



# Open Source Software and Open Data – What They Are and How to Use Them

October 23, 2020

Prepared by:

Carol Schweiger

**Schweiger Consulting LLC**



## **About N-CATT**

The National Center for Applied Transit Technology (N-CATT) is a technical assistance center funded through a cooperative agreement with the United States Department of Transportation's Federal Transit Administration (FTA). Operated by the Community Transportation Association of America (CTAA), the mission of N-CATT is to provide small-urban, rural and tribal transit agencies with practical, replicable resources that help them apply technological solutions and innovations. Among its activities, N-CATT produces a series of white papers, technical reports such as this document, and other resources, all of which can be accessed on-line at <https://n-catt.org>.

## **About this Document**

This document was prepared for CTAA by Carol Schweiger of Schweiger Consulting, LLC in October 2020 as part of the N-CATT cooperative agreement between CTAA and FTA. Opinions expressed or implied in this document are those of the authors. Nothing in this document is to be interpreted as position, policy or guidance from the United States Government. Incidental use of companies' names or the names of their products is made solely to facilitate discussion and should not be regarded as recommendations or endorsements.

## Table of Contents

<b>1</b>	<b><i>Introduction</i></b> .....	<b>2</b>
<b>2</b>	<b><i>Characteristics of Open Source Software</i></b> .....	<b>3</b>
<b>3</b>	<b><i>Benefits and Challenges of Open Source Software</i></b> .....	<b>4</b>
<b>4</b>	<b><i>Examples of Open Source Software</i></b> .....	<b>6</b>
<b>5</b>	<b><i>Costs Associated with Open Source Software</i></b> .....	<b>7</b>
<b>6</b>	<b><i>Open Source Software Case Studies</i></b> .....	<b>8</b>
6.1	OpenTripPlanner.....	8
6.2	RidePilot.....	9
<b>7</b>	<b><i>Characteristics of Open Data</i></b> .....	<b>11</b>
<b>8</b>	<b><i>Benefits and Challenges of Open Transit Data</i></b> .....	<b>12</b>
<b>9</b>	<b><i>Examples of Open Transit Data</i></b> .....	<b>14</b>
<b>10</b>	<b><i>Costs Associated with Open Transit Data</i></b> .....	<b>14</b>
<b>11</b>	<b><i>Open Transit Data Case Studies</i></b> .....	<b>15</b>
11.1	Worcester Regional Transit Authority (WRTA) .....	15
11.2	Open Transit Data Resources.....	17
<b>12</b>	<b><i>Conclusions</i></b> .....	<b>18</b>
<b>13</b>	<b><i>References</i></b> .....	<b>19</b>

## 1 Introduction

The use of open source software in the transit industry has been growing rapidly in order to provide an alternative to proprietary software, which has dominated the industry in the past. With open source software, transit agencies are not locked into one software vendor. The use of open transit data became widespread in the early 2010s in order to support and improve customer information that is provided via several dissemination media (e.g., real-time transit information signage and transit apps).

Leveraging both open software and open data can provide transit agencies with flexibility and tools to address certain challenges in their systems, either internally or with partnering entities. This white paper will define and explain what these two terms mean, what open source software and open data each can and cannot deliver for transit systems in rural, tribal, and small urban areas, and what early adopters of each of these types of technology are already finding.

“Open data and open source are phenomena that are often automatically grouped together, perhaps because they share the word ‘open.’ Although there are, indeed, elements they share through their openness, the ways in which they differ are significant.”<sup>1</sup> In fact, open source software and open data address very different issues and do not necessarily go together at all.

Open source software, as compared with proprietary software from only one vendor, is software with source code that anyone can inspect, modify, and enhance. Source code is the part of software that most computer users do not ever see. It is the code computer programmers can manipulate to change how a piece of software, also known as a program or application, works. Programmers who have access to a computer program's source code can improve that program by adding features to it or fixing parts that do not always work correctly.

The differences between open source software and open data are in the infrastructure that supports the original works. “The default practice in the open source world is that code will be published:

- In a public repository;
- With a complete version history (or at least versioning dating from its publication);
- In an environment that supports transparent reporting of issues, bugs and suggestions;
- In an environment that includes good documentation tools, such as a wiki; and
- Most importantly, in an environment that allows forks and improvements to be folded back into the original project.”<sup>2</sup>

On the open data side, typically<sup>5</sup>:

- Open data is published by single organization
- There is little insight into how data was curated, at best there is some documentation
- Data portals provide some infrastructure - for example issue reporting and documentation, but this is often limited in scope
- Data portals do not provide any support for encouraging collaboration or external contributions (except OSM)

“Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike. The full Open Definition<sup>3</sup> gives precise details as to what this means. To summarize the most important:

- **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
- **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavor or against persons or groups. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.”<sup>4</sup>

“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.”

Carol L. Schweiger, *Open Data: Challenges and Opportunities for Transit Agencies*, Transit Cooperative Research Program Synthesis 115, prepared for the Transportation Research Board, 2015, page 9.

## 2 Characteristics of Open Source Software

There are six principles of open source software as follows:

- **Open exchange.** “[A]uthors [of open software] make its source code available to others who would like to view that code, copy it, learn from it, alter it, or share it.”<sup>5</sup>
- **Transparency.** Providing access to information and materials necessary to understand the software and how it works.
- **Collaboration.** Anyone can participate in improving/modifying the software, which means that individuals do not need to address issues alone, but with the help of others.

- **Rapid prototyping.** This is an approach that accelerates making the most recent version available as well as pursuing specific solutions to problems.
- **Meritocracy.** As the definition of this word suggests, decisions about the software should be made by those who best know the field in which the software operates and the most about the software itself. “Including diverse perspectives in [the] conversations [ensures that] we've identified the best ideas.”
- **Community-oriented development.** Open source software is of interest to diverse people who want to use the software, want to help clients with the software and are subject matter experts. Communities form naturally as these diverse people work with the software and share ideas to improve the software or to add features to the software.

### 3 Benefits and Challenges of Open Source Software

There are several benefits of using open source software rather than proprietary software, as follows<sup>6 7</sup>:

- **Control.** Both programmers and software users have more control over open source software since they can recommend and make changes to the code, and can use the software in any way they wish.
- **Training.** Programmers and students benefit from open source software due to its community nature – people working with the software can learn from each other while improving and modifying the software.
- **Security.** It is thought that open source software is more secure than proprietary software because errors may be detected and fixed more quickly, and many more people can detect any issues regarding security because the software is publicly accessible.
- **Stability.** Open source software is considered more stable due to the use of open standards<sup>8</sup> and the community ensuring that software will be up-to-date and fully functional at all times.
- **Community.** As mentioned earlier open source software users and programmers work together, creating a community that program, test, use and update the software. The community shares the cost and value of improvements.

- **No License or Maintenance Fees.** “Annual increases in software license fees are the bane of financial officers, particularly those from small- to medium-sized transit entities. They justifiably reject the prohibitive expenses of annual fees and inevitable upgrade costs. With open source, agencies get a foundation that is free and unencumbered by restrictive and ongoing costly fees, unlike proprietary platforms that most agencies cannot afford.”<sup>9</sup>

Looking at open source software from a transit perspective, there are several key benefits, as follows<sup>10, 11</sup>:

- It allows multiple agencies to leverage the same resources by not (1) starting from scratch, moving forward with new projects faster; and (2) reinventing the wheel, meaning that agencies can utilize the resources you might have spent in other ways.
- It reduces the risk associated with vendor lock-in, which often happens when proprietary software is procured, because (1) any vendor can deploy and support an open-source solution; and (2) if a vendor does not perform, an agency can switch vendors and keep the software.
- It reduces the long-term risk of innovation failure since others familiar with the subject matter can improve upon past efforts.
- Enhancements made by other users are available free of charge for everyone.
- A transit/paratransit/human service agency can use its own information technology (IT) staff to manage and deploy the software. For example, more and more transit agencies are taking advantage of this Open Trip Planner, a widely used transportation-specific open source software platform.
- An agency can seek assistance from a competitive market of contractors/vendors who have the knowledge and skills to modify open source software.

“Developing open-source software is a complex undertaking given the competing demands of software code that is both (a) sufficiently generalizable to be useful to a diverse array of users, and (b) sufficiently tailored to any given end-users' needs. It is also challenging to coordinate open-source software development, especially when there are many users likely to be impacted by any given change.

The disadvantages of open source software are as follows<sup>1</sup>:

- It often takes more time to implement custom open source software than purchasing off-the-shelf software.
- Custom open source software implementations often cost more money than purchasing non-custom software.
- There are very few software developers who want to build and maintain open source software for limited public sector use cases, so agencies are often locked into vendors (consultants or companies) who are the only ones who want to bother continuing to develop custom code.

Regina Clewlow, “When Open Source Software Costs Cities More,” *Forbes*, January 30, 2020,

Public-sector transportation agencies can advance open-source software development using the following strategies, among others:

- Add open-source requirements to software procurements
- Add a budget line-item and/or project task for the "open-sourcing" and related documentation of newly developed code
- Clearly communicate successes and open-source software implementations
- Coordinate with peer agencies to identify list of shared priorities and share costs for larger 'capital' improvements to open-source code blocks."<sup>12</sup>
- Select the right contractor/vendor based on their approach to fulfilling the agency's needs, as well as the commitment to maintain quality and innovation. "The vendor should display the ability to generate a robust and reliable platform and commit to the principles of the open source community."<sup>13</sup>

#### 4 Examples of Open Source Software

There are many examples of open source transit software being used in the transit industry, including the following:

- **OpenTripPlanner.** Multimodal open source trip planner, with full suite of APIs and several open source front-end UIs available. Deployed by many US transit agencies and some state DOT's for regional or statewide trip discovery.
- **Real-time bus information for small agencies (OneBusAway, TheTransitClock).** OneBusAway is an open-source Customer Information System (with open source apps) to deliver real-time information. Could be coupled with low-cost wireless routers to enable real-time information for small bus operators. Send vehicle data to TheTransitClock prediction engine to generate high quality arrival predictions
- **1-Click.** Open-source mobility management program that enables users to discover eligibility-restricted transportation options
- **Link2Support.** A web-based application that helps people discover the essential human services they need and provides transportation options to help access them
- **RidePilot.** Scheduling system for door-to-door service including:
  - Ability to setup and automatically schedule recurring runs and trips
  - Tracking and generating federally-required reports
  - CAD/AVL component with a driver app that can be loaded onto Android tablets or phones
- **Rural Transit Assistance Program (RTAP) Website Builder.** Allows users to create and design websites that are hosted on the National RTAP server free of charge. In 2019, National RTAP released Website Builder 3.0 using customer feedback to make improvements. Website Builder 3.0 makes it easier for users to create and edit their sites by providing a basic template, simplified design functions, and new support resources. And those wishing to do more advanced work on their websites are still able to do so.



- **RTAP Cost Allocation Calculator.** As a requirement of reporting to the National Transit Database (NTD), public transit agencies must be able to provide expense information by different travel modes, jurisdictions, and service types. Transit agencies must also understand the costs of different services for managing federal and state grants as well as pricing and planning for new services. This Two-Variable Cost Allocation Calculator, developed by Texas A&M Transportation Institute (TTI) and RLS & Associates, Inc. for RTAP, is available as an MS Excel or MS Access application.
- **RTAP ProcurementPRO.** A free web-based application that guides rural and Tribal grantees and State DOTs through Federal Transit Administration (FTA) procurement procedures. Using basic project information entered by the user, the program provides the required FTA clauses and certifications that must be included in procurement documents. The application can also provide other helpful resources such as a procurement document template, check lists, and guidance for preparing procurement documents.
- **RTAP GTFS Builder.** A free Microsoft Excel-based web application which assists rural and tribal transit agencies to develop and generate fully valid GTFS for their bus routes. National RTAP partnered with Transnovation Inc. to develop GTFS Builder.
- **TransAM.** Open source asset performance management platform for transit and other transportation assets, focused on the connected asset management, grant management, and project planning life cycles and related reporting requirements
- **OpenStreetMap.** A map of the world, created by individuals and free to use under an open license

## 5 Costs Associated with Open Source Software

Open source software is not necessarily free – this is a misconception because people misinterpret the term “open” to mean that there is no charge for it. “Open source software programmers can charge money for the open source software they create or to which they contribute. But in some cases, because an open source license might require them to release their source code when they sell software to others, some programmers find that charging users money for ‘*software services and support*’ rather than for the software itself is more lucrative. This way, their software remains free of charge, and they make money helping others install, use, and troubleshoot it.”<sup>14</sup>

The costs associated with open source software vary a great deal depending on the function(s) that the software provides. You may experience costs in the following categories<sup>15</sup>:

- **Planning.** There is possibly a cost associated with making the deployment of the open source software as successful as possible. For example, if the software is designed to be used in a complex setting, this cost would help address and manage the complexity associated with a deployment. The cost would be based on how complex the implementation is.

- **Training.** This cost would depend on the type of deployment and the agency’s approach to training (i.e., train-the-trainer or training from outside the agency).
- **Configuration.** This cost is dependent on the complexity associated with making the software operable from a computer or server with specific rules (e.g., compliance with the Health Insurance Portability And Accountability Act (HIPAA) and data security)
- **Software development, if needed.** There could be costs associated with adding features to the software.
- **Deployment.** This is the cost associated with putting the software on an agency’s server. If the agency has the in-house expertise, they can do it themselves.
- **Periodic upgrades.** The cost if an agency deployed the software on its own server but does not have the in-house expertise and needs a new version of the software on the server. The cost could include scheduling the cut-over to the new version after hours, carrying out any data migrations, etc. This cost could be bundled with system support.

However, if the open source software is provided as Software as a Service (SaaS)<sup>16</sup> model, several of the aforementioned costs may not apply. SaaS may be open source software or proprietary, such as Adobe and Microsoft Excel. With SaaS some or all of the costs could be covered by an annual or monthly fee.

## 6 Open Source Software Case Studies

### 6.1 OpenTripPlanner

In 2017, the Vermont Agency of Transportation (VTrans) began a project to provide “flexible multimodal transit trip planning” (software for trip planning that seamlessly integrates demand-responsive modes of transit with fixed-route transit). Two years later, an emerging marketplace of providers [was established] that support routing similar to the routing that Vermont first launched. Most of the players in that market are directly engaged in working with the technologies developed through [this] project, and other transportation providers are researching how to collaborate with the open data approach pursued by VTrans. The VTrans [project] started from the problem statement “Trip planning software on the market does not work for residents of Vermont, because it is focuses on the needs of urban transit riders. To address that problem, the project team leveraged and augmented two existing technologies—GTFS-flex, an open data extension of the widely-used General Transit Feed Specification [GTFS] that describes demand-responsive transit services, and OpenTripPlanner (OTP), an open source software application used by more than 30 transit agencies around the world to provide automated itineraries for transit trips. Together, this data and software combination provided a uniform way to describe flexible transit services around Vermont and an application ready to be adapted to leverage that data, which already included most of the features VTrans sought. Thus, VTrans was able to create a system wholly owned by the State of Vermont without ongoing license fees. The outcomes of this project—GTFS-flex data sets, lessons for flex data contributed to a revised GTFS-flex v2, and the OTP adaptations—were contributed back to the community and are available for use by any party.”<sup>17</sup> GO VERMONT! Can be found at <https://www.connectingcommuters.org/>.

As summarized by Cambridge Systematics<sup>18</sup>, one of the project partners, OTP, the open source software used for this project has the following characteristics:

Description	<ul style="list-style-type: none"> <li>• Multimodal open source trip planner, with full suite of APIs and several open source front-end UIs available</li> <li>• Deployed by many US transit agencies and some state DOT's for regional or statewide trip discovery</li> </ul>
Users/Audience	Anyone planning a journey (but primarily involving transit)
Ease of Adoption	<ul style="list-style-type: none"> <li>• Requires hosting, technical configuration and parameter tuning for local service</li> <li>• May require extra work to fully integrate into existing website front end but can be easily linked</li> </ul>
Potential Value	With Flex (demand-responsive) routing, OTP can serve many small/rural agencies and support trip discovery of underutilized services

## 6.2 RidePilot

The Utah Transit Authority (UTA) implemented Computer Aided Scheduling and Dispatch (CASD) software system to meet the needs of small scale human service transportation agencies. They deployed RidePilot, an open source, web-based scheduling, reporting, and dispatch application to fill this need.

UTA participated in a joint software development project to build upon RidePilot's original functionality and expanded it to meet the needs of human service agencies in the Wasatch Front region. These enhancements included the ability to track driver's credentials, vehicle maintenance compliances, scheduling and tracking of daily trips, and a host of other details for accountability reporting. Further, in a subsequent phase, UTA, in conjunction with Cambridge Systematics (CS), incorporated new features and updates including; drag and drop scheduling, vehicle capacity tracking, driver availability matrix, subscription/recurring run and trip creation/scheduling, and new reporting options. Also developed was an Android mobile app that includes pre-trip inspection, driver manifest, and vehicle tracking (AVL).

- The total costs for UTA including, initial start-up and enhancements are as follows:
  - Phase 1 in 2015-16, which was funded by a VTCLI grant, made enhancements aimed at making system work in the UTA context. It cost UTA \$0<sup>19</sup> and project management time plus \$13,000 per agency (17 small agencies along the Wasatch Front Range (greater Salt Lake region) implemented the software.
  - Phase 2 in 2017-19 made enhancements aimed at making the system easier to use and more capable. It cost UTA \$350,000 and project management time plus \$50,000 per agency.
  - The cost of starting a new deployment has the following components:
    - Includes starting up RidePilot plus configuring/deploying Android driver app
    - Includes some training for System Admin and Provider Admins

- Operations & Maintenance costs include updating system software components, addressing any defects that arise, and secondary support
- Hosting costs include multiple servers for testing and redundancy
- Automatic vehicle location (AVL) and Google fees include costs for the estimated time of arrival (ETA) calculation in both RidePilot and the Android driver app
- The outcomes for providers and customers are as follows:
  - Providers: This was a single deployment by UTA with many providers that could cover any size region/state. UTA is the system administrator, there are 17 current providers, and UTA provides the FTA 5310 subrecipient oversight.
  - Customers: By increasing the number of providers using RidePilot, UTA enhanced the range of direct booking options available on its portal and streamlined the process of using specialized transportation service.

As summarized by Cambridge Systematics<sup>20</sup>, who offers RidePilot, this open source software has the following characteristics:

Description	Scheduling system for door-to-door service including: <ul style="list-style-type: none"> <li>• Ability to setup and automatically schedule recurring runs and trips</li> <li>• Tracking and generating federally-required reports</li> <li>• CAD/AVL component with a driver app that can be loaded onto Android tablets or phones</li> </ul> <i>Note: does not address fixed route transit</i>
Users/Audience	Very small demand-responsive operators may have previously relied on cumbersome tracking methods involving spreadsheets, whiteboards, or index cards
Ease of Adoption	<ul style="list-style-type: none"> <li>• Moderate hosting, configuration effort</li> <li>• Requires some training for operator staff</li> <li>• Lightweight CAD/AVL can be implemented with Android device</li> </ul>
Potential Value	<ul style="list-style-type: none"> <li>• System is designed to be centrally administered by a larger agency on behalf of many providers</li> <li>• Supports more efficient and competitive offerings by small operators</li> </ul>

## 7 Characteristics of Open Data

Open transit data is defined by Sound Transit, a Puget Sound, WA transit agency, as follows<sup>21</sup>:

Open Transit Data (OTD) is publicly-accessible transit data that is:

- Free and accessible – Sound Transit requires that users agree to their terms of service but beyond that users are free to use the data how and when they like.
- A community - people that love transit, people that like to work with data, people that help build things.

OTD is not:

- Static - at a minimum, most agencies issue a service change twice a year. Some are constantly tweaking routes, stops and schedules. Sound Transit strives to keep it current, but it is constantly evolving.
- Comprehensive – Sound Transit continue to add services and agencies, including expanding the real-time offerings.

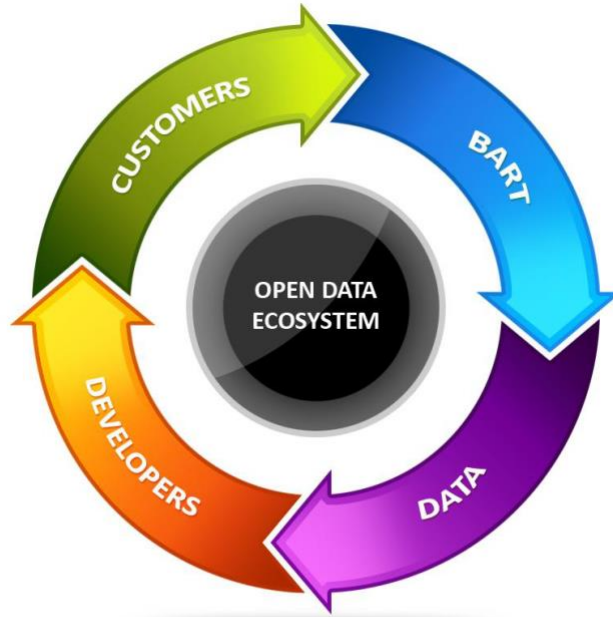
“The conditions of open data are defined as follows:

- *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)].”

Carol L. Schweiger, *Open Data: Challenges and Opportunities for Transit Agencies*, Transit Cooperative Research Program Synthesis 115, prepared for the Transportation Research Board, 2015, pg 9.

OTD is for anyone who has an interest in transit data, from students interested in maps and urban planning to developers looking to produce the next great app. Transit data is ever-evolving and as the data moves between systems, errors can occur. Identifying the problem's point of entry, fixing it, and repopulating the database(s) takes time. Also, improvements can be made to help the data flow faster and make the data more understandable.

In describing how Bay Area Rapid Transit (BART) creates value with open transportation data, Timothy Moore, former web services manager, used the graphic below 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). Another individual further defined these entities: the transit agency is the “discloser;” the developers are the “intermediaries;” and the customers are the “end users.”<sup>22</sup>



**Figure 1. Open Data Ecosystem<sup>5</sup>**

“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. 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 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.’”<sup>23</sup>

## **8 Benefits and Challenges of Open Transit Data**

“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.”<sup>24</sup>

Additional benefits of open transit data include<sup>25</sup>:

- Engaging application developers, other data users and customers accomplishes:
  - Obtaining feedback on data anomalies and data quality issues;
  - Ensuring that applications meet customers’ needs; and
  - Finding out more about how people want to use/ reuse data;
- Standards/formats facilitate the generation and use of open data; and
- Results in innovation that might not be accomplished within a transport agency.

The challenges associated with opening public transit data include the following<sup>26</sup>:

- Cost and developing a sound process for releasing data and maintaining oversight of its use.
- 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 Google Transit Feed Specification (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.
- 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. Therefore, when identifying a GTFS creation process, the maintenance and sustainability of the process must be considered.
- Data coverage, quality, privacy/confidentiality, accuracy, and timeliness: data may need significant amounts of “cleansing” or anonymizing before publication.
- 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.
- 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.
- Consider accessibility and equity.

The negative impacts of providing open data include the following<sup>27</sup>

- Development effort and maintenance, and generally, staff time
- Data quality:
  - Time required to do quality analysis on data, resulting in more pressure on staff

Additional challenges associated with providing open data can be found on page 52 of ***Open Data: Challenges and Opportunities for Transit Agencies.***

- 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
- More scrutiny because of increased visibility of data accuracy, including third-party users wanting zero downtime

## 9 Examples of Open Transit Data

Examples of open transit data can be found via the links in <https://transitfeeds.com/>. These include hundreds of transit providers. An example from a rural agency in Cumberland County, Maine is as follows.

TransitFeeds is now OpenMobilityData!

Home / Feeds / North America / US / ME / Cumberland County / Lakes Region Explorer / Lakes Region Explorer GTFS / Latest (6 February 2020)

### Latest (6 February 2020)

This feed is currently in service! (3 January 2020 - 29 April 2029)

Download Feed 0.1 MB

File	Size	Lines	Status
<a href="#">agency.txt</a>	243 B	1	OK <span>Agenicies: 1</span>
<a href="#">calendar.txt</a>	159 B	2	OK
<a href="#">calendar_dates.txt</a>	116 B	6	OK
<a href="#">fare_attributes.txt</a>	72 B	-	OK
<a href="#">fare_rules.txt</a>	55 B	-	OK
<a href="#">frequencies.txt</a>	42 B	-	OK
<a href="#">routes.txt</a>	136 B	1	OK <span>Routes: 1</span>
<a href="#">shapes.txt</a>	170.9 KB	4,633	OK
<a href="#">stop_times.txt</a>	4.0 KB	132	OK
<a href="#">stops.txt</a>	1.4 KB	21	OK <span>Stops: 21</span>
<a href="#">transfers.txt</a>	57 B	-	OK
<a href="#">trips.txt</a>	370 B	12	OK
<a href="#">feed_info.txt</a>	-	-	

## 10 Costs Associated with Open Transit Data



The costs associated with providing open transit data are shown in the following table<sup>28</sup>.

**Types of Costs Associated with Open Data (from survey responses)**

Types of Costs Associated with Providing Open Data	No. 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:		
<ul style="list-style-type: none"> <li>• Contract management</li> <li>• Cost to develop prediction software or use prediction Software as a Service (SaaS)</li> <li>• Everything above is already done for internal purposes and it is all automated</li> <li>• Investigation project agreement with the Faculty of Computing Sciences</li> <li>• Consultant time to build editing tool</li> <li>• License Routing service Mentz</li> <li>• No additional costs are incurred</li> </ul>		

The amount of time being spent on these activities by agencies that recognize their costs varies widely, as shown in the following table<sup>29</sup>.

**Labor Hours per Open Data Activity (from survey responses)**

Activity	No. 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

## 11 Open Transit Data Case Studies

### 11.1 Worcester Regional Transit Authority (WRTA)<sup>30</sup>

The Worcester Regional Transit Authority (WRTA) in Worcester, Massachusetts, is a transit authority in central Massachusetts with 41 fixed-route buses and 48 paratransit vehicles. After

the WRTA implemented technology on all of the agency's vehicles, they 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. In addition, the administrator of the agency at the time was 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 Chicago Transit Authority (CTA) to build apps for Android and iOS platforms because CTA had 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 their 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)

Two Android and one for iOS were developed and deployed for WRTA customers. In addition, WRTA learned from the "best practices" in open data employed by other transit agencies. WRTA used the following techniques to develop the agency's programs:

- Used a registration process to provide a key.
- Vetted 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,
- Used an official promotion "seal of approval" by the agency once a good relationship develops and there is proof of diligence by the developer.
- Relied on user reviews—let the market and comments be open, shared, and promoted by the agency.
- Created a web page for developers to formalize the process (this is under development).

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.

For smaller agencies, the WRTA suggests reusing work from other developers when possible to minimize the resources required to clean the data. 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 thought that their open data practice has high quality with low overhead. Some of the open data provided by the WRTA included:

- Branding— .css, colors, logos, etc.
- Interactive Voice Response (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 included 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/>).

## 11.2 Open Transit Data Resources

There are numerous resources for rural, small urban and tribal transit agencies to learn about open transit data and how to generate this data. First, the National Rural Transit Assistance

Program (RTAP) offers GTFS Builder, which is “a free Microsoft Excel-based web application which assists rural and tribal transit agencies to develop and validate GTFS for the services they provide. The GTFS Builder web application, along with a library of training videos and instructions, is available on the National RTAP website. The application is also fully supported through National RTAPs technical assistance center.”<sup>31</sup>

Second, the Center for Urban Transportation Research (CUTR) at the University of South Florida (USF) maintains a GitHub site that provides extensive information about open transit data, including a “Getting Started” page which is located at <https://github.com/CUTR-at-USF/awesome-transit>.

Third, the General Transit Feed Specification (GTFS) site maintains information that describes how to create a GTFS feed and to make a transit feed publicly available. This information can be found in <http://gtfs.org/getting-started/>. Another resource that is available at the GTFS website is GTFS data examples that show how rural transit agencies in Vermont and California describe specific operational information. These resources can be found in <https://gtfs.org/data-examples/>.

Finally, there is an “Open Transit Data Toolkit [that] provides ideas and instruction for making use of the growing amount of open data made publicly available by transit authorities across the US and around the world. The impetus for the development of the Transit Data Toolkit was the widening gap between growing data availability and the lack of accessible resources for generating knowledge and actionable insight from the data.”<sup>32</sup>

## 12 Conclusions

While transit software tends to be proprietary, there now is open source software that allows transit agencies to not be locked into a specific vendor’s solution. The benefits of open source software are numerous as described earlier in this paper, but it is not free. The cost of this type of software is shared among programmers, users, testers and other members of the community that works on open source software. Because of this cost-sharing, open source software tends to be less expensive over the life cycle of transit software than proprietary systems. There are several examples of open source software currently being used throughout the transit community including OpenStreetMap, RidePilot and 1-Click.

Open transit data drives many customer-facing systems, such as real-time transit information signage and smartphone apps. As mentioned earlier, many transit agencies across the world provide open data. Tools are available for agencies to produce open data such as those listed in Section 11.2. Further, as agencies contemplate new mobility services such as mobility as a service (MaaS), open data will play a critical role to ensure that the customers of these services have real-time information about their trips. Finally, the costs associated with providing open data primarily consist of the staff time that is necessary for validating the open data that is made available to the public.

## 13 References

In addition to the Endnotes, which contains many applicable references, the following documents can be used to obtain information about open source transit software and open transit data:

- Martin Catalá, **FTA Open Data Policy Guidelines**, prepared for the Federal Transit Administration, April 2016, FTA Report No. 0095, [https://www.transit.dot.gov/sites/fta.dot.gov/files/FTA\\_Report\\_No.\\_0095.pdf](https://www.transit.dot.gov/sites/fta.dot.gov/files/FTA_Report_No._0095.pdf)
- Ross MacDonald, **Mobility on Demand (MOD) Sandbox: Vermont Agency of Transportation (VTrans) Flexible Trip Planner**, prepared for Federal Transit Administration, Office of Research, Demonstration and Innovation, Final Report, FTA Report No. 0150, January 2020, <https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/research-innovation/146891/fta-report-no-0150.pdf>
- Bibiana McHugh, “FTA MOD Sandbox Grant: Multimodal Trip Planner,” TriMet Board of Directors Briefing, February 27, 2019, <https://trimet.org/meetings/board/pdfs/2019-02-27/presentation-2-27-19.pdf>
- Sean J. Barbeau, Ph.D., “Why Open-source Software?” January 15, 2019, <https://www.slideshare.net/sjbarbeau/why-open-source-software-128034178>
- Carol Schweiger, Sarah Anderson and Sean Barbeau, “Open Source Software and Open Data,” Webinar, May 21, 2020, <https://n-catt.org/tech-university/webinar-open-source-software-and-open-data/>
- Sean J. Barbeau, Ph.D., “Open Transit Data - A Developer's Perspective” May 28, 2020, <https://www.slideshare.net/sjbarbeau/open-transit-data-a-developers-perspective>
- Cecilia Viggiano and Glen Weisbrod with Shan Jiang, Emma Homstad, Melissa Chan and Sarah Nural, **Data Sharing Guidance for Public Transit Agencies – Now and in the Future**, Transit Cooperative Research Program (TCRP), TCRP Research Report 213, <https://www.nap.edu/download/25696#>
- Juho Lindman and Linus Nyman, “The Businesses of Open Data and Open Source: Some Key Similarities and Differences,” *Technology Innovation Management Review*, January 2014, [https://timreview.ca/sites/default/files/article\\_PDF/LindmanNyman\\_TIMReview\\_January2014.pdf](https://timreview.ca/sites/default/files/article_PDF/LindmanNyman_TIMReview_January2014.pdf)
- Connetics Transportation Group, Inc., **Maximizing Data Usage by Transit Agencies**, prepared for The Florida Department of Transportation - Freight, Logistics and Passenger Operations, September 2019, [https://fdotwww.blob.core.windows.net/sitefinity/docs/default-source/transit/documents/fdot-transit-agency-data-maximization-final-report.pdf?sfvrsn=6c831d55\\_2](https://fdotwww.blob.core.windows.net/sitefinity/docs/default-source/transit/documents/fdot-transit-agency-data-maximization-final-report.pdf?sfvrsn=6c831d55_2)
- Carol L. Schweiger, **Open Data: Challenges and Opportunities for Transit Agencies**, Transit Cooperative Research Program Synthesis 115, prepared for the Transportation Research Board, 2015, <https://www.nap.edu/catalog/22195/open-data-challenges-and-opportunities-for-transit-agencies>

- Sean J. Barbeau, Ph.D., “Open-source Transit Software” May 30, 2018, <https://www.slideshare.net/sjbarbeau/opensource-transit-software>
- Open Transit Software Foundation, <https://opentransitsoftwarefoundation.org/>
- OpenMobilityData, <https://transitfeeds.com/news/open-mobility-data>
- Catherine T. Lawson, Paul Tomchik, Alex Muro and Eric Krans, “Translation software: An alternative to transit data standards,” *Transportation Research Interdisciplinary Perspectives* 2 (2019) 100028, 20 August 2019, <https://reader.elsevier.com/reader/sd/pii/S2590198219300284?token=665DAB64743AD7886AD68265561D84AF13BA643F4823D10DFEEDD8113F9DCAC645BC616C9497B2BB7A7811468870C801>
- Open-source Software, [https://www.transitwiki.org/TransitWiki/index.php/Open-source\\_software](https://www.transitwiki.org/TransitWiki/index.php/Open-source_software)
- transitland, <https://transit.land/an-open-project/>
- Cambridge Systematics Open Source Software, <http://camsys.software/>

## Endnotes:

- 
- <sup>1</sup> Juho Lindman and Linus Nyman, “The Businesses of Open Data and Open Source: Some Key Similarities and Differences,” *Technology Innovation Management Review*, January 2014, page 1, [https://timreview.ca/sites/default/files/article\\_PDF/LindmanNyman\\_TIMReview\\_January2014.pdf](https://timreview.ca/sites/default/files/article_PDF/LindmanNyman_TIMReview_January2014.pdf)
  - <sup>2</sup> <https://blog.ldodds.com/2016/03/24/a-key-difference-between-open-data-and-open-source/>
  - <sup>3</sup> <https://opendefinition.org/od/2.1/en/>
  - <sup>4</sup> “What is Open Data?” <https://opendatahandbook.org/guide/en/what-is-open-data/>
  - <sup>5</sup> “What is open source?” <https://opensource.com/resources/what-open-source>
  - <sup>6</sup> Ibid, “What is open source?”
  - <sup>7</sup> Sarah Anderson, “Open Source Transit Software: The business case and tool examples,” N-CATT presentation, May 21, 2020, <https://n-catt.org/tech-university/webinar-open-source-software-and-open-data/>
  - <sup>8</sup> “Open standards are standards made available to the general public and are developed (or approved) and maintained via a collaborative and consensus driven process. Open Standards facilitate interoperability and data exchange among different products or services and are intended for widespread adoption.” [\[https://www.itu.int/en/ITU-T/ipr/Pages/open.aspx\]](https://www.itu.int/en/ITU-T/ipr/Pages/open.aspx)
  - <sup>9</sup> Sarah Anderson, “8 Open Source Benefits Every Agency Should Know,” *Metro Magazine*, September 13, 2017, <https://www.metro-magazine.com/10007433/8-open-source-benefits-every-agency-should-know>
  - <sup>10</sup> Sean Barbeau, “Why Open-source Software?” January 15, 2019, [https://www.slideshare.net/sjbarbeau/why-open-source-software-128034178?from\\_action=save](https://www.slideshare.net/sjbarbeau/why-open-source-software-128034178?from_action=save)
  - <sup>11</sup> Ibid, “8 Open Source Benefits Every Agency Should Know,”
  - <sup>12</sup> [https://www.transitwiki.org/TransitWiki/index.php/Open-source\\_software](https://www.transitwiki.org/TransitWiki/index.php/Open-source_software)
  - <sup>13</sup> Ibid, “8 Open Source Benefits Every Agency Should Know,”
  - <sup>14</sup> <https://opensource.com/resources/what-open-source>
  - <sup>15</sup> Email from Kevin Chambers to Carol Schweiger, July 1, 2020
  - <sup>16</sup> Software as a Service (SaaS) is defined as “a software distribution model in which a service provider hosts applications for customers and makes them available to these customers via the internet.” [\[https://www.infoworld.com/article/3226386/what-is-saas-software-as-a-service-defined.html\]](https://www.infoworld.com/article/3226386/what-is-saas-software-as-a-service-defined.html)
  - <sup>17</sup> Ross MacDonald, *Mobility on Demand (MOD) Sandbox: Vermont Agency of Transportation (VTrans) Flexible Trip Planner*, prepared for Federal Transit Administration, Office of Research, Demonstration and Innovation, Final Report, FTA Report No. 0150, January 2020, <https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/research-innovation/146891/fta-report-no-0150.pdf>
  - <sup>18</sup> Ibid, “Open Source Transit Software,” Slide 32

- 
- 19 Paul Sorensen, Cambridge Systematics; and Ryan Taylor and Clint Wilkinson, Utah Transit Authority, "Operational Impacts and Lifecycle Costs of Open Source RidePilot DRT Scheduling Software," presentation at TRB International Demand Response Conference, April 2019, Baltimore, MD
- 20 Ibid, "Open Source Transit Software," Slide 41
- 21 "Open Transit Data (OTD)," <https://www.soundtransit.org/help-contacts/business-information/open-transit-data-otd>
- 22 Carol L. Schweiger, *Open Data: Challenges and Opportunities for Transit Agencies*, Transit Cooperative Research Program Synthesis 115, prepared for the Transportation Research Board, 2015, page 16
- 23 Ibid, *Open Data: Challenges and Opportunities for Transit Agencies*, pages 13 and 14
- 24 Ibid, *Open Data: Challenges and Opportunities for Transit Agencies*, page 70
- 25 Ibid, *Open Data: Challenges and Opportunities for Transit Agencies*, page 71
- 26 Ibid, *Open Data: Challenges and Opportunities for Transit Agencies*, pages 34 and 35
- 27 Ibid, *Open Data: Challenges and Opportunities for Transit Agencies*, pages 49
- 28 Ibid, *Open Data: Challenges and Opportunities for Transit Agencies*, pages 49 and 50
- 29 Ibid, *Open Data: Challenges and Opportunities for Transit Agencies*, pages 49 and 50
- 30 Ibid, *Open Data: Challenges and Opportunities for Transit Agencies*, pages 65, 68 and 69
- 31 "Tools that Connect: Online Trip Planners and GTFS Builder," National RTAP, [http://nationalrtap.org/images/Tools that Connect Online Trip Planners and GTFS Builder.pdf](http://nationalrtap.org/images/Tools%20that%20Connect%20Online%20Trip%20Planners%20and%20GTFS%20Builder.pdf)
- 32 "Open Transit Data Toolkit," <http://transitdatatoolkit.com/>