NCMM
National Center for Mobility Management

Open-Source Software and Open Data: Key Concepts, Examples, and Opportunities for Mobility Managers

Open-Source Software and Open Data: Key Concepts, Examples, and Opportunities for Mobility Managers

Written by Kevin Chambers and Lara Bjork of
Full Path Transit Technology

For the National Center for Mobility Management
July 2021

About the National Center for Mobility Management (NCMM)

The National Center for Mobility Management is a national technical assistance center funded through a cooperative agreement with the Federal Transit Administration. The mission of the Center is to promote customer-centered mobility strategies that advance good health, economic vitality, self-sufficiency, and community. Content in this document is disseminated by NCMM in the interest of information exchange. Neither the NCMM nor the U.S. DOT, FTA assumes liability for its contents or use.

Table of Contents

Executive Summary

Photo by kilarov zaneit on Unsplash

This report compiles blog posts published on the National Center for Mobility Management website. The posts cover closely related topics: open-source software and open data. For both topics, the goals are similar: to define key terms and concepts, provide examples (both generalized and specific to public transit), and explore opportunities for mobility managers to further engage with these ideas.

The section on open-source software (OSS) focuses on:

- Defining "open-source software"
- Clarifying the nuances of "free" that are associated with this term
- Providing examples of different business models associated with OSS projects, including:
    - projects with significant corporate backing (Linux operating system and Firefox web browser),
    - a project driven by volunteers that has a decades-long history of competing with proprietary commercial counterparts (PostgreSQL database), and
    - a transit-specific example (OpenTripPlanner, also volunteer-supported)

- Describing one successful example of open-source software being used by mobility managers (at the Utah Transit Authority) as well as barriers to broad adoption of open-source software that is specific to the needs of mobility managers
- Outlining ways that mobility managers can participate in OSS projects

The section on open data focuses on:

- Defining "open data"
- Sharing examples of open data in our everyday lives
- Exploring examples of open data in the transit sector
- Forecasting what's on the horizon for open data and transit data standards
- Looking more deeply at the GTFS-flex data standard, what it can do for mobility managers, and how to get ready to open your agency's data

What Can Open-Source Software Do for Mobility Management?

When the topic of using technology in the public interest comes up, open-source software (OSS) is a frequent phrase on people's lips. But beyond thinking of it as a generally good thing in the world, few people outside of technology circles understand just what OSS is, what its costs and benefits are, or when to seek it as a solution. This type of software is already part of the larger transportation universe, although not in a context that is directly relevant to mobility management—at least, not yet.

In this section, we will:

1. Define what OSS is
2. Look briefly at how OSS has been used and compare it to proprietary software solutions
3. Find out about transit-specific examples of OSS
4. Examine reasons that there is not yet an OSS project specifically related to mobility management
5. List what mobility managers can do to be part of OSS efforts

What Is Open-Source Software?

To define open source, let's begin with the term "source." Software source code—which is what's meant by source—is the original human-readable set of instructions that produces working software. It's the rarely seen raw material you need to make changes to the inner workings of a piece of software.

When source code is "open," rarely does it mean that a developer has released their work with no strings at all. Instead, code becomes "open" when its owner releases it to the public under licensing terms that allow anyone to use it so long as they follow the terms of the license agreement. Those terms can vary a great deal. Some are very permissive with little more than a statement of no warranty—effectively, "don't sue us if you use this and something blows up." A standard, more restrictive licensing term requires that users of the code "share alike," or publicly release any changes they make to the code under the same conditions as the original work.

Licensing terms can get confusing very quickly, and it's possible to slap the "open" label on a software project but impose such onerous licensing restrictions that the source is effectively impossible to reuse legally. To address this, the Open Source Initiative (OSI) has developed a vetting process for licenses. Code released under an OSI-approved license can be trusted to be open with no "gotcha" clauses. Nonetheless, always read and understand the license when considering investing time and energy into an open-source software project!

The OSI logo

OSS is sometimes categorized as "free software," but this term brings risky ambiguity. The technology industry has come to distinguish the two meanings of "free" with handy catch-phrases—that a given tool is either "free as in beer" (closed source, with only the final software released to the public at no charge) or "free as in speech" (open source and also available free of charge). When considering a free software offering, be sure to clarify which kind of "free" is being offered. "Free as in beer" software carries additional risks, including malware, especially if you do not know the vendor well.

A Brief History of the Open-Source Movement

Publicly available software source code has been around since computers have been available to consumers. Organized advocacy for sharing code started in the 1980s, culminating in the creation of the GNU collaborative software project in 1984 and founding of the Free Software Foundation (FSF) in 1986. The operating system Linux made its first appearance in 1991 as a side project by then-college student Linus Torvalds, who released it under a share-alike license shortly after that. In 1998 Netscape kicked off what became known as the open-source movement when it publicly released the source code for its web browser, Navigator, which has since evolved into the Firefox web browser under the auspices of the Mozilla Foundation. The aforementioned Open Source Initiative was founded the same year.

These software projects and organizations, small and aspirational at their start, have reached astounding levels of success and social impact. Linux would not be a usable operating system without the tools that the GNU Project provides. Together, Linux and GNU software are now the dominant foundational tools of the "cloud" by a wide margin and serve as the internal plumbing for Android and ChromeOS. The FSF, OSI, and the organizations behind GNU, Linux, and Firefox all serve as models of transparent governance of technology. There are now dozens more foundations modeled on them that support other widely used software projects.

Open-source software and the underlying mechanisms for collaborative development have taken over the technology industry to such an extent that a return to purely

proprietary software is nearly unimaginable. Even Microsoft, for years a staunch opponent of OSS and whose former CEO once called Linux a "cancer," has completely reversed course is now a leading contributor to multiple influential OSS projects.

Not all OSS has the level of financial backing available to Linux and Firefox, though, as our next two examples of more obscure open-source projects will show. Our first example has stuck to a model of decentralized, collaborative management and development that competes with corporately supported products. Our second shows how transit leaders have leveraged OSS for an industry-specific success story. These two software success stories offer insight into what open source can achieve generally and within the field of transit. Those accomplishments have been possible because of committed individuals, solid governance structures, and sustainable sources of resources for development.

PostgreSQL

Pronounced "post-GRES-que-ell" and often referred to simply as "Postgres", this general-purpose relational database system has been around since 1986. It grew out of a federally funded project at the University of California at Berkeley, and has been under an OSI-approved open-source license since 1995 PostgreSQL version 13 was released in September 2021.

Of the hundreds of specialized and obscure open-source projects that keep the internet working every day, Postgres is a prime example of a success story. Twenty-five years after Postgres became open source, more than 500 people have contributed to the code over the years, and the project is sustainably managed and supported through donated funds, services, and volunteer time. DB-Engines ranks PostgreSQL among the most popular database products. It competes directly and successfully against proprietary industry heavyweights such as Microsoft and Oracle.

Postgres is able to compete and grow in popularity, David-versus-Goliath-style, without the stability of being owned by a single large corporation, in large part due to a rich ecosystem of committed developers and businesses. Some of its contributors are freelance programmers that can use their expertise to build their résumés, others are providers of hosting or consulting services for the database, while others yet are employees of large corporations that rely on Postgres as part of their core internal infrastructure. In other words, Postgres is not a utopian escapade fueled by the selfless acts of nerdy heroes (although there may be a bit of that). Rather, the project thrives because of a wide range of business models that all benefit from Postgres being a well-maintained shared resource. This has been called the cornucopia of the commons.

OpenTripPlanner

The world of public transit has its own OSS success story: OpenTripPlanner. It has some similarities to Postgres, but as a newer entrant into the tech world with a narrower function, it has a smaller reach and a different growth pattern.

OTP got its start with support from a grant from Oregon Metro, the Portland area's regional government agency. Metro awarded a grant to TriMet, the region's public transit agency, to develop a multimodal trip planner. TriMet hosted a three-day workshop in 2009 for transit and tech leaders, contracted with OpenPlans (which at the time had a software development team), and OTP was born.

OpenTripPlanner logo

At the time, it's unlikely that anyone would have predicted that 11 years later, OTP would be in use by at least 26 different government agencies in ten countries. More than 100 people have contributed to the OTP code. Version 1.0 was released in 2016, and work on version 2.0 is underway.

OTP is governed by the 12 members of the OpenTripPlanner Project Leadership Committee. Since 2013, OTP has been part of the Software Freedom Conservancy, an umbrella nonprofit organization that takes on the administrative functions of member projects. While not glamorous, having a decision-making body to guide the project and gain access to support for administrative, fundraising, and legal issues represents an important stage in OTP's growth and maintainability.

There are many more notable moments and technical developments in OTP's history, but we'll highlight three of them. In selecting these milestones, the goal is to emphasize how agencies have made choices that have helped meet their own needs while also contributing to the project's usefulness to the larger field, and to illustrate that major changes in the project have been closely tied to infusions of government funding.

Milestone #1: Development of Shared, Standards-Based Infrastructure

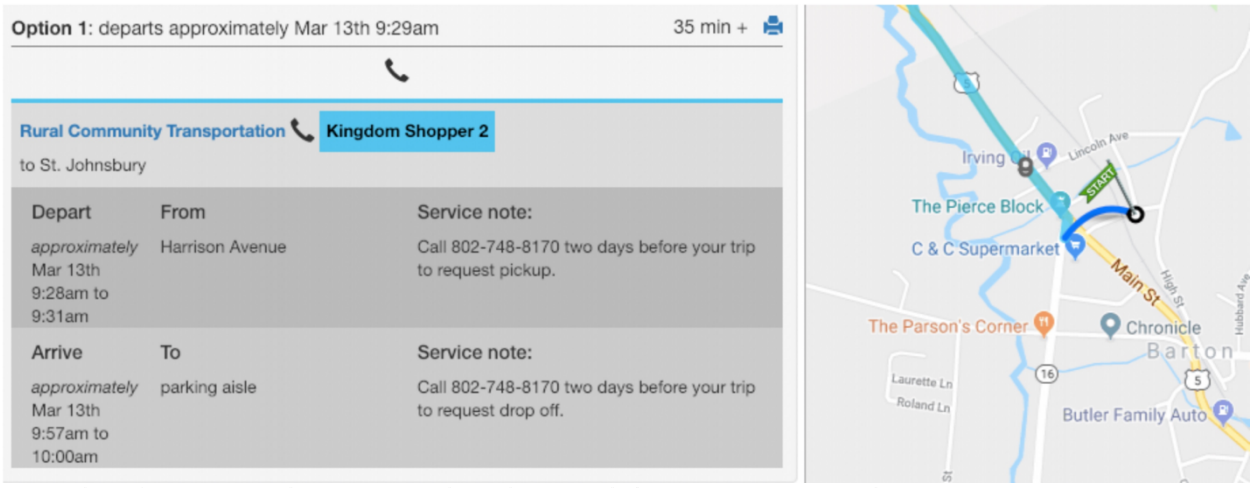
When TriMet wanted to get an improved trip planner with lower costs, they could have worked in isolation to do so; it would be reasonable for an agency to focus attention on meeting their own needs, after all. Instead, they pursued one-time funding that allowed them to initiate a collaborative development process. In effect, they set out to build not just their own solution, but a piece of IT infrastructure that's available to anyone. A further bonus: Transit agencies can use this powerful software without going through a potentially painful procurement process.

In addition to the software itself being open source, OTP has from the beginning relied on open data standards and open data (more on these topics in the second half of this report). Routes and schedule data can be loaded into the planner using the General Transit Feed Specification (GTFS). The default data source for maps and road networks is OpenStreetMap, a world-wide crowd-sourced GIS database available under an open-source license.

Milestone #2: Addition of GTFS-flex

OTP made it possible for riders to consider walking and biking options in addition to fixed-route transit as they planned their trips. But it wasn't possible to plan trips that included flexible transportation services such as taxis or ADA paratransit—in part because it wasn't possible for transit agencies to provide that information in a standardized computer-readable format. The aforementioned GTFS format only supported fixed routes and predefined schedules.

Fortunately, the world of transit standards continued to evolve: first proposed in 2013, GTFS-flex extends GTFS to include demand-response services. By 2017, the proposed spec was mature enough for the Vermont Agency of Transportation (VTrans) to bring this new standard into OTP for a statewide transit planning application that could account for more flexible forms of transit as well. VTrans received funding for this project from the Federal Transit Administration's Mobility on Demand Sandbox Program, which supports demonstration projects. Like TriMet's initial funding from Metro, the FTA grant to VTrans fueled the innovation but did not provide an ongoing source of funding.



Screenshot of VTrans Trip Planner Deviated Fixed Services link. Source: FTA Research Report No. 0150

Milestone #3: Going National

When it comes to transit, the differences between the United States and European countries are considerable, down to the basic terminology ("public transit" over "transport") and historical relationship to data standards, among other things. One area of commonality, though, is the use of OTP. The Netherlands, Finland, and Norway have all

worked with OTP at the national level, and we'll take a closer look at Norway as an example of open-source software's flexibility.

In 2016, railway reforms by the Norwegian government led to the creation of a publicly owned company called Entur. One of Entur's core mandates is nationwide trip planning for all public transport. To carry this out, Entur:

- Collects all public transport data for Norway (timetables, real-time data, and national registry of stops);
- Provides an open journey planning API;
- Maintains a national journey planner (as a supplement to local trip planners, rather than a competitor, with the intent of making national travel easier); and
- Acts as a national hub for open data.

Entur began working with OpenTripPlanner from the start, but they soon came to a choice point. OTP deployments relied on GTFS to load route and schedule data, but a different set of standards was more commonly used in Norway and much of Europe already: Transmodel, which uses the NeTEx standard for static data and SIRI for real-time data. Entur decided to pursue an OTP implementation that would be compliant with Transmodel.

If OpenTripPlanner had been a proprietary system, Entur wouldn't have had the flexibility to combine existing software with a set of standards that was a better fit for its needs. This adaptability is one of the clear benefits of open-source software.

This decision also set Norway up well to comply with a European Union mandate for each country to have a national access point for transit data. The goal is to ultimately support travel planning throughout Europe. Under that mandate, use of NeTEx is required and SIRI is recommended. Although Norway is not a member of the European Union, Entur's pursuit of a Transmodel-compliant planner sets them up to be able to connect to a continent-wide journey planning system. And because all of Entur's products are open source, their Transmodel-compliant version of OTP can be readily used by other European nations. You can find more details about Entur's work on OTP from documents shared from an April 2019 summit.

Considerations for Using Open-Source Software

Fundamentally, open-source software is software. It possesses no inherent special powers to solve problems or do good. An agency considering using open-source software should assess its value through the same sort of cost-benefit analysis that would apply to any other technology system. Even when a piece of software is "free" in both senses (available with a license that allows modification and also at no charge) and very well made, there will still be costs associated with its implementation. The more specialized the tool and the higher the impact of the system on the agency, the more scrutiny that should be brought to bear when considering an investment. And if any open-source project is to

remain sustainable for the long haul, it will need ongoing support from its committed user base, be it monetary or otherwise.

What can make open-source software a compelling choice are the human and institutional factors behind the software: The organizations that coordinate efforts and the business models of the private companies that carry out most of the development on open-source software projects.

Are There Already Open-Source Software Projects for Mobility Management?

The short answer: No. Although OSS is used in public transportation contexts, it's not really part of mobility management's small slice of that big pie. In addition to OpenTripPlanner, examples of current transit-related OSS projects include:

- OneBusAway, along with TheTransitClock, all of which provide real-time vehicle data
- OpenStreetMap, a community-driven effort to map the world
- RidePilot, a scheduling and dispatching system, originally conceived by this author, designed to meet the needs of a network of small demand-responsive transportation providers

For a detailed list that covers OSS and a multitude of transit-related technology, see this community-generated compilation from the Center for Urban Transportation Research at the University of Southern Florida.

While there are multiple organizations supporting *open data* for transit (like MobilityData), currently, the worlds of mobility management and OSS don't overlap much, even though there are many overlapping values, such as the mutual emphasis on collaboration and sharing of resources.

What's Getting in the Way?

The Utah Transit Authority (UTA) is one of the premier agencies in the United States that uses open-source tools to support mobility management. Ryan Taylor, UTA Coordinated Mobility Manager, describes mobility managers' challenges: "We're really fortunate in Utah. We have a large staff and we have an investment from our agency. And we have support with the direction that we've gone. That being said, I think that a lot of mobility managers are stretched really thin. I think they're asked to do a lot with very little, many are working on a shoestring budget. I think it's difficult for them to want to jump into an unknown world of open source as a mobility manager."

There are at least three barriers to connecting mobility managers with OSS.

- **Staff capacity:** Generally, mobility management is handled by one or two people at an agency, often along with other duties. Adding the responsibility of engaging in an OSS project's development or maintenance would be burdensome under most circumstances.

- **No OSS counterpart:** There is no open-source software development organization directly dedicated to supporting smaller-scale transit—in other words, the scale at which mobility management tends to operate.
- **Expertise:** Most mobility managers don't have a background in technology, so even if they had time in their days to take on OSS, the learning curve would be steep.

Capacity building could help overcome these barriers, but it would take a concerted effort to make a big impact.

Scale: A Key Ingredient

As hinted at in the overview of the Postgres and OpenTripPlanner projects, open-source solutions are most powerful when there is a critical mass of users and contributors. Ryan Taylor values open source for "the ability that it gives you to build upon and share without having to have multiple investments in the same thing." The most efficient way for open-source solutions to take hold is for a large funder (national or at least state-level) to provide seed funding for multiple implementations. That way, from the beginning a community of users can share the benefits as new features and customizations are developed.

What Can Mobility Managers Do Now?

There are ways to get involved with OSS, even before a large funder steps up to support the first OSS project dedicated to mobility management.

- **Use OSS as a consumer.** There are numerous open-source projects that are mature and ready to go regardless of your technology background. There's Firefox for browsing the web, Wordpress for creating websites, GIMP for editing images, and LibreOffice for productivity, all of which are designed to be easily installed (or hosted, in the case of Wordpress) and used. Go forth and download today.
- **Join an existing OSS community.** You don't need to be a programmer to participate in an open-source project. All the projects just listed above always need help with user testing, documentation, and language translation. Finally, there's governance; if you get involved with a project and are invested in its success, you may consider being a voice in charting its course for the long haul.
- **Expand your own and your organization's technology chops generally.** As stated previously, OSS is first and foremost software. If you work to understand the software landscape overall, you'll be better equipped to see the most realistic opportunities for OSS compared to proprietary software.
- **Advocate for OSS.** When your agency is procuring new software, support the exploration of OSS options. If there are OSS projects such as OpenTripPlanner that you think should be considered when making technology investments, work to make sure that formal procurements reflect that interest.
- **Emulate OSS principles.** Computer code isn't the only thing that can be collaboratively developed and maintained. Wikis like Wikipedia and TransitWiki allow users to contribute and edit content. Is there information that needs input from many people in your

organization that could be put into a wiki? Or do you have thoughts to add to an existing wiki? At a broader scale, emulating OSS principles can mean promoting government transparency.

Is It Time for an Open-Source Mobility Management Project?

At this time, it's unclear. There are precedents outside the transit/mobility realm for organized OSS development, such as the electronic medical record software available through OpenEMR. What would it mean for our sector to be "ready" to launch an open-source development effort paired with a nonprofit foundation? A good starting point could be identifying an unmet need for specific software.

What Can Open Data Do for Mobility Management?

Much like open-source software, open data is already playing an important role in transit, and the opportunities for more impact in the world of mobility management are growing.

In this section, we'll be exploring:

1. What "open data" means
2. What's possible in our everyday lives because of open data
3. What's possible in transit because of open data
4. What's on the horizon for open data and transit data standards
5. GTFS-flex, what it can do for mobility managers, and how to get ready to open your agency's data

Defining Open Data

The term open data is made up of two words that are so common and familiar that it's tempting to rely on an intuitive understanding of what they mean when combined into a single concept. However, like most technical terms, open data has a fairly specific meaning. To appreciate the power of open data, we need to start from the same definition.

The Open Knowledge Foundation presents a comprehensive definition of "openness" as it applies to open data and open content, summarizing their definition this way: "Open data and content can be freely used, modified, and shared by anyone for any purpose." It's not a coincidence that this definition sounds similar to that of open-source software: The open-source movement inspired further work to apply the same underlying principles to data.

What Makes a Dataset Open Data?

Another way to capture the essence of the open data concept is to describe open data as both "technically open and legally open," as presented in the California Open Data Handbook. The technical openness of data refers to how accessible it is; the legal openness of data refers to the license that gives users permission to use the data once they've accessed it.

Accessibility of Data

It's possible for data to be available to the public while still not being easy to acquire and analyze at a large scale. To be fully in the spirit of openness, data should be:

- **Machine readable**, without a requirement to purchase any particular type of software. To the degree possible, data should be presented in standard formats which themselves are open. Proprietary formats, such as those used by older versions of Excel, are by definition not open. For more information on machine readability and examples of open formats, see this primer from Data.gov.
- **Well documented.** At its best, technically open data has metadata ("data about the data") or other documentation that explains what it is, where it comes from, and how it's maintained.
- **Easily searchable.** Especially for large data sets, the holder of the data should provide some way to find the desired subset of the data so that a consumer does not need to download and sort through terabytes of files.
- **Readily accessible.** Users can access technically open data without paying money or logging in with a credential or account.

Permission to Use Data

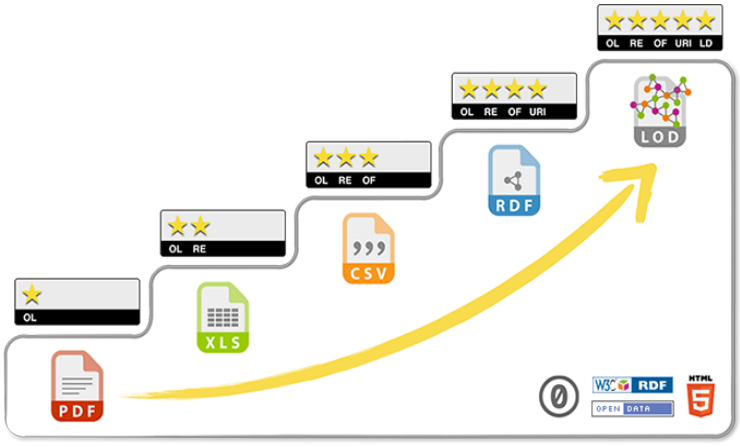
Legally open data has a license that clearly states that the data can be used with no restrictions, for commercial and non-commercial reasons.

Openness as a Continuum

It might seem like openness of data is a binary issue—either data is open or it's not. In practice, it's more like a window or a door, which can be ajar a little bit, open as far as it can possibly be opened, or anywhere in between.

Sir Tim Berners-Lee, who invented the World Wide Web, created a five-star system to mark specific points along a continuum of openness. To be considered "open" by his definition, data must earn at least 3 stars.

★	**Online:** Data is discoverable and available online, in any format
★★	**Standardized:** Online and accessible in a structured data format
★★★	**Open Format:** Online, standardized, *and* freely usable, in a nonproprietary format
★★★★	**Universal Reference:** Online, standardized, freely usable, *and* is anchored with a Uniform Resource Identifier (for example, a web address)
★★★★★	**Contextualized:** Online, standardized, freely usable, universal reference *and* linked to other datasets to provide context



The 5-star deployment scheme. Source: 5stardata.info

These examples provide a good overview of what's possible with data at each level of the five-star system.

Is Public Data the Same Thing as Open Data?

Data produced by government entities is quite often public data—but just because data is shared with the public, it does not mean that it is open data.

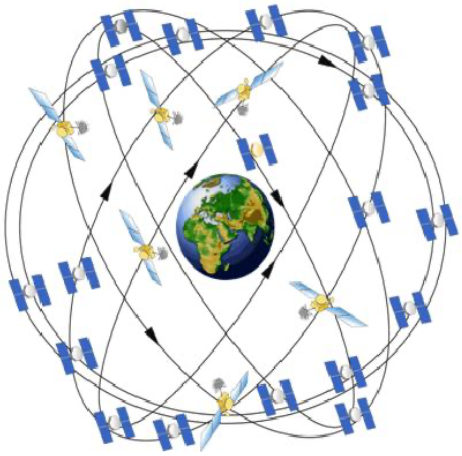
For example, posting a document in PDF format on a government website makes it publicly available. However, it's not *technically open* because a computer often can't reliably read the contents of a PDF. You can consult the document and draw conclusions from it, but if you want to take data from the PDF, reorganize it, analyze it, or otherwise manipulate it, you may have to manually extract the data or turn to sophisticated "scraping" software to convert it into something more accessible.

Open data doesn't just happen automatically uploading files to a website. Making data fully open requires a thoughtful effort that considers the particulars of the data itself, its potential consumers, and the barriers those consumers may encounter as they work to transform it into meaningful information. When such efforts are successful, the potential social gains become apparent.

Open Data in Our Everyday Lives

Open data is already a part of the everyday lives of most people in the United States. For example:

- **Public health:** The COVID pandemic dominates world news, and because of open public health data, researchers and reporters alike can analyze and disseminate the latest statistics and trends.
- **Weather:** In the 1970s, the National Oceanic and Atmospheric Administration (NOAA) began releasing weather data, which informed local weather reports on the nightly news. NOAA's open data now powers numerous weather apps.
- **Global positioning systems (GPS):** Long before GPS data provided directions to countless drivers, it was a military project, closely guarded by the United States government.
- **US Census:** Combing through handwritten census records when you're researching one of your ancestors for a family history is not terribly efficient, but it's often manageable. However, if you're trying to understand recent demographic trends, reviewing individual responses to the 10-year census would be



The GPS satellite system. Source: nasa.gov

almost impenetrable, at least if you wanted to work quickly. Thanks to the openness of US Census data, analyses of all kinds can be undertaken.

For end users, it's fairly straightforward to deal with open data when 100 percent of that data comes from a single source, as with NOAA and the US Census Bureau. Good data management practices internal to each agency ensure that the same type of information is available, and in the same format, on a consistent basis.

But what happens when users want to aggregate and analyze data that has been collected and stored by different agencies? The possibilities for variation abound, and if variations across data sets are too numerous, collective analysis may not be possible.

A solution to potential variation is the use of *data standards*, which are a set of shared expectations for communication between systems, much like the rules of a common language between people. Most of the time, computers exchange data in an established structure. A data standard defines the overall structure a data producer must adhere to, along with the types and formats of the data elements that fill that structure, with the goal that the data can be reliably interpreted by a data consumer.

When a data standard gets established and multiple data producers use that standard to structure their open data, new possibilities open up for anyone who wants to use that data. Public transportation in the US began its first significant foray into producing standardized open data in 2005 with the creation of what is now called the General Transit Feed Specification. The result has been a significant transformation in how riders engage transit, as well as how transit services are planned and operated.

How Open Data Has Already Changed Transit

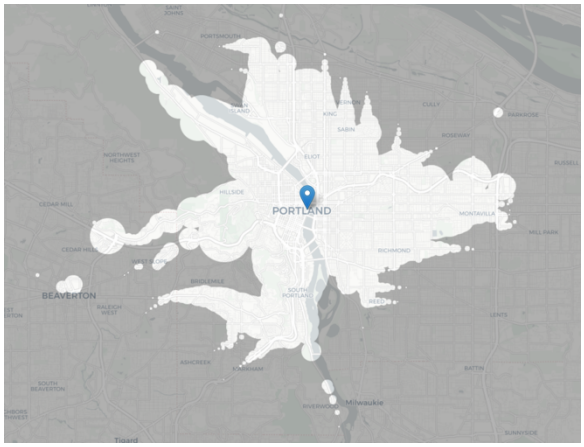
When you share a common language, a lot more becomes possible. Let's take a look at what standards and open data have done for transit and what's still to come.

Given that the most familiar data about transit for the public are timetables for fixed-route schedules, it's not surprising that open data's impact on transit began with this type of information. In 2005, Google and TriMet, the transit agency in Portland, Oregon, co-created what is now known as the General Transit Feed Specification, or GTFS. At first, it was called the Google Transit Feed Specification, but the name was changed in 2009 as the specification was embraced far beyond its initial developer.

GTFS describes fixed-route transit in terms of all the factors that affect schedules: agencies, routes, trips, stops, stop times, fares, and holidays. TriMet and Google worked together to develop this data standard so that the agency's data could be shared with Google and then incorporated into Google Maps. Other agencies were encouraged to provide their data in this format. With this data available, users can plan trips within what's now called the Google Transit application.

As transit information became available on Google Maps, other groups promptly saw opportunities. App developers, urbanists, planners, and policymakers began creating new tools that leverage open data. For example:

- Mapnificent, which "shows you the area you can reach with public transport from any point in a given time"
- WalkScore, which uses GTFS data to help evaluate a neighborhood's walkability
- Remix, Conveyal Analysis, and similar planning tools combine GTFS data with census data and other data sources to help agencies design fixed-route services
- OpenTripPlanner, an open-source trip planner, would not have been a viable project without a reliable open data format in which to load transit services
- Rome2Rio, which is one example of how GTFS data can be aggregated with other data sources (such as air travel information) to create long-distance itineraries that can string together many transportation providers



Mapnificent screenshot for Portland, Oregon.

More examples of GTFS-consuming applications can be found at the Transitwiki website.

In short, the initial partnership between Google and TriMet created a virtuous cycle. The development of tools that make data more useful results in greater demand for data and improvements to the specification. As more and better data becomes available, more and better software tools are developed. This is the ecosystem that open data can foster.

While fixed-route schedules were a compelling place to start standardizing transit data, it's far from the only such area where standards can bring value. Next, we'll explore how the ecosystem started by the GTFS is expanding and tackling increasingly complex data challenges.

What's on the Horizon for Open Data and Transit Data Standards

GTFS provided the starting point for standardized open transit data's impact on the transportation sector. Now that it has been in existence for more than 15 years, we can take a look at what's been added to GTFS, what other standards are out there that don't rely on GTFS, and important implications for data privacy.

As discussed previously, GTFS describes fixed-route transit, in terms of agencies, routes, trips, frequency of service, stops, stop times, fare rules and attributes, and calendar dates. Other aspects of fixed-route service and other types of service aren't well accounted for in the original GTFS standard. In some cases, the best option is to build on GTFS by creating

a GTFS extension (essentially, an additional set of files and fields to add to a GTFS dataset).

Some of the GTFS extensions already in use or in development include:

- GTFS-realtime, which creates the ability to share real-time information about vehicle locations, for instance
- GTFS-flex, which incorporates demand-responsive service (Because GTFS-flex is especially relevant to mobility managers, we'll discuss this extension in more detail later on.)
- GTFS-ride, which describes ridership (the number of people on a vehicle, as well as the number getting on or off at a particular stop)
- GTFS-eligibilities and GTFS-capabilities, two specifications currently in development, to capture what rider characteristics allow someone to access a specialized transportation service and what service characteristics support very specific needs of riders (Full disclosure: The authors of this report are under contract with the Oregon Department of Transportation to lead the development of a draft proposal for these two specifications.)
- General On-demand Feed Specification project, which will extend GTFS to include on-demand services, such as taxi cabs and ride-hailing services

GTFS has been widely adopted, and at first glance, it might seem like all transit data standards should stem from this core specification. In fact, as new data specifications are considered, it's important to assess whether GTFS is the right place to start. GTFS and its extensions work very well for discovering whether a trip is possible, under certain conditions (from point A to point B, at this specific time, for example). But the trip life cycle includes more than discovery, and that's where specifications that aren't GTFS extensions come into play.

- TCRP Report 210 presents a transactional data standard. This specification is aimed at simplifying the under-the-hood process for booking a range of demand-responsive services.
- E-fare standards, still very much in their infancy, can make it possible for a single payment instrument to be used across agencies, using different vendors.
- The Mobility Data Specification (MDS) covers shared transportation options like scooters, bikes, and cars. Created by the Los Angeles Department of Transportation, MDS supports communication between private-sector operators of these services and public agencies. This specification has been the subject of recent legal proceedings that are worth noting.

  A recent case in a federal court examined the privacy rights of e-scooter users in Los Angeles, where micromobility companies are required to share their service data in real time with the LADOT using the MDS standard. As of March 2021, the federal judge ruled in favor of the LADOT, finding that the data collected did not violate the plaintiffs' legal or privacy rights.

  This case raised the question of whether anonymized data can be "de-anonymized". For example, even though an individual's personal information (name, address, payment information, etc.) was not shared with the agency, could the shared data be used to infer an individual's behavior (e.g., a consistent trip from one location to another that could expose an individual's residence and/or workplace)? Although the judge's ruling in this

matter upheld the use of MDS, the questions raised in the case highlight the importance of considering the implications of opening data when the privacy of system users is at stake.

There are numerous benefits of open data: Easily shared information that allows new applications that support riders and new insights into transportation. It's also important to examine the potential downsides of opening data widely. MDS offers an example of open data colliding with privacy concerns.

GTFS-flex, What It Can Do for Mobility Managers, and How to Get Ready

As noted previously, transit's foundational data standard is the General Transit Feed Specification (GTFS). This data standard has prompted many transit providers to open up their data. GTFS was created to describe fixed-route transportation services, though, and so "extensions" to GTFS have been added to address gaps in the original GTFS.

GTFS-flex is one of the most important extensions for mobility managers because it incorporates demand-responsive service and other, more flexible approaches to transportation services. For a thorough look at the history, status, and uses of GTFS-flex, see this white paper on GTFS-flex written by Weston Shippy and Thomas Craig of Trillium Transit for the National Center for Applied Transit Technology (N-CATT).

Work on this extension started in 2013, and by 2017, the Vermont Agency of Transportation (VTrans) launched the first trip planner that incorporated GTFS-flex data. The Federal Transit Authority provided funding for this project via the Mobility On Demand Sandbox program.

Like many if not all data standards, refinements to GTFS-flex continue. There is currently a draft proposal for GTFS-flex Version 2, which is likely to be finalized in 2022, if not sooner. The proposal and approval process is being managed by MobilityData, a nonprofit organization, which plays this important role for many GTFS extensions, as well as the General Bikeshare Feed Specification.

With a new version of GTFS-flex on the horizon, there is great promise for trip planning tools that will give a more complete picture of specialized transportation options. Accordingly, it's a good time for mobility managers to assess their agencies' readiness to open their data via a GTFS-flex feed. We spoke with Carol Schweiger of Schweiger Consulting LLC, who recently wrote a white paper on open-source software and open data published by N-CATT, to get her perspective on the benefits of open data for mobility managers and steps to take before opening data.

Carol offered several benefits of open data:

- **Awareness:** "By putting open data out there, you can eventually increase awareness of your service."

- **Access:** "Opening your data results in better access—for individuals, other agencies you already work with, and potential partners."
- **Coordination:** Open data increases opportunities for coordination with other agencies and can make that coordination easier.
- **Transparency:** "Whether you're a public or private entity," open data "shows what agency spends money on, and shows how data is being used to meet transit goals."

When it comes to your first forays into open data, Carol recommends approaching the task carefully. "It's not an instant process. Be thoughtful, and take it slow. There's work that needs to be done prior to opening data." That work falls into two categories: data and organizational capacity.

On the data side:

- Look at your goals. Are you trying to coordinate your services with other providers? Create new partnerships? Something else? Whatever your goal is, keep it in mind as you look back at your data to identify gaps.
- Understand what data you have and what you're currently doing with it. For example, whatever you have for scheduling, make sure you understand what's in it.
- What's the quality of your data?

As far as organizational capacity:

- Build your data literacy by looking at the open data others are producing.
- You need to be able to understand data—but that doesn't mean that you have to become a programmer.
- If you don't already have in-house capacity for data-related work, consider whether you will train, hire, or bring in a consultant who has data skills.
- Plan for time to monitor your data; you'll need to fix errors and maintain how you're putting data out into the world. Keeping up with data monitoring isn't easy for a small organization. Monitoring is a hidden cost of open data, and even if you outsource this task, you have to have a liaison to that person.
- Collaboration with other agencies about your data also takes time and represents another hidden cost.

See Carol's white paper for more information on the benefits and challenges of open data, examples and case studies, and associated costs. These topics are also covered for open source software.

In addition to Carol's guidance, we offer one final caveat for any data-driven project: the more data you have available to you, the more tempting it is to measure everything. But a more targeted approach guided by specific goals and questions will help you get meaningful data—and will help you balance the workload required to maintain that data.

Conclusion

Open-source software and open data are powerful concepts that are worth understanding, even if they are not yet fully part of the day-to-day work of mobility managers. As with so many aspects of technology, neither will magically solve problems, once and for all. But with a solid grounding in the benefits, limitations, and potential costs associated with projects that use open-source software and/or that release transit agency data as open data, mobility managers will be better positioned to make good decisions about related technology.