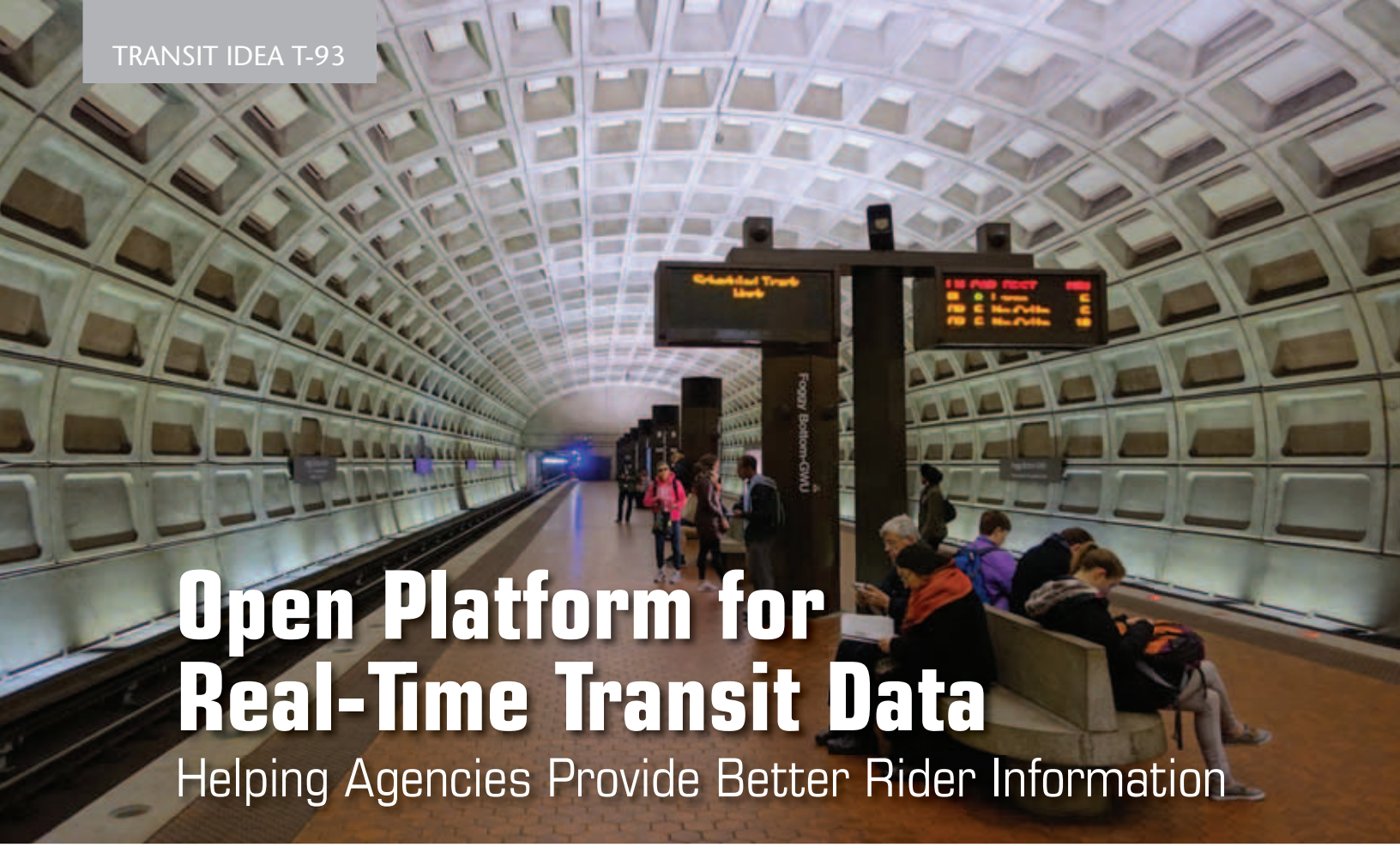

Open Platform for Real-Time Transit Data

Helping Agencies Provide Better Rider Information

Photo: Daderot, Wikimedia Commons

DREW DARA-ABRAMS

The author is a principal at Interline Technologies in San Francisco, California.

Above: Passengers wait for their train to arrive at the Foggy Bottom–GWU (George Washington University) Metro Station in Washington, D.C. In cities all over, transit riders get real-time transit information via smartphone apps that—among other benefits—helps them manage their travel time. But such an advantage can be thwarted when information is incorrect or when transit agencies fail to monitor their data.

Pull out a smartphone, open an app, and see when the next bus is estimated to arrive. What was once available in only a handful of cities has now become commonplace, if not expected. Real-time transit information has brought measurable benefits to riders and agencies but also has introduced new complexities for agencies and their staffs. Researcher Ian Rees at Interline Technologies in San Francisco, California—along with Sean Barbeau at the Center for Urban Transportation Research (CUTR) at the University of South Florida—collaborated as part of a team that recently completed a TRB Transit IDEA (Innovations Deserving Exploratory Analysis) project to develop an open web-based platform that validates the quality of real-time transit data and helps agency staff to understand issues.[1]

[1] For the full project report, see http://www.trb.org/Main/Blurbs/181415.aspx.

Real-Time Transit Data: Benefits and Challenges

Researchers have found that when transit agencies provide real-time transit information to their riders, there are measurable gains, including:

1. Shorter perceived and actual wait times. Wait times are *actually* shorter because riders are able to consult real-time information in advance, decide when to arrive at a stop or station, or decide to go to an alternative stop or station. Wait times are *perceived* to be shorter because riders have more information once they have arrived at a stop or station (*1*).
2. A more welcoming experience for new riders (*2*).
3. An increased feeling of safety (e.g., at night). Riders may be able to choose to wait at an alternative location before going to a stop or station, or they may have more confidence in their safety

Photo: Piqsels

When passengers have access to reliable real-time information, transit agencies receive a direct benefit: an equally reliable increase in ridership.

once they have arrived at a stop or station (*3*–*4*).

4. Increased ridership as a direct result of the previously listed effects and as the benefits compound (5–6).

An additional plus for transportation agencies is that real-time transit data is—compared with many other potential operational or capital improvements to bus or rail service—an affordable and efficient means of increasing ridership.[2]

However, there are two key challenges to realizing the benefits of real-time information:

1. Providing real-time transit data can backfire if that information is not up to date and accurate. Incorrect real-time transit data have been found to have a negative effect on ridership, the rider's opinion of an agency, and the rider's satisfaction with mobile transit apps (*4*).
2. A limited number of transit agencies monitor the reliability and accuracy of their traveler information systems.

Lacking this capability, many agencies provide real-time information of unknown quality. Because of this uncertainty, some agencies instead try to limit distribution of real-time data (*7*–*8*).

How can agency staff actually measure the quality of their real-time transit data? This is a simple question to ask but a more complex one to answer. GTFS (general transit feed specification) and GTFS Realtime are closely related data specifications that agencies use to disseminate transit information. The technical architectures used to create and distribute GTFS and GTFS Realtime data feeds are often complex, with different agency departments or vendors responsible for different components (Figure 1).

Solution: An Open Platform for GTFS Realtime Validation

The research team's goal for the Transit IDEA T-93 project was to build and test a web-based platform that transit agency staff could use to assess their GTFS Realtime feeds. They began this project with two open-source components already in hand:

1. **Transitland**—an open-data platform that aggregates GTFS data from thousands of transit providers. Transitland provides a directory of

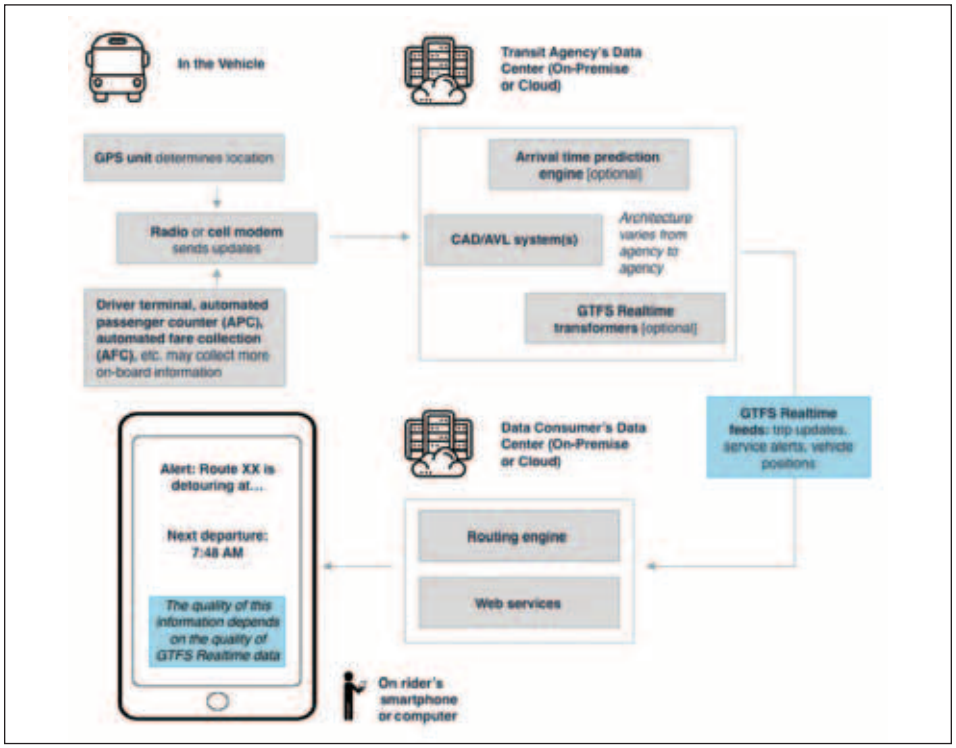

FIGURE 1 GTFS Realtime feeds are typically produced and consumed in an overall technical system architecture. (Note: CAD/AVL = computer-aided dispatch/automatic vehicle location.)

[2] The *APTA 2017 Fact Book* notes: "The growth of automatic vehicle location systems, which improves the operation of bus fleets, as well as the availability of information on bus arrival times, has made public transit systems more efficient and data more accessible."

these feeds and a set of application programming interfaces for querying the feeds and their contents.[3]

2. **The GTFS Realtime Validator**—a prototype created by CUTR that can be run in a variety of contexts to evaluate a given GTFS Realtime feed and produce a report on the feed's data quality.

The researchers' goal was to combine the breadth and ease of use of Transitland—accessed using any web browser—with the depth and power of the GTFS Realtime Validator, which requires some expertise to run and tune.

The team's combined platform collects GTFS Realtime data from each feed, runs the validation process, and produces a report on any detected errors. Each report shows the counts of data entities, the percentage with errors, and a brief text description of any errors (Figure 2). Links take users to additional documentation about each error type. Some errors also provide further contextual information in maps and tables to assist users as they try to determine root causes (Figure 3).

User Research and Platform Testing

The project team tested the platform by preparing validation reports for seven public transit agencies and reviewing the results in the platform user interface with agency staff members. In these user-testing sessions, the team collected information from agency staff about how GTFS and GTFS Realtime data are currently created at each agency, all known issues, and any open goals. After being given a tour through the platform and its interface, agency staff reviewed the reports for their own GTFS Realtime feeds. Using a standardized question list, agency staff were asked to provide input on both the specific quality checks and the overall presentation and approach used by the platform.

Key findings included the following:

1. GTFS and GTFS Realtime are often the responsibility of separate groups or departments within an agency. For agencies that operate both bus

[3] See https://www.transit.land.

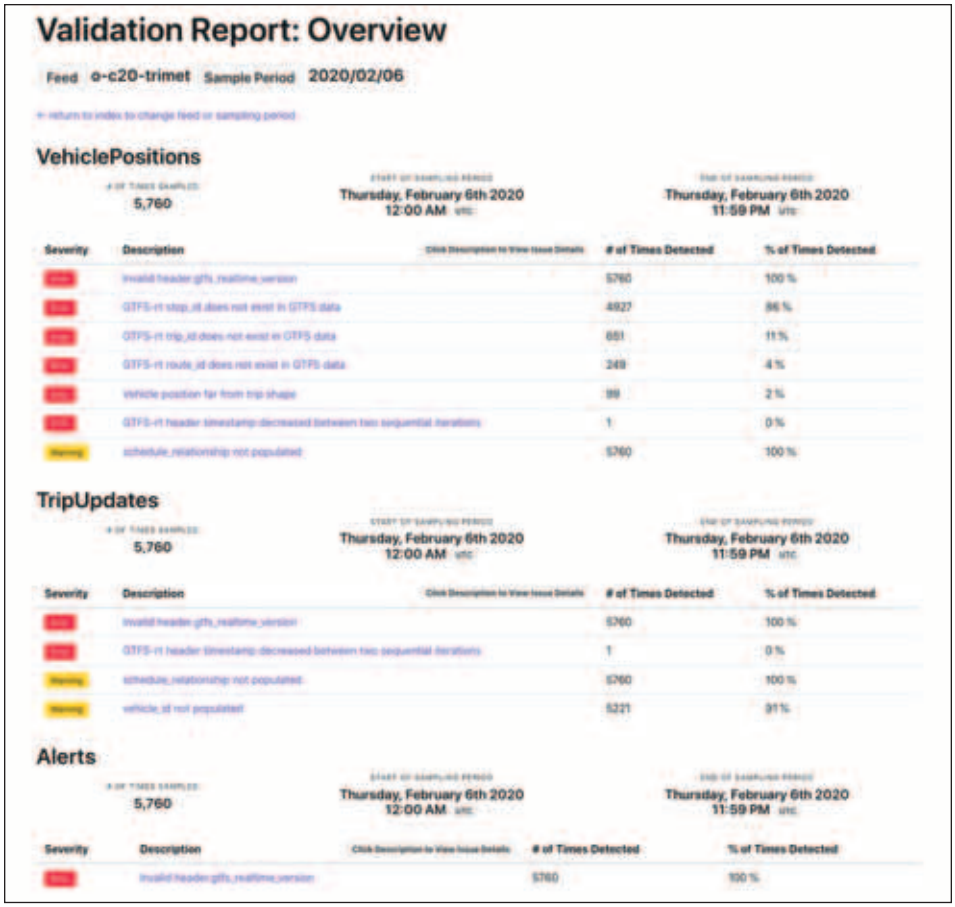

FIGURE 2 The validator library run against a day's worth of GTFS Realtime data displays the new user interface in a web browser.

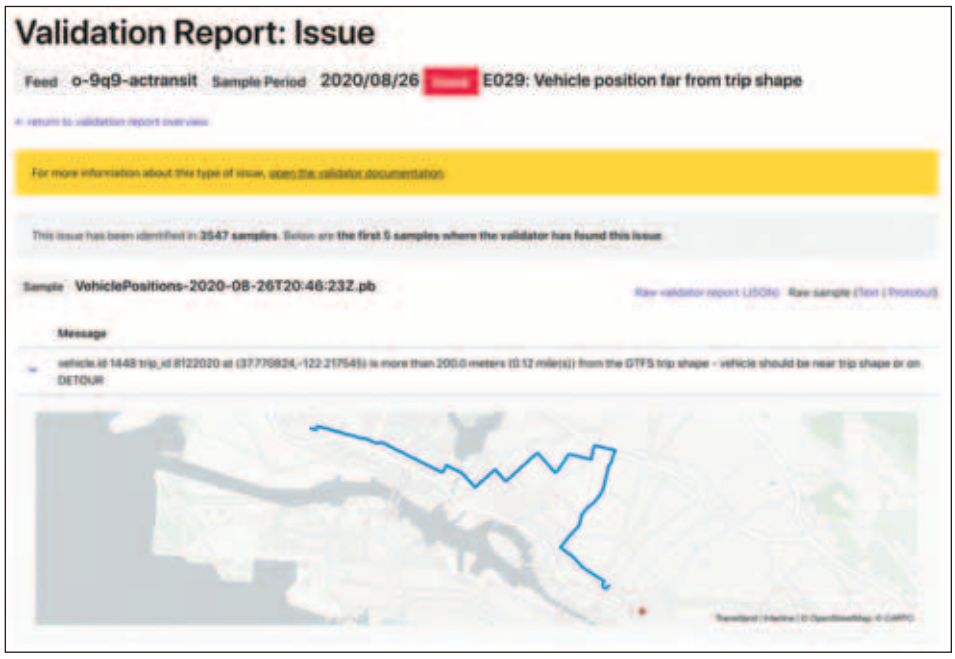

FIGURE 3 A screenshot of the validator platform shows contextual information from both a GTFS Realtime feed and its associated static GTFS feed. The E029 error indicates that a vehicle position is too far from its associated trip shape. The map shows the bus position as a red dot and its scheduled route shape–alignment in blue.

and rail service, separate groups or departments often are responsible for the data systems associated with each mode. A few agencies have formed working groups with cross-cutting responsibilities for static and real-time data across bus and rail.

2. Most agencies validate their static GTFS feeds, but most have not validated the contents of their GTFS Realtime feeds.
3. Vendor systems are "black boxes" with "no visibility into data generation." Agency staff are aware of what types of information are sent to the vendor systems and what data are output, but they are not provided with access to see what processing and transformation happens within the systems. Without this understanding, agency staff may not be able to identify the source of data quality issues.
4. Improvements to GTFS Realtime feed generation often must happen at the same time as new computer-aided dispatch/automatic vehicle location hardware is procured, so overall requests for proposals may be complex to write and evaluate.
5. Several agencies voiced the opinion that vendors rather than agency staff may be more responsive to fixing errors flagged by the validator, as the validator would be seen as an objective tool.

All agencies found the experience of exploring the warnings and errors useful, to the point that most of the sessions ran long. Video calls turned into miniature consulting sessions, with the conversation between agency staff and the research team often ranging from specific GTFS Realtime data fields to system-level architecture concerns. On the one hand, this shows the power of the platform as a way to surface useful information for such wide-ranging investigations. Agency staff reported that some of this information was available through other sources but not aggregated in one place, while other information was previously unknown. On the other hand, these consultations showed how unique each GTFS Realtime system is and how wide a range of information and functionality is necessary for the platform to serve all agencies' potential needs.

Next Steps

The research team welcomes transit agencies with existing GTFS, GTFS Realtime feeds, or both to add them to Transitland's open directory or to provide more information.[4] Registering feeds will make them available to a wider range of users (Figure 4).[5] Registered feeds also will be available through the validation platform when it is rolled out for self-serve use by agency staff. Based on the initial round of user testing, the research team has identified ways to simplify the validation reports so that they will be easy for self-serve use.

Finally, based on the user-testing feedback and the findings from validating a representative sample of real-time data feeds, the team has started to sketch a certification process for GTFS Realtime data feeds. This certification process would provide a common set of minimum expectations, recommended best practices for achieving those expectations, and the hosted web-based platform for assessing feeds.

The researchers are currently pursuing sponsorship for this next round of work, and they welcome feedback and collaboration.

[4] E-mail hello@transit.land for more information.
[5] The map is available at https://www.transit.land/map.

REFERENCES

1. Brakewood, C., and K. Watkins. A Literature Review of the Passenger Benefits of Real-Time Transit Information. *Transport Reviews*, Vol. 39, No. 3, 2018, pp. 327–356. https://doi.org/10.1080/01441647.2018.1472147.
2. Cluett, C., S. Bregman, and J. Richman. Customer Preferences for Transit ATIS. Federal Transit Administration, U.S. Department of Transportation, 2003.
3. Ferris, B., K. E. Watkins, and A. Borning. OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. Presented at the Proceedings of the 28th International Conference on Human Factors in Computing Systems, Atlanta, Georgia, 2010.
4. Gooze, A., K. E. Watkins, and A. Borning. Benefits of Real-Time Transit Information and Impacts of Data Accuracy on Rider Experience. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2351, 2013, pp. 95–103. https://doi.org/10.3141/2351-11.
5. Tang, L., and P. Thakuriah. Ridership Effects of Real-Time Bus Information System: A Case Study in the City of Chicago. *Transportation Research Part C: Emerging Technologies*, Vol. 22, 2012, pp. 146–161.
6. Brakewood, C., G. Macfarlane, and K. Watkins. The Impact of Real-Time Information on Bus Ridership in New York City. *Transportation Research Part C: Emerging Technologies*, Vol. 53, 2015, pp. 59–75.
7. Schweiger, C. L. *TCRP Synthesis 91: Use and Deployment of Mobile Device Technology for Real-Time Transit Information.* Transportation Research Board of the National Academies, Washington, D.C., 2011.
8. Schweiger, C. L. *TCRP Synthesis 115: Open Data: Challenges and Opportunities for Transit Agencies.* Transportation Research Board of the National Academies, Washington, D.C., 2015.

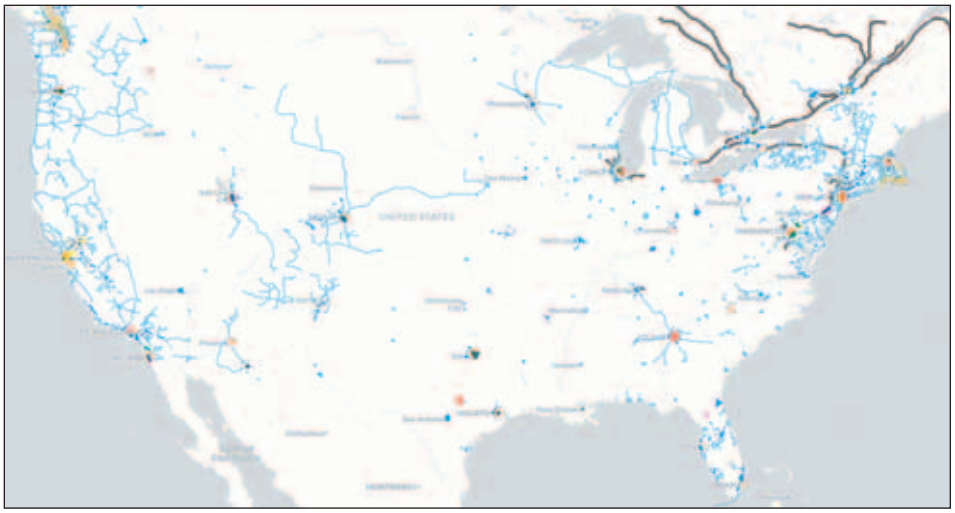

FIGURE 4 Transitland's global transit map shows its coverage of the continental United States. All routes are covered in static GTFS feeds, and some are covered in GTFS Realtime feeds.[4] Users may click on routes to learn more about available data sources. Agencies have the option to specify route colors, and many agencies that operate rail service have provided color codes.