

The Data Transit Riders Want



**A Shared Agenda for
Public Agencies and Transit
Application Developers**

The Data Transit Riders Want

A Shared Agenda for Public Agencies and Transit Application Developers

TransitCenter works to improve public transit in ways that make cities more just, environmentally sustainable, and economically vibrant. We believe that fresh thinking can change the transportation landscape and improve the overall livability of cities. We commission and conduct research, convene events, and produce publications that inform and improve public transit and urban transportation. For more information, please visit www.transitcenter.org.

TransitCenter Board of Trustees

Rosemary Scanlon, Chair

Eric S. Lee

Darryl Young

Emily Youssouf

Jennifer Dill


Clare Newman

Christof Spieler

Publication Date: December 2018

1 Whitehall Street, 17 Floor, New York, NY 10004

www.TransitCenter.org

 @transitcenter

 [Facebook.com/transitctr](https://www.facebook.com/transitctr)

Contents

Acknowledgements	1
Introduction	2
Setting the Stage	6
Public agencies	10
Transit application developers	13
Challenges and opportunities	14
An Agenda for Better Transit Data	16
1. Data Management and Policy	17
2. Data Quality	24
3. Data Specifications	28
Conclusion	36
Endnotes	39
Appendix	40



TransitCenter



Acknowledgements

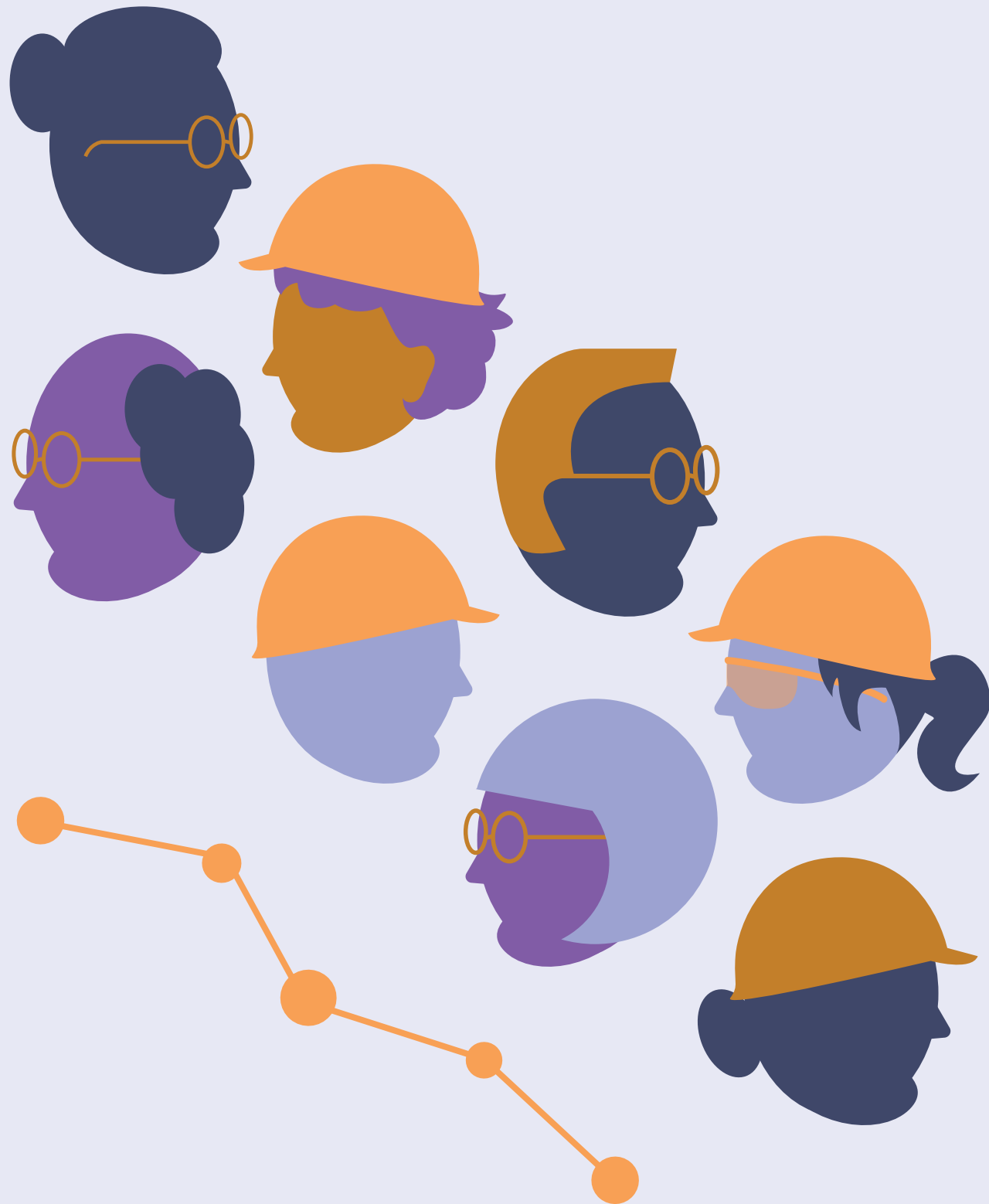
This paper was inspired and informed by the participants of the 2017 Interoperable Transit Data Workshop¹, held at TransitCenter's office in New York City on October 18-19, 2017. The workshop, developed and hosted jointly by TransitCenter and Rocky Mountain Institute (RMI), featured over 35 speakers and attendees from transit agencies across the United States and private firms hailing from North America, Europe, and Australia. Participants were brought together for the two-day workshop to identify shared challenges and opportunities to improve transit data practices across the industry, yielding more reliable and convenient travel experiences for transit riders around the world.

The Data Transit Riders Want is authored by Zak Accuardi, Chris Pangilinan, Aaron Antrim, and Greg Rucks, with contributions from Alyssa Wright. The authors are grateful for thoughtful reviewer input from Léo Frachet, Ruth Miller, Kurt Raschke, Jonathan Wade, RMI's Kelly Vaughn, and TransitCenter's David Bragdon, Tabitha Decker, Kirk Hovenkotter, Jon Orcutt, Hayley Richardson.

TransitCenter is a New York City based foundation that works to improve public transit in ways that make cities more just, environmentally sustainable, and economically vibrant. Rocky Mountain Institute, an independent nonprofit with offices in Boulder and Basalt, CO; New York City; Washington, D.C.; and Beijing, China, works to transform global energy use to create a clean, prosperous, and secure low-carbon future. TransitCenter and RMI share the belief that improving public transportation is fundamental to creating livable, sustainable cities, and that better transit requires interoperable transit data.

This report contains a number of links to valuable outside resources. They are noted as purple text. Please visit TransitCenter.org/publications/transit_data to read the report online and access those links

Introduction



High quality transit data that digitally depicts, in real-time, schedules, route networks, vehicle locations, fare structures, and ridership patterns makes life easier for transit riders, public agencies, and application developers.

With access to improved transit data feeds, riders are able to make better-informed travel decisions; agencies preserve and grow ridership and improve their planning and operations capacity; and third-party trip-planning apps are able to offer a more seamless travel experience to an expanded user base (Google Maps alone has more than a billion daily users²).

Each of these outcomes is a win for riders, yet despite this, data programs at most public agencies lag behind industry best practice. Implementation of existing data specifications is inconsistent, with many agencies releasing low-quality data that makes travel more difficult for transit riders in every app. Absent champions in senior leadership, data departments lack the resources necessary to do more than maintain the same data the agency has been relying on for years, while the rest of the transportation industry is undergoing rapid, iterative technological change.

With this change comes opportunity, and strong alignment between public and private sector goals creates opportunities for symbiotic collaborations. This alignment is embedded in the history of transit data specifications, as the General Transit Feed Specification (GTFS) and its GTFS-realtime extension have been built on a rich history of public-private collaboration.³ These specifications have evolved over time, but not fast enough to meet the needs of transit riders and practitioners in both public and private sectors.

Data producers (transit agencies) and transit application developers (private companies) need to develop a set of actionable, shared priorities that will enable transit data specifications and their implementations to match the needs of transit riders today. This was the motivation behind an October 2017 workshop hosted jointly by TransitCenter and Rocky Mountain Institute (RMI), which brought

High quality transit service requires high quality data

Agencies must not only invest in their physical infrastructure, but their data infrastructure

together a diverse group of public and private stakeholders to start setting a new agenda for transit data specifications and compiling best practices for their implementation. RMI and TransitCenter, in collaboration with transit data producers and application developers, aim to continue to carry this agenda forward through a combination of digital and in-person convenings, peer-reviewed research, and development of improved data specifications and open-source software.

After reviewing notes from the workshop's proceedings, as well as ongoing conversations with transit data practitioners in the public and private sectors, TransitCenter and RMI compiled our findings in this summary report. The report's recommendations are targeted at public transportation agencies and transit application developers, and are intended to yield improved and expanded transit data capabilities that can deliver an improved experience for transit riders in North America and beyond. Those recommendations are organized under three headline themes:

I. Data management and policy

Producing high quality, publicly available data must be a priority for transit agencies that seek to improve their service for riders. The first step in doing so is for agencies to recognize the importance of their data and make data infrastructure a priority.

II. Data quality

Comprehensive and widely available data is only valuable if it is accurate and timely. If it is not, riders suffer from poor trip planning information, agencies are unable to rely on their own data to improve planning and operations, and application developers as well as agencies can quickly lose rider trust. Transit agencies should invest in their data hardware and software systems to ensure they are able to produce high quality data, and application developers should work with each other and with transit agencies to provide feedback on data structures.

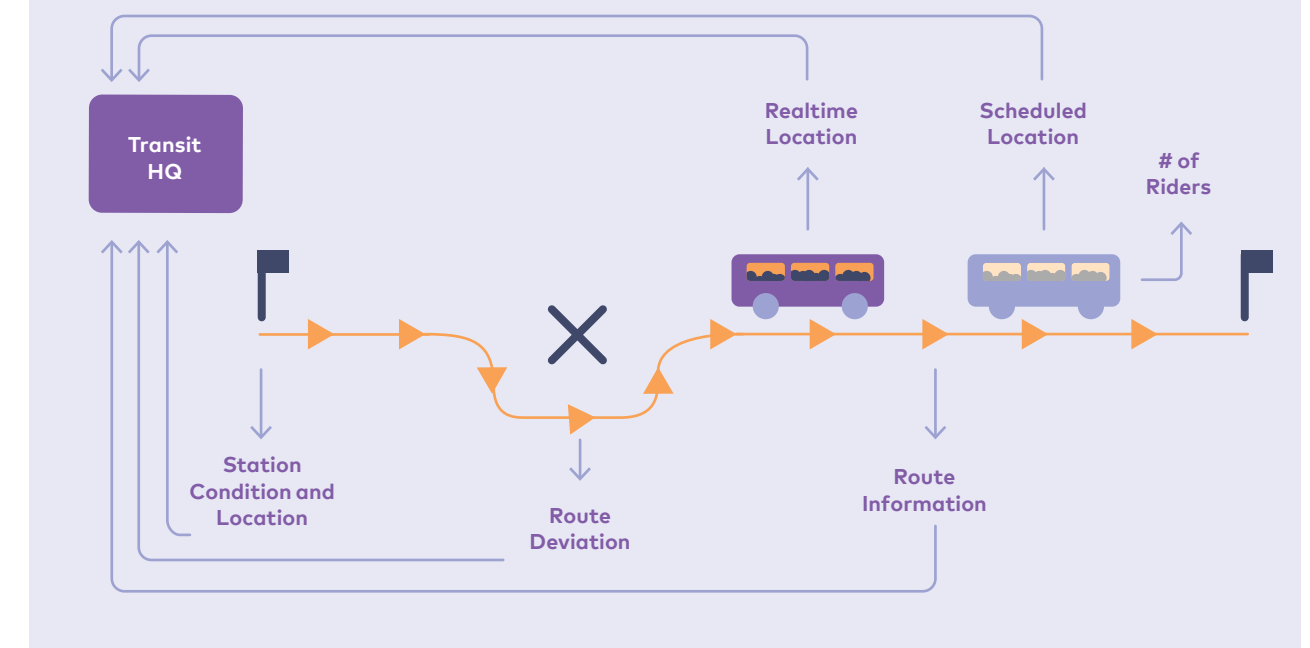
III. Data specifications

Since the release of GTFS in 2006 and GTFS-realtime in 2011, online trip planning is now nearly ubiquitous and multiple third-party application developers have flourished by bringing accurate

Agencies and developers must speak the same language for their data to be useful

transit routing information to millions of riders. Agencies have also benefited by having more data available to inform planning and operational decisions. Expectations for information have only increased, and this has put pressure on the existing data specifications to evolve. Agencies and application developers will need to continue working together to expand on GTFS to be able to bring riders and agencies themselves more information, such as in-station routing and temporary changes to transit service.

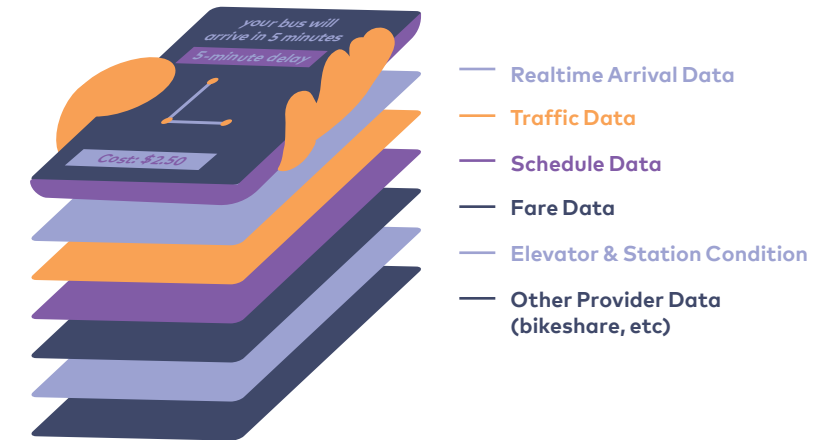
Figure 1. What is Transit Data? A 101



Setting the Stage

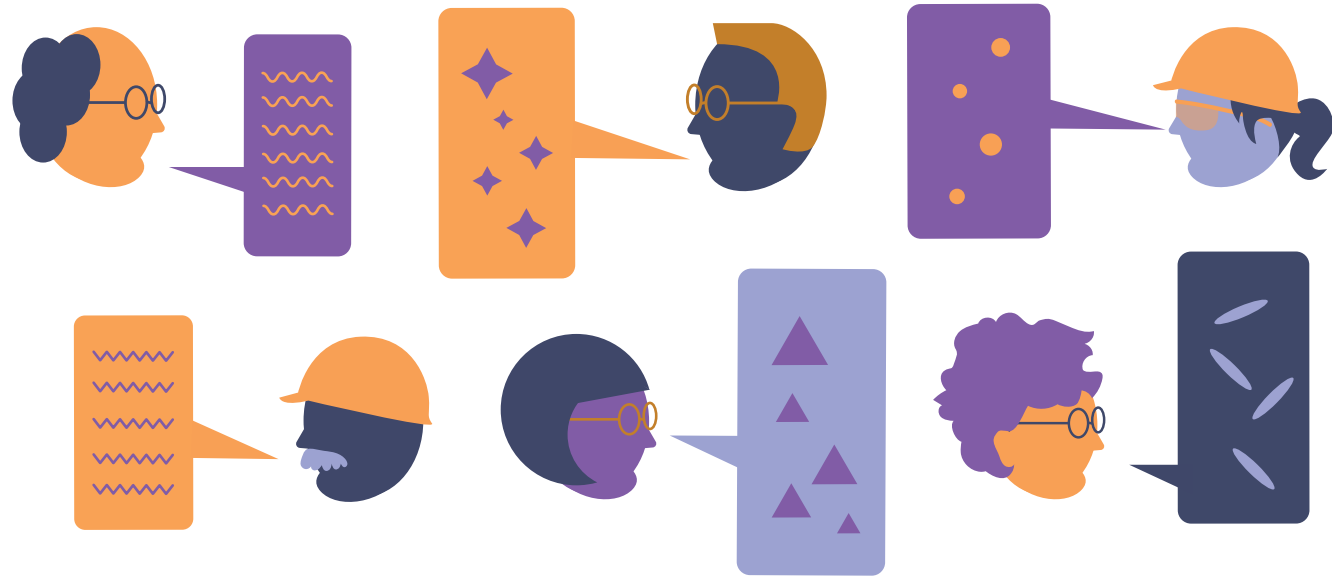


Data is the new essential infrastructure for transit agencies who want to attract and keep riders. Millions of transit riders rely on third-party and agency-managed apps, which in turn rely on high-quality schedule and real-time information. Riders interact with these apps multiple times daily, making open data the most important customer communication channel agencies offer to the public. Agencies should strive to provide the most current and accurate data to transit riders, no matter which app is displaying those data.



High-quality GTFS data make it easy for riders to find their bus stops and to know when the bus or train should come. Inaccurate GPS stop coordinates can cause frustration as a would-be rider watches their bus drive by as they scramble to find a stop they haven't visited before—a bad first impression rather than a warm welcome to transit.

Real-time vehicle information tells riders whether there's a bus just around the corner, whether they have two more minutes to finish drinking their coffee before walking out the door to the bus stop, or whether an unusually large gap between buses might mean an alternate route will be faster. Providing real-time information has been shown to increase ridership by approximately two percent^{4,5} via a combination of reduced real and perceived passenger wait times,⁶ improved rider satisfaction,⁷ and improved perception of safety at transit stops.⁸ Other aspects of interoperability (like predictive analysis and fare payment integration) could boost ridership an additional 3-7 percent.⁹ Conversely, riders might lose



Providing real-time information can increase ridership by two percent

faith in real-time information and shift to other modes if they have one-too-many experiences where agency data say the bus will arrive in two minutes and it does not show up for ten (or vice-versa).

Riders could also expect to see fewer large gaps between buses if agencies are able to harness those real-time data to build real-time dispatch tools to actively manage headways and keep frequent transit service more evenly spaced. No one likes to wait for 20 minutes only to see three buses arrive one after another.

With special events, maintenance, and urban construction regularly disrupting existing service patterns, transit riders can be confused when transit routes are not operating as they normally do. Riders can better respond to planned and unplanned changes when customer alerts and updates to existing transit feeds help communicate what’s going on with the transit network in real time. Good data build trust — which builds ridership.

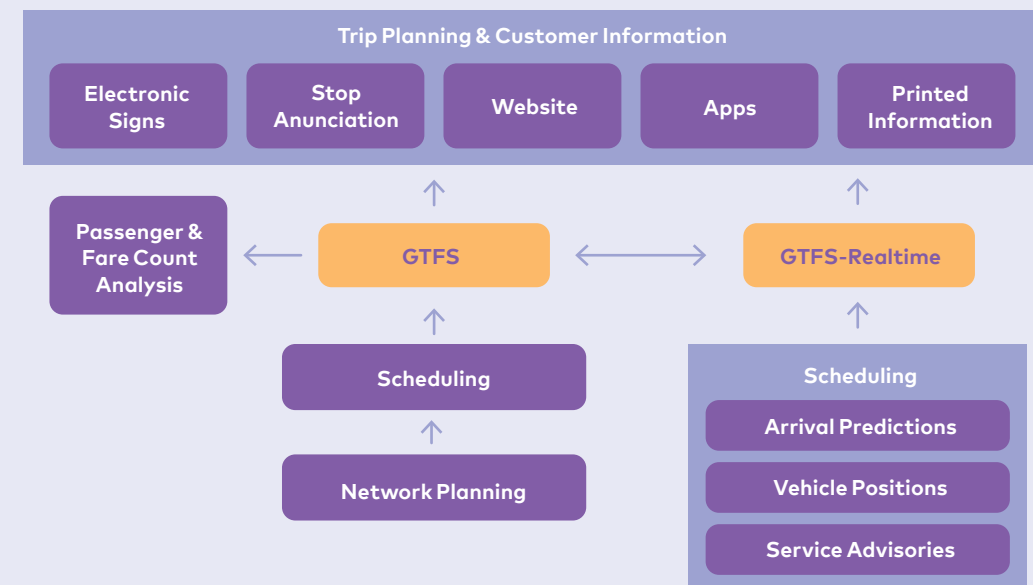
Senior transit agency leadership, performance managers, and planners rely on performance statistics that are only as accurate as the data they are built on. Civic advocates also rely on these data to hold public agencies accountable to delivering on their promises to provide and improve service over time. The accuracy of the data accessible to agency staff and advocates can have substantial impacts on the quality of transit service for transit riders in any region. Using the same data sources to feed customer-facing applications, inform

Good data build trust — which builds ridership

real-time operations decisions, and calculate internal performance statistics increases accountability for data accuracy, simplifies agency work-flows, and strengthens the case for improving those centralized, underlying data sources, rather than data spread across a patchwork of separate systems.

Many people across different transit agency departments (and across organizations, when consultants or other agencies are involved) are involved in generating GTFS data, and even more different people depend on it. Figure 2 provides a simplified version of the processes that generate GTFS schedule and real-time data, as well as a sampling of rider-facing outputs that depend on those data. Transit agency staff plan routes and set schedules, and real-time data are generated by agency-procured hardware and software installed aboard transit vehicles. Those data sources are combined and/or displayed in trip-planning apps, printed and digital signage and materials, on-board announcements, and used in agency operations and performance management.

Figure 2. Transit data pipelines and rider-facing outputs



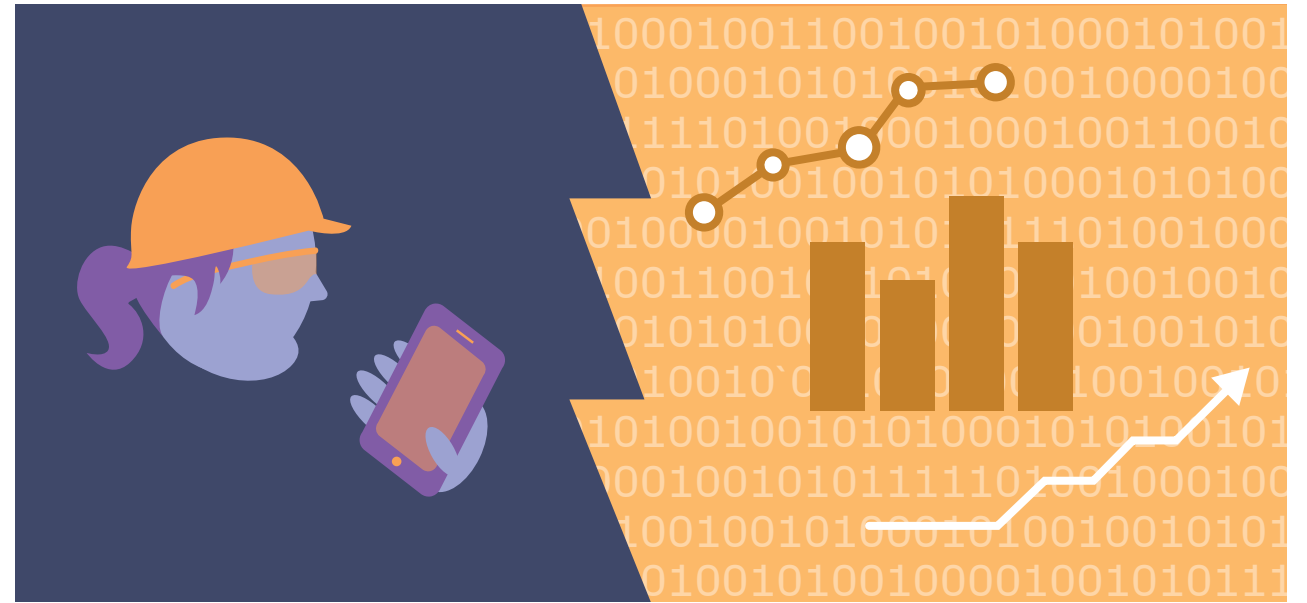
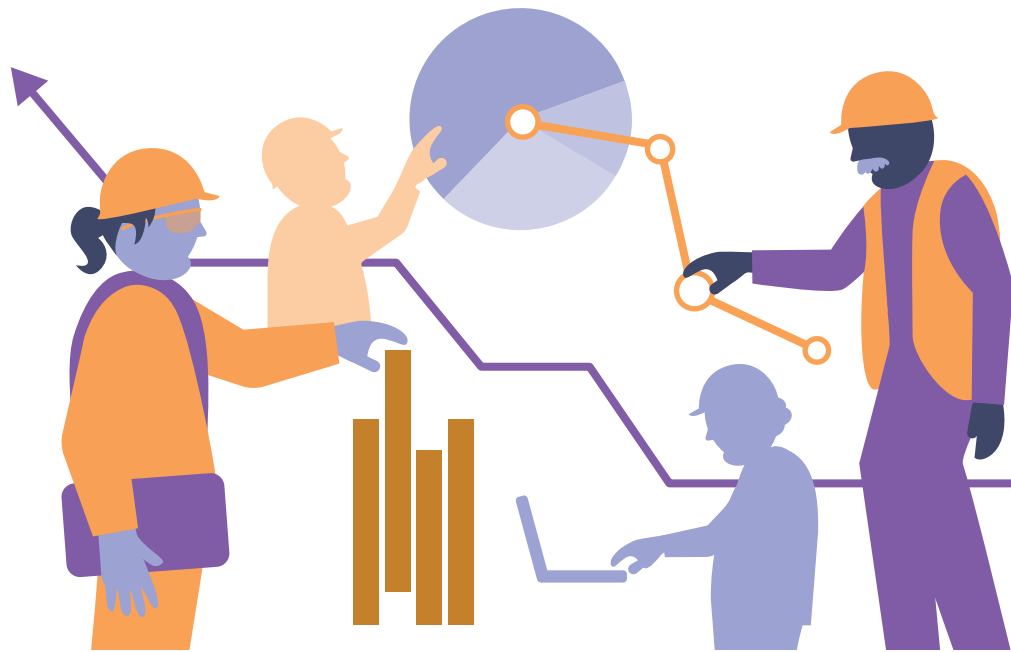
Inconsistent data quality frustrates riders planning trips and hinders staff dispatching buses in real time.

Public agencies

High-quality transit data begins with the public agencies who produce and publish the data. Agency leadership are responsible for ensuring that potential riders have access to the information they need to make transit their preferred choice, and that agency staff are equipped with the data necessary to make informed planning and operational decisions.

All transit agencies attending TransitCenter and RMI’s October 2017 transit data workshop collect and publish similar data.¹⁰ Each agency produces and publishes schedule data in GTFS format, as well as real-time vehicle location data, primarily but not exclusively using the GTFS-realtime data specification. All but one agency also have access to automatic fare collection data, and all have at least a portion of their fleet equipped with automatic passenger counter units, with some having full fleet coverage.

Data *quality* is not as consistent, however, with agencies using a broad spectrum of hardware, software, and internal business practices. The resulting accuracy shortcomings frustrate riders planning trips and staff dispatching buses in real time.



Unlike physical infrastructure, there is no “state of good repair” mindset around improving data infrastructure. Safety, smooth operations, capital projects, and successful maintenance tend to command agency leadership’s urgent attention, while data infrastructure remains neglected by comparison. With limited resources, agencies struggle to hire data-savvy staff who can appropriately leverage the data that agencies generate. When agencies do manage to hire good staff, those staff can be hampered by a lack of access to important data ‘owned’ by another siloed department. And even with access to all the right data, outdated agency management or planning processes may limit the effective-ness of applying those data in practice. When one or more of these institutional barriers stands in the way of implementing good transit data practices, service quality and rider experience suffer.

Transit agencies also lack important insight into how widely used their data is. Much of the end-user behavior and usage metrics for third-party applications is held by private firms. Third-party app usage, an important indicator of the customer-derived value of open data, is therefore often unavailable to transit agencies. [Transit App’s partner program](#) presents an opportunity for transit agencies to gain visibility into rider behavior; other trip-planning apps have shared

Iterative feedback loops involving developers and agencies results in accurate information for transit riders

usage information with transit agencies under limited circumstances. Absent access to data from third parties directly, transit agencies can also survey their customers directly regarding where they receive transit service information.

It is not necessarily a bad thing to have third-party companies acting as agencies' primary customer interface. Private transit application developers have dedicated user experience designers, in-house expertise, and benefits of scale (i.e., operating in dozens or hundreds of cities) that many public agencies may not feel they can afford and/or justify. Competition is a powerful motivator. These application developers have an incentive to attract as many users as possible and, as a result, to be responsive to their needs and demands. Smaller agencies in particular can benefit from leveraging third-party trip-planning apps, especially given the investment required to develop them and the substantial risk that agency-developed software will fall short of its third-party equivalents.

Nonetheless, the quality of user experience delivered by application developers relies most fundamentally on the quality of data provided by transit agencies, regardless of the aesthetic appeal of whatever front end is on it. All parties benefit—transit riders, agencies, and private companies alike—when transit agencies prioritize publishing high-quality data. This remains true regardless of agencies' approach to trip-planning software.

Even if an agency prioritizes improving data infrastructure, its ability to do so may be limited by hardware and software vendors' capacity. Several prominent transit software vendors' hardware and software lack support for open data specifications, which limits agencies' flexibility to update or adapt those vendors' systems to their specific needs. Vendors have a strong incentive to lock transit agencies into closed and proprietary systems, and so it is important for transit agencies to protect their interests by including specific requirements for interoperability. Large technology projects can take years to implement, leading to the implementation of systems that can be out of date by the time they are installed. Interoperable data specifications support modular systems, which increase flexibility and reduce vendor lock-in. Agencies need well-defined data specifications and best practices for implementation so that they can procure interoperable systems confidently and hold their vendors accountable.

Transit application developers

Private companies offering trip-planning apps increasingly serve as de facto front ends for transit agencies. Apps like Google Maps, Apple Maps, Transit App, Citymapper, and others deliver transit agencies' data to transit riders, seeking to grow their user base by providing the best user experience. While these companies typically rely on transit agency-provided data, they also commonly invest significant resources in improving and/or augmenting those data to conform to their quality standards and even to add new features to the datasets to meet rider expectations. All application developers that attended the TransitCenter-RMI workshop rely on both GTFS and GTFS-realtime feeds in particular, and many would also use real-time vehicle occupancy information if it were available.

Other consulting and software firms conduct analysis using the same transit agency-provided data, with the agency as the audience. In both analysis and trip-planning use cases, the quality of transit data determine the ultimate utility of these services to agencies and transit riders, respectively.

With the private sector having a vested interest in the quality of these agency data, the ecosystem of transit data has started to evolve to reflect a shared interest in improved data quality among agencies and application developers. Agencies have increasingly opened channels of communication with data consuming (or packaging)

Figure 3. Apps, whether made by a public agency or private company, are transit's primary customer interface

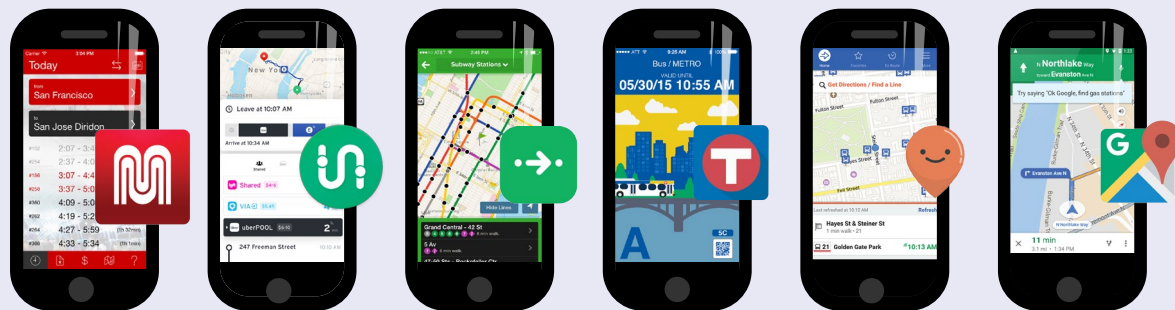


Figure 4. Best Practices for High Quality Transit Data Feeds

- Datasets should be published at a public, permanent URL
- GTFS data is published in iterations so that a single file at a stable location always contains the latest official description of service
- Use consistent id fields for stop_id, route_id, and agency_id
- One GTFS dataset should contain current and upcoming service valid for at least the next 7 days
- Remove old services from the feed
- If a service modification will go into effect in 7 days or fewer, express this service change through a GTFS-realtime feed GTFS data should be configured to correctly report the file modification date

Source: Rocky Mountain Institute's GTFS Working Groups' GTFS Best Practices Document

organizations, who can provide feedback on data quality, for example by flagging inaccurate stop locations, route shapes, or opportunities to adopt best practices and improve quality. What used to be a one-way relationship of agencies publishing and private companies consuming data has evolved to include iterative feedback loops. The end result is better information for transit riders.

Challenges and opportunities

Implementing iterative feedback loops is easier said than done. Third-party data improvements are typically made in company-based silos, and agencies' ability to benefit from application

An improved data specification would include vehicle information, fare data integration, station amenities, and temporary service changes

developers' insights is dependent on having the staff or consultant resources to incorporate that feedback on an ongoing basis.

When data producers, i.e. transit agencies, publish data that are already accurate and 'complete', from application developers' perspective, those application developers save time and money that they might have spent editing and augmenting those data. By providing high-quality and complete data, transit agencies in turn benefit by ensuring that travelers see the best possible data regardless of which app or service is ultimately delivering it.

A key barrier to accurate and complete transit data is a lack of clear guidelines—for both application developers and agencies—defining what a high-quality data feed looks like. In the hopes of taking the first step toward providing these guidelines, RMI convened a GTFS working group consisting of 19 transit data stakeholders and, based on their input, published [GTFS Best Practices](#) in February 2017 at [gtfs.org/best-practices](#). Still, not all application developers or agencies are aware of or adhering to these guidelines. In-person convenings can help close this gap.

The Best Practices, which offer guidance to transit agencies on how to make their GTFS files more readable to humans and machines, do not currently address all GTFS data challenges. Remaining needs include complete best practices for GTFS-realtime, common practices around open data licensing, and standardized update frequencies and management and validation practices among transit agencies.

With respect to GTFS-realtime, a "2.0" specification was developed by the Center for Urban Transportation Research (CUTR) at the University of South Florida. As with the GTFS Best Practices, awareness and adoption of this improved specification among agencies and application developers is relatively low. Practices (e.g., polling rates) and quality standards (e.g., GPS accuracy) related to translation of data resulting from automatic vehicle location (AVL) technology also vary widely among agencies.

A number of other traveler needs are not currently addressed in either the GTFS or GTFS-realtime specifications, such as vehicle information; fare data integration; station amenities; and temporary service changes, i.e., due to planned maintenance, inclement weather, or major events. The inability of existing data specifications to incorporate these factors limits application developers' (and agencies') ability to meet transit riders' needs.

An agenda for better transit data

1 Data management and policy



The ability of public and private stakeholders to implement technical improvements to GTFS, GTFS-realtime, and other data specifications will depend on their ability to advance high-quality transit data as a shared priority

Many agencies transit data would benefit from revised organizational charts and streamlined interdepartmental coordination. Public and private stakeholders can also work together to reduce the technical barriers to creating high-quality transit data. Civic, independent actors like RMI and TransitCenter will also play an important role in preserving the momentum of these shared efforts. Opportunities to improve transit data management and policy can accelerate transit data improvements and must be pursued in order to close the gap between current practice and industry potential. We begin this section with opportunities to improve internal agency operations, and move on to practices that both agencies and application developers can adopt to streamline collaboration.

Key opportunities include:

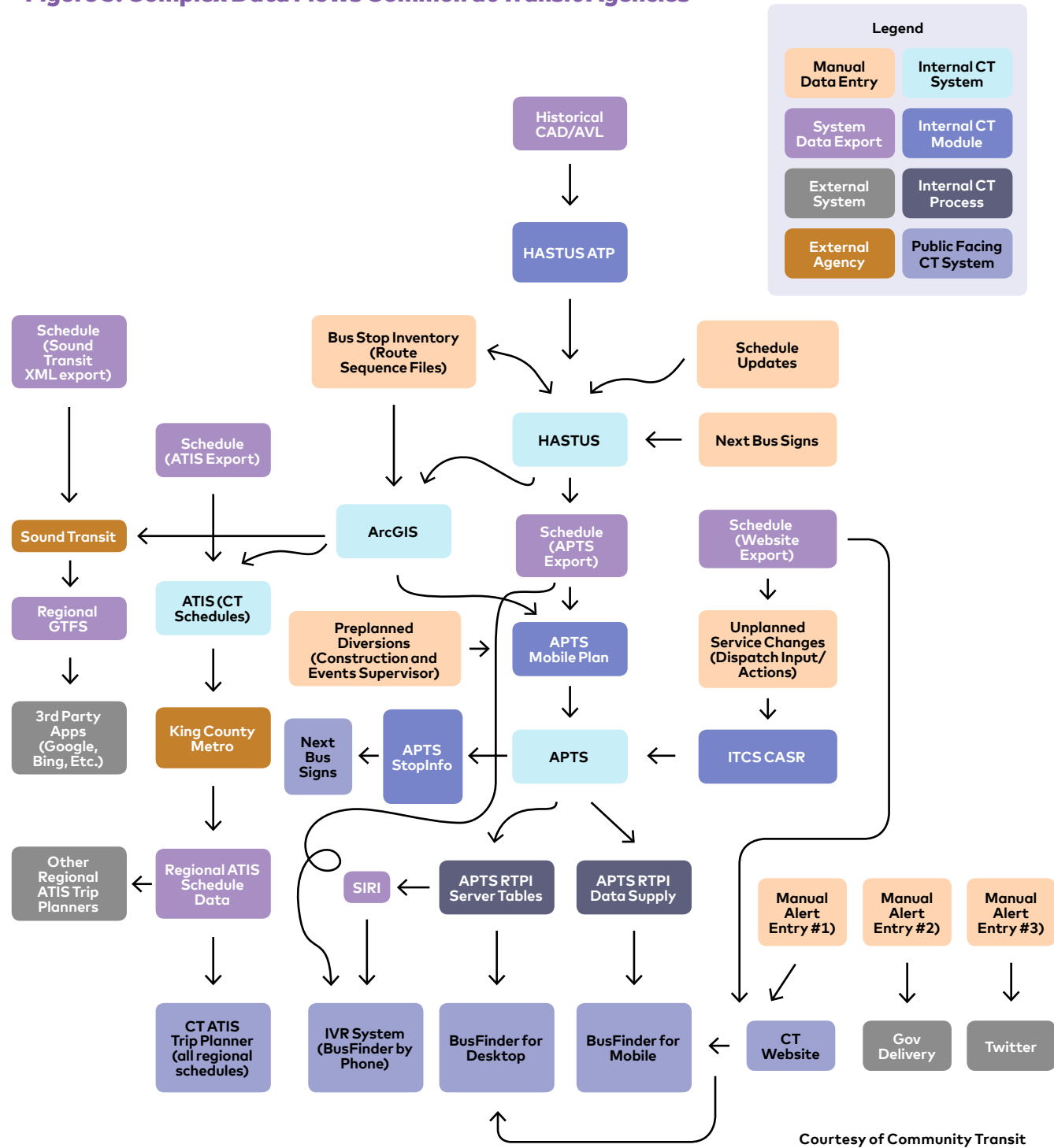
- 1. Agencies should embrace data infrastructure as an agency priority, adequately fund and staff the departments and groups responsible for maintaining and improving transit data, and ensure that agency staff have access to data relevant to their work**
- 2. Agencies should take inventory of existing data sources and how they are created via an agency-wide 'data audit', with support from the private sector as appropriate: who is responsible for managing transit data, who else in the agency relies on each data source, and to what end?**

- 3. Agencies should use the same data feeds that are provided to the public for internal management, planning, and performance analysis**
- 4. Agencies should set data quality targets and adopt best-practice data specifications as a matter of policy, and include those targets and specifications as clear vendor requirements during relevant procurements**
- 5. Agencies and application developers should share emerging data challenges and new knowledge publicly to help advance shared tools and specifications across the transit industry**
- 6. Agencies and application developers should identify and make available to each other a point-person or single monitored email account that provides a clear path for feedback, requests, and questions**

Talk to staff managing your data to identify what their needs are and how to adequately fund and staff them

Agency staff can advocate internally for investment in better transit data, making the business and public interest cases for the data prerequisites to providing high-quality transit service. SFMTA leadership made data quality improvements an agency priority once technical staff helped them understand that public datasets are an essential customer communication tool. This can be illuminated by mapping flows of data through the organization, to answer questions like, where do the data originate? Who is responsible for maintaining them? Who else uses them, and for what purpose? Snohomish County's Community Transit worked with IBI Group to answer these questions and map their intra-agency data processes, using the results of that "data audit" to develop a strategic plan for transit data. Their results showcase the complexity of these internal data flows, even at a relatively small transit agency (see chart on following page).

Figure 5. Complex Data Flows Common at Transit Agencies



The public and developers deserve the same quality of data staff use to make decisions

Agencies can also support each other by sharing emerging challenges and new knowledge publicly

This kind of audit will inevitably highlight the diverse organizational functions whose success ultimately depends on high-quality transit data—from scheduling to operations to IT to planning and analysis and, critically, to customer communications. Through any similar audit process, it is critical that agencies seek to understand the accuracy and limitations of transit data at their furthest upstream source, as any issues there will propagate throughout the agency’s workflow and cannot be repaired further downstream.

Agencies can create powerful internal incentives to maintain and improve data quality by committing to use the same data feeds that the agency publishes for use by application developers and, by extension, the general public. This practice (also known as “dog-fooding”) ensures that the agency remains accountable to providing high-quality information to riders, no matter where they are finding information about transit. Application developers are often at the front lines of transit rider communication and by promoting their own transit trip-planning services they are also, in effect, marketing on behalf of transit agencies. As a result, agencies should work to ensure that they are publishing the most accurate information possible.

With the resources that public agencies have on hand, another major lever they control is data and technology procurements. By including data quality and specification requirements in requests for proposals, agencies can accelerate improvements and even contribute to improving the specifications themselves.

Agencies can also advance data specifications through work with on-call contractors or through their day-to-day work. MTA New York City Transit is currently working to implement subway station navigation as part of its data offerings, which has required an experimental extension of the GTFS specification as well as improvements to *OpenTripPlanner* that could be used by other agencies around the world. Both contractors and agency staff have worked on the project.

Agency practice also varies considerably when it comes to internal data management practices. RTD-Denver publishes updates to its GTFS schedule data weekly, typically with 1-3 changes, a process that requires discipline and active coordination across several departments and data systems, but which provides the

Application developers stand to benefit directly by making their data format requirements transparent and easily accessible

Make it as easy as possible for application developers to get in touch with dedicated agency staff

benefit of an always up-to-date schedule feed. Some large agencies update their public GTFS schedules as infrequently as three times per year, even when small service changes are made more frequently. The GTFS update process can be complex—for example, SFMTA manages 24 distinct internal systems that rely on GTFS and related data feeds—so improving update frequency is not a trivial task.

Agencies can also support each other by sharing emerging challenges and new knowledge publicly. Portland’s TriMet has long been a leader in this respect, with their efforts leading to the creation of both the GTFS specification and the open-source trip-planning engine *OpenTripPlanner*.¹¹ Knowledge-sharing is also part of the mission of the Federal Transit Administration’s Mobility on Demand Sandbox Program, which includes multiple data-oriented grants (including [grants to TriMet](#) and the [Vermont Agency of Transportation](#), both of which will further develop OpenTripPlanner) whose findings will be distributed widely to the industry. Formal and/or informal coordination with other agencies can help use public investments in data and technology to improve shared tools and data specifications.

Agencies can also benefit from making it as easy as possible for application developers to get in touch with dedicated agency staff. Application developers can benefit from making it easy for agencies to get in touch with them as well. Many agencies lack a clear point of contact, making it difficult for third party companies or organizations to provide feedback that could benefit the agency and the riders it serves. New York’s MTA created and actively monitors its own Google Group for developers, and Boston’s MBTA established generic contact email addresses that remain constant even if staff turns over, but which are automatically forwarded to specific points of contact at any given time. A [current proposal to add operator contact information](#) to GTFS feeds would make it easier for application developers to direct feedback to relevant transit agency staff.

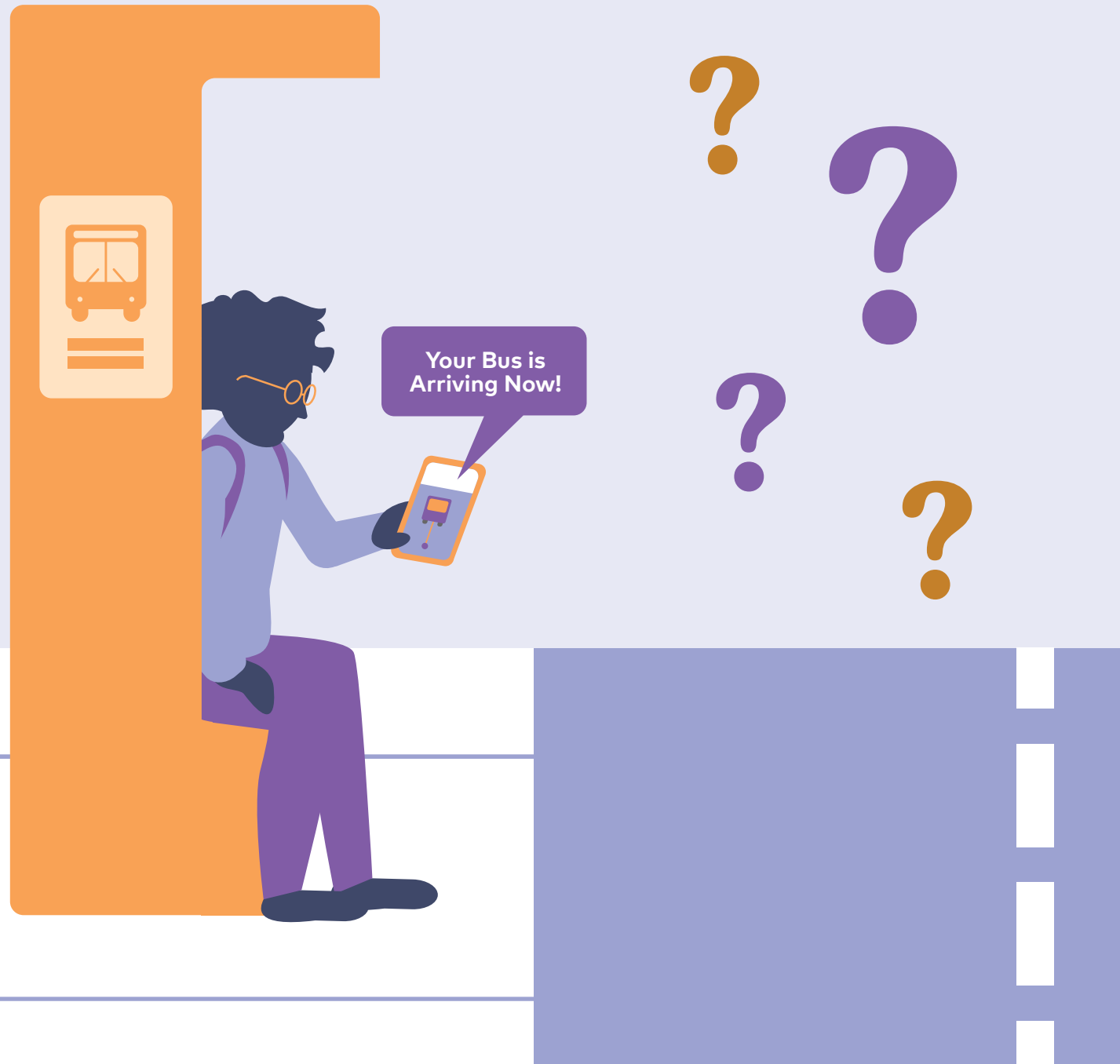
Private application developers can be allies to agency staff in building the case for better data, by providing statistics or other analysis that supports the need for improved data quality. Trip-planning companies frequently flag issues in public GTFS feeds as those companies observe errors in those feeds. IBI Group, which

consults with several transit agencies on data management and performance analysis, has presented to agency leadership on the benefits of improved data quality.

Because application developers often work with transit data feeds from a diverse list of public agencies, they can also provide valuable technical perspective on best practices and emerging needs across the industry. Given the nature of their shared reliance on public datasets, private companies may also benefit from sharing their input, expertise, and even software more publicly than they might be used to. Many application developers provide information about their practices and quality standards to transit agencies and other data suppliers, providing a clear path to agencies and data suppliers who wish to have their data included in any given app. Application developers stand to benefit directly by making their requirements transparent and easily accessible, but in doing so they can also accelerate adoption of best practices and facilitate data specification improvements in the long-term. For example, Transit app provides links to the GTFS and to GTFS Best Practices at <https://transitapp.com/developers>. Google also references the GTFS Best Practices, and provides public documentation for optional “Google Transit Extensions to GTFS” at <https://developers.google.com/transit/gtfs/reference/gtfs-extensions>. These extensions support additional functionality in Google Maps but are available for adoption by other apps, and they have in the past become the basis for official changes to the core GTFS.

Independent actors have an opportunity to play an important role maintaining momentum on shared priorities identified by public and private sector stakeholders. Coordination across public and private sectors has proved challenging since the creation of the GTFS more than a decade ago, pointing toward the need for third parties like TransitCenter, RMI, Open Transport Partnership (who, with NACTO, manages the [SharedStreets](#) project), and others to help set goals and create sustainable structures for collaboration. Momentum can be preserved through further peer education opportunities, project management of specific specification improvements, documenting best practices, and expanding publicly available tools for managing and improving data, like CUTR’s GTFS-realtime validator and Transit.land.

2 Data Quality



All the data in the world is of no use to transit riders, transit agencies, or application developers if it is not accurate. Agency investments in new apps or analytical tools can be wasted if the underlying data are shaky.

Whether riders plan their trips in an app built by the public-sector or private-sector, poor data quality can cause riders to miss their bus or add significant time to any given trip.

Specific opportunities to improve transit data quality include:

- 1. Agencies should upgrade hardware to improve real-time source data, e.g., AVL polling rates and GPS accuracy**
- 2. Agencies and application developers should develop their software to improve customer-facing real-time information, e.g., data latency and arrival time prediction reliability**
- 3. Application developers should share their validation tools and techniques with others in the industry in a replicable and/or easily implementable way**
- 4. Agencies should use validation tools that check for adherence to best practices (not just baseline specification compliance) and actively solicit application developer input to improve transit data quality and accuracy on an ongoing basis**

Some data quality issues are driven by hardware limitations. While large transit agencies generally have full coverage of automated

Agency investments in new apps or analytical tools can be wasted if the underlying data are shaky

vehicle location (AVL) systems in their vehicle fleets, GPS polling rates and accuracy varies from one implementation to another. High polling rates and greater accuracy improve wait time estimates for riders, enable more targeted real-time dispatch and operational controls, and make performance analysis more accurate.

Large agencies also typically have full fleet coverage of automatic fare collection (AFC) systems, but not necessarily of automatic passenger counting (APC) systems. These systems, when agencies are able to collect and view data in real-time (both a hardware and software problem), augment real-time dispatch possibilities and offer the potential for improved rider experience, for example by allowing agencies and trip-planning apps to display vehicle crowding information.

Data quality can also be improved through software implementation. Research from SFMTA shows that inaccurate arrival time estimates frustrate transit riders and make them more likely to use other travel modes.¹² Improved algorithms can make arrival times more accurate and improve rider experience. The open-source [TransitClock project](#) offers one approach; Citymapper's wait time estimates are augmented by real-time traffic data,¹³ and Transit App uses both real-time data and historical performance analysis in partnership with Swiftly.¹⁴

Several transit application developers already use in-house validation tools and could benefit from sharing their techniques with each other and with transit agencies, who could implement similar validation steps upstream to save themselves (and application developers) considerable time and effort. For this to be effective, however, application developers would need to ensure that relevant transit agency staff would be able to easily implement the published tools and/or methodology.

New and improved validator tools and best practices are under development. *GTFS Best Practices* developed by RMI are under continual revision and have already been enhanced to reflect stakeholder feedback since their release. GTFS-realtime best practices are also under development and will define minimum data requirements, set expectations with respect to AVL-to-GTFS-realtime translation (including minimum requirements for polling rate and GPS accuracy), indicate how service changes might provide producers an easily implementable stepping stone toward full provision of real-time data, and provide definition of a universal

Data forums provide the opportunity for agencies and developers to advance data quality together

update frequency for both static and real-time transit data feeds. The Center for Urban Transportation Research (CUTR) at the University of South Florida is developing an updated GTFS-realtime validator tool that will provide a common basis for evaluating feeds against *GTFS-realtime 2.0*.

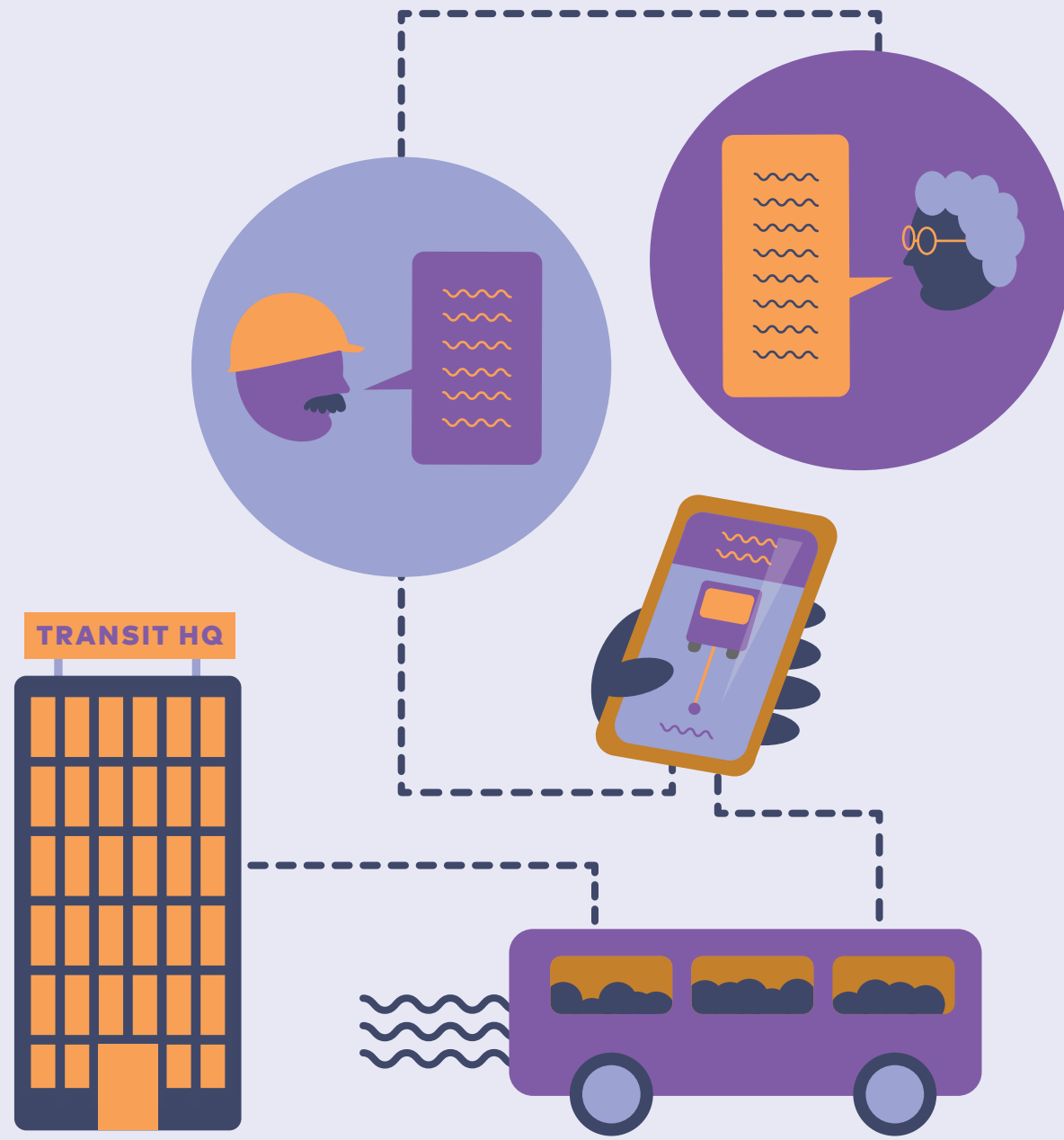
A consolidated catalog of GTFS feeds (1) allows agencies to ensure they are reaching the broadest possible audience of application developers, (2) makes it more efficient for application developers to find and access data, (3) ensures all application developers have access to the most up-to-date data, (4) provides a shared view of data practices from which to inform data spec development and governance, and (5) makes it easier to apply validation tools to improve data quality. For example, the [Transit.land](#) platform offers a consolidated and shared approach to GTFS feed aggregation, which could ease coordination and simplify agency-application developer feedback loops.

Information on the most up-to-date tools and transit data improvement initiatives can be accessed via [GTFS.org](#). Agencies and application developers can also stay involved and up to date by joining the [transit-developers](#), [GTFS-changes](#), and [GTFS-realtime](#) forums.

Some data quality issues cannot be improved by software or hardware, but rather require improved geospatial data, for example bus stop location data. These data typically require some level of manual validation, whether in the field or digitally to ensure trip-planning apps will send riders to exactly the right location.

Improving data quality is fundamental to unlocking the potential of transit data for riders, agencies, and application developers alike. Not only do transit riders rely on these data to plan their transit trips on a daily basis, but real-time data have become essential to transit agency operations and planning, enabling performance analysis that informs service planning and agency policy, as well as the creation of real-time operational tools (e.g., for field staff, maintenance staff, and dispatchers).

3 Data Specifications



Opportunities to advance specifications include describing stop amenities, service changes, fare payment, and on-demand transit information.

Improved management practices and data quality are needed to improve adoption of today’s transit data best practices, but improvements to the underlying data specifications can expand what is possible. The GTFS format documentation was released in 2006 and GTFS-realtime arrived in 2011. As transit industry hardware, software, and data-driven operational and planning practices have evolved during that time, new needs and potential use cases have become apparent. Near-term opportunities to advance these data specifications include (but are not limited to) the capability to describe stop and station amenities, represent planned and unplanned service changes, integrate fare schedules and payment options, and provide schedule and/or real-time information for on-demand transit services.

There are three primary ways that agencies and application developers can advance transit data specifications:

- 1. Agencies and private transportation providers should invest in extending existing specifications as part of their project development and/or procurement processes when opportunities arise**
- 2. Agencies and application developers should publish documentation for data feeds, focusing on experimental or unique features to help others use those features and facilitate further specification development. For example:**
 - MBTA GTFS Documentation
 - Trillium GTFS Documentation
- 3. Agencies and application developers should stay up-to-date and actively engage in specification development processes through available discussion venues. For example:**
 - GTFS Google Group and GitHub
 - GTFS-realtime Google Group
 - GTFS-flex Google Group and Github

Applications can make navigating complex transit stations easier, only if that data specification is widely adopted

This section, rather than focusing on these recommendations per se, will explore some concrete near-term opportunities to improve and extend the primary transit data specification in use in North America, GTFS.

Station amenities

Transit riders benefit from understanding the facilities available at a given bus stop or rail transit station, especially riders with disabilities or other mobility impairments that make it hard or impossible to navigate stairs and steep inclines, and that make it especially important to have a place to sit down while waiting for the bus. The existing GTFS specification does not offer guidance on how to represent these details in a machine-readable format.

Transit agencies across the country are working on different aspects of this problem. Bay Area Rapid Transit (BART) is developing a real-time escalator feed to represent whether escalators are functional and which direction they are running. WMATA already uses GIS-based maps of its stations, including elevator locations. NYC Transit offers a real-time feed describing elevator status, but it is not yet integrated with GTFS. NYC Transit is also working with Cambridge Systematics and Trillium to develop a [draft data specification](#) to enable station navigation, which could in turn be

Figure 5. GTFS Pathways Proposal

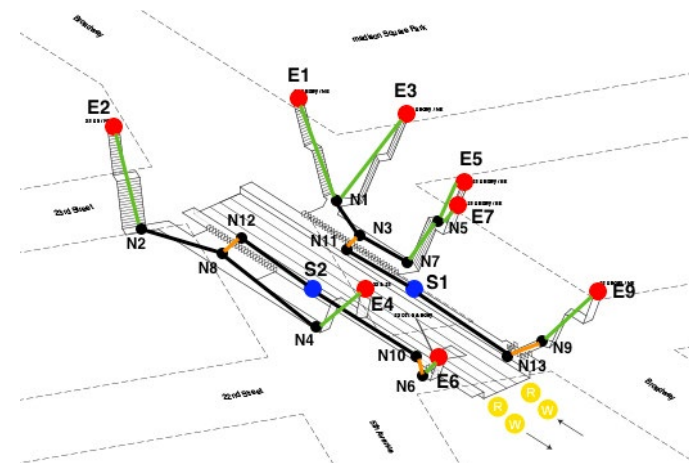


Diagram from the "GTFS Pathways Proposal", with credit to Candy Chan and Project Subway NYC

Whether planned or unplanned, service detours and changes must be reflected in the transit data that riders rely on

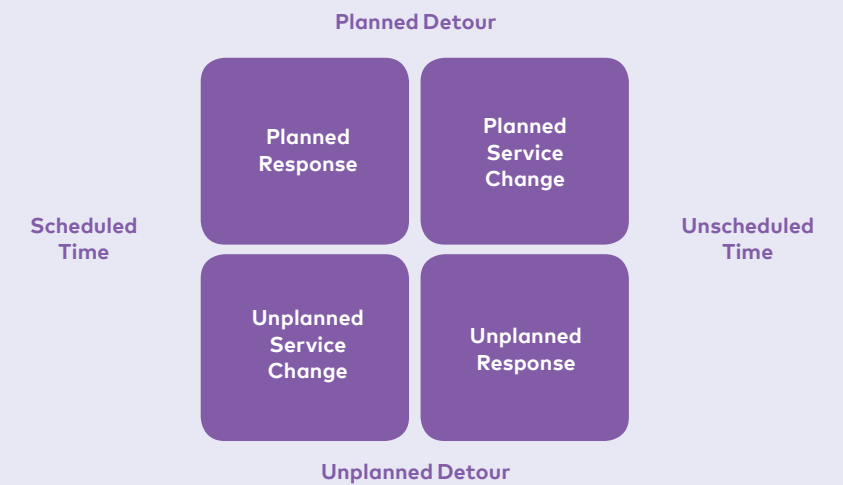
used by any agency. SFMTA is developing a GTFS-integrated real-time elevator and escalator status feed. Maryland MTA is conducting a system-wide inventory of ADA-accessible sidewalks near bus stops.

GTFS currently lacks the capability of fully modeling transit stop access through stations, station entrances, elevators, escalators, and pathways. The "Pathways (and levels) proposal" is based on work and input from Google, Transit App, Trillium, Cambridge Systematics, MBTA, and NYC Transit to model station facilities. NYC Transit is planning to utilize parts of this data format for a new official trip planner. The proposal demonstrates a working model where shared problems are identified and idiosyncratic project-specific solutions are synthesized into a shared specification.

Service changes and "detours"

Both transit providers and application developers have a growing responsibility and interest to offer riders accurate information on when a schedule will be disrupted from expected scheduling, due to planned or unplanned service changes. Printed announcements and website or social media updates are commonly used but do not help riders unless they have seen the announcement directly. Agencies could instead update schedules and/or real-time data to

Figure 6. Four Types of Transit Service Changes



A seamless trip planning and fare payment experience will depend upon applications and fare payment systems speaking the same language



reflect service changes, enabling transit riders planning their trips to access up-to-date information on getting from A to B.

Service changes can be divided into four categories (See Figure 6):

1. Planned service detour at a planned time (e.g. concert) > common
2. Planned service detour at a unplanned time (e.g. snowstorm) > common, though less common
3. Unplanned service detour at a planned time (e.g. construction scope change) > most rare
4. Unplanned service detour at an unplanned time (e.g. burst underground pipes) > rare

Agencies across the board identified this as a major gap in current practice, and a few are actively working to fill this gap. Austin’s Capital Metro has a “GTFS Detours” spec currently under development, for example, and Transit App led a review of GTFS service changes best practice that could inform real-time service change specification development at other agencies.¹⁵ Application developers have developed additional tools for both internal and external use, some of which remain unknown to agencies and would benefit from wider dissemination. In the meantime, Transit App

applies MTA New York City Transit (NYC Transit) service changes using a partially-automated review of published changes, with “human oversight”.¹⁶ Citymapper uses natural language processing to convert text-based service change announcements published by NYC Transit into actual routing changes for its users in New York City.¹⁷ This considerable investment on the part of third party applications indicates the value of service change information to transit riders, and the inefficiency (and potentially inaccuracy) of redundant private efforts to provide transit service information.

GTFS-flex

GTFS-flex (gtfsflex.com) is a proposed extension to the GTFS to describe the availability of on-demand transit for discovery in trip planners and other applications. An in-progress FTA Mobility on Demand Sandbox-funded project in Vermont has resulted in the first reference implementation, in the form of GTFS-flex support in OpenTripPlanner.¹⁸ GTFS-flex is immediately applicable to existing on-demand services like RTD-Denver’s “Call-n-Ride” and AC Transit’s “Flex” service, and will become increasingly important as transit agencies expand their on-demand offerings. GTFS-flex data could also be published by companies offering private on-demand transportation services, expanding publicly-available trip-planning offerings to these transportation modes.

Fare data integration

There are currently no standard data formats with the capability of describing the wide variety of transit fares. The GTFS fare model is very limited, providing the capability of describing single-trip zone- and route-based fares with transfers, but not regional transit fares, pass products, and other pricing schedules such as time-of-day and distance-based fares. This means that fare structures need to be encoded differently for different applications, and also that third-party trip planner software often lacks complete fare information. Brian Ferris (an early OneBusAway project member, and GoogleMaps engineer) previously established the [GTFS Fare Working Group](#) and provided outputs that could inform revisions of the GTFS fare model:

- [Fare systems of the world](#)
- [GTFS fare model proposal](#)

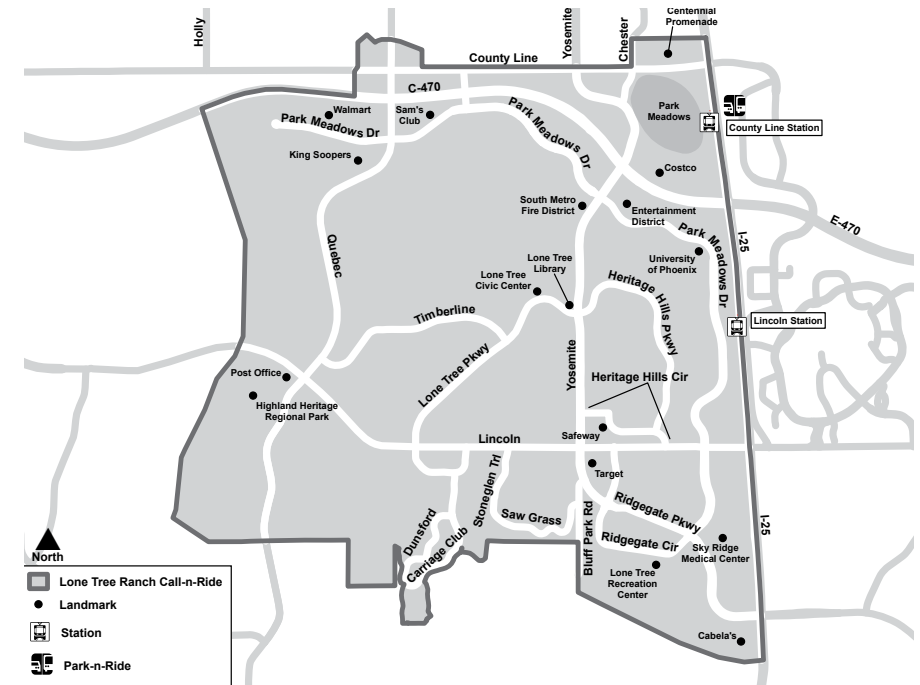
Readily available, timely, and accurate transit data is critical for agencies to stay competitive

There are no standardized interfaces for transit trip planners and third-party applications to interact with ticketing and fare payment systems, which poses a barrier to a seamless customer experience from trip planning to purchasing a transit ticket. There are at least two possible bridges to seamlessness for users and application interoperability:

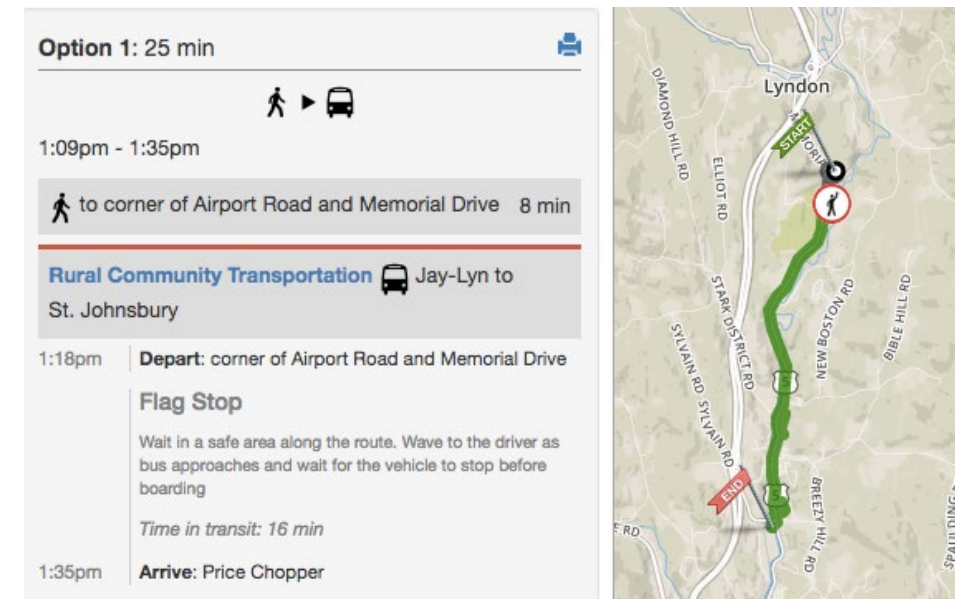
1. Least complicated but least integrated: Add ticketing app field to GTFS to enable deeplinks from trip planning applications to a mobile ticketing app.
2. More complicated with full integration: Develop an open payments API or SDK to third-party mobile applications to sell transit tickets without requiring users to leave the app. This experience would be similar to how travelers can hail an Uber ride without leaving Google Maps.¹⁹ The API that TriMet has developed for the HOP pass could provide the foundation for a shared specification.

Transit vehicle information

Transit riders are sometimes disappointed when a full bus pulls up to a stop and they are unable to board. It can also be helpful to riders to know what color the bus is—particularly in regions where buses come in multiple colors—and to know about other features a vehicle might have, like onboard WiFi, bicycle-loading amenities (and bicycle occupancy), electrical outlets, or air conditioning. This type of transit vehicle information is uncommon domestically (transit operators in Sydney, Seoul, and Singapore offer real-time bus occupancy information), but U.S. agencies are beginning to expand their offerings. Some international markets would also support information such as gender restrictions on certain vehicles. For example, BART plans to offer real-time vehicle occupancy information on its new trains, and Capital Metro is working to add vehicle capacity information to its real-time feed.



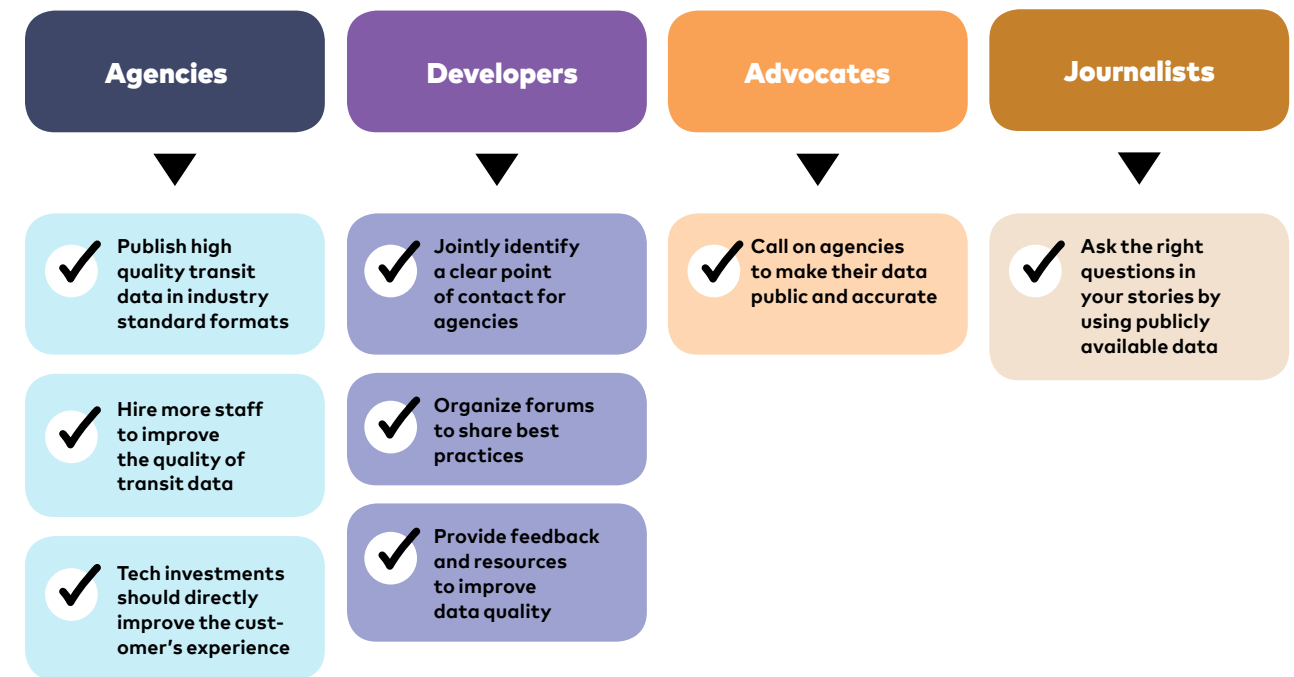
Source: Regional Transportation District, Denver



Source: Trillium Solutions

Conclusion

Figure 7. Your role in making transit data that riders want



For transit agencies seeking to remain competitive in an increasingly crowded transportation landscape — to retain existing riders and attract new ones — producing readily available, timely, and accurate transit data is critical for their success. By doing so, transit riders receive dual benefits. The first is a more seamless trip planning experience, and the second is a higher quality trip.

Any time riders use an app, web-based interface, or even a paper schedule to plan and conduct their travel on public transportation, they are relying on transit data to make their travel choices. These data originate in public transportation agencies but are increasingly massaged and shared with riders through private sector stakeholders, who share agencies' investment in providing a positive user experience. Access to timely and accurate data before and during a trip has been shown to increase ridership and improve the satisfaction of existing riders.

Transit data can also serve as the basis for internal management and monitoring tools to help agencies become more efficient and effective at delivering their service. Planning can be more responsive to riders' transportation needs, and operations can be faster and more reliable. Together, this makes for a higher quality trip that will be more attractive to existing and new riders alike.

For transit to reap these benefits and be the backbone of physical urban mobility networks, its data must be strong enough to anchor the digital urban mobility system. The persistence of poor-quality transit data would isolate transit agencies and limit transit's utility as rider expectations and behavior evolve to demand increasingly accurate and sophisticated travel information, and higher quality transit service.

To continue improving these data, transit agencies must prioritize data internally as a new, essential infrastructure that permeates agency practice from operations to policy and planning to customer communications. Agencies and transit application developers must also seek out opportunities to work together, ideally in the open, both to improve data quality and expand data specifications. Through this shared pursuit, transit riders throughout North America and beyond will benefit, making it easier to choose transit as a direct result.

Endnotes

1. Rocky Mountain Institute also hosted the 2015 Interoperable Transit Data Workshop <https://d231jw5ce53gcq.cloudfront.net/wp-content/uploads/2017/03/Mobility-InteroperableTransitData-Report.pdf>.
2. <http://www.businessinsider.com/gmail-has-1-billion-monthly-active-users-2016-2>.
3. <https://trilliumtransit.com/2017/12/31/life-of-data-standards/>.
4. https://www.alaskapublic.org/wp-content/uploads/2012/06/bis_transit_chicago.pdf.
5. <https://www.sciencedirect.com/science/article/pii/S0968090X1500297>.
6. <https://www.sciencedirect.com/science/article/pii/S0965856411001030>.
7. <http://pugetsound.onebusaway.org/xwiki/bin/download/Main/Research/TRB-2011-OneBusAway-Changes-Paper.pdf>.
8. <http://trrjournalonline.trb.org/doi/abs/10.3141/2351-11>.
9. https://www.rmi.org/wp-content/uploads/2017/03/consortium_approach_to_ITD_report2016.pdf.
10. These agencies overwhelmingly represented large urban centers, and thus are not representative of the transit industry as a whole; for a full list of attending agencies and companies, see Appendix
11. <http://docs.opentripplanner.org/en/latest/History/>.
12. <https://trid.trb.org/view/1495956>.
13. <https://medium.com/citymapper/using-data-to-figure-out-traffic-d6b1262589>.
14. <https://medium.com/transit-app/la-metro-gets-better-real-time-transit-data-bf694bb82218>.
15. "GTFS-rt Service Changes [Editor's Draft]" by Léo Frachet: <http://bit.ly/gtfs-servicechanges>.
16. <https://medium.com/transit-app/how-were-bringing-real-time-countdowns-to-nyc-s-lettered-lines-482d3b8f9899>.
17. <https://www.citylab.com/transportation/2017/08/the-transit-app-that-can-see-subway-outages/537987/>.
18. <https://trilliumtransit.com/2016/10/21/announcing-the-flexible-transit-trip-planner/>.
19. <https://www.uber.com/newsroom/googlemaps/>.

Appendix

Participating organizations

Alameda County and Contra Costa County Transit (Oakland, California)

Bay Area Rapid Transit (San Francisco Bay Area)

Capital Metro (Austin, Tex.)

Chicago Transit Authority

Dallas Area Rapid Transit

King County Metro (Seattle)

Massachusetts Bay Transportation Authority (Boston)

Metro Los Angeles

Metro Transit (Minneapolis-St. Paul)

MTA New York City Transit

Nashville Metropolitan Transit Authority

Regional Transit District (Denver, Colo.)

San Francisco Municipal Transit Agency (“Muni”)

Tri-County Metropolitan Transportation District of Oregon (“TriMet”) (Portland, Oregon)

Washington Metropolitan Area Transit Authority

Cambridge Systematics

Conveyal

Google

Kisio Digital

IBI Group

Mapzen

Remix

Swiftly

Transit App

Trillium Solutions

University of South Florida

Workshop Format and Methodology

The 2017 Interoperable Transit Data Workshop was designed to identify collaborative solutions to shared transit data challenges. The two-day, actively-facilitated workshop employed a combination of educational “lightning talks” from industry experts, small-group “breakout sessions” aimed at identifying actionable next steps within specific problem categories, and large-group “plenary” sessions geared toward synthesizing and sharing small-group and collective insights. Participants indicated shared challenges and needs via a pre-workshop survey. Some breakout session topics were identified in advance via participant interviews and the pre-workshop survey, while others were identified and voted on in real time by participants at the workshop. Participants then completed a post-workshop survey, indicating which aspects of the workshop’s format, content, and facilitation they found helpful and/or in need of improvement.

For organizations or individuals wishing to host a collaborative workshop of their own, RMI and TransitCenter invite further questions and would be happy to share details on the workshop agenda as well as anonymized survey questions and results.

Published by

TransitCenter

1 Whitehall Street, 17 Floor

New York, NY 10004

Design and Illustrations L+L Design (landl.us)

Printing Shapco

Typefaces Mark by Hannes von Döhren, Christoph Koeberlin and FontFont Type Department; Lyon by Kai Bernau

© 2018 TransitCenter

