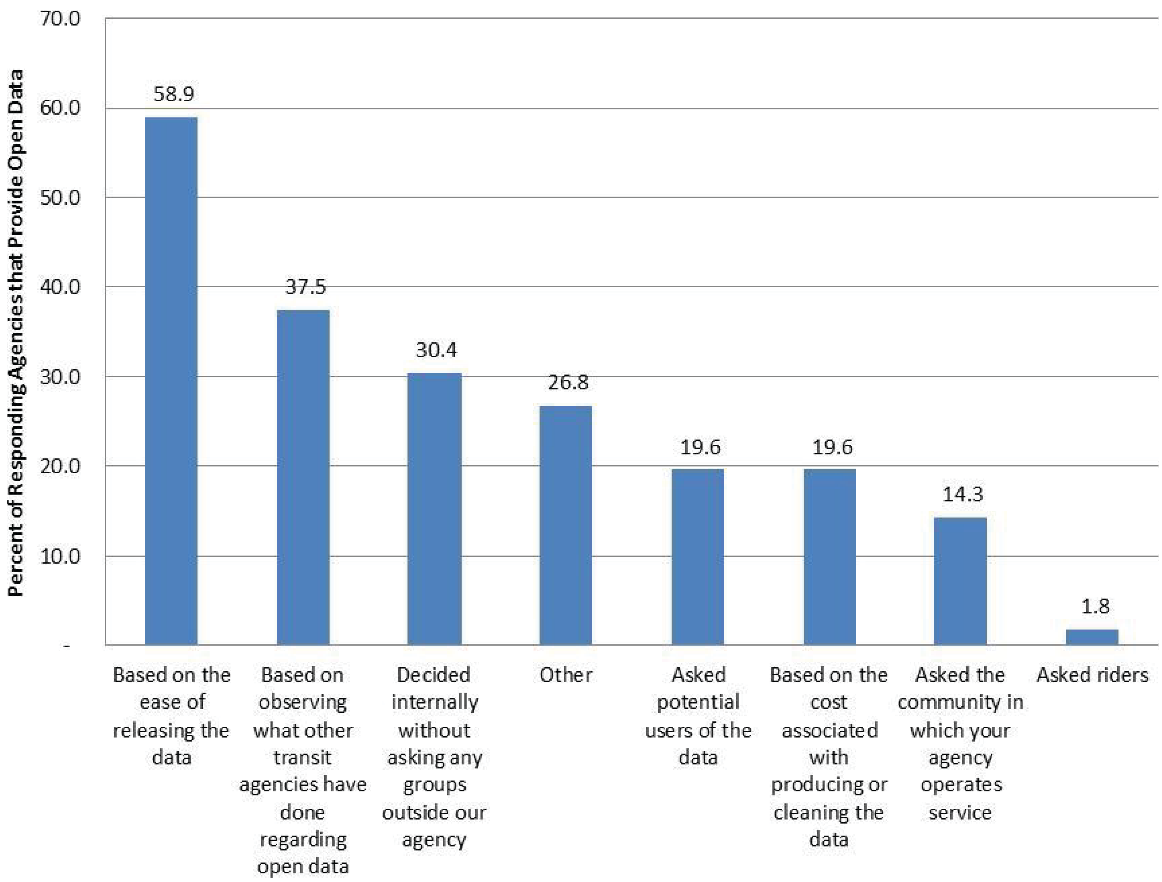


FIGURE 27 Factors in deciding which data to open (from survey responses).

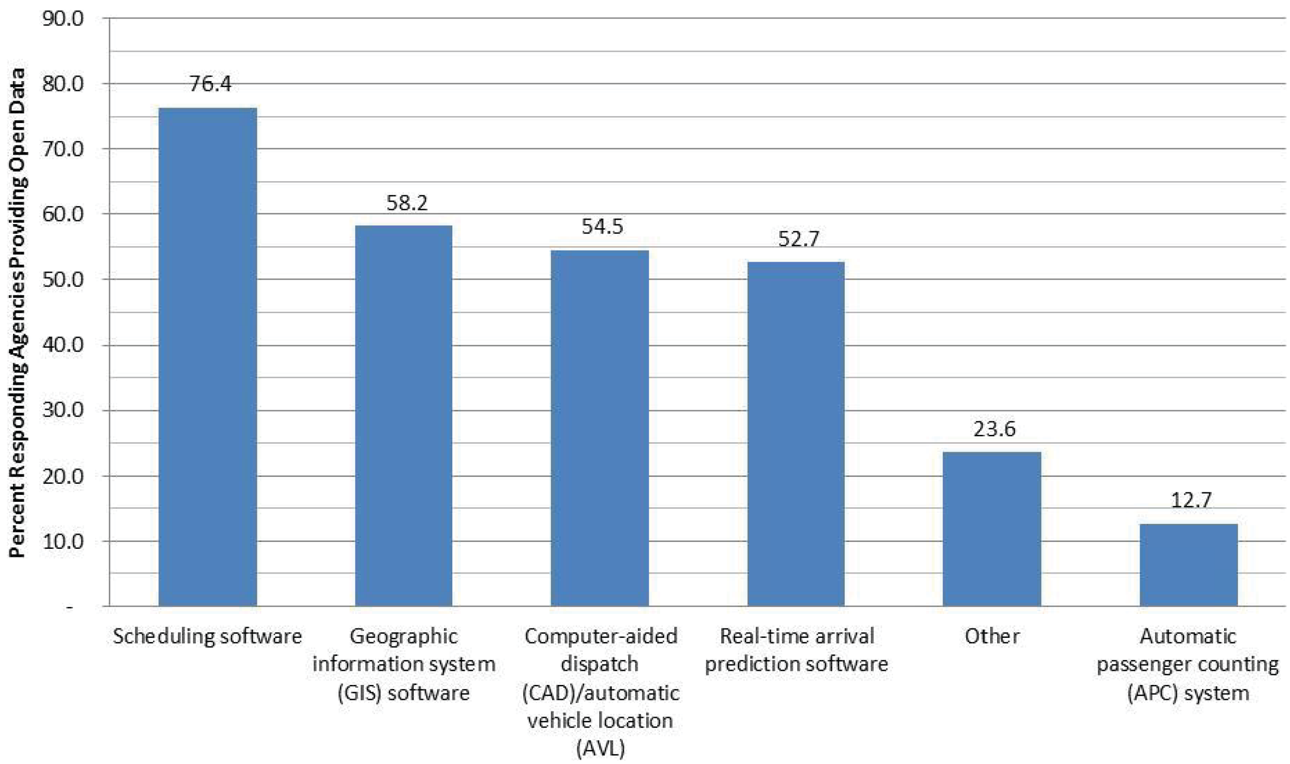


FIGURE 28 Underlying technologies that generate open transit data (from survey responses).

TYPES AND STANDARDS/FORMATS OF OPEN DATA

The types of open transit data provided by survey respondents are shown in Table 8.

The frequency with which the open data are updated or modified is shown in Figure 29. The following responses are in the "other" category:

- Three times a year
- As needed
- At least every quarter
- According to the Board period
- Major service changes (three times a year) plus periodic, significant bi-weekly adjustments
- On demand
- On request
- Real-time data are "real time"; route updates are every 2 weeks
- Schedules are updated when changes occur
- On average bimonthly
- Every time there is a change in the data based on its significance. For example, the tiniest changes might not be updated every time they are noticed.

TABLE 8
TYPES OF OPEN TRANSIT DATA

Types of Information	Number of Respondents	Percent
Route data	51	89.5
Schedule data	50	87.7
Station/stop locations	49	86.0
Real-time information	33	57.9
Park-and-ride locations	17	29.8
Fare media sales locations	14	24.6
Ridership data	14	24.6
Other	12	21.1
Budgetary data	10	17.5
Performance data	8	14.0

From survey responses.

The degree to which the data are open was examined from four different perspectives, as shown in Figure 7. The survey results for each of these four characteristics are shown in Figures 30 through 33.

GTFS is the format most commonly used to provide open transit data. The survey results indicate that a number of other standards and formats are being used, as shown in Table 9.

The agency's website is the outlet used most frequently through which open transit data are provided. The GTFS Exchange website is the next most commonly used, followed by APIs. All survey responses to this question are shown in Table 10.

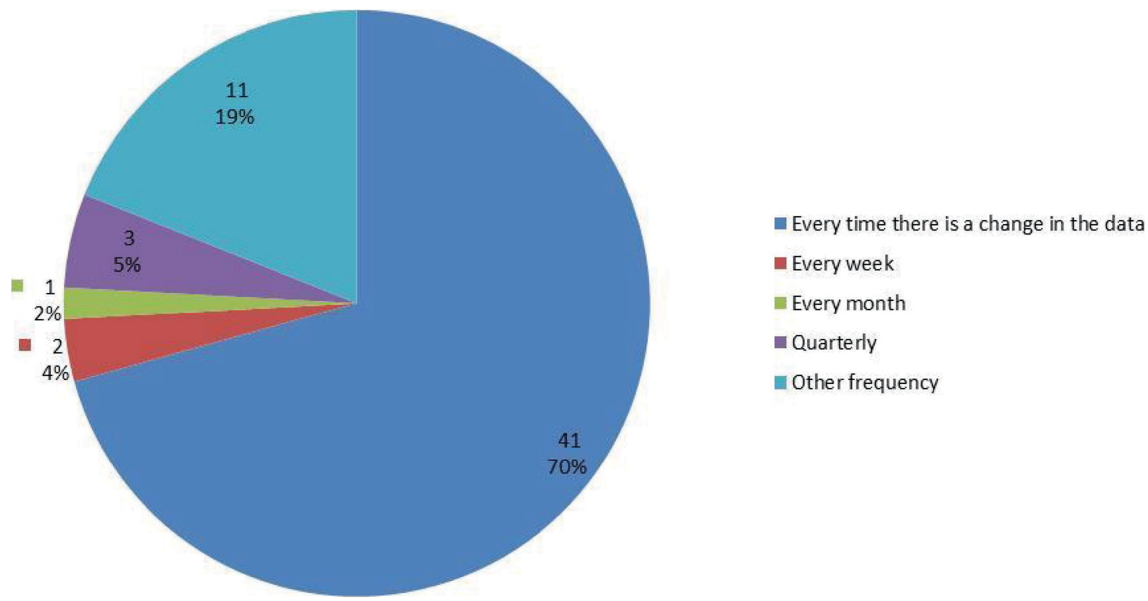


FIGURE 29 Frequency with which open data are updated or modified (from survey responses).

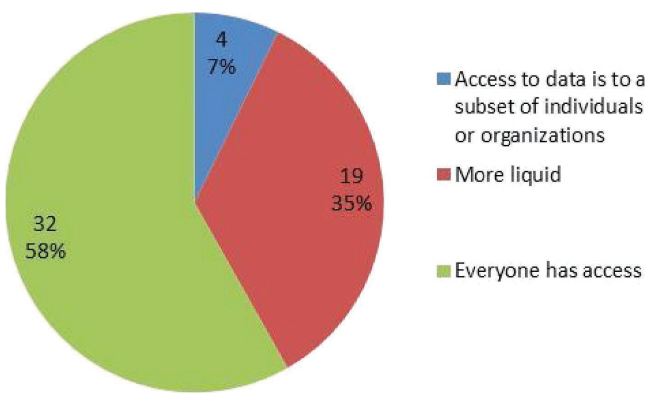


FIGURE 30 Degree of access of open data for percentage of survey respondents (from survey responses).

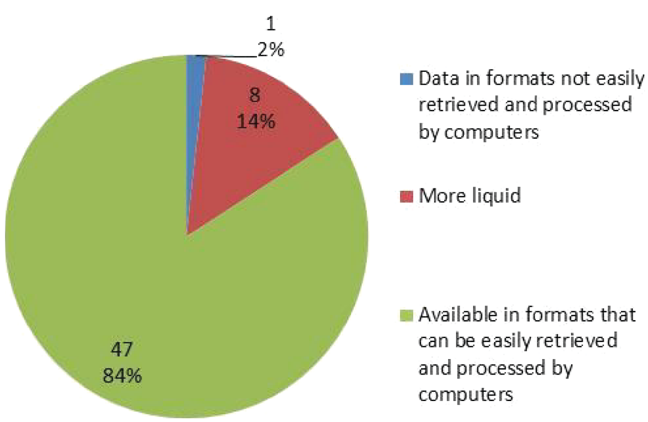


FIGURE 31 Machine readability of open data for percentage of survey respondents (from survey responses).

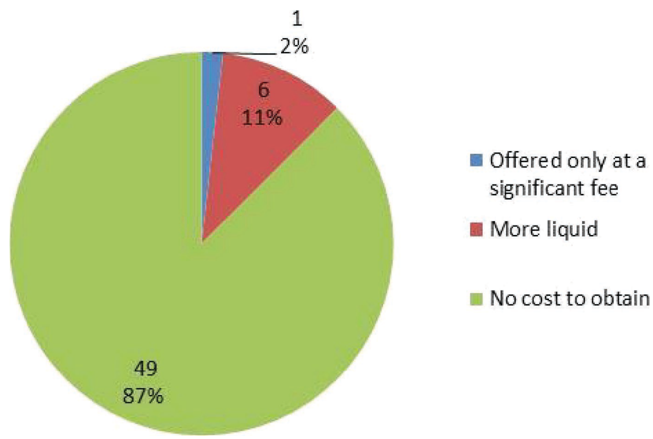


FIGURE 32 Cost of open data for percentage of survey respondents (from survey responses).

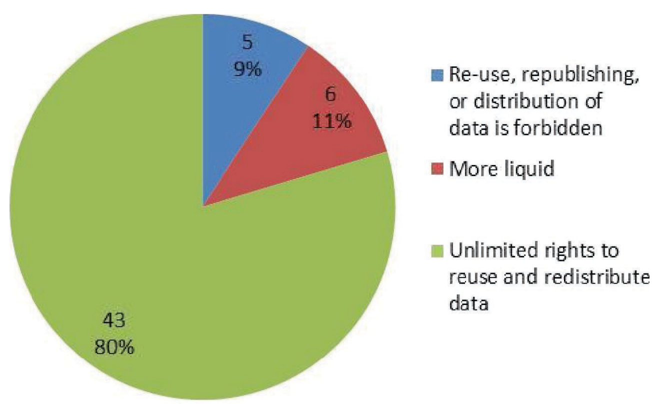


FIGURE 33 Rights to open data for percentage of survey respondents (from survey responses).

TABLE 9
STANDARDS AND FORMATS USED BY SURVEY RESPONDENTS

Standards and Formats	Number of Respondents	Percent
General Transit Feed Specification (GTFS)	47	83.9
eXtensible Markup Language (XML)	26	46.4
Comma Separated Values (CSV)	18	32.1
General Transit Feed Specification-realtime (GTFS-realtime)	15	26.8
Geographic information system (GIS) software	14	25.0
Keyhole Markup Language (KML)	10	17.9
Service Interface for Real Time Information (SIRI)	10	17.9
TXT	7	12.5
DATEX2	2	3.6
Transit Communications Interface Profiles (TCIP)	2	3.6

Other:

- Developed our own—real-time owing to performance required.
- Excel
- GIRO HASTUS proprietary format
- IN-TIME (Intelligent and efficient travel management for European cities) Platform, Traffic Management Data Dictionary (TMDD)
- INIT–Trapeze special interface
- JavaScript Object Notation (JSON)
- JSON—service-oriented architecture (SOA) services
- MS Word and Excel
- Network Exchange (NeTEx)
- Scheduled information only
- Web services
- XML and Simple Object Access Protocol (SOAP)
- Web services/APIs
- Linked Data, Resource Description Framework (RDF), GeoJSON (Geospatial data interchange format based on JSON.

From survey responses.

TABLE 10
WHERE OPEN TRANSIT DATA ARE MADE AVAILABLE

How Agencies Make Data Available	Number of Responses	Percent
Via your agency's existing website	39	69.6
GTFS Data Exchange website	18	32.1
As an application programming interface (API)	18	32.1
Via a separate agency website	12	21.4
Via a third-party site	11	19.6
Via a single regional centralized site	9	16.1
Public Feeds Wiki page on Google Transit Data Feed Google Code project	8	14.3
Providing data in bulk	7	12.5
Via an ftp site	3	5.4

Other:

- CD Bustools API
- IN-TIME platform (Co-Cities project)
- OGD Vienna
- Really Simple Syndication (RSS) and XML Feeds
- XML files through agency firewall
- Printed schedule as well as website
- Private web service.

From survey responses.

CHAPTER FOUR

SURVEY RESULTS: LEGAL AND LICENSING ISSUES AND PRACTICES

Just more than half (29 or 50.9%) of the survey responses show that agencies require a license or agreement to use the agency's open data. Almost 60% of the respondents (16 responses of 27) said that they require acknowledgment of a license agreement before a third party accesses the open data. Just more than 83% of the respondents (25 responses of 30) note that they do not require another type of registration before a third party accesses and uses the data.

The respondents' license agreements cover a variety of items, the most common of which is the right to use the agency's data. The next most common item is the limitation on data availability (nonguarantee of data availability), accuracy or timeliness, and the third most common is the liability limitations for missing or incorrect data. The remaining items and the frequency of their occurrence are shown in Table 11.

Almost all of the survey respondents (54 or 98.2%) have experienced no legal issues resulting from the release of their data to the public.

Figure 34 shows the steps taken by the survey respondents to disclose their data publicly. Those who responded that they took "other" steps to disclose their data reported the following:

- Accepted Google's request
- All of above (meaning all of the choices for this question in the survey) were done for corporate reasons, but enabled open data publication
- Developers page on agency website
- Open data are available through a local development environment (ITS Factory)
- Developed license agreement
- Work with developers
- Our data were already accurate and used via web services developed for internal purposes. Internal purposes turned out to be similar needs for external developers, just a few minor tweaks based on comments that actually improved it
- Only provide GTFS and website information to public. All other data are restricted to the Metropolitan Planning Organization (MPO), government agencies and research entities by request.

The survey asked what agencies do if they discover irresponsible users. Among the responses were 15 indicating that this situation has not occurred. Several responders indicated that they have a policy in place to handle this situation but have never had to exercise it.

- Limit or terminate/cancel access (some by revoking the key that developers receive when they register to use the data).
- Block the data.
- Contact/follow-up directly with the publisher and user, and try to resolve the issue.
- Limit access and monitor the public sites or API.
- If it is an incorrect time/broken app, do nothing and let the market sort itself out through reviews and other means.
- Ignore them.
- Probably would let the active developer community publicize the offender.
- We identify our and data consumers' responsibilities in the terms of use. We take no responsibility if someone violates those terms.
- No misuse of data to date, but periodically ask new developers to please comply with the terms of use if they are not. The few that were encountered apologized and corrected immediately.
- 1. Verbal warning to stop. 2. Remove their access to the feed.
- We issue a license key for each user however so we would simply revoke the key of any irresponsible user.
- If a trademark or copyright violation, it would be referred to legal for review.
- For real-time data access can be blocked and the key can be revoked.

TABLE 11
ITEMS COVERED IN LICENSE AGREEMENTS

Items Covered in License Agreement	Number of Respondents	Percent
Right to use the agency's data	20	71.4
Nonguarantee of data availability, accuracy, or timeliness	19	67.9
Liability limitations for missing or incorrect data	18	64.3
Agency's right to alter data without notice or liability	17	60.7
Data ownership	15	53.6
Use and placement of copyrighted logos and images	14	50.0
Indemnity from technical malfunctions due to users' use of data	13	46.4
Limitations on use of data	13	46.4
Indemnity from legal actions against data users	12	42.9
Indemnification	9	32.1
Termination	5	17.9
Licensing fees and royalties	3	10.7
Quality control	3	10.7
Other: • Applicable law • Caveat to be provided when data published • Do not overwork servers by requesting more data than necessary • License is implicit with the use of the data • We use acceptance of terms of use • Differs for web services and static data sets. It is very liberal. • http://www.transitchicago.com/developers/terms.aspx • No license in place yet, only agreement with app developers, MA DOT provides GTFS access without license.		

From survey responses.

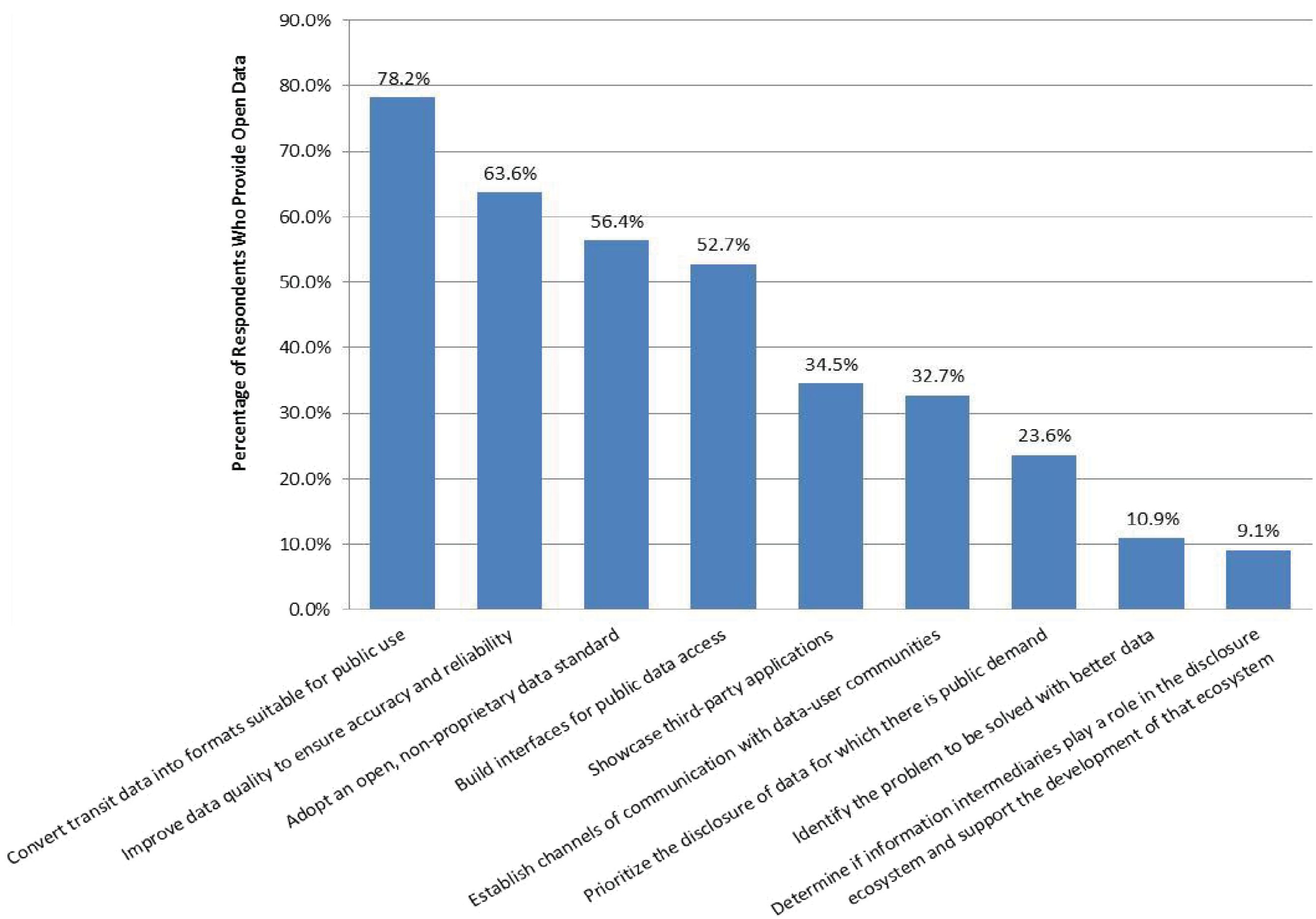


FIGURE 34 Data disclosure practices (from survey responses).

CHAPTER FIVE

SURVEY RESULTS: USES OF OPEN DATA

APPLICATIONS

The survey contained several questions regarding how open data were being used in terms of customer applications, decision-support tools used by the agency itself, and non-transit applications. Further, there were two questions regarding how agencies monitor use of the open data.

Customer applications are the most prevalent use of open transit data. As shown in Table 12, trip planning is the most common use of open data, followed by mobile applications and real-time transit information.

The types of decision-support tools that use the open data are shown in Table 13. Data visualization is the most common tool, followed by service planning and evaluation, and route layout and design.

Survey respondents reported numerous applications that used their open data. The applications reported to be used most frequently are:

- Google Maps
- Google Transit
- HopStop
- OneBusAway
- Open Trip Planner
- Rome2Rio
- RouteShout
- TimeTable Publisher
- WalkScore

Although the other applications reported by respondents are too numerous to list and are, for the most part, locally developed, the survey responses indicate that even the smallest agencies have more than one application that uses open data.

Almost two-thirds (33 or 63.5%) of respondents stated that they were not aware of other uses of their agency's open data. Further, the same number of respondents do not track usage of their open data. Table 14 shows the methods used by the agencies that do track usage.

DOWNLOADS

When survey recipients were asked to estimate how much data are being used or downloaded over a certain time frame, there was a wide variety of answers. Of the 31 responses to this question, just less than half (15) reported that they either do not know or cannot estimate how much data are being used or downloaded. Those who could estimate the volume reported the following:

- Per day
  - 1,800,000 queries per day
  - 2,000,000 API calls per day
  - About 250,000 unique user accesses daily
  - Data download average is 4 gigabytes per day
  - 100,000 API requests per day
  - Number of daily transactions approximately between 20,000 and 35,000 per day
  - Approximately 85,000 requests per day
- Per month
  - Less than 30 megabytes per month
  - 250 megabytes per month
  - Three terabytes per month
- Per year
  - 7,213 downloads in the past year (about 40 gigabytes)
  - For the real-time feed, 18,045 gigabytes in the last 12 months
  - Approximately 100 megabytes per year, with a GTFS file slightly less than 1 megabyte
- Per download: 100 megabytes per download.

Obviously, the volume is based on several factors, including the amount of data being accessed, the size of the agency, and the number of applications accessing the data.

The types of applications reported by survey respondents were as follows:

- Thirty-seven respondents (88.1% respondents) who indicated that they have mobile applications counted a total of 764 applications;
- Twenty-eight respondents (66.7%) who indicated that they have web-based applications counted 191 of them; and

TABLE 12
TYPES OF CUSTOMER APPLICATIONS USING OPEN TRANSIT DATA

Customer Applications	Number of Respondents	Percent
Trip planning	41	75.9
Mobile applications	38	70.4
Real-time transit information (arrival/departure times, delays, detours)	32	59.3
Maps	31	57.4
Data visualization	21	38.9
Timetable creation	17	31.5
Interactive voice response (IVR)	14	25.9
Accessibility	12	22.2
Other	8	14.8
Ridesharing	8	14.8
Crowdsourcing	5	9.3

From survey responses.

TABLE 13
TYPES OF DECISION-SUPPORT TOOLS USING OPEN TRANSIT DATA

Decision-support Tools	Number of Respondents	Percent
Data visualization	17	51.5
Service planning and evaluation	13	39.4
Route layout and design	11	33.3
Performance analysis	11	33.3
Travel time and capacity analysis	10	30.3
Spatial analysis	7	21.2
Regional transit analysis	7	21.2
Other	7	21.2
Demand modeling	4	12.1
Temporal analysis	3	9.1
Financial analysis	3	9.1
Cost/benefit analysis	2	6.1
Energy consumption	2	6.1
Safety analysis	2	6.1

From survey responses.

TABLE 14
WAYS TO MONITOR OPEN TRANSIT DATA USAGE

How Respondents Monitor Data Usage	Number of Respondents
Monitor data downloads	10
Keep track of applications developed	9
Other	5
Temporal analysis	3
Track number of downloads per application	2
Performance analysis	2
Service planning and evaluation	2
Travel time and capacity analysis	1

From survey responses.

- Those indicating "other" (31%) types of applications counted 46 of them.

The types of platforms running these applications are shown in Table 15. Of the respondents, 91.9% have Android and 91.9% have iOS applications. The next most commonly used platform is Windows Mobile.

Approximately one-third (18 or 36%) of the respondents indicated that they have a web location where potential application customers can review available applications.

Twenty-seven of the respondents estimate that 6,438 developers are using their open data. Sixteen respondents indicate that they get 265,858,333 API calls per month, as of the beginning of 2014.

VISUALIZATIONS

Several visualizations were noted by survey respondents. Those mentioned by the MBTA, BART, and TfL are included as part of their case examples in chapter seven. Others were found in the literature and described in chapter two. One mentioned in response to a survey question is the distribution of rent and transportation cost burdens in the Denver area, which is shown in Figure 35 (http://www.denverregionalequityatlas.org/, accessed on March 28, 2014).

TABLE 15
TYPES OF MOBILE PLATFORMS USING OPEN TRANSIT DATA

Platforms for Mobile Applications	Number of Respondents	Percent
Android	35	91.9
iOS (Apple)	35	91.9
Windows Mobile	11	29.7
Blackberry	7	18.9
Nokia	6	16.2
Mobile Linux	1	2.7

Other:

- Text messaging app (Dabnab)
- HTML5
- Jolla
- Palm WebOS
- Pebble
- Short message service (SMS)
- Windows 7
- OSX
- Mobile web apps
- Spotbros.

From survey responses.

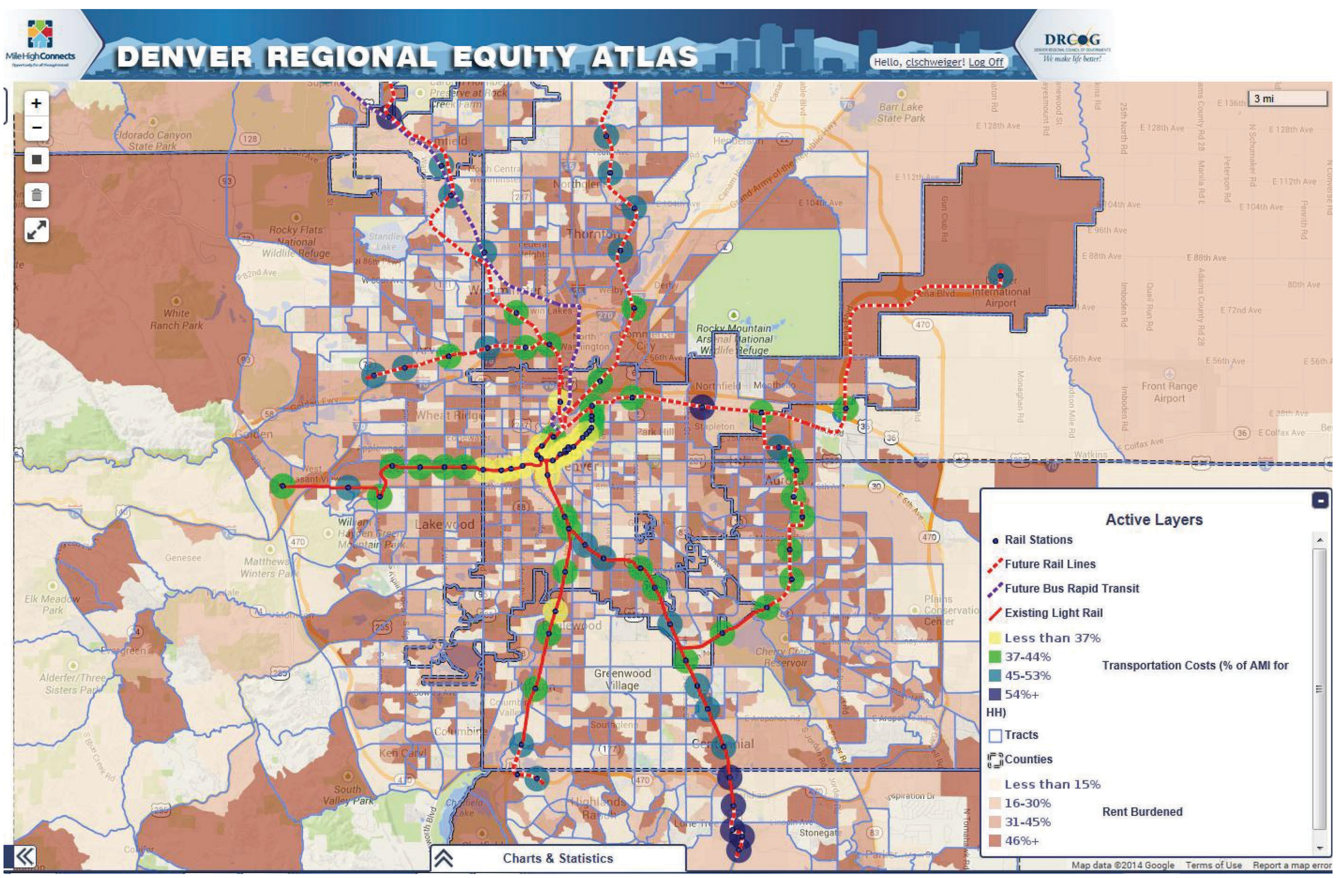


FIGURE 35 Sample Denver regional transportation costs versus rents.

CHAPTER SIX

SURVEY RESULTS: COSTS, BENEFITS, CHALLENGES, AND OPPORTUNITIES

COSTS

Survey respondents reported several types of costs associated with providing open data, as shown in Table 16. Just more than three-quarters of the respondents reported that staff time is required to update, fix, and maintain the data, and almost 70% reported that internal staff time is required to convert the data to an open format. The third most frequently reported type of cost is the staff time needed to validate and monitor the data for accuracy.

Almost 90% (43 or 89.4%) of survey respondents indicated that they cannot quantify how much time is spent on any of these activities. This finding is consistent with the author's experience in exploring the actual costs of providing open data—because much of the costs are internal, many agencies do not track staff costs, so they may not be aware of the exact cost of the activities associated with opening their data. Further, there are support costs associated with software that generates open data. It is rare that agencies account for these costs when assessing the costs of open data.

The amount of time being spent on these activities by the several agencies that do know their costs varies widely, as shown in Table 17.

About 95% (46) of the respondents cannot identify the actual costs associated with any of these activities. Only one agency reported that the monthly cost associated with the web service for hosting data is $1,500.

BENEFITS

The survey asked about the benefits that the agency experienced as a result of providing open data. More than 75% of the respondents reported that it increased the awareness of their services. About 75% reported that it empowered their customers and encouraged innovation outside of the agency. The other benefits reported by respondents are shown in Table 18.

DATA USER AND REUSER ENGAGEMENT

More than two-thirds (33 or 69.6%) of the respondents stated that they engage or have a dialogue with existing and potential data users and reusers. Table 19 shows the reasons for the engagement.

The techniques used by the respondents to engage existing and potential data users and reusers are shown in Table 20. The most frequently used techniques are face-to-face events, followed by conferences and meetups.

OPPORTUNITIES AND IMPACTS

The impacts of providing open data were explored through the survey. This open-ended question yielded a variety of responses, generally in the categories of positive and negative impacts, as follows:

- Positive Impacts
  - Increased transparency (most frequently mentioned impact–mentioned by five respondents)
  - Labor reallocation
  - Increased return-on-investment (ROI) from existing web services
  - Improved market reach and time to market
  - Better and more accurate internal data
  - Better and more accessible transit information for transit users, empowering customers in terms of choice and competition, and improving public rider perception and awareness
  - Better visibility of provided public services
  - Easier use of transit system
  - Fosters innovation
  - Improved relationships/coordination with MPO, DOT, research institutions, and neighboring agencies
  - Reduction in the requests for data from other government and private (consultant) users
  - Allows the agency to realign its information services and delivery in a more cost-effective and customer-focused manner
- Negative Impacts
  - Development effort and maintenance, and generally, staff time
  - Data quality
    - △ Time required to do quality analysis on data, resulting in more pressure on staff
    - △ Increased awareness of data quality issues, requiring more resources to devote to data consistency
  - Requires cross-checks using disparate information sources
  - Obligated to provide up-to-date data
  - Additional workload when dealing with developers who may not fully understand transit data

TABLE 16
TYPES OF COSTS ASSOCIATED WITH OPEN DATA

Types of Costs Associated with Providing Open Data	Number of Respondents	Percent
Staff time to update, fix, and maintain data as needed	38	76.0
Internal staff time to convert data to an open format	35	70.0
Staff time needed to validate and monitor the data for accuracy	28	56.0
Staff time to liaise with data users/developers	25	50.0
Web service for hosting data	23	46.0
Publicity/marketing	12	24.0
Consultant time to convert data to an open format	11	22.0

Other:
- Contract management
- Cost to develop prediction software or use prediction Software as a Service (SaaS)
- Everything above is already done for internal purposes and it is all automated
- Investigation project agreement with the Faculty of Computing Sciences
- Consultant time to build editing tool
- License Routing service Mentz
- No additional costs are incurred.

From survey responses.

TABLE 17
LABOR HOURS PER OPEN DATA ACTIVITY

Activity	Number of Respondents	Range of Labor Hours per Month
Internal staff time to convert data to an open format	4	3–40
Staff time needed to validate and monitor the data for accuracy	4	1–10
Staff time to update, fix and maintain data as needed	3	2–20
Publicity/marketing	3	0.1–2
Staff time to liaise with data users/developers	2	0.25–6
Consultant time to convert data to an open format	2	20
Web service for hosting data	1	1

From survey responses.

TABLE 18
BENEFITS OF OPEN DATA

Benefits	Number of Respondents	Percent
Increased awareness of our services	39	78.0
Empowered our customers	37	74.0
Encouraged innovation outside of the agency	37	74.0
Improved the perception of our agency (e.g., openness/transparency)	33	66.0
Provided opportunities for private businesses	24	48.0
Encouraged innovation internally	21	42.0
Improved our market reach	18	36.0
Become more efficient and effective as an agency	11	22.0
Increased our return-on-investment from existing web services	10	20.0
Experienced cost savings	5	10.0
Been able to reassign staff	3	6.0

From survey responses.

TABLE 19
REASONS FOR ENGAGING DATA USERS AND REUSERS

Reasons for Engagement	Number of Respondents	Percent
Obtain feedback on data anomalies and data quality issues	25	75.8
Find out more about how people want to use/reuse your data	21	63.6
Expose your data to a wider audience	21	63.6
Provide technical support	20	60.6
Announce updates, modifications, etc.	19	57.6
Find out more about the demand for our data	18	54.5
Suggesting features to improve the functionality of applications	17	51.5
Find out more about prospective users/reusers	15	45.5
Enable existing and prospective users/reusers to find out more about your data	14	42.4
Explain transit jargon and definitions	12	36.4
Solicit requests for future data	11	33.3
Enable prospective and existing users to meet each other	7	21.2

From survey responses.

  - More scrutiny because of increased visibility of data accuracy, including third-party users wanting zero downtime
- Neutral Impacts
  - Thinking about data reuse versus public policies
  - Public awareness of what agencies are doing and how they are doing it.

The impacts on the public sector (e.g., riders, community citizens) of providing open data also were explored. The following were reported:

- Creating and improving access to additional and higher quality public services, including more and free applications
- Providing better, more accessible, and more timely public information and tools
- Resulting in improved and sustainable mobility
- Improving transparency and accountability
- Improving customer service, customer satisfaction, and public perception/image of transit, including service reliability
- Empowering the public
- Making transit more competitive (reducing “costs” of trip related to customer uncertainty) and easier to use
- Providing more visual information
- Providing more innovative applications that government agencies may not be able to provide
- Providing a better transit experience
- Increasing ridership
- Increasing competition for transit riders
- Providing better regional coordination
- Encouraging the development of third-party tools and applications

TABLE 20
DATA USER AND REUSER ENGAGEMENT TECHNIQUES

Engagement Techniques	Number of Respondents	Percent
Face-to-face events	24	60.0
Conferences	18	45.0
Meetups	12	30.0
Hackathons	10	25.0
Application competitions	7	17.5
Unconferences/BarCamps (conferences with no set agenda—the agenda is set at the time of the conference by the participants)	5	10.3
Speed Geek events (participation process used to quickly view a number of presentations within a fixed period of time)	2	5.1

Other:
- Local Open Data Advocacy Group
- Email
- Various online industry and developer forums

From survey responses.

The impacts on the private sector (e.g., developers) were reported by survey respondents as follows:

- Providing business/commercial and development opportunities, including new and expanded companies that could create a new eco-system of private entrepreneurs
- Enabling innovation and the creation of applications
- Providing data to cover new needs
- Decreasing the need for agency to develop apps on a multitude of differing platforms, which would be costly to do internally or to outsource
- Providing more visual information
- Providing a broader reach for customers
- Adding value to existing services
- Private sector interacting with transit more comfortably because they know more about transit
- Adding data by large trip planning services (Google, Bing, HopStop)
- Improving access with high potential for untapped growth areas that the public sector cannot fund or have access to
- Creating interest in agency and desire to show off coding ability

Many of the impacts to the agencies, public sector, and private sector are repeated, proving that opening transit data has a significant value for all three groups.

CHALLENGES

The survey asked about the challenges associated with providing open data and how the challenges were overcome. The survey responses include the following:

- Resources and organizational issues
  – Limited dedicated resources (both time and staff) responsible for managing open data (including data conversion/cleaning and validation)
  – The process/philosophy is still not fully understood
  – Securing management support
  – Agency coordination challenges
  – The lack of technical know-how within the agency
  – Challenging internal parties who believe that we should be charging for the release of data
  – Helping internal groups see the benefit/value of participation and demonstrating how this can reduce manual publication load
  – Closer attention to change management
  – Internal fear that we should not do it because not all predictions would be accurate and we would be criticized for that
- Data quality and timeliness
  – Ensure/preserve data quality, completeness and equity, and timely release of data
  – Necessary to clean the data
  – Interaction with regional systems and inconsistencies with what data are used in which of their systems
  – Improve/preserve rider perception of accuracy of arrival predictions, given operational impacts such as adverse weather, reroutes, construction
  – Ensure safety/security of files and information disseminated
- Standards and formatting
  – Standards help overcome formatting issues
  – Better organization of the marketing of available information to public
  – Managing the evolution of internal data model
  – Data scalability
- Marketing
  – Making the data uniform resource locators (URLs) known
  – Partnering with an organization (e.g., Mobility Lab) to publicize availability
  – Initially, resistance because of branding issues
- Technical issues
  – Tracking users and developers
  – Process of making the data available when new schedules are released
  – Finding ways to represent some of the unique aspects of rural transit (such as deviations, or stops not always reached in the same order) in a standardized format
  – Ensuring the route changes are reported in a timely manner to the individual responsible for maintaining the data feed
  – The ability to get the data out to developers at the speed they want
  – Local development environment
  – Developers want a wide array of features
  – Slow data retrieval
  – Allowing direct access to data through agency firewall
  – How to provide large amounts of data in a timely manner

LESSONS LEARNED

The lessons learned noted by survey respondents cover four major areas as follows:

- *Data quality and accuracy* are critical to the success of an open data program. Respondents mentioned the following:
  – Put quality checks in place when opening data;
  – Be as open as possible, but test the data before releasing it to developers;
  – Start small in terms of the amount of open data offered and then grow that when confident of data quality of new sources/data sets;
  – It is important to have good, clean data—things that you understand internally as a transit agency don't always translate well to people who are less familiar with your operations; and
  – Data must be compatible with or identical among the different formats in which they are made available.

- *Open data are not free.*
  - If you do not have staff to support open data (planning, engineering and maintenance, especially), do not implement such a program;
  - Providing open data requires a lot of technical understanding when establishing it. Think of the costs that come with providing such data;
  - Use standards to make it easier to provide open data; and
  - Select a technology vendor that supports open data or require it in the contract with the vendor.
- Recognize that *opening data will create changes within and external to the agency.* Respondents summarized this point as follows:
  - There is a shift that agencies have to get comfortable with—from providing solutions to providing data;
  - Open data will not solve every customer requirement, and agencies will still have to stay in the game (e.g., SMS, accessible services);
  - Customers are smart—they can tell which third-party services are best, and they will not hold the agency responsible for third-party services that are poor;
  - It is important that agencies not interfere with the market to ensure that the benefits of competition can be realized;
  - Further, an open data program should be supported by a project champion;
  - Carefully assign staff roles and skills;
  - Having buy-in from coordinating agencies is crucial;
  - Considering open data is a fundamental part of the overall information system; and
  - Ensure that data reuse complies with public policies.
- *Engagement and developing relationships with developers* is key to success as well. Respondents mentioned the following in relation to engaging data users and reusers:
  - Early engagement with potential users is key. Find out what they want and how they want it. Try and track who is developing what, particularly to understand the successes and failures;
  - Respond quickly to opportunities. Developers work on much shorter schedules than planners;
  - Developers will know the latest mobile platforms and can utilize these with your data;
  - Make it as easy as possible for developers to access your data, and make the license understandable and intimidating; and
  - Developers will help you determine the quality of the data if you provide a forum for this type of feedback.

CHAPTER SEVEN

CASE EXAMPLES

Several of the transit agencies and organizations that responded to the synthesis survey were interviewed by telephone to obtain more detailed information on their deployment of electronic signage. The results of the interviews are presented in this section as case examples.

MASSACHUSETTS BAY TRANSPORTATION AUTHORITY (BOSTON)

In July 2009, the Executive Office of Transportation (now Massachusetts Department of Transportation) began an open transportation data program by creating a developer's web page that contained a variety of transportation information, including route and schedule data for the Massachusetts Bay Transportation Authority (MBTA) and Massachusetts regional transit authorities (*109*). On November 14, 2009, the First Annual MassDOT Developers Conference was held at the Massachusetts Institute of Technology (MIT) to encourage the development of applications based on the newly opened transportation data. Two important announcements were made at this conference: (1) the winners of the 2009 MassDOT Developers Challenge were announced; and (2) the MBTA announced the availability of real-time bus information for selected routes (*110*). Laurel Ruma referred to the MBTA opening its data at Ignite 2010, where she presented "Better than Winning the World Series: Boston Opens Real-Time Transit Data" (http://igniteshow.com/videos/better-winning-world-series-boston-opens-real-time-transit-data). Her speech covered the eight steps that led to the successful opening of the MBTA's data: (1) build a community, (2) learn the lingo, (3) open data, (4) hold a contest, (5) be prepared to be blown away, (6) award unique prizes (e.g., subway pass for a year), (7) tell the world, and (8) repeat.

According to David Barker, there were three primary reasons the MBTA opened its data (D. Barker, manager of operations technology for the MBTA, personal communication, March 10, 2014):

1. The MBTA wanted to get its schedule and real-time data out to customers, realizing that this information is the ultimate in advertising MBTA services;
2. Customers wanted real-time information; and
3. There was enthusiasm for open government ideas and for trying something new (the MBTA wanted to be on the forefront).

The MBTA independently opened its schedule data and some real-time information, but initially did not share the real-time information. They wanted to put the real-time information in a data feed on a trial basis before opening it to the public. Further, the market was right for developers—they could use open data for web applications, phone apps, and electronic signs. Before the open data revolution, there were far fewer options available to developers. The MBTA got a positive response after opening its data, and that provided the momentum for continuing the open data program.

The choice of standards used for the open data was based on what was needed within the API and what was available in the marketplace. When the MBTA reviewed potential standards, agency personnel wanted to support GTFS-realtime and wanted to be on Google. (The MBTA was one of the first transit agencies in the United States to provide real-time information on Google Transit.) They looked at SIRI but found it to be verbose and somewhat complicated, so they decided against using it. However, if SIRI does become more prevalent among developers, they could support it. TCIP was well-suited for communications within an agency, but not for communications with developers, so it was not selected for use with the open data.

The MBTA developed an API in order to retrieve smaller sets of information than what is contained in GTFS-realtime (GTFS-RT) and include some information that is not in GTFS-RT. The MBTA selected XML format for its API because it is an industry standard for APIs. The documentation describing the MBTA's real-time open data can be found in MBTA and IBI Group (*111*).

MBTA's engagement with developers consists of doing a survey of developers and having a newsgroup in which developers can ask and answer questions. The MBTA received a better response from the survey than the newsgroup. In addition, the MBTA suggests running an event (developer's conference) when an agency has something to announce that is related to open transit data.

Although the MBTA has not conducted a survey that directly addresses the effects of the agency's open data program, a survey was conducted by the MIT examining the impacts of real-time information displayed on the electronic signs in MBTA subway stations. The data displayed on the signs, the arrival times of the next two subway trains, are driven by the same open data available to developers (*112, 113*). The conclusions of the impact study were as follows.

> Countdown signs have significantly altered how passengers view their wait for public transit in Boston. Passengers reduce estimates of wait time on average by 0.85 minutes. After controlling for service disruptions (when the countdown predictions are less accurate), passengers reduce wait time estimates by 1.3 minutes. This corresponds to a reduction of approximately 17% in total passenger wait time estimation, and a 50% reduction in wait time overestimation. This coincides with improved wait satisfaction for headways lower than 5 minutes, but decreased for headways greater than 9 minutes. Wait time overestimation remained high at 34% even with countdown timers (*112*, p. 9).

As of early April 2014, the MBTA's next steps were integrating with Twitter and taking steps to reduce message volume to customers who subscribe to T-Alerts. The agency considered moving from e-mailing alerts to providing them by means of text messages (SMS), but the cost was too high. The improvements in message formatting, which have been implemented, were:

- The long subject lines will be replaced with auto-generated information from meta-data that summarizes the subject; and
- A message needs to fit within Twitter length requirements—automated abbreviations.

Further, the MBTA is interested in conducting another contest or hackathon. In addition, the agency wants to leverage its API internally. For example, the agency is interested in placing a real-time information sign in one station showing the time until the next bus(es). Currently, there is a sign like this in Ruggles Station, but the data come directly from a vendor's product rather than the MBTA's API. The agency wants to replace this sign, and place signs in other stations, with signs that use the MBTA's own feeds and include alerts for services leaving that station, including real-time bus and subway times, and elevator outages. The more the MBTA uses its own API, the more the agency will discover improvements that developers might not think of and the more it will reduce vendor lock-in.

In a discussion about opportunities and challenges related to open transit data, Mr. Barker mentioned the following:

- There is less development in the open data area currently. The original open data products are being maintained.
- Adding staff to handle open data is challenging. One way this has been overcome is being able to refer to "good press"—that is helping to get support to add staff.

In terms of costs and benefits, the NextBus contract costs $20 per bus per month to make and publish predictions. Further, the public address/two-display countdown signs have to be maintained. However, the ROI is "fantastic," and many customer applications are free.

The MBTA reports six lessons learned regarding open data:

1. Start small and iterate;
2. Listen to and interact with developers;
3. Know that when you first release the data, there will be a grace period during which you will discover issues;
4. Plan for how you will sustain an open data program;
5. Capitalize on good press. Although it is hard to measure improvements resulting from open data, positive press helps to demonstrate the value of open data; and
6. It is important to maintain the momentum and demonstrate an open data project's performance.

In addition to numerous customer applications, several visualizations have been developed using MBTA open data. One visualization, shown in Figure 36, "shows a bit more

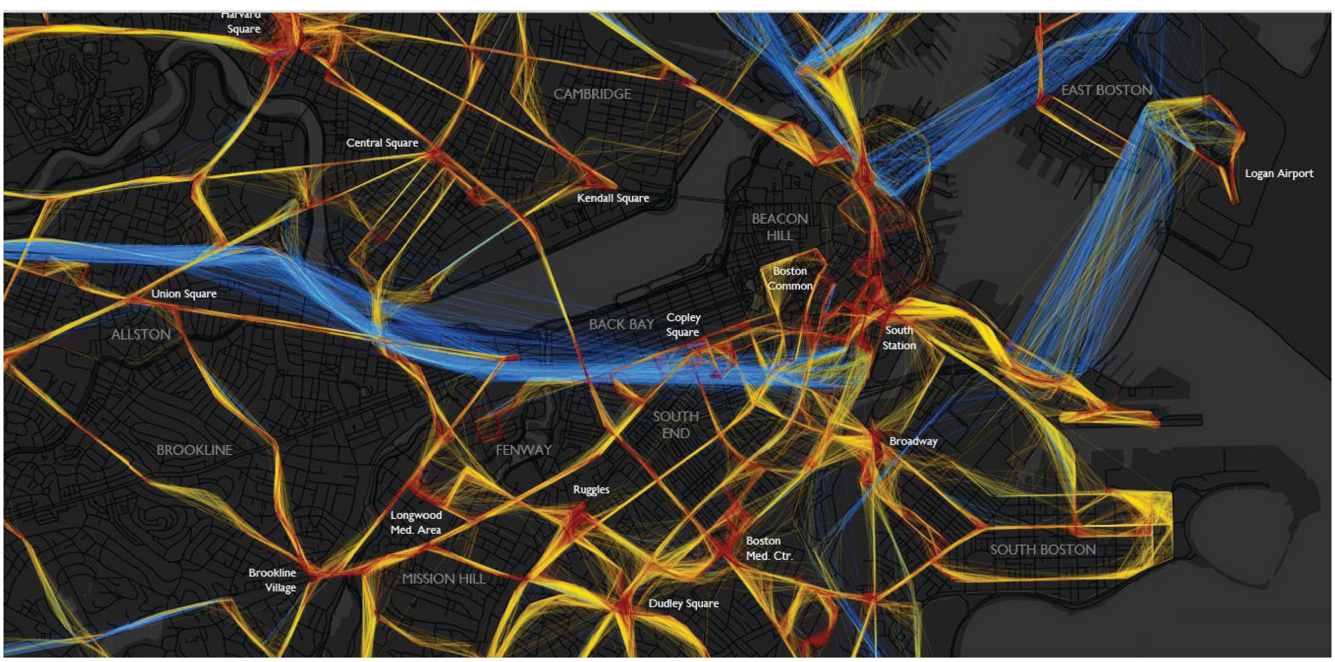


FIGURE 36 MBTA bus speeds (http://bostonography.com/wp-content/uploads/2011/11/mbta_1104_labels.jpg).

than 24 hours' worth of bus location data [on November 4, 2011] with colored lines representing the speed of each vehicle. Red indicates speeds less than 10 miles per hour, yellow is 10–25 mph, and blue is faster than 25 mph. It's drawn from 2,058,574 data points in all" (*114*).

Bostonography.com contains several visualizations based on MBTA open data, as shown in Figures 37 through 40. Figure 40 shows the MBTA Orange Line as a 24-hour clock. Each ring is a train station, and the thickness of the line represents the amount of people on the train at that time.

Another visualization (Figures 41 and 42) shows "24 hours of MBTA ticket swipes. The color represents the train line and the thickness represent[s] the amount of people riding at that hour" (http://thunderhead.esri.com/readonlyurl/MBTA/MBTA1.html). A visualization shown in Figures 43 and 44 displays the patterns of MBTA commuters across all subway lines and stations, or for a particular line or station over a single day.

In June 2014, another visualization of MBTA data was published in http://mbtaviz.github.io/. This visualization is an in-depth analysis of how MBTA subways operate using a Marey diagram and a heat map that shows the average number of people who enter and exit subway stations for every hour throughout the month of February 2014.

TRANSPORT FOR LONDON—LONDON, UNITED KINGDOM

As discussed in *TCRP Synthesis 91* (*1*), Transport for London (TfL) began providing open data in June 2010. According to Phil Young, Head of Online at TfL (P. Young, personal communication, Feb. 25, 2014), the chronology of events leading to providing open data through early April 2014 is:

- 2007—Launched embeddable "widgets" for live travel news, map, and Journey Planner
- 2009—Special area for developers launched on TfL website
- 2010
  - Additional real-time feeds launched, with hundreds of developers registered
  - Greater London Authority (GLA)—digital advisory board, mayoral and deputy mayor support, mayor's live event, London Datastore

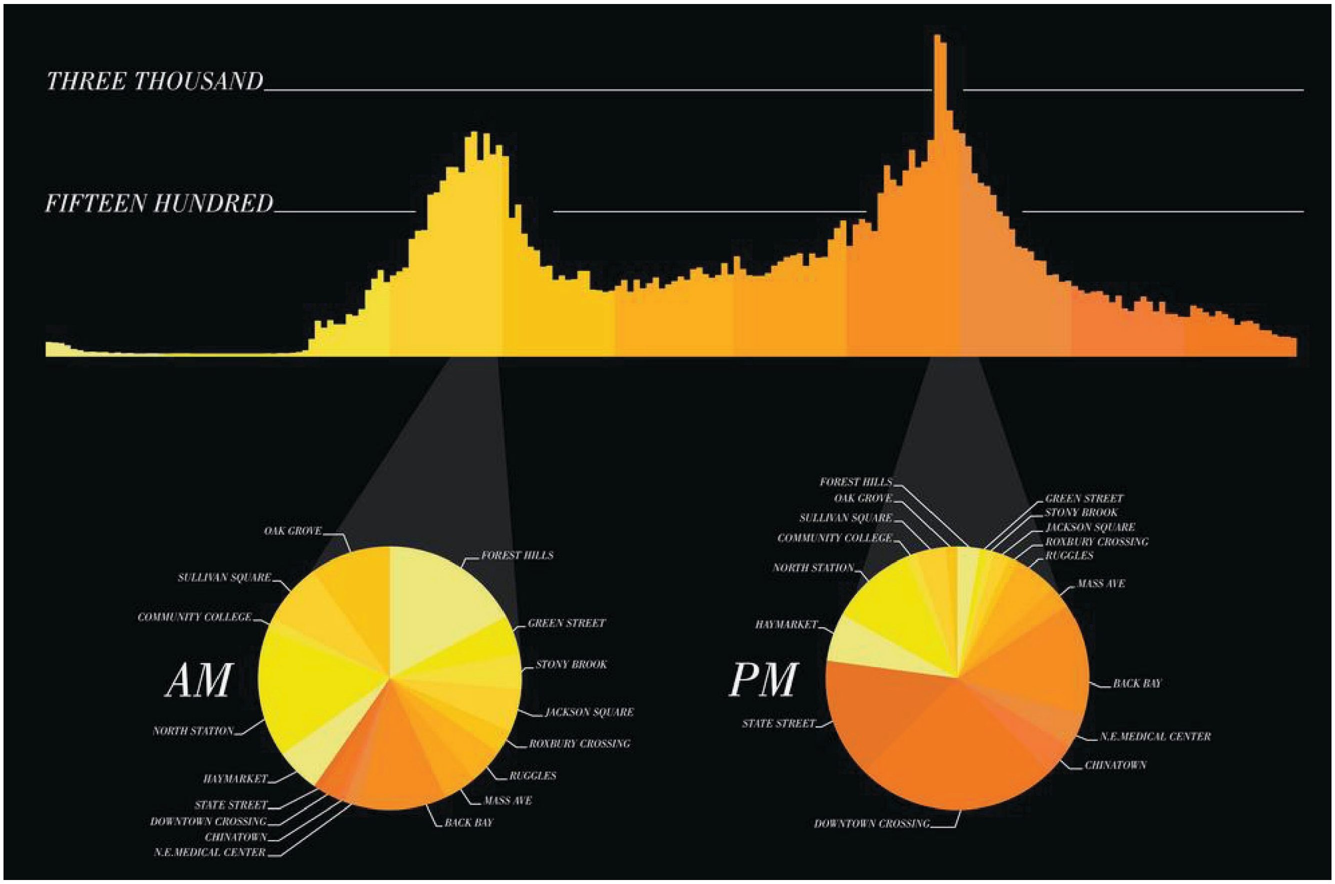


FIGURE 37 Visualization of MBTA Orange Line ridership for August 12, 2009 (http://vanderlin.cc/projects/mbta_visualizations: accessed on April 3, 2014).

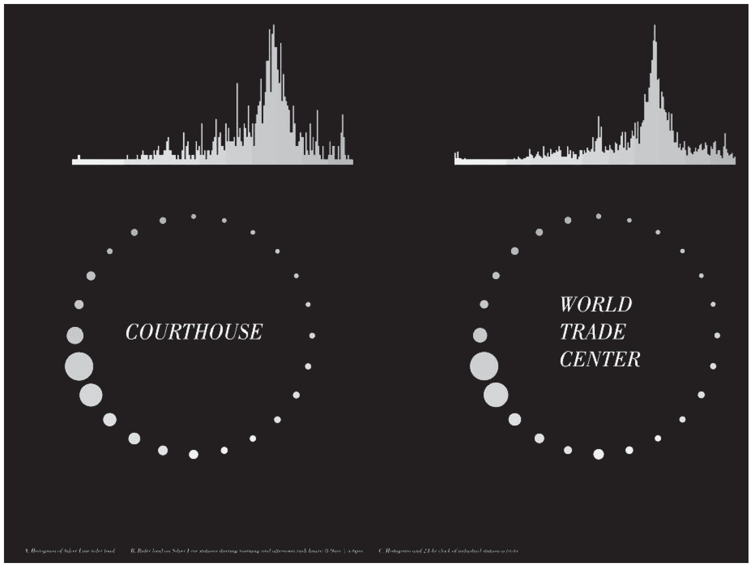


FIGURE 38 Visualization of MBTA Silver Line ridership for August 12, 2009 (http://vanderlin.cc/projects/mbta_visualizations: accessed on April 3, 2014).

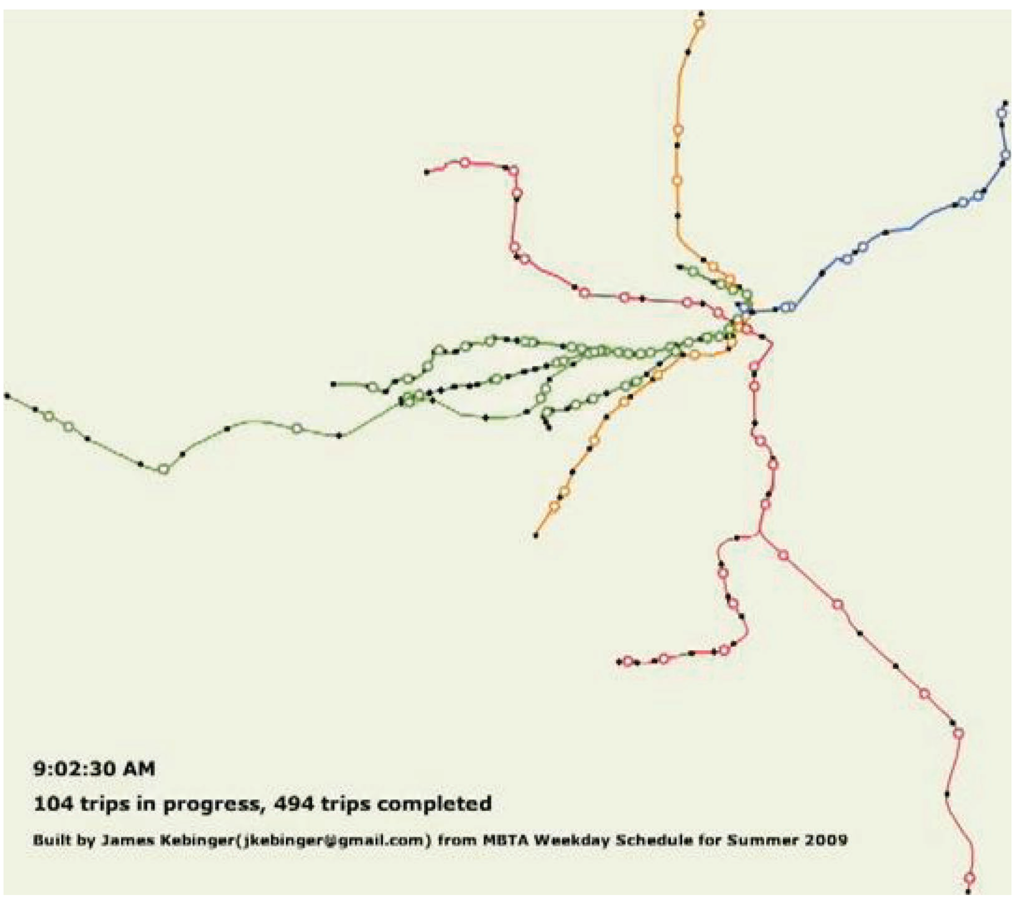


FIGURE 39 Snapshot of MBTA subway traffic for 24 hours in 5 minutes (James Kebinger “MBTA in Motion,” August 30, 2009: http://www.youtube.com/watch?v=0tuzjxEBto4).

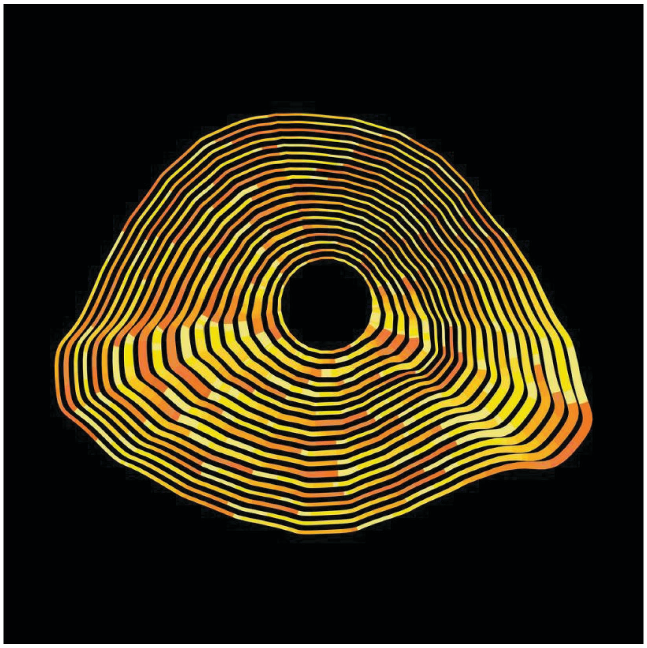


FIGURE 40 MBTA Orange Line over 24 hours (http://vanderlin.cc/projects/mbta_visualizations).

  - U.K. government—Public Sector Transparency Board 2010—central government launch of data.gov.uk to drive forward the release of "machine readable" data
- 2011
  - London Underground train location and Journey Planner APIs launched. Registered developers rise to more than 1,000.
  - Launch of U.K. Open Data Institute 2011—U.K. government-funded body to promote the social and economic benefits of open data
  - EU Open Data Strategy 2011—establishing a principle of open data as a rule across the EU
- 2012—Bus departures API launched, full London 2012 Olympic and Paralympic Games transport data portal. More than 4,000 developers registered.
- 2013—More than 5,000 developers, 30 data feeds, and hundreds of apps on the market serving millions of customers. New accessibility and road feeds added. At the same time, demand on the TfL website grows: 8 million unique users a month, 250 million annual visits, 2 billion annual page views

TfL's overall strategy is

> 'free and open data by default.' [TfL] works proactively with developers so that a range of apps are developed by the market where possible. Apps will only be developed by TfL itself where the market cannot deliver business objectives or required levels of security. We make all of our information feeds freely and openly available to more than 5,000 apps developers who then produce products—more than 200 in the case of London—which we would never have produced ourselves (P. Young, personal communication, Feb. 25, 2014).

TfL's Digital Strategy covering 2011 to 2014 describes how (*15*, p. 3)

- TfL will achieve an integrated, coherent, relevant presence across digital media, including
  - Online presence—web, tablet, mobile
  - On-system digital information
  - Open data
  - Social media
  - Digital marketing.

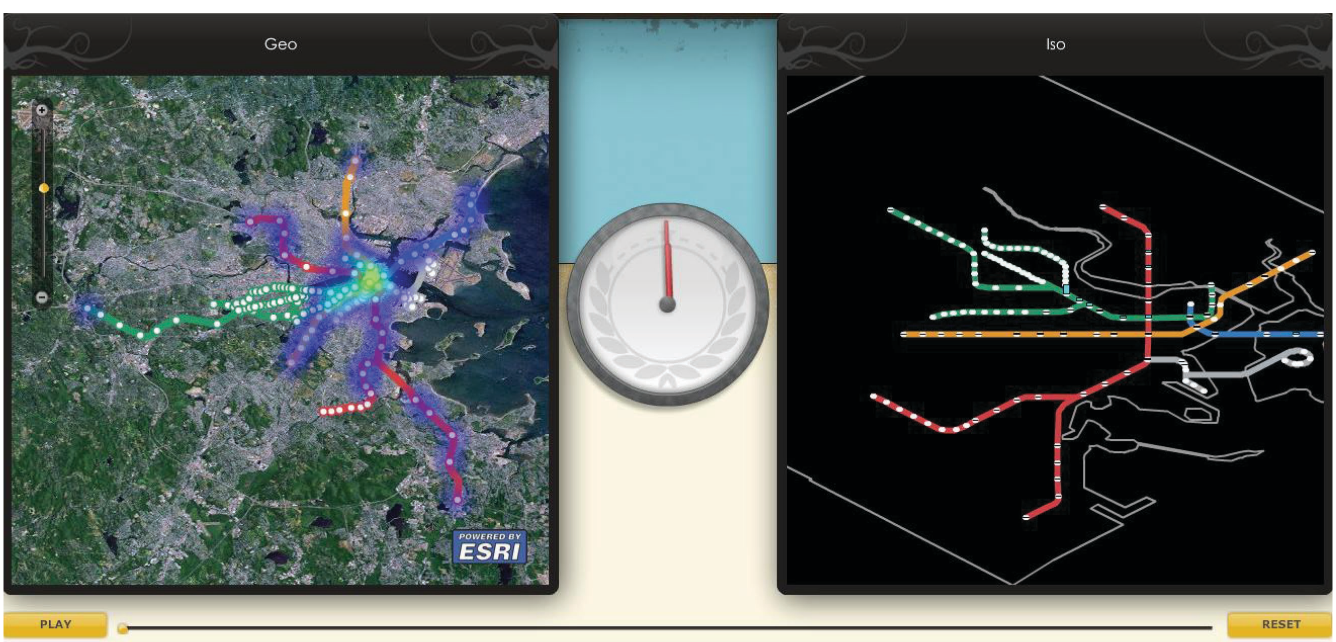


FIGURE 41 MBTA ticket swipe visualization at midnight.

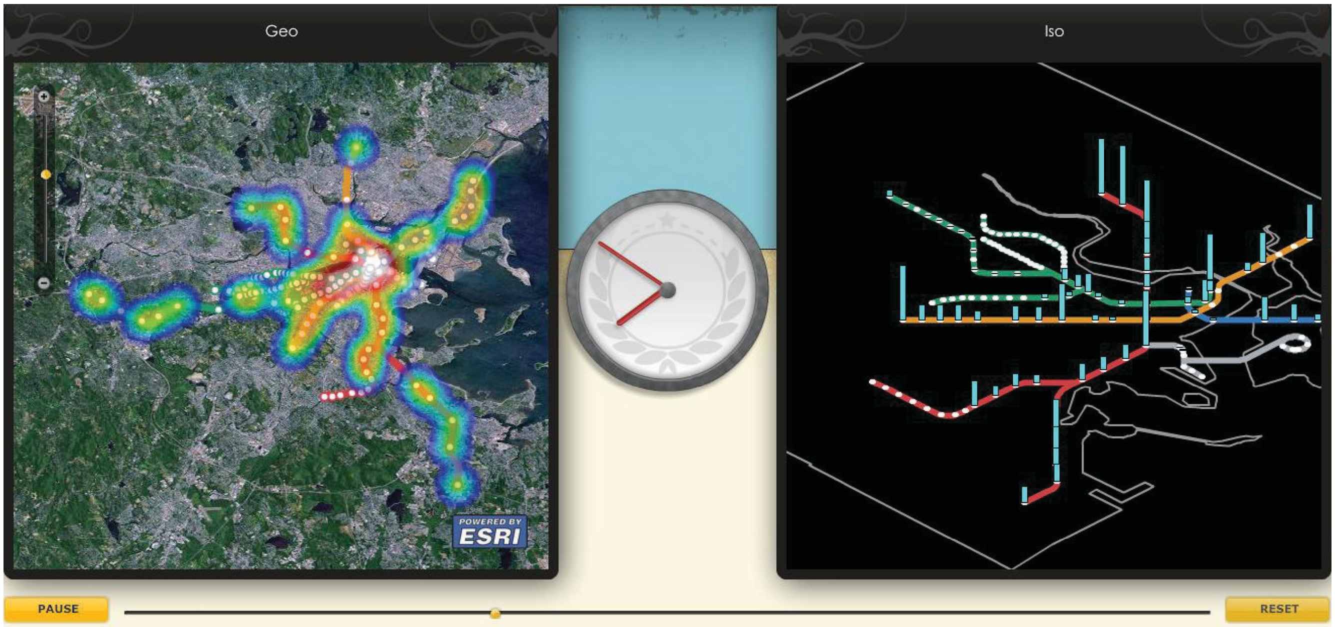


FIGURE 42 MBTA ticket swipe visualization at 7:50 a.m.

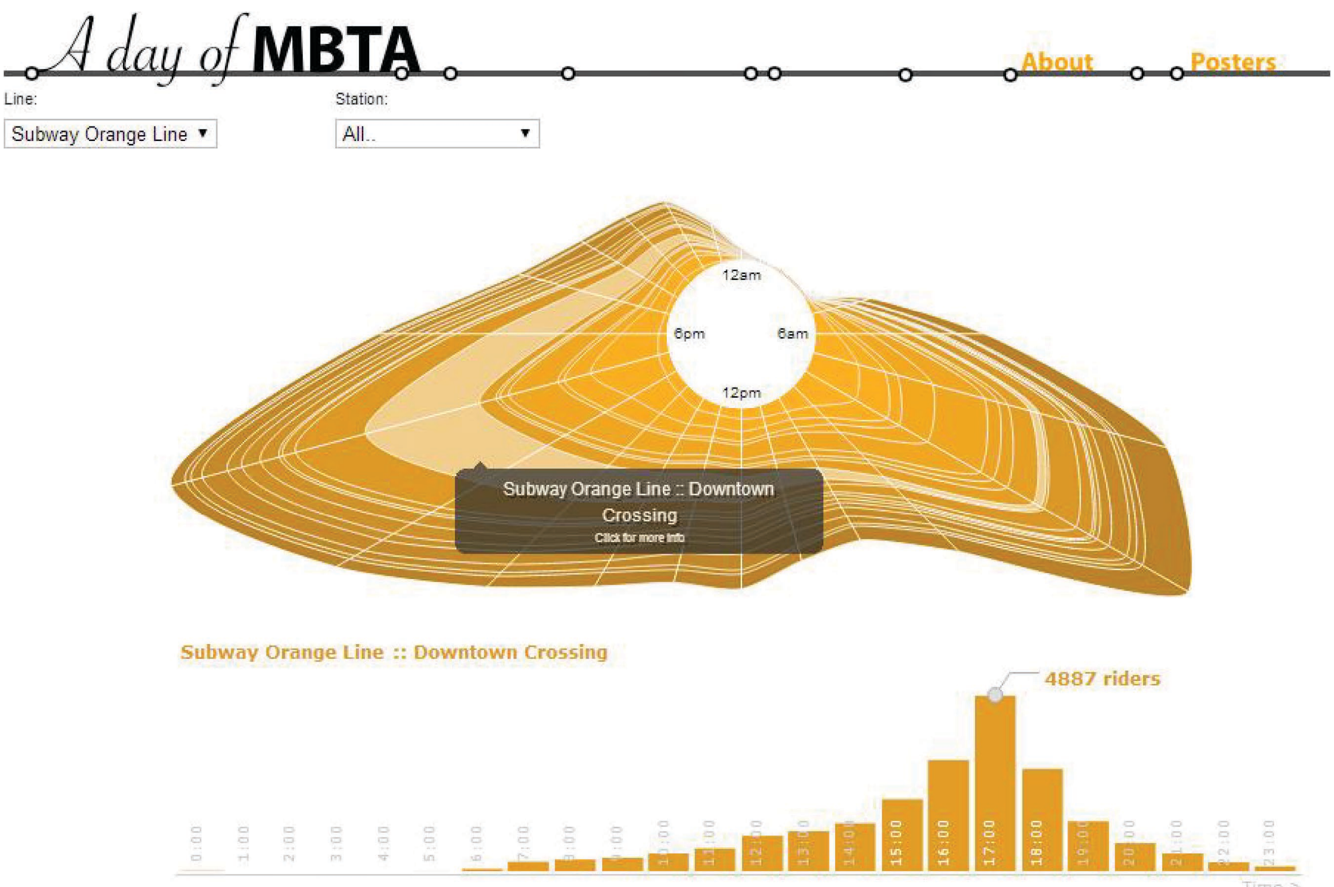


FIGURE 43 Visualization for MBTA Orange Line over 24 hours (“A Day of MBTA” http://adayofmbta.com/).

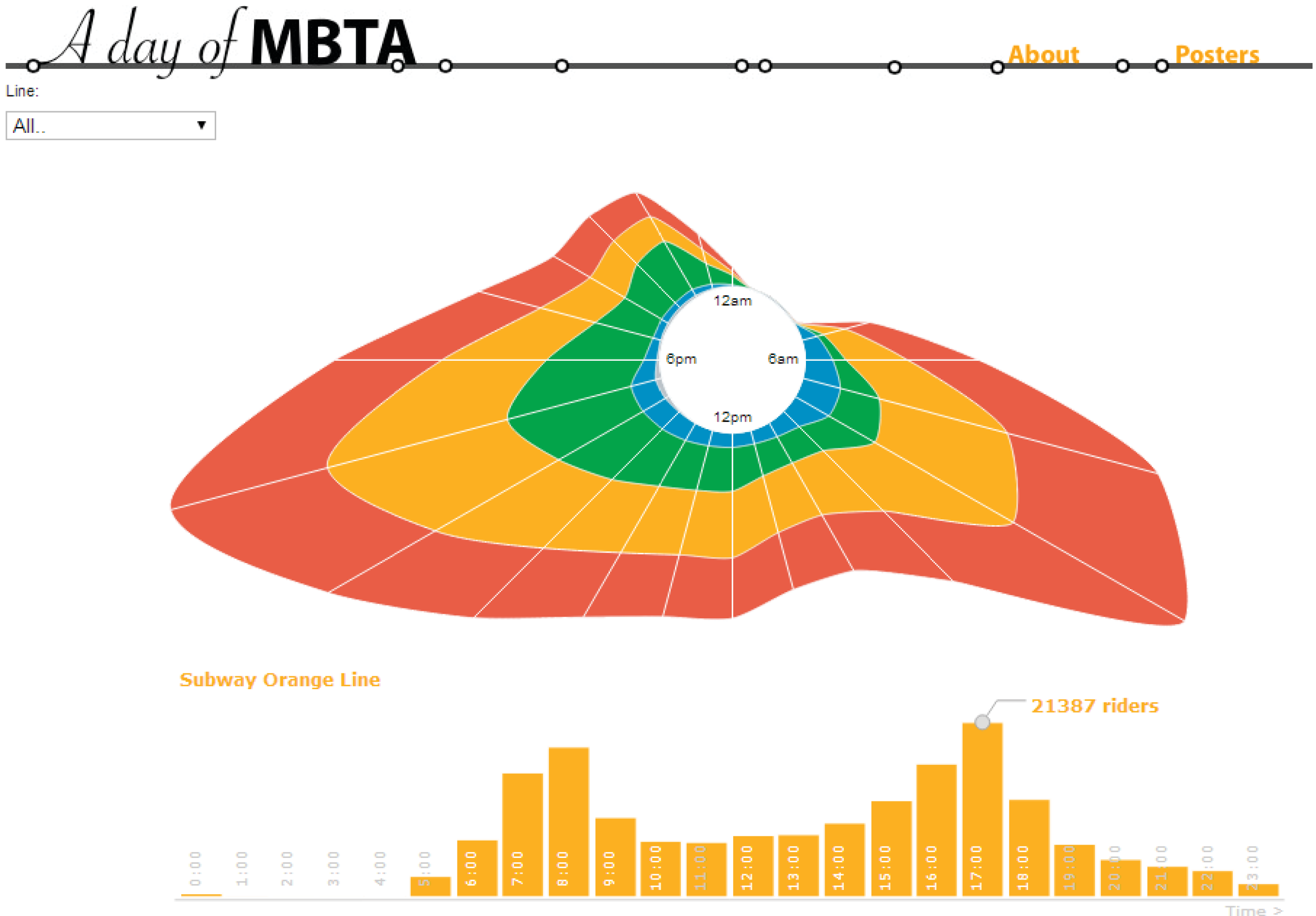


FIGURE 44 Visualization for all MBTA subway lines over 24 hours ("A Day of MBTA" http://adayofmbta.com/).

- For open data, TfL will deliver "Data openly syndicated to third parties, where commercially, technically and legally feasible, while TfL engages developers where necessary to meet our business objectives."

Examples of developers' products include the following:

- *London's Nearest Bus*—allows users to find the nearest buses and live departure times from their location. Users can also set individual bus alerts to trigger when a bus is due;
- *Station Master*—offers detailed accessibility information for every London Underground station;
- *Tube Tracker*—a multiservice app that finds the nearest station to the user with directions. Provides automatically updated live departure information, a journey planning function, first/last Tubes, and Tube status alerts;
- *Colour Blind Tube Map*—This displays the London Underground map in various formats for easier viewing by people with all forms of vision impairment.

Young reports that a global survey of transport companies by the International Association of Public Transport (UITP) shows that 54% of transport operators are now making their data openly available, with only 6% levying any charge to developers. The industry is also seeing it as an important way in which to demonstrate openness and transparency to customers and stakeholders.

At the same time, TfL is experiencing continued and rapid growth in visits to the agency's other information services. For example, the agency's website is getting 8 million unique users a month, and the number of people following the agency's Twitter feeds has risen to more than 1 million.

The reasons TfL opened its data are as follows:

1. *Public data*—as a public body funded by fares and taxpayers, the agency's transport information is seen to be owned by the public.
2. *Reach*—the agency's goal is to ensure persons needing travel information can get it wherever and whenever they wish in any way they wish. Open data allows them to extend the reach of their information.
3. *Optimal use of transport network*—by enabling customers to make more informed choices, TfL makes the most efficient use of the capacity of the transport network.

4. *Economic benefit*—open data saves customer time (up to £58 million per annum, according to a recent study) and facilitates the growth of small and medium technology companies, generating employment and a highly skilled workforce.
5. *Innovation*—by having thousands of developers building applications, services, and tools with TfL data and APIs, the agency stimulates innovation through competition.

Each developer must register with TfL to gain access to feeds and APIs. This ensures that the agency can maintain a relationship with them and provide information about changes, maintenance, or new services. In summary, the TfL license states the following.

- Data are free to access and use, including for commercial purposes.
- Applications can be created from the data and commercialized.
- The developer must not pretend to be TfL or use the TfL brand.
- The developer must not "screen-scrape" login-based services.

TfL's open data are presented in three main ways: (1) Static data files (which rarely change); (2) feeds (data files refreshed at regular intervals); and (3) API, which enable a query from an application to receive a bespoke response. Data are presented as XML when possible. In addition, TfL is moving toward the use of U.K. standards NaPTAN and TransXChange.

Young describes five lessons learned related to open data.

1. *Transport data are in great demand*—particularly real-time information, such as bus and train status and arrivals/departures.
2. *Developers need support*—processes and resources are needed to support, answer queries, and engage to deliver improvements to data and services.
3. *Something is better than nothing*—developers are highly creative and would prefer the feeds and data to be released early even if not perfect. They can still do great things with it.
4. *Open data can improve data quality*—By opening up data, feedback is produced, which refines the data quality and can be used to improve source systems.
5. *Open data are good value*—emerging research indicates that open data provide ROI of up to 50 times the sums invested in terms of customer benefit.

TfL's next steps in open data are as follows:

- *New TfL website in early 2014*—built upon the principle of APIs, making integration, build, and open data better and easier to deliver.
- *Single TfL API*—A single TfL API consolidating all existing and new APIs into a single normalized model.
- *App Garden*—to showcase applications "powered by" TfL data. This will give consumers choice and ensure apps meet minimum standards. Aspects of the Garden are:
  - Draw in customer ratings from app stores;
  - Improve branding guidelines so customers can see which are "powered by TfL";
  - Improve control of app developer access through a new access portal; and
  - Minimum standards check before applications are added to the App Garden.

A wealth of TfL open data is available from the London Datastore (http://data.london.gov.uk/taxonomy/organisations/tfl), as shown in Appendix H.

Many applications have been developed using TfL open data. One of the earliest applications was a live map of the London Underground, London's subway system. This app, which won second place at the 2011 Open Data Challenge competition that featured 430 entries from 24 European Union member states, is shown in Figure 45 (*115*). The same app developer created a similar real-time map showing bus locations (see Figure 46) (*116*). The app uses OpenStreetMap and open data from TfL.

Another visualization, shown in Figure 47, is described by Hargreaves: "It takes live position data from the 100s of buses that travel through the TFL network. The lines are the routes and the dots are the individual buses. The contours of the landscape have been exaggerated to emphasise the sense of space within the visualisation" (*117*).

Another visualization uses smartcard (called the Oyster card) data from TfL to show a day's worth of transactions, as shown in Figure 48.

> Green indicates the number of passengers in the transit system, whether on a bus or in one of several rail modes. Blue indicates the presence of riders prior to their first transaction of the day or after their last: it is assumed that the location of a rider's first or last transaction approximates their place of residence. Red indicates cardholders who are between transit trips, whether transferring, engaging in activities, or traveling outside the transit system (*118*).

Several additional visualizations using TfL data are shown in Reades (*119*).

BAY AREA RAPID TRANSIT—OAKLAND, CALIFORNIA

BART's open data initiative was first reported by TRB (*1*). In 1998, BART was approached by students from the University of California at Berkeley who wanted to develop one website with schedules for all 26 transit agencies in the Bay Area. At the time, there was no single source of transit

FIGURE 45 Real-time train locations in London (http://traintimes.org.uk/map/tube/).

information. BART knew that it would be virtually impossible to manage schedules for all agencies, so this request was a compelling value proposition. The students could have "scraped" the schedules. So BART gave the students schedules in .csv format. The site built by the students eventually grew and became the Bay Area's 511.org. It was positive for BART to support a third-party initiative that led to a sustainable regional initiative.

Another request for data came when the Westfield Mall above Powell Street wanted to install a touch-screen kiosk with BART schedules on it. They requested the schedules in a specific format. BART agreed to provide the schedule data. At this point, BART did not consider making anyone sign a contract to access the data on the kiosk or use schedule data in general. This point is critical to understanding BART's simple use agreement that does not need to be signed.

Because other people began to ask for schedules, fulfilling all of these requests would have been "one-offs" if BART had not opened their data. For example, trying to do an embedded quick planner or electronic displays was not the agency's core business. Then BART saw GTFS and recognized that using GTFS was a way to solve these problems. The thinking was that if the agency published data in this format, it could take care of requests. So it solved the challenges associated with responding to the data requests.

BART was one of the few agencies to pilot GTFS-realtime, but the agency had already been providing real-time feeds in XML format (the first real-time feed was provided in 2008). This XML feed was updated every 30 seconds and has since been retired. Currently, the use of the agency's API allows more granular calls that provide data such as car lengths. Several data items of interest are not in GTFS or GTFS-RT, such as station information and load factors. There are more requests about entries and exits with as much granularity as possible.

The overall chronology of BART's open data program, which has been in existence more than 15 years, is as follows (*28*, p. 6):

- 1998—Schedules provided in .csv format
- 2005—Embedded Quick Planner (iframe format) (retired November 2013)
- 2006—DIY display [in HyperText Markup Language (html) format] (retired in November 2013)
- 2007—Delay, elevator advisories (RSS)
- 2008—Real-time estimated times of arrival (ETAs) (XML format) (retired in November 2013)
- 2010—Trip plans, station info (API)
- 2011—Real-time ETAs, advisories (GTFS-realtime)
- 2012—App Map + Geospatial (KML format)

One key aspect of BART's open data program, according to Timothy Moore (interview, March 20, 2014), is that BART used an organic process to open the agency's data. BART's universe of data is manageable because of the agency's size and because there is only one pick (operator signup) per year.

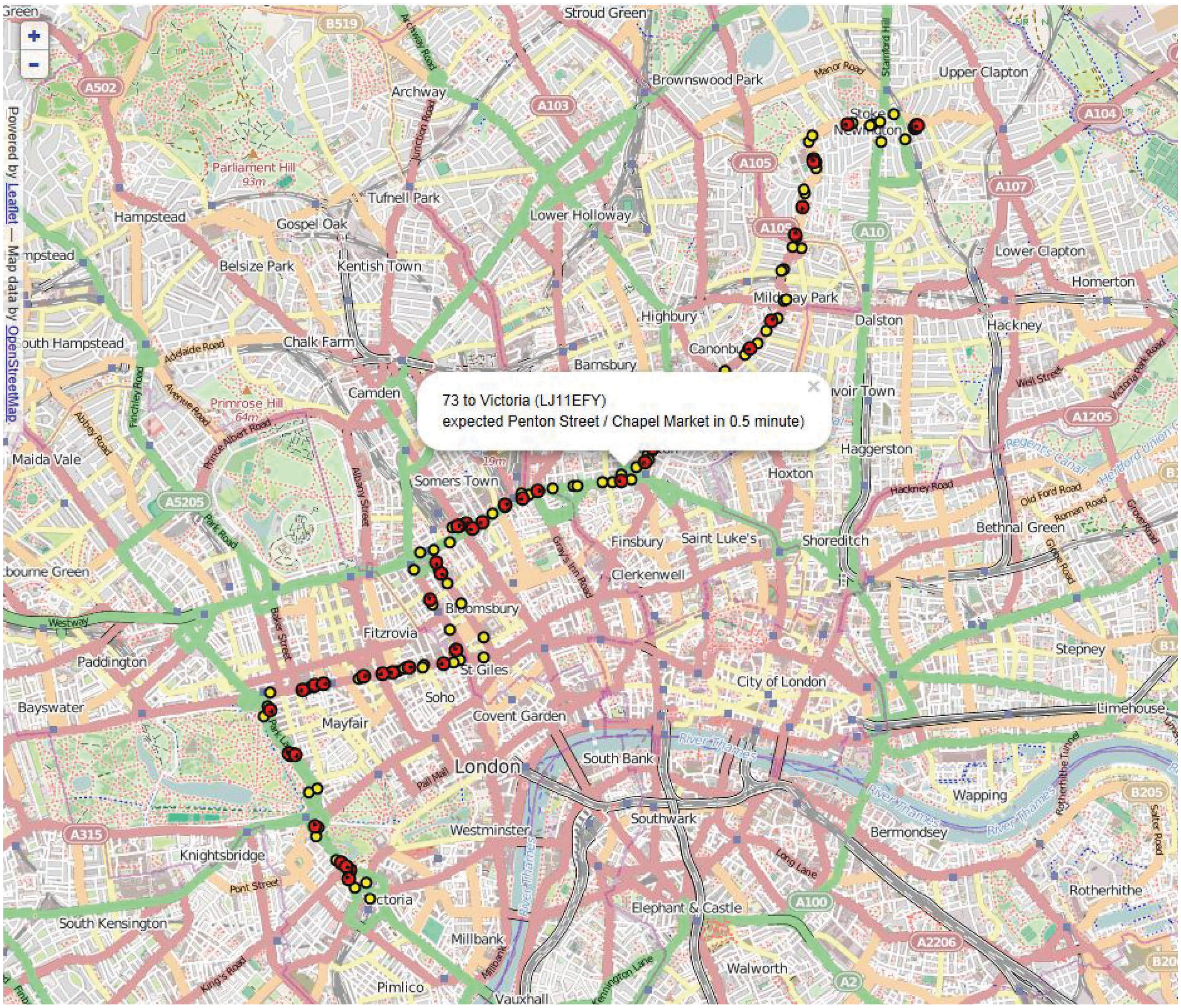


FIGURE 46 Live London bus map.

FIGURE 47 Snapshot of video visualization of London buses.

FIGURE 48 Visualization using Oyster Card Data over a 24-hour period.

As of December 2013, BART had 115 apps and open data services, 2,700 developers, and 40 million monthly API calls. Moore reported that the value of open data to developers is as follows (*28*, p. 7):

- Focus on new functionality
- Easier to combine/mash-up
- Add value to an existing play
- Some revenue potential (beyond BART's reach)
- Pride, fun, competition, and experience.

Moore also described the value for BART (*28*, p. 8):

- Cost savings
- Labor reallocation (one BART full-time equivalent)
- Increased ROI from existing web services
- Scale, improved market reach/time to market
- Empowered customers (choice, competition)
- Innovation and "trickle up"
- Increased awareness of BART's service
- Positive perception: openness/transparency.

BART's lessons learned regarding open data include the following:

- Releasing the data and adopting standards that gain traction are critical.
- Make the license readable by non-lawyers, and do not require that the license be signed.
- Create multiple paths into using the data:
  - Provide simple tools to the casual user
  - Provide RSS feeds to the medium-level user
  - Provide GTFS-RT and geospatial and the agency's API to the advanced user.
- Open data are to be documented and released using a transmission format that is accessible.
- Do not play favorites in the market. Do not interfere in determining the most effective apps—the public should be trusted to do that.
- Stay responsive to customer needs and developer needs.
- Provide information to developers to ensure that customers' needs are being met—synchronizing customer needs and developer skills are critical.
- Promoting data and apps by using free advertising and car cards promoting that data are available.
- Open data needs to be part of the agency's culture. It helps customers understand that BART is not responsible for every app.
- Hackathons can be good for launching an open data program, but they may not be effective on an ongoing basis. Now BART participates in more organized events, such as TransportationCamp West, and uses Google Groups and an e-mail list to stay engaged with developers.

Of the transit agencies in major metropolitan areas, BART has the smallest number of riders per app, as shown in Figure 49.

	BART	Portland	Boston	Chicago	NY
Avg. Weekday Riders (000's)	421.8	322.3	1,314.7	1,716.9	12,071.1
Number of Apps	115	109	66	41	109
App / Rider (000's)	1/3.6	1/5.7	1/19.9	1/41.8	1/110.7

Sources: APTA 3O2013 Public Transportation Ridership Report and agency websites, November 2013.
Note: iOS and Android apps are counted once (not once per platform).

FIGURE 49 Apps per rider (*28*, p. 5).

BART will be releasing a new multiplatform, free, smartphone app to report crimes and take pictures of and report suspicious items or hazards. Riders have been asking for a safe, silent and discreet way to communicate with BART when they are on a train or in a station. BART will be the first transit agency to offer both Spanish and Simplified Chinese options for the app. The app will feature a silent photo and flash-free feature and is GPS enabled.

Several visualizations have been developed using BART open data. One example, which was funded by the Knight Foundation, is shown in Figure 50 (http://barthood.news21.com/). The source for this visualization was mainly BART's station profile study (the raw data can be found at http://www.bart.gov/about/reports/profile).

Several data visualizations related to the BART strike in 2013, including ridership and regional traffic impacts (http://enjalot.github.io/bart/). Figure 51 shows one of the visualizations related to ridership by station. Figure 52 shows the passenger trips between Balboa Park and other stations. Figure 53 shows ridership from a different perspective at BART stations.

WORCESTER REGIONAL TRANSIT AUTHORITY—WORCESTER, MASSACHUSETTS

The Worcester Regional Transit Authority (WRTA) in Worcester, Massachusetts, is a transit authority in central Massachusetts with 48 fixed-route buses and 50 paratransit vehicles. Beginning in 2009, the WRTA embarked on a program to implement technology on all of the agency's vehicles, resulting in, among other things, providing real-time information to the public. Once the technology implementation was completed, the agency's information technology consultant, Christopher Hamman, began working on opening the agency's data. The reason the WRTA decided to open data was to show transparency and encourage developers to create new and better applications for their riders (Interview with Christopher Hamman on March 20, 2014). In addition, the administrator of the agency is visionary and fully supported opening the data to the public to get applications in the hands of customers. The WRTA pursued developers who had created apps for the

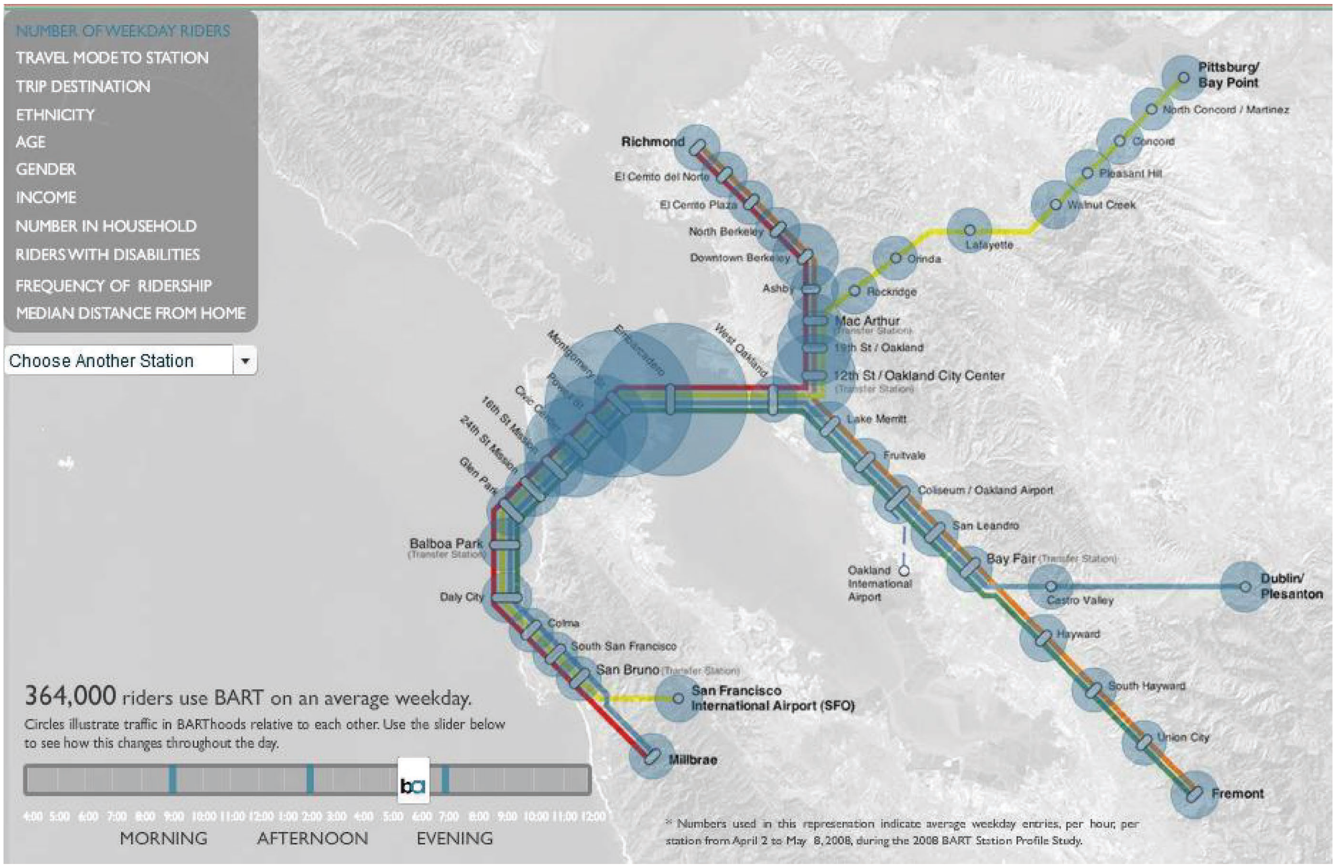


FIGURE 50 Visualization of ridership by BART station.

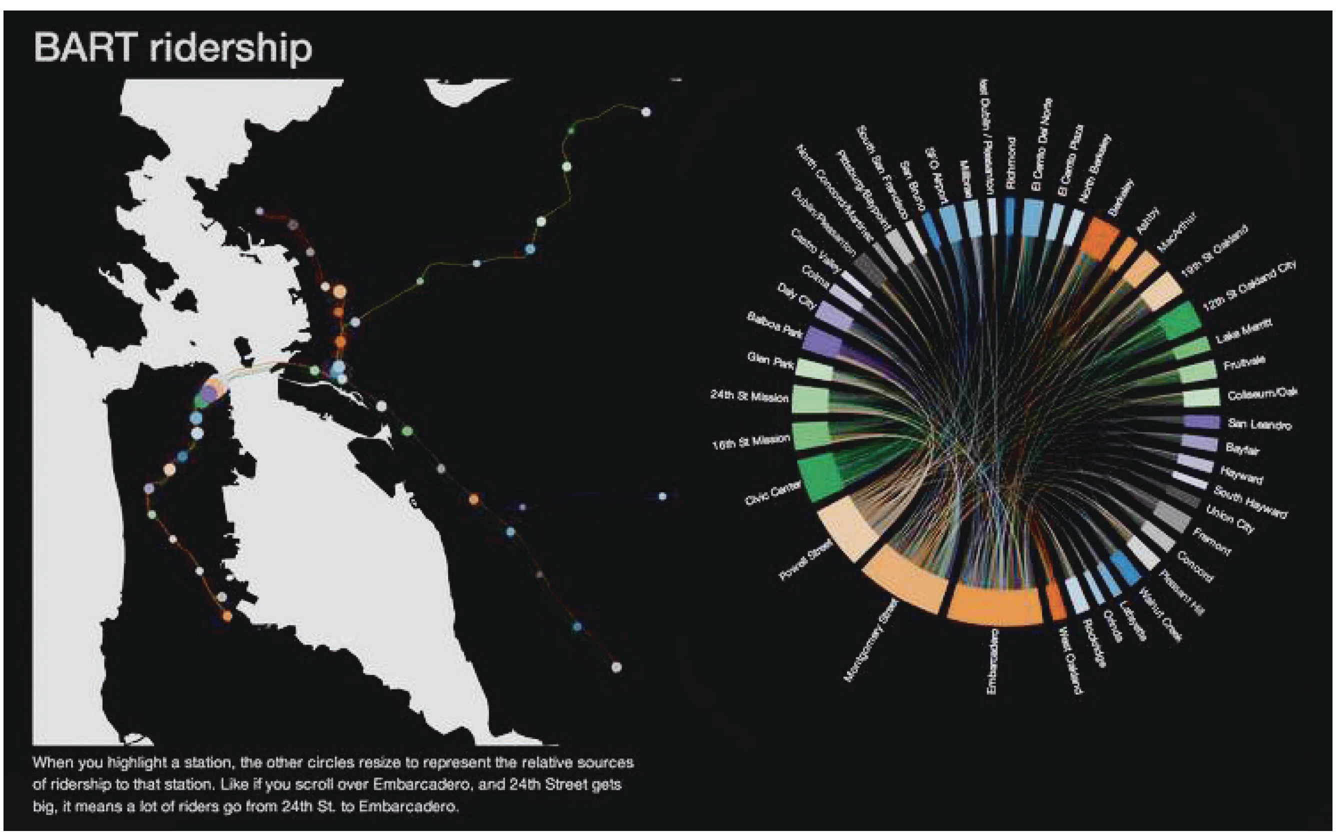


FIGURE 51 BART ridership visualization (http://enjalot.github.io/bart/#chapter-05).

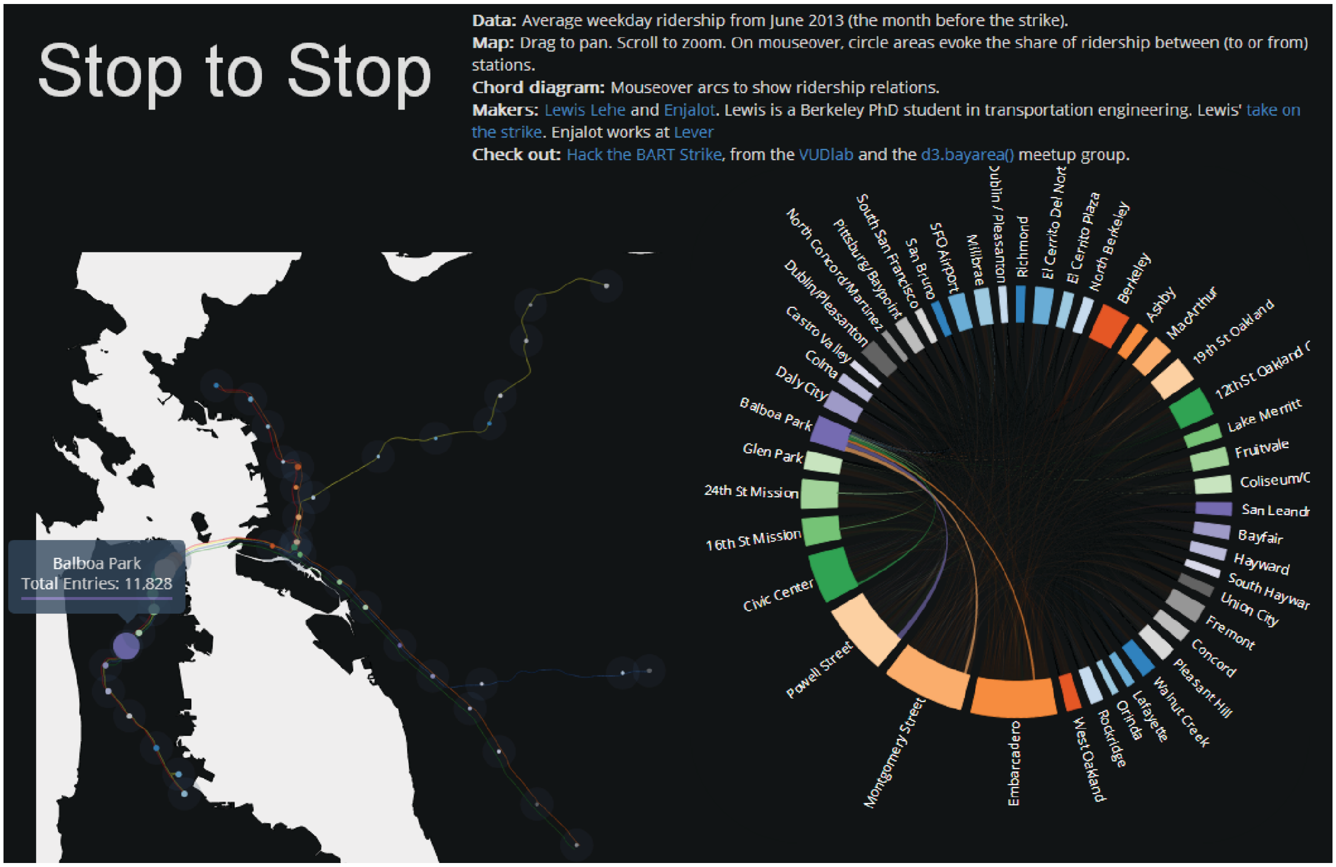


FIGURE 52 BART ridership to/from Balboa Park Station (http://vudlab.com/bart/).

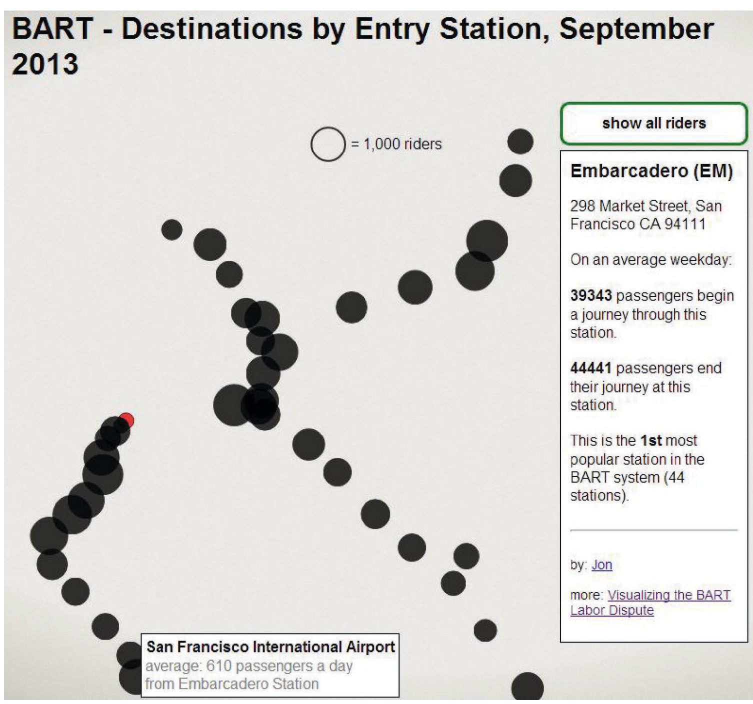


FIGURE 53 BART ridership by station (http://enjalot.github.io/bart/#chapter-05).

Chicago Transit Authority (CTA) to build apps for Android and iOS platforms because CTA has the same technology vendor. This pursuit consisted of the following steps:

- Conducting a survey of what CTA did (the agency has the same real-time system);
- Down-selecting and contacting the top five developers for Android and top five for Apple;
- Because no funding was available, eliminating from discussion developers who requested funding to move forward were eliminated from discussion;
- Finding that several developers were willing to extend their product (developed for the CTA) to the WRTA (in addition to making changes for CTA); and
- Continuing development of CTA apps means so that they are rolled out to WRTA afterward.

WRTA's initial research to identify developers started with a spreadsheet developed by the ITS consultant. This spreadsheet contained the following fields:

- App name
- Transit agency
- Platforms (e.g., iPhone, SMS, Android, Website)
- App cost
- Link Android (e.g., to Google Play)
- Link iTunes
- Developer name (company or individual)
- Company's contact (e.g., e-mail address, website)

Now there are two Android apps [WRTA Bustracker and Just another transit app (JATA) and one for iOS (TransitStop)].

In addition, WRTA learned from the "best practices" in open data employed by BART, MTA, and CTA. WRTA used the following techniques to develop the agency's programs:

- Use a registration process to provide a key.
- Vet a developer by trying out the app and testing it in-house before releasing it to the public, and conduct a question-and-answer session with the developer,
- Use an official promotion "seal of approval" by the agency once a good relationship develops and there is proof of diligence by the developer.
- Rely on user reviews—let the market and comments be open, shared, and promoted by the agency.
- Create a web page for developers to formalize the process (this is under development).

The WRTA is using Clever Devices BusTime Developer's API. WRTA staff think that CTA's markup of the Developer's API is an excellent product, so they are using this instead of the vendor's document (http://www.transitchicago.com/assets/1/developer_center/Bus Time_Developer_API_Guide.pdf). WRTA staff are waiting for a leader to emerge among these standards: SIRI, TCIP, NextBus, Clever Devices, and GTFS-realtime. Eventually, they would like to create a truly universal vendor-agnostic translator to make it easier for people to get the data they want. Further, WRTA staff would like to see a service that could translate between vendor APIs to GTFS-realtime.

Although the WRTA has not surveyed developers or customers about their open data, they are hoping to increase developer activity and create a rewards/incentive system for developers that will, in turn, foster increased options for consumers. In addition, they would like the riding community to promote, test, use, and comment concerning which apps are most effective. In the future, the WRTA personnel are thinking about using these techniques:

- Regular customer surveys would include a technology piece that might ask about the use of apps.
- WiFi connectivity—require that a user provide demographic information to connect for the first time on a WRTA bus. Then the WRTA would monitor this media access control address (no personal information would be collected) to determine usage and travel patterns.
- For app development, developers would be required to provide customer survey questions as part of their license.

One challenge associated with initiating and maintaining the open data program is that a system with frequent updates requires constant revision and updates to the open data. The updating process has become more stabilized over time—it is now predictable. It is challenging when schedules are volatile. The WRTA started a "developer's center," but this will be modified because it has not been published or promoted. For smaller agencies, the WRTA suggests reusing work from other developers when possible to minimize the resources required to clean the data.

Currently, the labor time needed to maintain the open data is 5 to 15 hours for Mr. Hamman and an occasional intern each time a schedule is changed. In addition to using developers to indicate if there is a problem with the data, the WRTA has Ridecheck to match with Hastus data, to match with GTFS data, to match with data from the agency's CAD/AVL vendor, so it is easy to check for issues.

In terms of best practices, the WRTA recommends providing as much information as possible in as many ways as possible while balancing the amount of effort required. With most of the agency's database update processes being automated, agency staff think that their open data practice has high quality with low overhead. Some of the open data provided by the WRTA includes:

- Branding—.css, colors, logos, etc.
- IVR prompts—every street name, intersection, stop ID, and so forth, which can be reused by developers to make their apps more accessible
- GTFS schedule data provided on the MassDOT developer's website, GTFS exchange, and other sites

- Excel timetable format for each route
- Microsoft Access table
- Images and maps
- Schedules in portable document format (pdf).

Other techniques used by the WRTA include the following:

- Using source control, such as GitHub, to provide quality assurance and control for each release.
- Modifying the website to publish to RSS, Twitter, Facebook, and Wordpress automatically. These free broadcast media expand the reach of real-time updates to riders.
- Using developer's API.
- Implementing a do-it-yourself (DIY) kiosk program with nine community partners, including schools and social services (e.g., Quinsigamond Community College, Family Health Center, etc.). The WRTA licenses the asset (e.g., electronic sign showing both WRTA and partner's information) to each community partner. This allows partners to show their internal stakeholder information and WRTA bus times and transit-related information. Further, it is a less expensive way of getting WRTA information disseminated. So far, 15 kiosks have been deployed.
- WRTA's Open Checkbook initiative (http://www.therta.com/about/open-checkbook/).

> The Worcester Regional Transit Authority is committed to providing citizens with open and transparent government. As part of this proactive approach to civic engagement, the WRTA has developed this Open Checkbook webpage. Open Checkbook is meant to be a window into the authority and to provide the public with access to the authority's spending information. Open Checkbook will detail vendor payments, identifying who was paid and when, how much was paid, and what was the purpose of the payment.

CHAPTER EIGHT

CONCLUSIONS

SUMMARY OF PROJECT SCOPE

The primary purpose of this synthesis is to determine transit's experience with open data, how agencies have opened their data, and the uses of the data. A survey was used to collect key information about open transit data and was sent to 67 transit agencies around the world. There was a 100% response rate. Of the 67 surveys received, three were from Canadian agencies and 14 from European agencies.

The project examined and documented the state of the practice in open data using the following five elements:

- Characteristics of open transit data
  - Reasons for choosing to provide open data
  - Standards and protocols for providing open data
  - Underlying technology used to generate the open data
- Legal and licensing issues and practices
  - Legal and licensing issues
  - Public disclosure practices
- Uses of open data
  - Applications
  - Decision-support tools
  - Visualizations
- Costs and benefits of providing open data
- Opportunities and challenges
  - Techniques for engaging users and reusers of data
  - Challenges associated with providing open data
  - Impacts on transit agencies and the public and private sectors

The project was conducted in four major steps as follows:

- Literature review;
- Survey to collect information on a variety of factors;
- Analysis of survey results; and
- Interviews conducted with key personnel at agencies that have experience with open transit data.

This section of the report contains the project's findings, lessons learned, and conclusions.

PROJECT FINDINGS

Key statistics from the study are as follows:

- Fifty-seven or almost 83% of the survey respondents provide open data; and
- The top four reasons for not providing open data are:
  - Too much effort to produce the data/not enough time or people to do the work required;
  - Too much effort to clean the data;
  - Concern that the agency cannot control what someone will do with the data; and
  - Concern regarding the accuracy of the data.

KEY FINDINGS

In summary, based on the literature review, the responses to the questionnaire, and the case examples, there are four key findings of this synthesis project:

1. Although the costs of providing open data are not well understood, the benefits to the agency, public, and community strongly support open transit data. The availability of open transit data encourages innovation that could not be accomplished solely by agency staff. The rapid creation of new mobile and Internet platforms, requiring new information technology (IT) development, would create a strain on typically limited agency resources. By focusing the limited resources on providing accurate, reliable, and timely open data, an agency can cost-effectively provide its information to the public, relying on third parties (e.g., application developers) to create customer applications and conduct data analyses.

   The overall benefits experienced by survey respondents included the following:
   - Increased awareness of the agency's services;
   - Empowerment of customers;
   - Encouragement of innovation outside of the agency;
   - Improvement in the perception of the agency (e.g., openness/transparency);
   - Provision of opportunities for private businesses;
   - Encouragement of innovation internally;
   - Improvement in our market reach;
   - Greater efficiency and effectiveness as an agency;
   - Increased return on investment (ROI) from existing web services;
   - Cost savings; and
   - Ability to reassign staff.

   The Moscow case study shows the power of open data to change decision processes.

2. Engaging application developers, other data users, and customers is an approach that can accomplish several critical tasks, including:
   - Obtaining feedback on data anomalies and data quality issues;
   - Ensuring that some portion of the applications developed by third parties meet the needs of customers; and
   - Learning more about how people want to use/reuse agency data.

   There are several ways to engage developers and customers. Results of the survey indicated that the most effective methods are conducting face-to-face events, conferences, and "meetups." Meetups are informal meetings to discuss particular topics, such as app development. For example, Mobility Lab in Arlington, Virginia, hosts meetups to discuss transportation issues and support programmers who are interested in transit; biking and walking; and open data, data visualization, and mapping.
3. The results of the literature review and survey indicate that standards and commonly used formats are to be used to facilitate the generation and use of the open data. The literature discusses how standards are used to generate the open data, such as the case of many scheduling software packages providing schedule data in General Transit Feed Specifications (GTFS) format. Further, using standards makes it easier to transfer applications from one agency to another, which was the case when Worcester Regional Transit Authority (WRTA) was looking for applications; it was easy to take applications that were developed for Chicago Transit Authority (CTA) and adapt them for WRTA because of the standards used. Without standards, planning and operations analyses, such as those described by Wong (*19*) and Catalá et al. (*75*), could not be accomplished easily.
4. Opening transit data results in innovation that could not be accomplished within a transit agency. That is not to say that the intellect does not exist in a transit agency—it is an issue of having sufficient resources to develop applications and conduct analyses at the scale that can be done in an open market. Stephen Goldsmith, in the article "Open Data's Road to Better Transit" (*102*), mentions that "some members of the American Public Transportation Association believe that open data initiatives have catalyzed more innovation throughout the industry than any other factor in the last three decades."

FINDINGS BASED ON FIVE ELEMENTS

Specific findings based on the aforementioned elements are as follows.

Characteristics of Open Transit Data

As expected, the top three types of open data are routes, schedules, and station/stop locations. This result correlates directly with the high use of GTFS, which requires these data elements, among others. Further, these data types are required to perform trip planning, which is the subject of many customer applications developed using open transit data. The next most common type of open data are real-time information, which is provided by more than half of the survey respondents. This corresponds to the use of either GTFS-realtime or Service Interface for Real Time Information (SIRI) by almost half of the survey respondents.

The most prevalent underlying technologies that produce open data are scheduling software, geographic information system (GIS) software, computer-aided dispatch (CAD)/automatic vehicle location (AVL), and real-time arrival prediction software. This finding is expected, given the types of open data reported by survey respondents.

The overwhelming reasons for opening transit data are related to customer information—increasing access to this information, and improving the information and customer service. This result corresponds with almost all of the survey respondents indicating that providing open data is a way to maintain or increase ridership. Improving perception of the transit system and fostering innovation were the next most frequently reported reasons for opening data.

The factors that went into the decision about what data to open were driven primarily by the ease of releasing the data (more than half of the survey respondents indicated this). The next two most prevalent decision factors were observing what other transit agencies have done regarding open data and deciding internally without asking any groups outside the agency.

A variety of standards and formats are being used, including GTFS (47 or 83.9% of respondents), Extensible Markup Language (XML) (26 or 46.4%), and comma-separated values (.csv) (18 or 32.1%), followed by GTFS-realtime (15 or 26.8%). The degree of openness in the four categories mentioned is as follows:

- Thirty-two or 57.4% of the respondents reported that the data are completely open (everyone has access).
- Forty-seven or 83.6% reported that the data are available in formats that are easily retrieved and processed.
- Forty-nine or 87.3% reported that there is no cost for the open data.
- Forty-three or 79.2% reported that there are unlimited rights to use, reuse, and redistribute data.

Legal and Licensing Issues and Practices

Twenty-nine or 50.9% of the survey respondents reported that their agency requires a license or agreement to use the open data. The top three elements that license agreements cover are the right to use the agency's data; nonguarantee of data availability, accuracy, or timeliness; and liability limitations for missing or incorrect data. Almost 60% (16 responses of 27) of respondents indicate their agency requires acknowledgment of a license agreement before data can be accessed. Only one respondent reported agency legal issues resulting from the release of open data to the public.

According to the respondents, the top three steps that agencies took to publicly disclose data are to (1) convert transit data into formats suitable for public use; (2) improve data quality to ensure accuracy and reliability; and (3) adopt an open, nonproprietary data standard.

Uses of Open Data

The top five types of customer applications that have been developed as a result of providing open data are (in descending order of frequency) trip planning, mobile applications, real-time transit information (arrival/departure times, delays, detours), maps and data visualization. The top five decision-support tools that have been developed are data visualization, service planning and evaluation, route layout and design, performance analysis, and travel time and capacity analysis.

Almost two-thirds (33 or 63.5%) of respondents reported their agencies do not track usage of open data. The two most prevalent methods of tracking are to monitor data downloads and to keep track of applications developed. For mobile applications, an equal number of respondents reported Android and iOS applications. Sixteen respondents reported a total of almost 266 million API calls per month.

Costs and Benefits of Providing Open Data

The top five types of costs associated with providing open data are staff time to update, fix, and maintain data as needed; internal staff time to convert data to an open format; staff time needed to validate and monitor the data for accuracy; staff time to liaise with data users/developers; and web service for hosting data. Almost 90% (43 or 89.4%) of respondents could not quantify how much time is spent on any of these activities.

Although activities required to provide open data were identified by respondents, resource requirements varied widely. There was limited information regarding the actual labor required from specific staff in the organization or the costs associated with open data.

Finally, the top three benefits experienced by survey respondents are (1) increased awareness of their services, (2) empowerment of customers, and (3) encouragement of innovation outside of the agency. Almost 70% (33 or 69.6%) of the respondents engage or have a dialogue with existing and potential data users and reusers. Twenty-five or 75.8% of the respondents engage data users and reusers to obtain feedback on data anomalies and data quality issues. Twenty-four or 60% of the respondents use face-to-face events to engage these groups.

Opportunities and Challenges

In terms of impacts on the agency and the public and private sectors, the majority of impacts reported by respondents were positive. The organizational impacts on the agency ranged from increased transparency to better and more accurate internal data to lower costs to provide information. Impacts on the customer were numerous, including better and more accessible information for customers; better perception, visibility, and awareness of services; and improved customer satisfaction. The majority of negative impacts were related to resources needed to maintain an open data program.

In terms of impacts on the public, creating and improving access to additional and higher quality public services was mentioned, along with improving public perception/image of transit, making transit more competitive, providing better regional coordination of services, encouraging innovation, and providing a better transit experience.

The impacts on the private sector are primarily providing business/commercial and development opportunities, including new and expanded companies (e.g., creating a new ecosystem of private entrepreneurs), enabling innovation and the creation of applications that may not have been created by the public sector, and adding value to existing public services.

Challenges were noted by survey respondents in five areas, as follows:

- Resources and organizational issues, which largely consist of limited resources and securing support for an open data program;
- Data quality and timeliness issues, which largely describe having to ensure data quality, completeness, timeliness, accuracy, and equity;
- Standards and formatting issues;
- Marketing issues relating to making the open data known and addressing branding issues;
- Technical issues, which consist of tracking users, including who has built what apps and how successful they have been; complying with developers' wishes; how to provide large amounts of data in a timely manner; and having a process in place to make the data available when new schedules are released.

LESSONS LEARNED

Respondents cite the following lessons learned:

- Data quality and accuracy
  – Put quality checks in place when opening data
  – Test the data before releasing it to developers

  - Start small in terms of the amount of open data offered and then grow that when confident of data quality of new sources/data sets
  - Ensure data are compatible with or identical among the different formats in which they are made available
- Cost of open data
  - Staff to support an open data program is needed to implement such a program
  - Use standards to make it easier to provide open data
  - Select a technology vendor that supports open data or require it in the contract with the vendor
- Organizational and institutional effects, including changes within and external to the agency
  - Agencies have to get comfortable with providing data when they are accustomed to providing only transit service.
  - Open data will not solve every customer requirement.
  - Customers will recognize which third-party services are most effective, and they will not hold the agency responsible for poor third-party services.
  - It is important that agencies not interfere with the market to ensure that the benefits of competition can be realized.
  - An open data program should be supported by a project champion.
  - Staff roles must be carefully assigned.
  - Buy-in from coordinating agencies is crucial.
  - Open data are a fundamental part of an overall information system.
  - Agencies must ensure that data reuse complies with public policies.
- Developing relationships with and engaging data users and reusers:
  - Early engagement with potential users is key. Find out what they want and how they want it. Try and track who is developing what, particularly to understand the successes and failures.
  - Respond quickly to opportunities. Developers work on much shorter schedules than planners.
  - Make it as easy as possible for developers to access data, and make the license understandable.
  - Developers will help determine the quality of the data if the agency provides a forum for this type of feedback.
  - Developers will know the latest mobile platforms and can use these with the data.

CONCLUSIONS

Several conclusions can be drawn from the results of the synthesis.

- The benefits of providing open transit data far outweigh the costs. Benefits accruing to the agency itself, customers, and the public and private sectors are far-reaching. Several of the survey respondents discussed using open data as a way to improve their agency's ability to conduct analyses internally and the perception of public transit within the community. In addition, the agency's transparency as a result of open data has had more positive than negative effects. In a time when public agencies are being scrutinized more than ever, providing data about their operations and internal processes reflects overcoming the old thinking that data should not be released beyond providing paper schedules.
  - The impacts of open transit data on customers and the general public are significant. Now customers (and those who have not yet taken transit) have free tools that essentially break down the barriers to using transit, such as interpreting a paper schedule or map. Further, real-time information makes it easier to plan trips. In addition, the tools resulting from open data satisfy the desire many people have for obtaining travel information almost instantaneously. However, one important factor in assessing the customer and public impact is ensuring that the tools being developed actually fulfill the customers' needs.
  - The impacts on the private sector have been encouraging over the past several years. Applications and visualizations that perhaps could not have been conceived or developed by a transit agency have been created. These apps have changed the nature of travel, where in some cases, the public transit option is more prominent and understood. Further, this has resulted in businesses being established that may have not existed if not for open transit data. Finally, developers are creating innovative ways in which to analyze the data, resulting in potential improvements in service.
- The legal fears often thought to be barriers to opening transit data have not been realized. The survey results show that only one agency responding to the survey experienced any legal issues resulting from the release of open data to the public. The literature, survey results, and case examples indicate that simple agreements with data users and reusers can accomplish what is needed to ensure proper use and distribution of the data, along with rules regarding the use of logos, images, and so forth. In addition, as stated in the survey responses, having a plan in place to handle irresponsible users is critical. For example, several techniques for managing irresponsible users included contacting developers and discussing the problem, and limiting or terminating access to the data.
- Standards greatly facilitate the use of open transit data. Although this sometimes requires additional effort in producing the open data, it makes it much easier for the data to be used. Clearly, from the literature review and survey, GTFS has become a de-facto standard, with at least 726 agencies using it. Further, it is being used, as reported, in a number of agencies that are just beginning to open their data, particularly those that have provided only paper-based data (e.g., schedules) for fixed-route services (and in some case, no information at all about other services). In addition, the use of standards has facilitated traditional planning and analysis

of transit data, as reported extensively in the literature. Further, even vendors with proprietary products have developed "translators" that reformat the data within their software to one of the standard formats. Finally, standards are still evolving. Open standards, such as GTFS, OpenStreetMap, and OpenTripPlanner, have led the way in the transit industry and are being used extensively to create new applications.

- Engaging with data users and reusers has the potential to increase the value of the applications and visualizations that are developed. Engaging with developers and the public will ensure that developers are taking customers' needs into consideration. Further, there are many different ways to engage users. In addition, the survey responses indicate that methods of engagement might be based on the sophistication of the agency in terms of open data.
- Several factors lead to a successful open data program.
  - Obtaining and maintaining management-level support for such a program and avoiding bureaucratic delays. This factor speaks to embracing transparency, realizing that transit will be more visible in the community and that there is the potential to improve the perception of transit as a result of providing open data.
  - Recognizing the need for the appropriate level of resources needed to provide and maintain open data.
  - Establishing ways to monitor data accuracy, timeliness, reliability, quality, usage, and maintenance is a key component of an open data program. Making a decision as to whether each application based on the open data will be tested is part of this factor. Some agencies let the market decide if an application is good or not, and others test each application.
  - Creating and maintaining licensing or registration to ensure that if a data user or reuser is misusing the data, action can be taken with minimal effort. As suggested by Bay Area Rapid Transit (BART) and Massachusetts Department of Transportation (MassDOT), a license or registration should be simple, conveying the basic principles associated with using the data.
  - Having an ongoing dialogue with developers and customers regarding the open data program has been shown to increase the value of the data and products that are based on the data.

SUGGESTIONS FOR FUTURE STUDY

Based on the survey results and literature review, the following areas are suggested for future study:

- *Using open data to support performance measurements.* Although the literature covers visualizations that examine transit performance, there is no guidance for transit agencies regarding effectively using open data to perform these types of analyses.
- *Amount of staff time that is required.* This will vary by the volume of data opened and depends on the system architecture. A study that examines just how much time is required by various departments and staff and a discussion of the actual costs associated with an open data program could be helpful. For example, if a tool is used to export the data from a scheduling system, this one-time investment facilitates reuse afterward (effectively lowering ongoing costs). Such study should include examining how much time and resources are spent on public records requests, which sometimes are made when open data are not available.
- *Guidance describing each step in setting up an open data program.* Such a document would contain sections relating to the factors mentioned in the final conclusion and detail about the process that many agencies use to issue open data when there is a change to their data (e.g., when there is a schedule change) and when they provide new data elements. In addition, this guidance could contain items describing various types of engagements with data users and reusers.
- *Guidance to use GTFS to depict nonfixed-route transit services.* Although work is being conducted by organizations, such as The World Bank and Ride Connection in Portland, Oregon, more guidance would be helpful in this area, given the large number of demand response systems in the United States.
- *Guidance to create accessible applications.* Guidance for application developers could include the key elements of an accessible application. Because this issue has not been directly addressed by the Americans with Disabilities Act (ADA), it is a topic that could be helpful to developers as they are conceptualizing their products.
- *Open fare system data.* Most open transit data are operational in nature, but such data do not always contain fare information because of the sensitive nature of the data. With the proliferation of electronic fare collection, particularly smartcard and mobile fare payments, developing and using open fare data is of interest to transit agencies.
- *Changing the corporate culture.* There is a gap in knowledge regarding how to show that open transit data have value to the corporate culture within the transit industry. Further, describing practices that encourage the development and dissemination of open transit data, rather than hindering them (e.g., bureaucratic processes), would be helpful.
- *Open transit data from the developers' perspective.* This report is written primarily from the transit industry experience with open transit data. More information is needed from the developers to better understand their needs and concerns.
- *Policies on app centers.* Transit agencies could use guidance on the policies regarding making applications available on their websites; this is another gap in the literature. It would be helpful if this guidance included information about how to determine which apps are included in an app center, disclaimers to consider, and so forth.

- *Visualization and other open transit data tools.* There is an evolution of tools that use open data to visualize important aspects of transit operations (e.g., performance), but there is limited information about these tools.
- *Open transit data ROI.* Although the literature contains information discussing the value of open transit data, there is a gap in describing how to calculate ROI. If open transit data have value, it is helpful to quantify it to factor into decision making about which data sets to open and how to open them. For example, the Open Government Portfolio Public Value Assessment Tool (PVAT), described earlier in the report, can help agencies determine the public value of their open government initiatives.
- *How to use metadata.* Metadata, which describes the characteristics of data, is a critical part of any data set, including open transit data. The use of metadata in open transit data is not covered in the current literature. For example, Metropolitan Transportation Authority (MTA) in New York is conducting a travel pattern project in which mobile phone metadata are being analyzed to understand trip flows.
- *Crowdsourcing to combine open transit data with other data and open source software evolution.* This is another gap in the current literature. For example, as mentioned in the report, TriMet created a new open source, multimodal trip planner (OpenTripPlanner) that uses OpenStreetMap (OSM), a crowdsource open data set designed for routing. Guidance to agencies regarding the use of crowdsourced data/information and open source software would be extremely valuable as it would help agencies move away from proprietary solutions.
- *Open transit data as an element of enterprise architecture.* Information on this topic is lacking in the literature and would be most helpful to transit agencies for accommodating open data as they build or rebuild their IT infrastructure. This includes identifying automation (e.g., automated generation of open data from scheduling software), relational database management systems and use of a cloud-based IT framework to facilitate inclusion of open transit data.
- *Procurement processes that support open transit data and open source software.* The literature and transit industry practices are lacking regarding how to procure solutions that support open data and open source software. Guidance for agencies to procure "best value" solutions that support open data would be helpful to move the industry away from proprietary solutions.

ABBREVIATIONS AND ACRONYMS

ADA	Americans with Disabilities Act
APC	Automatic passenger counter
API	Application programming interface
AVL	Automated vehicle location
BART	Bay Area Rapid Transit
CAD	Computer-aided dispatch
CCTV	Closed-circuit television
CEN	European Committee for Standardization
CSV	Comma-separated values
CTA	Chicago Transit Authority
CUMTD	Champaign-Urbana Mass Transit District
DMS	Dynamic message sign
DVRPC	Delaware Valley Regional Planning Commission
GIS	Geographic information systems
GPS	Global positioning system
GTFS	General Transit Feed Specification
GTFS-realtime	General Transit Feed Specification-realtime
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
IFOPT	Identification of Fixed Objects in Public Transport
IP	Internet Protocol
IT	Information technology
ITS	Intelligent transportation systems
ITSA	Intelligent Transportation Society of America
IVR	Interactive voice response
JSON	JavaScript Object Notation
KCATA	Kansas City Area Transportation Authority
KML	Keyhole Markup Language
KSF	Key success factors
LA	Local authority
MassDOT	Massachusetts Department of Transportation
MBTA	Massachusetts Bay Transportation Authority
MTA	Metropolitan Transportation Authority (New York City)
NeTEx	Network Exchange
NYCT	New York City Transit
NTCIP	National Transportation Communications for ITS Protocol
ODI	Open Data Institute
OS	Operating systems
PCB	Professional capacity building
PT	Public transport
QR	Quick response
RATP	Régie Autonome des Transports Parisiens
RDF	Resource Description Framework
REST	Representational state transfer
RFID	Radio-frequency identification
RIAS	Remote Infrared Audible Signage
RSS	Really Simple Syndication
RTC	Regional Transportation Commission of Washoe County
RTIG	Real Time Information Group
RTPI	Real-time Passenger Information
SEPTA	Southeastern Pennsylvania Transportation Authority
SIRI	Service Interface for Real Time Information
SMS	Short Message Service

SOAP	Simple Object Access protocol
TAD	Travel Assistant Device
TCIP	Transit Communications Interface Profiles
TfL	Transport for London
TFT	Thin-film transistor
TIDE	Transport Innovation Deployment for Europe
TIL	Traveline Information Limited
TRID	Transport Research International Documentation
TriMet	Tri-County Metropolitan Transportation District of Oregon
TxC	TransXChange
U.K.	United Kingdom
WAP	Wireless Application Protocol
WMATA	Washington Metropolitan Area Transit Authority
W3C	World Wide Web Consortium
XML	Extensible Markup Language

REFERENCES

1. Schweiger, C.L., *TCRP Synthesis 91: Use and Deployment of Mobile Device Technology for Real-Time Transit Information*, Transportation Research Board of the National Academies, Washington, D.C., 2011 [Online]. Available: http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_syn_91.pdf.
2. "What Makes Data," Open Data Institute, London, U.K. [Online]. Available: http://theodi.org/guides/what-open-data [accessed Mar. 11, 2014].
3. Peterson, A., "Access to Data and Information—A Business Between Stakeholders and a Democratic Issue," *Proceedings of 2012 ITS World Congress*, Paper No. EU-00196, Vienna, Austria, Oct. 14–18, 2012, p. 3.
4. Orszag, P.R., "Memorandum for the Heads of Executive Departments and Agencies: Open Government Directive," Number M10-06, Dec. 8, 2009 [Online]. Available: http://www.whitehouse.gov/open/documents/open-government-directive.
5. de Vreeze, M., "The Future for Open Data," presented at the 20th ITS World Congress, Tokyo, Japan, Oct. 14–18, 2013.
6. Schwegmann, C., "Open Data in Developing Countries," Topic Report No. 2013/02, European Public Sector Information Platform, Feb. 2012 [Online]. Available: http://www.epsiplatform.eu/sites/default/files/127790068-Topic-Report-Open-Data-in-Developing-Countries.pdf.
7. "Open Government Plan—Chapter 3: DOT's Open Government Initiatives and Activities for 2012–2014" [Online]. Available: http://www.dot.gov/open/plan-chapter3 [accessed Dec. 27, 2013].
8. Eros, E., S. Mehndiratta, C. Zegras, K. Webb, and M. Catalina Ochoa, "Applying the General Transit Feed Specification (GTFS) to the Global South: Experiences in Mexico City and Beyond," prepared for the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2014, p. 2.
9. World Bank Open Transport Team, "An Overview of Open Transport in East and Southeast Asia," brochure published by The World Bank, 2 p. [Online]. Available: http://siteresources.worldbank.org/INTURBANTRANSPORT/Resources/340136-1395424136020/EAP-Open-Transport-Services.pdf.
10. "The Open Data Movement" [Online]. Available: http://visual.ly/open-data-movement [accessed Sep. 26, 2011].
11. Rojas, F.M., "Transit Transparency: Effective Disclosure through Open Data," Transparency Policy Project, Ash Center for Democratic Governance and Innovation, Taubman Center for State and Local Government, Harvard Kennedy School, Cambridge, Mass., June 2012, 9 pp. [Online]. Available: http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8%2028%202012.pdf.
12. Beasley, S., "Transport Innovation Deployment for Europe [TIDE]," Reading Borough Council, presented at the 2013 Annual Polis Conference, Brussels, Belgium, Dec. 4–5, 2013.
13. Heaton, B., "Open Government Initiatives Evaluated by New Assessment Tool," *Government Technology*, May 13, 2011 [Online]. Available: http://www.govtech.com/policy-management/Open-Government-Initiatives-Assessment-Tool.html.
14. Watkins, K., "Open Transit Data: State of the Practice," presented at the 2013 ITS World Congress, Tokyo, Japan, Oct. 14–18, 2013, pp. 17–18.
15. Young, P., "TfL Open Data, Jan '13," Head of TfL [Online]. Available: http://www.rtig.org.uk/web/portals/0/PhilYoung_Jan13.pdf.
16. McHugh, T., "Opening TriMet," presented at the APTA 2012, CTO, p. 3 [Online]. Available: http://www.apta.com/mc/annual/previous/2012/presentations/Presentations/McHughT-Opening-TriMet.pdf.
17. "GTFS Data Exchange: Transit Agencies Providing GTFS Data" [Online]. Available: http://www.gtfs-data-exchange.com/agencies#filter_official [accessed Feb. 23, 2014].
18. "All US Transit Agencies with Open Data," and "All USA Transit Agencies with Open Data," City-Go-Round [Online]. Available: http://www.citygoround.org/agencies/us/?public=all [accessed Mar. 29, 2014].
19. Wong, J., "Leveraging the General Transit Feed Specification (GTFS) for Efficient Transit Analysis," presented at the 92nd Annual Meeting of the Transportation Research Board, Jan. 2013, p. 10–13.
20. "Global Open Data Index: Survey," Open Knowledge Foundation, Cambridge, U.K. [Online]. Available: https://index.okfn.org/ [accessed Mar. 30, 2014].
21. "Timetables," Open Knowledge Foundation, Cambridge, U.K. [Online]. Available: https://index.okfn.org/country/dataset/timetables [accessed Mar. 30, 2014].
22. "Transit," Open Knowledge Foundation, Cambridge, U.K. [Online]. Available: http://us-city.census.okfn.org/dataset/transit [accessed June 15, 2014].
23. Kaufman, S.M., "Getting Started with Open Data: A Guide for Transportation Agencies," Rudin Center for Transportation Policy and Management, Robert F. Wagner Graduate School of Public Service, New York University, May 1, 2012, p. 6 [Online]. Available: http://wagner.nyu.edu/files/rudincenter/opendata.pdf.
24. Lee, Y., "Transit Open Data & Transit Information Apps," presented at 20th ITS World Congress, Tokyo, Japan, Oct. 14–18, 2013.
25. Barbeau S.J., "Open Transit Data—A Developer's Perspective," May 2, 2013 [Online]. Available: http://www.cutr.usf.edu/wp-content/uploads/2013/05/CUTR-Webcast-Handout-05.02.13.pdf.

26. Manyika, J., M. Chui, P. Groves, D. Farrell, S. Van Kuiken, and E. Almasi Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information," McKinsey Global Institute, McKinsey Center for Government and McKinsey Business Technology Office, Oct. 2013, p. 3 [Online]. Available: http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information.
27. "Applying the 5 Star Open Data Model to Your High Value Public Data," ICT.govt.nz [Online]. Available: http://ict.govt.nz/programmes/open-and-transparent-government/toolkit-agencies/applying-5-star-open-data-model-your-high-value-pu/ [accessed Feb. 23, 2014].
28. Moore, T., "The Open Data Ecosystem," presented in "Learn from the Experts: Open Data Policy Guidelines for Transit—Maximizing Real Time and Schedule Data Use and Investments," Talking Transportation Technology (T3) Webinar, ITS Professional Capacity Building (PCB) Program, Dec. 5, 2013 [Online]. Available: http://www.pcb.its.dot.gov/t3/s131205/s131205_open_data_presentation_moore.asp.
29. Nobbe, H., "How Open Data Will Bring Traffic Information to the Next Level in the Netherlands," presented at the 2013 ITS World Congress, Tokyo, Japan, Oct. 14–18, 2013.
30. Lee, K., M. Bae, K. Kim, and S.R. Park, "A Study on the Connecting Methods of Traffic Information Using OPEN-API," Preprint No. IS04-11, prepared for the 2010 ITS World Congress, Busan, South Korea, Oct. 2010.
31. [Online]. Available: http://transportation-camp-dc-2013.wikispaces.com/Transit+Data+Standards [accessed Mar. 11, 2014].
32. Reed, L., Thesis, Georgia Institute of Technology, Dec. 2013, Copyright © Landon Turner Reed 2013 [Online]. Available: https://smartech.gatech.edu/bitstream/handle/1853/50218/REED-THESIS-2013.pdf?sequence=1.
33. Wong, J., L. Reed, K. Watkins, and R. Hammond, "Open Transit Data: State of the Practice and Experiences from Participating Agencies in the United States," Preprint No. 13-0186, presented at the 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2013.
34. "Comma Separated Values in Technology," Dictionary.com [Online]. Available: http://dictionary.reference.com/browse/comma+separated+values [accessed June 15, 2014].
35. "National Public Transport Access Nodes," GOV.UK [Online]. Available: http://www.dft.gov.uk/naptan/ifopt/ [accessed Mar. 11, 2014].
36. "Network Exchange, Department of Transport [Online]. Available: http://user47094.vs.easily.co.uk/netex/ [accessed Mar. 11, 2014].
37. "What is GTFS-Realtime," Google Developers [Online]. Available: https://developers.google.com/transit/gtfs-realtime/ [accessed June 15, 2014].
38. "Resource Description Framework," W3C Semantic Web [Online]. Available: http://www.w3.org/RDF/ [accessed June 15, 2014].
39. "Representational State Transfer," Technopedia [Online]. Available: http://www.techopedia.com/definition/1312/representational-state-transfer-rest [accessed June 15, 2014].
40. "What Is RSS: RSS Explained" [Online]. Available: http://www.whatisrss.com/ [accessed June 15, 2014].
41. "Soap," TechTerms.com [Online]. Available: http://www.techterms.com/definition/soap [accessed June 15, 2014].
42. Transmodel [Online]. Available: http://www.dft.gov.uk/transmodel/ [accessed Mar. 11, 2014].
43. "TransXChange," GOV.UK [Online]. Available: https://www.gov.uk/government/collections/transxchange [accessed Mar. 11, 2014].
44. "TransXChange Overview," GOV.UK [Online]. Available: https://www.gov.uk/government/publications/transxchange-overview [accessed Mar. 11, 2014].
45. Wong, J., "The OpenPlans Guide to GTFS Data," Aug. 2, 2012 [Online]. Available: http://openplans.org/2012/08/the-openplans-guide-to-gtfs-data/.
46. *Transit IDEA Project 58: Google Transit Data Tool for Small Transit Agencies*, Transportation Research Board of the National Academies, Washington, D.C., Aug. 2011 [Online]. Available: http://onlinepubs.trb.org/Onlinepubs/IDEA/FinalReports/Transit/Transit58_Final_Report.pdf.
47. McHugh, B., "Beyond Transparency: Open Data and the Future of Civic Innovation," In Chapter 10, *Pioneering Open Data Standards: The GTFS Story*, B. Goldstein and L. Dyson, Eds., Code for America Press, San Francisco, Calif., © Code for America, 2013.
48. "General Transit Feed Specification Reference," Google Developers [Online]. Available: https://developers.google.com/transit/gtfs/reference [accessed Mar. 10, 2014].
49. GTFS-realtime Reference, Google Developers [Online]. Available: https://developers.google.com/transit/gtfs-realtime/reference [accessed Mar. 10, 2014].
50. *SIRI Handbook & Functional Service Diagrams*, Version 0.13, 2008/01/10, Njsk Kizoom, DRAFT, ©Copyright 2007, 2008 Kizoom Limited, London, U.K. [Online]. Available: http://user47094.vs.easily.co.uk/siri/schema/1.3/doc/Handbook/Handbookv15.pdf.
51. "APTA Surveys Transit Agencies on Providing Information and Real-Time Arrivals to Customers," Policy Development and Research, American Public Transportation Association, Washington, D.C., Sep. 2013 [Online]. Available: http://www.apta.com/resources/reportsandpublications/Documents/APTA-Real-Time-Data-Survey.pdf [accessed Feb. 23, 2014].
52. "OneBusAway Application Suite," Github [Online]. Available: https://github.com/OneBusAway/onebusaway-application-modules/wiki [accessed Mar. 30, 2014].

53. Iley, J., "Moving America on Transit—Innovation in Real-time Transit Information," Center Identification Number 79017-00, Mar. 6, 2012 [Online]. Available: http://www.nctr.usf.edu/2012/03/moving-america-on-transit-innovation-in-real-time-transit-information/.
54. Hazarika, H., "Viability of Data Exchange Standards in Public Transport for the UK," Preprint No. 20108, presented at the 15th ITS World Congress, New York, N.Y., Nov. 16–20, 2008.
55. Chambers, K., "Following the Global South's Lead: Improving Flexible Transit Technology in the U.S.," *GTFS in the World*, Jan. 15, 2014.
56. Klopp, J., J. Mutua, D. Orwa, P. Waiganjo, A. White, and S. Williams, "Towards a Standard for Paratransit Data: Lessons from Developing GTFS Data for Nairobi's Matatu System," presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2014, p. 16.
57. Open Knowledge Foundation, *Open Data Handbook Documentation, Release 1.0.0*, Nov. 14, 2012, p. 15 [Online]. Available: http://opendatahandbook.org/pdf/OpenDataHandbook.pdf.
58. Gomi, Y. and R. Weiland, "An Open Platform for Telematics," *Proceedings of 11th World Congress on ITS*, Nagoya, Japan, Oct. 17–22, 2004.
59. McCann, L., "Empowering the Open Data Dialogue: Creating Resources for Talking About the Benefits of Open Data," Oct. 22, 2013 [Online]. Available: http://datasmart.ash.harvard.edu/news/article/empowering-the-open-data-dialogue-330 [accessed Dec. 27, 2013].
60. Antrim, A. and S. Barbeau, "The Many Uses of GTFS Data—Opening the Door to Transit and Multimodal Applications," presented at the ITS America Annual Meeting, Nashville, Tenn., Apr. 22–24, 2013, p. 6.
61. McCann, L., "Reasons Not to Release Data," Sunlight Foundation, Washington, D.C., Sep. 2013 [Online]. Available: http://sunlightfoundation.com/blog/2013/09/05/reasons-not-to-release-data/ [accessed Mar. 30, 2014].
62. Moore, T., "Leveraging Open Data to Reach Mobile Customers in Real Time," presented at the APTA ITS Best Practices Workshop, Oakland, Calif., Apr. 17, 2012 [Online]. Available: http://www.apta.com/mc/its/previous/2012a/presentations/Presentations/Leveraging-Open-Data-to-Reach-Mobile-Customers-in-Real-Time-Timothy-Moore.pdf.
63. Hartonen, S. and K. Hiltunen, "Finnish Transport Agency's View on Data Policies," presented at 20th ITS World Congress, Tokyo, Japan, Oct. 14–18, 2013.
64. Frumin, M., "Bus Customer Information Systems," presented to APTA TransITech, Mar. 31, 2011 [Online]. Available: http://www.apta.com/mc/fctt/previous/2011tt/schedule/Presentations/session%2011%20Frumin%20Bus%20Customer%20Information%20Systems.pdf.
65. "Bus Customer Information Systems" [Online]. Available: http://www.apta.com/mc/annual/previous/2012/presentations/Presentations/QueallyD-Bus-Customer-Information-Systems.pdf.
66. "Samtrafiken: APIs Providing a Path to Innovation in Public Transportation," Samtrafiken, London, U.K. [Online]. Available: http://apigee.com/about/customers/samtrafiken-apis-providing-path-innovation-public-transportation [accessed Nov. 26, 2013].
67. Group 7: Pessoa, Reed, Tzegaegbe, Wong, Yan, "Enabling Transit Solutions: A Case for Open Data," prepared for Georgia Institute of Technology CE 6602—Transportation Planning (Fall 2011), Dec. 14, 2011, p. 11 [Online]. Available: http://www.prism.gatech.edu/~lreed3/opendata.pdf.
68. McHugh, B., "The OpenTripPlanner Project," Final Report, Prepared for Metro 2009–2011 Regional Travel Options Grant, Aug. 31, 2011, p. 10 [Online]. Available: https://github.com/openplans/OpenTripPlanner/wiki/Reports/OTP%20Final%20Report%20-%20Metro%202009-2011%20RTO%20Grant.pdf.
69. Schweiger, C.L., *TCRP Synthesis 104: Use of Electronic Passenger Information Signage in Transit*, Transportation Research Board of the National Academies, Washington, D.C., 2013.
70. "Open Data Gives New Lease of Life to Public Travel Information Screens," first published in *ITS International*, Jan.–Feb. 2014, as "Apps Aren't the only Answer" [Online]. Available: http://www.itsinternational.com/sections/transmart/features/open-data-gives-new-lease-of-life-to-public-travel-information-screens/?utm_source=Adestra&utm_medium=email&campaign_id=536&workspace_name=ITS%20International&workspace_id=3&project_name=E-newsletters&link_url=http%3A%2F%2Fwww.itsinternational.com%2Fsections%2Ftransmart%2Ffeatures%2Fopen-data-gives-new-lease-of-life-to-public-travel-information-screens%2F&link_label=Read%20more..&campaign_name=ITS%20International%206th%20March%20eNewsletter.
71. Elepano, M., "Public Open Data Feeds," presented at the Redmon Group Inc., Alexandria, Va. [Online]. Available: http://www.vatransit.com/wp-content/uploads/2013/05/Redmon_VTA_Conference-2.pdf.
72. "Innovation Toolbox: 15 Inspiring Transport Measures That Can Change Your City," TIDE (Transport Innovation Deployment for Europe), pp. 24–25[Online]. Available: http://www.tide-innovation.eu/en/upload/Results/TIDE-InnovationToolbox-ENG-lite.pdf.
73. "Visualization," WhatIs.com [Online]. Available: http://whatis.techtarget.com/definition/visualization [accessed Apr. 1, 2014].
74. "OTP Analyst," Open Trip Planner [Online]. Available: http://www.opentripplanner.org/analyst/.
75. Catalá, M., S. Downing, and D. Hayward, "Expanding the Google Transit Feed Specification to Support Operations and Planning," prepared for Florida Department of Transportation and Research and Innovative Technology Administration, U.S. Department of Transportation, Project numbers FDOT BDK85 #977-15 and DTRT07-

G-0059, Nov. 15, 2011, p. 37 [Online]. Available: http://www.nctr.usf.edu/wp-content/uploads/2012/02/77902.pdf.

76. "Maps That Show Time," Mapumental [Online]. Available: https://mapumental.com/ [accessed Apr. 1, 2014].
77. "Accessibility of Welsh Schools by Public Transport—Visualised," Myfanwy, Oct. 2, 2013 [Online]. Available: http://www.mysociety.org/2013/10/02/accessibility-of-welsh-schools-by-public-transport-visualised/.
78. Schade, M., "Capital Bikeshare Hackers Pedal Their Wares at Mobility Lab," Dec. 11, 2013 [Online]. Available: http://mobilitylab.org/2013/12/11/capital-bikeshare-hackers-pedal-their-wares-at-mobility-lab/.
79. Hillsman, E.L. and S.J. Barbeau, *Enabling Cost-Effective Multimodal Trip Planners through Open Transit Data*, Report No. USF 21177926, May 2011 [Online]. Available: http://www.nctr.usf.edu/wp-content/uploads/2011/06/77926.pdf.
80. Badger, E., "This Map Wants to Change How You Think About Your Commute," *The Atlantic Cities*, Jan. 2014 [Online]. Available: http://www.theatlanticcities.com/commute/2014/01/map-wants-change-how-you-think-about-your-commute/8197/ [accessed Jan. 28, 2014].
81. Schade, M., "Techies Work to Merge Data From Multiple Transit Agencies," Feb. 18, 2014 [Online]. Available: http://mobilitylab.org/2014/02/18/techies-work-to-merge-data-from-multiple-transit-agencies/ [accessed Feb. 23, 2014].
82. "Transit Tech Initiative," Department of Environmental Services, Arlington, Va. [Online]. Available: http://mobilitylab.org/tech/transit-tech-initiative/ [accessed Feb. 23, 2014].
83. Raschke, K., "OneBusAway Demo App Offers Best Transit Info Yet for DC Users," Oct. 25, 2013 [Online]. Available: http://mobilitylab.org/2013/10/25/onebus away-demo-app-offers-best-transit-info-yet-for-dc users/ [accessed Dec. 27, 2013].
84. "GoogleTransitDataFeed Open Source Software Project" [Online]. Available: https://code.google.com/p/googletransitdatafeed/, site offers tools to convert to and from GTFS format [accessed Feb. 23, 2014].
85. "Google Maps: Transit: Cities Covered" [Online]. Available: https://www.google.com/landing/transit/cities/index.html.
86. Schweiger, C.L., "Electronic Signage for Public Transport: Is it History?" *Proceedings of 2013 ITS World Congress*, Session TS006: Traveler Information (1), Tokyo, Japan, Oct. 15, 2013.
87. Diez Sarasola, M. and L. Bardaji, "Interconnection Model on the Cloud as a Base for a Global Public Transport Data Interconnected System: Moveuskadi Project" *Proceedings of 2012 ITS World Congress*, Paper No. EU-00089, Vienna, Austria, 2012.
88. McGurrin, M.F. and D.J. Greczner, "Performance Metrics: Calculating Accessibility Using Open Source Software and Open Data," Preprint No. 11-0230, presented at the 90th Annual Meeting of Transportation Research Board, Washington, D.C., Jan. 2011.
89. Jariyasunant, J., E. Mai, and R. Sengupta, "Algorithm for Finding Optimal Paths in a Public Transit Network with Real-Time Data," Preprint No. 11-3791, presented at the 90th Annual Meeting of Transportation Research Board, Washington, D.C., Jan. 2011.
90. Puchalsky, C., D. Joshi, and W. Scherr, "Development of a Regional Forecasting Model Based on Google Transit Feed," Preprint No. 12-0779, presented at the 13th TRB Planning Application Conference, Reno, Nev., May 2011.
91. Lee, S.G., M. Hickman, and D. Tong, "Stop Aggregation Model (SAM): Development and Application," Preprint No. 12-1287, presented at the 91st Annual Meeting of Transportation Research Board, Washington, D.C., Jan. 2012.
92. "Highlights of Seamless Transport: Making Connections," International Transport Forum: 2012 Annual Summit, Leipzig, Germany, May 2–4 2012 [Online]. Available: http://www.internationaltransportforum.org/pub/pdf/12Highlights.pdf.
93. Tran, K., E. Hillsman, S. Barbeau and M.A. Labrador, "GO_SYNC—A Framework to Synchronize Crowd-Sourced Mapping Contributions from Online Communities and Transit Agency Bus Stop Inventories," Preprint No. IS06-1354, presented at the 2011 ITS World Congress, Orlando, Fla., Oct. 16–20, 2011.
94. Hægstad, A., "Energy Saving in NSB," n.d.
95. Maltby, P., Oct. 29, 2013 (14:42), comment on "What did open data ever do for us?" Feb. 23, 2014 [Online]. Available: http://data.gov.uk/blog/what-did-open-data-ever-do-us.
96. Thomas, L.W., *Legal Research Digest 37: Legal Arrangements for Use and Control of Real-Time Data*, Transportation Research Board of the National Academies, Washington, D.C., June 2011.
97. Williams, J., "Open Data: A Traveline Perspective—Delivering High Quality Mobility Information for the Passenger," *Traveline Information Limited*, p. 11 [Online]. Available: http://www.rtig.org.uk/web/Portals/0/JulieWilliams_Jan13.pdf.
98. Polis Traffic Efficiency & Mobility Working Group, "The Move Towards Open Data in the Local Transport Domain," *Polis Position Paper*, June 2013.
99. Lewis, K., "Open Data Presents Opportunity, Challenge for Public Transit Systems," *Passenger Transport*, June 28, 2013 [Online]. Available: http://passengertransport.apta.com/aptapt/issues/2013-06-28/29.html.
100. Barbeau, S.J., "Open Transit Data—A Developer's Perspective," May 2, 2013 [Online]. Available: http://www.apta.com/mc/fctt/presentations/Presentations/The%20Many%20Uses%20of%20GTFS%20Data%20%E2%80%93%20Opening%20the%20Door%20

to%20Transit%20and%20Multimodal%20Applications%20-%20Barbeau.pdf, and http://www.cutr.usf.edu/wp-content/uploads/2013/05/CUTR-Webcast-Handout-05.02.13.pdf.

101. "Project Open Data: Open Data Policy—Managing Information as an Asset" [Online]. Available: http://project-open-data.github.io/ [accessed Mar. 1, 2014].
102. Goldsmith, S., "Open Data's Road to Better Transit," *Government Technology*, Sep. 19, 2013 [Online]. Available: http://www.govtech.com/transportation/Open-Datas-Road-to-Better-Transit.html.
103. Raschke, K., "Building Momentum with Open Data, Open Source, and Open Architecture," *Raschke on Transport*, Nov. 16, 2013, ©2011 Kurt Raschke [Online]. Available: http://transport.kurtraschke.com/2013/11/building-momentum.
104. Fossey, B., "The Circulator Dashboard: Making Bus Performance Data Accessible, Transparent, and Effective," presented at the APTA TransITech Conference, Miami, Fla., Mar. 29–31, 2011 [Online]. Available: http://www.apta.com/mc/fctt/previous/2011tt/schedule/Presentations/session%204%20Fossey%20The%20Circulator%20Dashboard.pdf.
105. Hoadley, S., "Opening up Transport Data," presented at the Polis conference, Dec. 3–4, 2014.
106. Hartonen, S. and K. Hiltunen, "Finnish Transport Agency's View on Data Policies," presented to the 20th ITS World Congress, Tokyo, Japan, Oct. 14–18, 2013.
107. Marples, D., C. Bresser, and P. van Koningsbruggen, "Approaches to Practical and Realizable Open Mobility Information Services," *Proceedings of the 2012 ITS World Congress*, Paper No. EU-00627, Vienna, Austria, Oct. 22–26, 2012.
108. Williams, J., "Traveller Information in the UK: A Traveline Perspective," p. 13 [Online]. Available: http://www.delfi.de/sites/default/files/Traveller%20Information%20in%20the%20UK_Julie_Williams.pdf.
109. "MassDOT Developers Conference Saturday," *MassDOT Blog*, Nov. 9, 2009 [Online]. Available: http://blog.mass.gov/transportation/uncategorized/change-is-coming-to-transportation-and-we-want-you-to-be-a-part-of-it-youve-probably-heard-about-the-reform-bill-which-merg/.
110. "Developers Challenge Winners, Real-Time Bus Data," *MassDOT Blog*, Nov. 14, 2009 [Online]. Available: http://blog.mass.gov/transportation/mbta/developers-challenge-winners-realtime-bus-data/.
111. MBTA and IBI Group, "MBTA-Realtime Developer Documentation (V 1.0.4)," Nov. 5, 2013 [Online]. Available: http://realtime.mbta.com/Portal/Content/Documents/MBTA-realtime_DeveloperDocumentation_v1.0.4_2013-11-05.pdf.
112. Chow, W., D. Block-Schachter, and S. Hickey, "Impacts of Real Time Passenger Information Signs in Rail Stations at the Massachusetts Bay Transportation Authority," to TRB 93rd Annual Meeting Compendium of Papers, Paper 14-0194, Transportation Research Board of the National Academies, Washington, D.C., 2014, 18 pp.
113. Chow, W., "Impacts of Real-Time Passenger Information Signs at the MBTA," presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 13, 2014.
114. "An MBTA bus-iness day," Flowing City [Online]. Available: http://flowingcity.com/visualization/mbta-bus-speeds/.
115. Murphy, I.B., "Live London Underground Map Shows Possibilities of Free Data API," by July 9, 2012 3:02 p.m., *Data Informed* [Online]. Available: http://data-informed.com/live-london-underground-map-shows-possibilities-free-data-api/.
116. Chalabi, M., Oct. 2, 2013 (0645), comment on "Live London bus tracker," *theguardian.com* [Online]. Available: http://www.theguardian.com/news/datablog/interactive/2013/oct/02/live-london-bus-map.
117. Hargreaves, S., "TFL Bus Data Visualization," Sep. 27, 2012 [Online]. Available: https://www.youtube.com/watch?v=FKs7DND1kkg.
118. Sherry, S., "Real Beauty in Visualization," Feb. 28, 2013 [Online]. Available: http://www.bigdatarepublic.com/author.asp?section_id=2642&doc_id=259738&f_src=bigdatarepublic_sitedefault.
119. Reades, J., "Open Data Visualisation," UCL Centre for Advanced Spatial Analysis, [Online]. Available: http://www.casa.ucl.ac.uk/simulacra/resources/open_data_and_visualisation.pdf.

APPENDIX A

Survey Questionnaire

TCRP Synthesis SA-34: Open Data: Challenges and Opportunities for Transit Agencies

Respondent Information

Purpose of this Survey: Transit agencies are making schedule and real-time operational data available to the public. This "open data" provides opportunities for agencies to inform the public in a variety of ways about a transit agency's services. Also, open data are being used to create internal decision-support tools that can help to optimize operations, improve maintenance and inform capital programs/planning. However, while open data presents opportunities, it also presents challenges for agency operations and other business functions. This survey focuses on collecting information about providing open data, use of open data and the impact of open data on the public and private sector. Once the survey results are reviewed, key agencies that are providing open data will be selected for telephone interviews to gather more in-depth information. All survey responses will be confidential. The final results of the survey will be synthesized into a report that will be published by the Transportation Research Board (TRB). **Thank you for taking the time to complete this survey!**

If you have any questions, please email or call Ms. Carol Schweiger, Vice President, TranSystems Corporation at clschweiger@transystems.com or 857-453-5511.

1. Date *

2. Name and Title of Respondent: *

Transit/Public Transport/Transportation Agency Name *

Address Line 1 *

Address Line 2

City *

State/Province *

Country *

Afghanistan
Albania
Algeria
Andorra
Angola
Antigua and Barbuda
Argentina
Armenia
Australia
Austria
Azerbaijan
Bahamas, The
Bahrain
Bangladesh
Barbados
Belarus
Belgium
Belize
Benin
Bhutan
Bolivia
Bosnia and Herzegovina
Botswana
Brazil
Brunei
Bulgaria
Burkina Faso
Burma

Burundi
Cambodia
Cameroon
Canada
Cape Verde
Central African Republic
Chad
Chile
China
Colombia
Comoros
Congo, Democratic Republic of the
Congo, Republic of the
Costa Rica
Cote d'Ivoire
Croatia
Cuba
Curacao
Cyprus
Czech Republic
Vietnam
Yemen
Zambia
Zimbabwe

Phone number *

Respondent's Email Address *

Introduction

3. Has your agency provided open data? *
The term "open data" refers to data that is free for anyone to use, reuse, and redistribute with an attribution or share-alike requirement.

- Yes
- No

If agency does not provide open data

Why? *

- ☐ Too much effort to produce the data/we do not have the time or people to do the work required
- ☐ Do not see the benefit(s)
- ☐ Our data could be mis-used or misinterpreted
- ☐ There is a lack of interest internally
- ☐ It will put a strain on our systems
- ☐ Proprietary vendor contracts preclude us from sharing data with third parties
- ☐ Too much effort to clean the data
- ☐ Our agency is too small
- ☐ Our agency will be liable if erroneous data is provided to the public
- ☐ We cannot control what someone will do with our data
- ☐ It is too expensive
- ☐ We cannot legally do that
- ☐ We do not know the accuracy of our data
- ☐ There is a risk-averse culture within our agency
- ☐ Our agency does not know how to open our data
- ☐ Other (please specify) ________ *

Transit System Characteristics

4. Which modes does your agency either directly operate or operate using a contractor? Please check all that apply.

- ☐ Fixed-route bus
- ☐ Paratransit
- ☐ Heavy rail/subway
- ☐ Light rail/streetcar
- ☐ Bus rapid transit
- ☐ Commuter rail
- ☐ Ferry

- [ ] Other (please specify) ________ *
- [ ] We do not operate service - we are a Metropolitan Planning Organization, State DOT or other transportation agency that does not provide transit service

5. How many total riders did your system carry in 2012?

____

6. Do you see providing open data as a way to maintain or increase ridership?

- ( ) Yes
- ( ) No

Characteristics of Open Data

7. In what year did you start providing your data openly?

____

8. What information does your agency release as open data? Please check all that apply.

- [ ] Schedule data
- [ ] Route data
- [ ] Station/stop locations
- [ ] Fare media sales locations
- [ ] Ridership data
- [ ] Real-time information (e.g., arrival/departure times, vehicle location)
- [ ] Budgetary data
- [ ] Park-and-ride locations
- [ ] Performance data
- [ ] Other (please specify) ________ *

9. What are the reasons why your agency decided to provide open data? Please check all that apply.

- ☐ Agency transparency
- ☐ Improve upon existing customer information and customer service or create new customer information services
- ☐ Foster a more positive perception of transit/encourage more people to try public transit
- ☐ Provide ways to better understand and use transit information within our agency
- ☐ Increase information access to transit riders
- ☐ Facilitate information sharing within the agency, and with partners and customers
- ☐ Provide incentives for others to help maintain datasets, reducing the maintenance cost for the agency
- ☐ Participate in the latest trend in the transit industry
- ☐ Improve effectiveness of the agency and its services
- ☐ Increase customization for customer information
- ☐ Availability of data standard(s) for transit information (e.g., GTFS)
- ☐ Will cut costs to our agency
- ☐ Foster/encourage innovation around the agency's data or help third parties develop skills and services (e.g., with which the agency can contract)
- ☐ Help achieve other agency goals (e.g., by providing a wider audience for published information)
- ☐ Improve or provide new private products and services
- ☐ Measure the impact of transit on the community(ies) that we serve
- ☐ An information gap existed that could be bridged by better public data
- ☐ There was demand for us to open our data/we were requested to provide open data
- ☐ Other (please specify) ________ *
- ☐ Other (please specify) ________

10. Did your agency use any evaluation measures to assist you in deciding to open your data?

- ☐ No
- ☐ Yes and the first measure was ________ *
- ☐ Yes and the second measure was ________
- ☐ Yes and the third measure was ________

11. How did your agency decide which data to open? Please check all that apply.

- ☐ Asked the community in which your agency operates service
- ☐ Asked potential users of the data
- ☐ Based on the ease of releasing the data
- ☐ Asked riders
- ☐ Based on the cost associated with producing or cleaning the data
- ☐ Based on observing what other transit agencies have done regarding open data
- ☐ Decided internally without asking any groups outside our agency
- ☐ Other (please specify) ________ *

12. If your agency is an MPO, state DOT or other transportation agency, does your agency release compilations of open data (e.g., regional feed that combines feeds from regional transit agencies)?

- ○ Yes
- ○ No

13. What technologies does your agency use to produce the open data? Please check all that apply.

- ☐ Computer-aided dispatch (CAD)/automatic vehicle location (AVL)
- ☐ Scheduling software
- ☐ Geographic information system (GIS) software
- ☐ Real-time arrival prediction software
- ☐ Automatic passenger counting (APC) system

☐ Free tool (please specify)

☐ Other (e.g., outsourced to a vendor) - Please specify *

14. What standard(s) does your agency use for the open data?

☐ General Transit Feed Specification (GTFS)

☐ General Transit Feed Specification-realtime (GTFS-realtime)

☐ Service Interface for Real Time Information (SIRI)

☐ TransXchange (UK nationwide standard for exchanging bus schedules and related data)

☐ DATEX2

☐ Comma Separated Values (CSV)

☐ TXT

☐ Keyhole Markup Language (KML)

☐ Extensible Markup Language (XML)

☐ Transit Communications Interface Profiles (TCIP)

☐ Geographic information system (GIS) software (please specify)

☐ Other (please specify) *

15. How does your agency make data available? Please check all that apply.

☐ Via your agency's existing website

☐ Via a separate agency website

☐ Via a third-party site

☐ Via an ftp site

☐ Via a single regional centralized site

☐ Providing data in bulk

☐ As torrents (protocol supporting the practice of peer-to-peer file sharing that is used to distribute large amounts of data over the Internet)

☐ As an application programming interface (API)

☐ GTFS Data Exchange website

- ☐ Public Feeds Wiki page on Google Transit Data Feed Google Code project
- ☐ Other (please specify) *

16. How often does your agency update or modify the static open data (e.g., schedule data)?

- ☐ Every time there is a change in the data (e.g., schedule or fare change)
- ☐ Every week
- ☐ Every month
- ☐ Other frequency (please specify) *

17. To what degree is your agency's data open? This question deals with the degree of access. Please use the sliding scale to indicate the degree of openness.

Access to data is to a subset of individuals or organizations		Everyone has access

To what degree is your agency's data open? This question deals with machine readability. Please use the sliding scale to indicate the degree of machine readability.

Data in formats not easily retrieved and processed by computers		Available in formats that can be easily retrieved and processed by computers

To what degree is your agency's data open? This question deals with cost. Please use the sliding scale to indicate the cost of the open data.

Offered only at a significant fee — No cost to obtain

To what degree is your agency's data open? This question deals with rights. Please use the sliding scale to indicate the rights to the open data.

Re-use, republishing or distribution of data is forbidden — Unlimited rights to reuse and redistribute data

Legal and Licensing Considerations

18. Do you require a license or agreement to use your agency's open data?

- Yes
- No

19. Does your agency require another type of registration before accessing and using the data?

- Yes
- No

Legal and Licensing Considerations (continued)

20. What does your license agreement cover?

- ☐ Use and placement of copyrighted logos and images
- ☐ Right to use the agency's data
- ☐ Agency's right to alter data without notice or liability
- ☐ Non-guarantee of data availability, accuracy or timeliness
- ☐ Liability limitations for missing or incorrect data
- ☐ Indemnity from technical malfunctions due to users' use of data
- ☐ Indemnity from legal actions against data users
- ☐ Data ownership
- ☐ Licensing fees and royalties
- ☐ Limitations on use of data
- ☐ Termination
- ☐ Indemnification
- ☐ Quality control
- ☐ Other (please specify) ________ *

21. Does your agency require acknowledgement of a license agreement before accessing data?

- ○ Yes
- ○ No

22. Has your agency experienced any legal issues resulting from the release of open data to the public?

- ○ Yes
- ○ No

Legal and Licensing Considerations (continued)

23. Please briefly describe the issue and its resolution.

24. Which steps did your agency take to publicly disclose your data? Please check all that apply.

- ☐ Convert transit data into formats suitable for public use
- ☐ Improve data quality to ensure accuracy and reliability
- ☐ Build interfaces for public data access
- ☐ Showcase third-party applications
- ☐ Identify the problem to be solved with better data
- ☐ Prioritize the disclosure of data for which there is public demand
- ☐ Determine whether information intermediaries play a role in the disclosure ecosystem and support the development of that ecosystem
- ☐ Adopt an open, non-proprietary data standard
- ☐ Establish channels of communication with data-user communities
- ☐ Other (please specify) *

25. What does your agency do if you discover irresponsible users?

Uses of Open Data

26. What customer applications have been developed as a result of your agency providing open data? Please check all that apply.

- ☐ Trip planning
- ☐ Maps
- ☐ Ridesharing
- ☐ Timetable creation
- ☐ Mobile applications
- ☐ Data visualization
- ☐ Accessibility
- ☐ Interactive voice response (IVR)
- ☐ Real-time transit information (arrival/departure times, delays, detours)
- ☐ Crowdsourcing
- ☐ Other (please specify) ________ *
- ☐ Other (please specify) ________ *

27. What decision-support tools have been developed as a result of your agency providing open data? Please check all that apply.

- ☐ Demand modeling
- ☐ Data visualization
- ☐ Temporal analysis
- ☐ Performance analysis
- ☐ Service planning and evaluation
- ☐ Travel time and capacity analysis
- ☐ Financial analysis
- ☐ Cost/benefit analysis
- ☐ Spatial analysis
- ☐ Regional transit analysis
- ☐ Energy consumption
- ☐ Route layout and design

- ☐ Safety analysis
- ☐ Other (please specify) ______ *
- ☐ Other (please specify) ______ *
- ☐ Other (please specify) ______ *

28. Are you aware of other uses of your agency's open data?

- ○ Yes
- ○ No

Uses of Open Data (continued)

29. Please specify the application name and briefly describe the application

- ☐ Application 1 ______ *
- ☐ Application 2 ______
- ☐ Application 3 ______
- ☐ Application 4 ______
- ☐ Application 5 ______
- ☐ Application 6 ______

30. Does your agency track usage of the open data?

- ○ Yes
- ○ No

31. How?

- ☐ Monitor data downloads
- ☐ Keep track of applications developed

- ☐ Track number of downloads per application
- ☐ Temporal analysis
- ☐ Performance analysis
- ☐ Service planning and evaluation
- ☐ Travel time and capacity analysis
- ☐ Financial analysis
- ☐ Other (please specify) *
- ☐ Other (please specify)
- ☐ Other (please specify)

Uses of Open Data (continued)

32. Can you estimate how much data is being used or downloaded over a certain timeframe? Please provide a data volume in number of units (e.g., 100 terrabytes) and specify the timeframe within which this volume of data is being used/downloaded (e.g., by month).

33. Please indicate the number of applications you are aware of that have been developed based on your open data

- ☐ Mobile applications
- ☐ Web-based applications
- ☐ Other applications

34. Please indicate the number of mobile applications by platform that you are aware of

- ☐ Android
- ☐ Bada
- ☐ Blackberry
- ☐ iOS (Apple)
- ☐ Mobile Linux

- ☐ Nokia
- ☐ Windows Mobile
- ☐ Other platform (please specify name and quantity) *

35. Does your agency have a web location where potential application customers can review available applications

- ○ Yes
- ○ No

Uses of Open Data (continued)

36. If Yes, please indicate the web address for your list of applications

37. How many developers currently use your data?

38. How many API calls do you get currently per month?

Costs, Benefits, Challenges and Opportunities

39. What are the types of costs associated with your agency providing open data? Please check all that apply.

- ☐ Internal staff time to convert data to an open format
- ☐ Staff time to update, fix and maintain data as needed
- ☐ Staff time needed to validate and monitor the data for accuracy
- ☐ Consultant time to convert data to an open format
- ☐ Web service for hosting data

- ☐ Staff time to liaise with data users/developers
- ☐ Publicity/marketing
- ☐ Other (please specify) ______ *

40. Can you quantify how much time is spent on any of these activitites?

- ○ Yes
- ○ No

Costs, Benefits, Challenges and Opportunities (continued)

41. Please indicate the number of labor hours per month by activity.

- ☐ Internal staff time to convert data to an open format ______
- ☐ Staff time to update, fix and maintain data as needed ______
- ☐ Staff time needed to validate and monitor the data for accuracy ______
- ☐ Consultant time to convert data to an open format ______
- ☐ Web service for hosting data ______
- ☐ Staff time to liaise with data users/developers ______
- ☐ Publicity/marketing ______
- ☐ Other (please specify) ______ *

42. Can you identify the cost associated with any of these activities?

- ○ Yes
- ○ No

Costs, Benefits, Challenges and Opportunities (continued)

43. Please indicate the monthly cost by activity.

- ☐ Internal staff time to convert data to an open format
- ☐ Staff time to update, fix and maintain data as needed
- ☐ Staff time needed to validate and monitor the data for accuracy
- ☐ Consultant time to convert data to an open format
- ☐ Web service for hosting data
- ☐ Staff time to liaise with data users/developers
- ☐ Publicity/marketing
- ☐ Other (please specify) *

44. What benefits has your agency experienced as a result of providing open data?

- ☐ Experienced cost savings
- ☐ Been able to reassign staff
- ☐ Become more efficient and effective as an agency
- ☐ Increased our return-on-investment from existing web services
- ☐ Improved our market reach
- ☐ Empowered our customers
- ☐ Encouraged innovation internally
- ☐ Encouraged innovation outside of the agency
- ☐ Increased awareness of our services
- ☐ Improved the perception of our agency (e.g., openess/transparency)
- ☐ Provided opportunities for private businesses
- ☐ Other (please specify)
- ☐ Other (please specify)

45. Do you engage or have a dialogue with existing and potential data users and reusers?

- Yes
- No

Costs, Benefits, Challenges and Opportunities (continued)

46. Why do you engage these groups?

- Find out more about prospective users/reusers
- Find out more about how people want to use/reuse your data
- Enable prospective and existing users to meet each other
- Announce updates, modifications, etc.
- Obtain feedback on data anomalies and data quality issues
- Suggesting features to improve the functionality of applications
- Find out more about the demand for our data
- Enable existing and prospective users/reusers to find out more about your data
- Expose your data to a wider audience
- Provide technical support
- Solicit requests for future data
- Explain transit jargon and definitions
- Other (please specify) ____
- Other (please specify) ____

47. What techniques do you use to engage existing and potential data users and reusers? Please check all that apply.

- Face-to-face events
- Conferences
- Unconferences/BarCamps (conferences with no set agenda - the agenda is set at the time of the conference by the participants)
- Meetups

- ☐ Speed Geek events (participation process used to quickly view a number of presentations within a fixed period of time)
- ☐ Hackathons
- ☐ Application competitions
- ☐ Other (please specify)
- ☐ Other (please specify)

48. What are the impacts on your agency of providing open data?

49. What are the impacts of providing open data on the public sector (e.g., riders, community citizens)?

50. What are the impacts of providing open data on the private sector (e.g., developers)?

51. What challenges has your agency faced in providing open data and how have you overcome them?

52. Please describe any additional "lessons learned" that would benefit transit agencies that are considering providing open data.

53. Are there other agencies that you know of that we should speak to regarding "best practices" in terms of providing open data? If so, please provide contact information.

APPENDIX B

Agencies and Staff Titles Responding to the Survey

Project Manager and CTO
Alameda-Contra Costa Transit District
Oakland, California

IT Manager
Ann Arbor Area Transportation Authority
Ann Arbor, Michigan

Public Transit Technology Manager
Arlington Transit (ART)
Arlington, Virginia

Responsible for IT department
AtB AS
Trondheim, 7413 Norway

Transportation Technologist
Atlanta Regional Commission
Atlanta, Georgia

Senior Transportation Planner
Bangor Area Comprehensive Transportation System
Brewer, Maine

Web Services Manager, Office of the CIO
Bay Area Rapid Transit (BART)
Oakland, California

Councillor
Bilbao City Council
Bilbao, Bizkaia—Basque Country, Spain

ITS Manager
Blacksburg Transit
Blacksburg, Virginia

Operations Technology Specialist
Chittenden County Transportation Authority
Burlington, Vermont

Manager of Database Services
Capital Metropolitan Transportation Authority
Austin, Texas

Senior ITS Developer
Central Florida Regional Transportation Authority
Orlando, Florida

Director of Planning
Central New York Regional Transportation Authority
Syracuse, New York

Software Developer
Champaign-Urbana Mass Transit District
Urbana, Illinois

CTO
Charlotte Area Transit System (CATS)
Charlotte, North Carolina

Manager Transit Systems Support
Chicago Transit Authority
Chicago, Illinois

Consorcio Regional de Transportes de Madrid
Madrid, Spain

IT Manager
DART
Des Moines, Iowa

Transit Planner
Delaware Transit Corporation
Wilmington, Delaware

IT Manager
Empresa Municipal de Transportes de Madrid, S.A.
Madrid, Spain

Planning Manager
Fairfax County DOT/Fairfax Connector
Fairfax, Virginia

Acting Manager Transit Development
Grand River Transit (Region of Waterloo)
Kitchener, Ontario, Canada

Transportation Planner
Greater Bridgeport Transit
Bridgeport, Connecticut

Manager: Operations Analysis, Research and Systems
Greater Cleveland Regional Transit Authority
Cleveland, Ohio

Transport Analyst
Helsinki Regional Transport Authority
Helsinki, Finland

IT Manager
Interurban Transit Partnership
Grand Rapids, Michigan

Director of IT
Kansas City Area Transportation Authority
Kansas City, Missouri

Project/Program Manager
King County Metro
Seattle, Washington

Director Information Management, Transit Applications
Los Angeles County Metropolitan Transportation Authority
Los Angeles, California

Transit Manager
Manatee County Area Transit
Bradenton, Florida

Manager of Operations Technology
Massachusetts Bay Transportation Authority
Boston, Massachusetts

Executive Technical Advisor, Policy, Planning, and Innovation
Metrolinx
Toronto, Ontario, Canada

Director of Development and Regional Coordination
Metropolitan Atlanta Rapid Transit Authority
Atlanta, Georgia

Chief Information Officer
Metropolitan Transportation Authority
New York, New York

Program Coordinator
Metropolitan Transportation Commission (MTC)
Oakland, California

Director of IT
Monterey-Salinas Transit District
Monterey, California

SmartBus Program Manager and Director, Passenger Communication Technology
NJ Transit
Newark, New Jersey

Senior Principal Engineer
NPRA
Trondheim, Sør-Trøndelag Norway

Administrator Bureau of Rail & Transit
New Hampshire DOT
Concord, New Hampshire

Chief Technology Officer
North County Transit District
Oceanside, California

Chief Operating Officer
Norwalk Transit District
Norwalk, Connecticut

Department Manager
Orange County Transportation Authority
Orange, California

Transit Network Coordinator
Oregon DOT Rail + Public Transit Division
Salem, Oregon

Technology Services Manager
Pennsylvania Public Transportation Association
Harrisburg, Pennsylvania

Deputy Executive Director, Strategic Services
Pace Suburban Bus
Arlington Heights, Illinois

Director of Information Technology
Pinellas Suncoast Transit Authority
St. Petersburg, Florida

Director of IT
Roaring Fork Transportation Authority
Aspen, Colorado

Senior Transit Planner
Regional Transportation Commission of Washoe County
Reno, Nevada

Service Development Manager
Regional Transportation District
Denver, Colorado

IT Director
Stark Area Regional Transit Authority
Canton, Ohio

IT Manager
Suburban Mobility Authority for Regional Transportation
Detroit, Michigan

Samenwerkingsverband Regio Eindhoven (SRE)
Eindhoven, Netherlands

Relations Internationales & Affaires Européennes
Syndicat des transports d'Île-de-France (STIF)
Paris, France

Director, Planning and Resources
Organization of Urban Public Transportation of Thessaloniki
Thessaloniki, Greece

ITS Project Manager
Urban Transport Administration
City of Gothenburg
Gothenburg, Sweden

Deputy Chief–Information Systems Manager
Utah Transit Authority
Salt Lake City, Utah

Manager IT
San Mateo County Transit District
San Carlos, California

Innovation and Technology Advisor
Société de transport de Laval
Laval, Quebec, Canada

Supervisor, Research & Technology Program
Sound Transit
Seattle, Washington

Public Transportation Engineer
Tampere City Public Transport
Tampere, Finland

Head of Technical Services Group, London Buses
Transport for London
London, United Kingdom

IS Programme Manager
Transport for Greater Manchester
Manchester, United Kingdom

IT Manager of Geographic Information Systems
TriMet
Portland, Oregon

Director of Administration
Votran
South Daytona, Florida

Wiener Linien
Vienna, Austria

IT Consultant
Worcester Regional Transit Authority
Worcester, Massachusetts

Manager
York Region Transit, Toronto, Canada
Richmond Hill, Ontario, Canada

APPENDIX C

Transportation Organizations and Conferences Discussing and Promoting Open Transit Data

Many transportation organizations are discussing and promoting open transit data. They include:

- Mobility Lab in Arlington, Virginia, sponsors events like Transportation Techies (a group for programmers interested in transit, biking and walking) Hack Nights, in which applications and visualizations using open transit data are created and discussed. For example, on April 3, 2014, "Bus Hack Night" was held at Mobility Lab to discuss bus technologies and ways to use data that is collected from buses (http://www.meetup.com/Transportation-Techies/, accessed on April 1, 2014).
- The subject of open public transport data was prominent in IT-TRANS, an international conference and exhibition on Information Technology (IT) Solutions for Public Transport organized by Karlsruher Messe- und Kongress-GmbH (KMK) and the International Association of Public Transportation (UITP) that was held from February 18–20, 2014.
- TransportationCamp, a series of unconferences that started in 2011 and have been held in Washington, D.C.; San Francisco, California; New York, New York; Montreal, Quebec, Canada; Cambridge, Massachusetts; and Atlanta, Georgia, has contained many sessions related to open transit data.
- Several transit agencies, including the Massachusetts Bay Transportation Authority (MBTA) and Metropolitan Transportation Authority (MTA) in New York City, have held challenges and contests for application developers since 2009.
- Between February 22 and April 30, 2014, U.S. DOT sponsored a Data Innovation Challenge, which promoted and recognized the creation of the most innovative applications, tools and visualizations of publicly-available transportation data.
- Open transport is now the subject of conferences such as the Open Data Transport day organized by a French association of transport authorities and held on June 3, 2013 (*C1*).
- Data Days 2014 conference, which was held February 17–19, 2014, in Ghent, Belgium, covered the idea that open transport is creating a path to more efficient multimodal mobility. "Personal urban mobility is changing. Cars are no longer '# 1' anymore, and commuters are less exclusive in the way they move around in their daily life. Indeed, they use bike-share systems, ride a bus to the train station, or take a taxi back home when working late. They might also do something entirely different the following day depending on such day's particular circumstances like weather, traffic, or personal engagements. In this new context, having accurate information and data has become increasingly useful. Thus, it is good news that as never before, we can now benefit from so many ready-to-use technologies [smartphones, global positioning system (GPS), mobile Internet, etc.]. Major transport companies (national railway, local transport or taxi companies) have begun implementing apps to help people find their way around and to plan the best route possible. However, the best solution is yet to be developed: one that would meet every single need. Maybe it is not their responsibility anymore because our needs are getting more individual and unique. Or maybe, simply because the 'ultimate' solution will come from the users themselves or from the infinite creativity of the internet business world" (*C2*).
- OpenDataLab contest sponsored by Régie Autonome des Transports Parisiens (RATP) (http://www.tom.travel/2013/05/opendatalab-la-ratp-recompense-les-applications-voyageurs/) to promote the creation of new applications for travelers. This event had 110 participants in 12 teams. There were three winning projects.
- Tampere (Finland) region's organizations, specialists and developers are committed to an open data approach. There has been four "Open Data Tampere Meets" since early 2013 (*C3*).
- ITS Innovation Stockholm Kista is an "innovation competition which is organised by The City of Stockholm, Swedish Transport Administration, Stockholm Public Transport, Swedish ITS Council and Kista Science City, and financed by Sweden's Innovation Agency. [It] is arranged as a precommercial procurement, the first ever in Sweden. The challenge for competitors is to develop innovative solutions that meet the demand of more effective travels and transports to and from the outer Stockholm district Kista. In the long run, solutions are to be scalable and equipped with proficient business models so that they will serve citizens in the larger Stockholm region after competition closure. In order to facilitate data access, the competition organisers have developed a data market where competitors get free access to some forty datasets through one API" (*C4*).

REFERENCES

C1. "A review of the 'Open Data Transport' day," June 7, 2013 [Online]. Available: http://www.sustainable-mobility.org/getting-around-today/public-transport/a-review-of-the-open-data-transport-day.html [accessed Feb. 23, 2014].

C2. "Open Transport: The Shortest Way to Efficient Multi-modal Mobility?" *Data Days 2014*, De Bijloke, Ghent, Belgium, Feb. 17–19, 2014 [Online]. Available: http://www.datadays.eu/session/open-transport/.

C3. Kulmala, M., and A. Lumiaho, "Open Data as Enabler for ITS Factory," presented at 20th ITS World Congress, Tokyo, Japan, Oct. 14–18, 2013.

C4. Löfgren, J., "ITS Innovation Stockholm Kista—Stimulating Innovative ITS Solutions in Sweden's First PCP," presented at 20th ITS World Congress, Tokyo, Japan, Oct. 14–18, 2013.

APPENDIX D

Total Annual Ridership for Each Responding Agency

This section contains the total annual ridership for each responding agency.

Agency Name	Annual Ridership (Annual Unlinked Passenger Trips*)
Alameda–Contra Costa Transit District (AC Transit)	54,396,776
Ann Arbor Area Transportation Authority (AAATA)	6,606,838
Arlington Transit (ART)	2,630,421
AtB AS	N/A**
Atlanta Regional Commission	N/A
Bangor Area Comprehensive Transportation System	N/A
Bay Area Rapid Transit (BART)	118,674,764
Bilbao City Council	N/A
Blacksburg Transit	3,516,869
Chittenden County Transportation Authority	2,755,250
Capital Metropolitan Transportation Authority	35,512,338
Central Florida Regional Transportation Authority	29,250,069
Central New York Regional Transportation Authority	10,385,770
Champaign–Urbana Mass Transit District	11,118,500
Charlotte Area Transit System (CATS)	28,243,662
Chicago Transit Authority (CTA)	545,577,917
Consorcio Regional de Transportes de Madrid	N/A
Des Moines Area Regional Transit Authority	4,577,547
Delaware Transit Corporation	11,594,262
Empresa Municipal de Transportes de Madrid, S.A.	N/A
Fairfax County DOT/Fairfax Connector	10,895,833
Grand River Transit (Region of Waterloo)	N/A
Greater Bridgeport Transit	5,951,650
Greater Cleveland Regional Transit Authority	48,234,103
Helsinki Regional Transport Authority	N/A
Interurban Transit Partnership	11,980,418
Kansas City Area Transportation Authority	16,517,706
King County Metro	119,952,268
Los Angeles County Metropolitan Transportation Authority	464,875,164
Manatee County Area Transit	1,849,279
Massachusetts Bay Transportation Authority	401,616,849
Metrolinx	N/A
Metropolitan Atlanta Rapid Transit Authority	134,889,690
Metropolitan Transportation Authority (MTA)	2,612,523,069
Metropolitan Transportation Commission (MTC)	N/A
Monterey-Salinas Transit District	4,522,313
NJ Transit	266,823,218
Norwegian Public Roads Administration (NPRA)	N/A
New Hampshire DOT	N/A

Agency Name	Annual Ridership (Annual Unlinked Passenger Trips*)
North County Transit District	12,081,329
Norwalk Transit District	2,116,755
Orange County Transportation Authority	55,211,248
Oregon DOT Rail + Public Transit Division	N/A
Pennsylvania Public Transportation Association	N/A
Pace Suburban Bus	35,399,701
Pinellas Suncoast Transit Authority	14,018,425
Roaring Fork Transportation Authority	N/A
Regional Transportation Commission of Washoe County	8,277,735
Regional Transportation District	98,518,888
Stark Area Regional Transit Authority	2,658,900
Suburban Mobility Authority for Regional Transportation	11,379,228
Samenwerkingsverband Regio Eindhoven (SRE)	N/A
Syndicat des transports d'Île-de-France (STIF)	N/A
San Mateo County Transit District	13,422,728
Société de transport de Laval	N/A
Sound Transit	28,540,694
Tampere City Public Transport	N/A
Transport for London	N/A
Transport for Greater Manchester	N/A
Tri-County Metropolitan Transportation District of Oregon (TriMet)	103,218,538
Organization of Urban Transportation of Thessaloniki	N/A
Urban Transport Administration	N/A
Utah Transit Authority	42,365,346
Votran	3,876,417
Wiener Linien	N/A
Worcester Regional Transit Authority	3,938,857
York Region Transit	22,700,000

*Figures come from the National Transit Database, Annual Transit Profiles (available at http://www.ntdprogram.gov/ntdprogram/data.htm), or directly from survey responses.
**N/A means that total ridership was not available.

APPENDIX E

Example Developer License Agreements and Terms of Use

This section contains examples of agency developer license agreements and terms of use.

- Bay Area Rapid Transit (BART): http://www.bart.gov/schedules/developers/developer-license-agreement
- Champaign-Urbana Mass Transit District (CUMTD): https://developer.cumtd.com/terms-of-use
- Chicago Transit Authority (CTA): http://www.transitchicago.com/developers/terms.aspx
- City of Madison Metro Transit: http://www.cityofmadison.com/metro/Apps/terms.cfm
- East Japan Railway Company: http://www.jreast.co.jp/e/termsofuse/index.html
- GO Transit: http://www.gotransit.com/public/en/schedules/goapps/web/goapps.aspx
- King County Transit: http://www.kingcounty.gov/transportation/kcdot/MetroTransit/Developers/TermsOfUse.aspx
- Massachusetts Department of Transportation (MassDOT): https://www.massdot.state.ma.us/DevelopersData.aspx
- Metropolitan Atlanta Rapid Transit Authority (MARTA): http://www.itsmarta.com/developers/data-sources.aspx
- Metropolitan Transportation Authority (MTA) (New York City): http://web.mta.info/developers/
- New Jersey Transit (NJT): https://www.njtransit.com/mt/mt_servlet.srv?hdnPageAction=MTDevLoginTo
- Southeastern Pennsylvania Transportation Authority (SEPTA): http://www2.septa.org/developer/
- Transport for London (TfL): http://beta.tfl.gov.uk/corporate/terms-and-conditions/transport-data-service
- TriMet: http://developer.trimet.org/
- VIA Metropolitan Transit: http://www.viainfo.net/Opportunities/DevLicense.aspx
- Washington Metropolitan Area Transit Authority (WMATA): https://www.wmata.com/rider_tools/license_agreement.cfm
- York Region Transit (YRT): http://yrt.ca/en/aboutus/GTFS.asp

APPENDIX F

Examples of Open Transit Data Applications

This section contains examples of open transit data applications.

- BART: http://www.bart.gov/schedules/appcenter
- CUMTD: http://www.cumtd.com/maps-and-schedules/app-garage
- CTA: http://www.transitchicago.com/apps/
- City of Madison Metro Transit: https://www.cityofmadison.com/metro/apps/
- East Japan Railway Company: https://play.google.com/store/apps/developer?id=East+Japan+Railway+Company+ICT (Android only)
- GO Transit: http://www.gotransit.com/public/en/schedules/goapps/web/goapps.aspx
- King County Transit: http://www.kingcounty.gov/transportation/kcdot/MetroTransit/Developers/AppCenter.aspx
- MARTA: http://www.itsmarta.com/developers/app-station.aspx
- MBTA: http://www.mbta.com/rider_tools/apps/Default.asp
- MTA: http://web.mta.info/apps/
- SEPTA: http://appsforsepta.org/apps
- Transport for London (TfL): http://data.london.gov.uk/search/node/TfL
- TriMet: http://trimet.org/apps/
- VIA Metropolitan Transit: http://www.viainfo.net/BusService/Mobile.aspx
- WMATA: http://developer.wmata.com/Application_Gallery

APPENDIX G

Examples of Applications from Transport Innovation Deployment for Europe (TIDE) Project

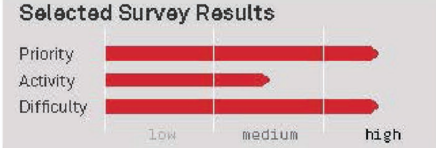
Cluster 3:

Advanced network and traffic management to support traveller information

Measure 1:

Open data server for applications–based traveller information

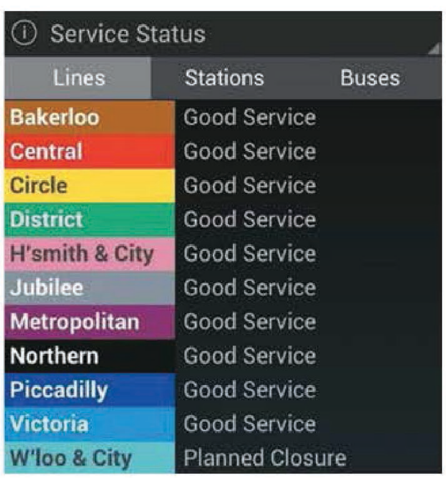
Selected Survey Results



Good Practice

Transport for London (TfL) publishes a wide range of open data to support and encourage developers to devise the best, most accurate apps, widgets and online tools for information dissemination. The aspiration is to publish TfL data in a format which maximises its accessibility and usefulness to developer partners, wherever legally, commercially and technically possible. The data available in the open server includes: live bus arrivals, tube departure boards, Barclays cycle hire availability, Oyster card journey information, live traffic camera images (CCTV), etc.

TfL has developed clear guidance to support and encourage the use of data in the open server. It includes: data guidelines giving details of the available data; branding guidelines containing do's and don'ts; common questions in general, technical and legal areas. To access data, a developer needs to register to the site and accept the terms and conditions. A number of applications have been developed that use the data from TfL's open server.



Service updates board

Photo: www.tfl.gov.uk

Key Characteristics

Multi-modal traveller information is generated from a wide range of data sources including those related to congestion, car park occupancy, incidents, roadworks, CCTV images, variable message sign messages and real-time public transport information. The provision of open access data allows any third party to freely use such traffic data to provide traveller information services for a variety of user groups. This measure focuses on the management and manipulation of the various data sources in order to provide information to the open access server that is relevant to third party applications and web developers. This approach enables the private sector to deliver more and better travel information to the end user. This allows travellers to make more informed decisions and should encourage the choice for more sustainable modes of travel. It should also support local authorities' traffic management strategies to reduce congestion and pollution.

Key Benefits

Open-access data for traveller information:

- stimulates the private sector app developers to innovate and provide apps according to user needs;
- provides an option to disseminate multi-modal traveller information in a cost-effective way;
- allows travellers to make more informed decisions and could encourage the choice of more sustainable modes of travel;
- shows the willingness of the local authority to honour its obligations under "freedom of information";
- avoids developers needing to apply (and pay) for data to create applications — as long as they follow the terms of agreement.

Web Links:

> TfL
www.tfl.gov.uk/businessandpartners/syndication/default.aspx
> Toronto
www.toronto.ca/open
> Reading
www.reading-travelinfo.co.uk

FIGURE G1 TfL open data (*72*, p. 24).

Key Aspects for Implementation

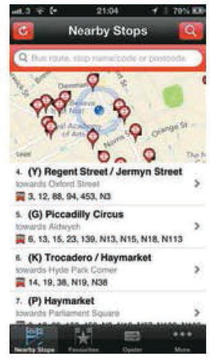
Checklist

City size	No restrictions
User needs	• Local authorities are looking for innovative ways to disseminate traveller information – covering different modes of transport; – covering private and public transport; – utilising the traffic information collected in a better way; – helping to address traffic management strategies; • Travellers need timely, accurate information on different aspects of their journey (e.g. real time bus arrival time, incidents/accidents, congestion levels, etc.); • Apps developers need easy access to accurate and reliable traffic data.
Costs	• Modest capital investment needed to set up; • Support costs typical of other data management applications.
Time horizon	Planning of scheme and preparation of materials within a few months
Key stakeholders involved	• Local authorities • Transport managers • Apps developers • General public
Crucial factors	• Commitment for open data server provision; • Clear policy about the provision and use of data; • Cost effectiveness; • Enthusiastic apps developers; • Apps users.
Excluding factors	None

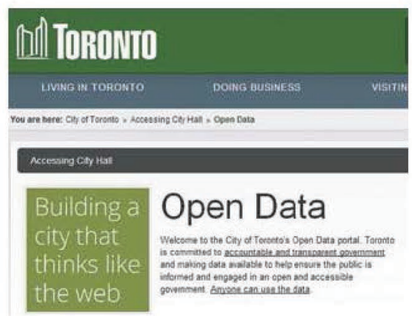
TIDE Contact:

> Rob McDonald
RMcDonald@peterbrett.com

> Birendra Shrestha, TRG
b.p.shrestha@soton.ac.uk



App developed using TfL's open data
Source: www.computeractive.co.uk



Toronto's open data portal
Source: www.toronto.ca/open

FIGURE G1 (*continued*) (*72*, p. 24).

APPENDIX H

TfL Open Transit Data Available from the London Datastore

This section contains open transit data available from the London Datastore (http://data.london.gov.uk/taxonomy/organisations/tfl).

- TfL Live Traffic Cameras—Live traffic camera feeds of London's streets
- TfL Station Locations—Locations and facilities of London's Underground, Overground, and DLR stations
- Number of Bicycle Hires—Total number of hires of the Barclays Cycle Hire Scheme, by day, month, and year
- Public Transport Journeys by Type of Transport—Number of journeys by TfL reporting period, by type of transport. Data are broken down by bus, underground, DLR, tram, and Overground.
- London Underground Performance Reports—London Underground periodic performance reports, and TfL's key London Underground performance measures
- Vehicles entering c-charge zone by month—Total number of vehicles that entered the Congestion Charging Zone during charging hours
- Cycle Flows on the TfL Road Network—Cycle flows on the Transport for London Road Network (TLRN)
- Transport Crime in London—Number of crimes and crime rate by type of public transport, including bus, LU/DLR, London Overground, and London Tramlink
- Bus Use and Supply Data 1999–2022—Technical analysis for London Assembly Transport Committee report on bus services in London (October 2013)
- Number of Buses by Type of Bus in London—Number of buses in the TfL fleet by type of bus in London
- Key Performance Indicators on the TfL Road Network—Number of hours of Serious and Severe Disruption on the road network by planned and unplanned status, journey time reliability, total number of works undertaken on the road network, and number of cycle hires with average hire time
- Killed and Seriously Injured Casualties, London—Killed and seriously injured casualties in Greater London 1994–2012
- Public Transport Accessibility Levels—Transport for London's 2008 Public Transport Accessibility Levels (PTALs)
- GLA Group Land Assets—Data showing the location of the GLA Group's land and property holdings and development opportunities
- Low Carbon Generators—Listing of low carbon energy generators installed on GLA group properties
- TfL Rolling Origin and Destination Survey—Rolling programme to capture information about journeys on the London Underground network
- TfL Complaints Reports—The first TfL Complaints Report for 2011/12
- TfL Live Bus Arrivals API—Developer's API for the Live bus arrivals data
- Travel Patterns and Trends, London—A summary of the key travel patterns and trends relating to the TfL network and Airports
- Tube Network Upgrade Data from Transport Committee Report—Data on the Tube upgrade programme between 2003/04 and 2010/11, requested by the London Assembly
- Tube Network Performance Data from Transport Committee Report—Data on the performance of the Tube between 2003/04 and 2010/11, requested by London Assembly
- Journey Planner API Beta—Developer's API for the TfL Journey Planner
- Cycle Hire availability—The TfL feed contains details of Barclays Cycle Hire docking stations including the number of available bikes and docking points.
- TfL Expenditure Over £500—Transport for London expenditure over £500
- Tube Departure Boards, Line Status, and Station status—Train prediction and London Underground station and line status
- TfL Station Facilities—A Geo-coded KML feed of most London Underground, DLR and London Overground stations.
- Dial a Ride Usage Statistics—Quarterly Report, by London Borough, of Dial-a-Ride usage
- Coach Parking Locations—Details of coach parking facilities and other useful information for coach drivers in London
- Estimated Number of Londoners with Reduced Mobility—Estimated number of Londoners with reduced mobility in 2010, 2018, and 2031
- TfL Live Roadside Message Signs—Information on the location and live message on every sign currently displaying information in London.
- TfL Live Traffic Disruptions—Information about the location, nature, impact and timing of a range of disruptions being monitored by TfL's 24/7 traffic control centre
- TfL Timetable Listings—Multimodal working timetable data from the TfL Journey Planner, including Tube, Bus, and DLR
- TfL Cycle Hire Locations—Name and coordinates of the Barclays Cycle Hire docking station locations launched in Summer 2010
- TfL Bus Stop Locations and Routes—Locations of more than 19,000 bus stop and details of all of TfL's contracted bus routes

- Oyster Ticket Stop Locations—Locations of more than 3,800 Oyster Ticket Stop outlets across London
- London Underground Signals Passed at Danger—The number of signals passed at danger by London Underground or on London Underground's infrastructure
- TfL Pier Locations—Details of TfL pier locations
- River Boat Timetables—Timetables for river boat crossings
- Accessibility of London Underground Stations—Detail relating to the physical accessibility of London Underground stations
- TfL Investment Programme 2009/10 to 2017/18—Data incorporated in Transport for London's Investment Programme Reports 2009/10
- TfL Business Plan 2009/10 to 2017/18—Data incorporated in Transport for London's Business Plan

Abbreviations used without definitions in TRB publications:

A4A	Airlines for America
AAAE	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
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
HMCRP	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
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