U.S. Department of Transportation
Federal Railroad Administration

RESEARCH RESULTS
FRA OFFICE OF RESEARCH, DEVELOPMENT & TECHNOLOGY

RR 22-02 | **January** 2022

EVALUATING CROWDSOURCED DATA FROM WAZE FOR RAIL SAFETY

SUMMARY

The U.S. Department of Transportation (U.S. DOT) John A. Volpe National Transportation Systems Center (Volpe), under the direction of the Federal Railroad Administration (FRA) Office of Research, Development, and Technology (RD&T), analyzed Waze traffic data and its potential application to driver safety at and near highway-rail grade crossings.

In 2020–2021, the Volpe research team accessed the Waze dataset through the Waze Connected Citizens Program, available throughout the U.S. The team assessed the feasibility of using that data to identify blocked crossings and crossings where there may be heavy vehicular congestion.

The Volpe team determined that Waze alerts data does contain meaningful information, including user-reported traffic jams. However, due to the nature of crowdsourced data, there are limitations to using this dataset, including reliability of the reports as well as discrepancies between where incidents occur versus where users report them.

BACKGROUND

Several railroads have started using the Waze navigation platform to deliver safety messages to drivers approaching highway-rail grade crossings. In 2018, the Long Island Rail Road (LIRR) worked with Waze to integrate a railroad crossing warning feature to alert users at grade crossings on the LIRR system [1]. In 2019, Norfolk Southern (NS) partnered with Waze "to increase driver awareness around railroad crossings in the City of South Fulton and some areas of Clayton County," in Georgia [2]. Other railroads are also exploring the use of navigation apps to deliver rail safety messages.

In April 2020, Waze rolled out a new feature consisting of a grade crossing alert to its navigation application nationwide [3]. Figure 1 shows an example of the Waze grade crossing alert. A similar feature has also been recently implemented in Google Maps. However, this technology's potential effectiveness in increasing highway-rail crossing safety has not yet been studied.

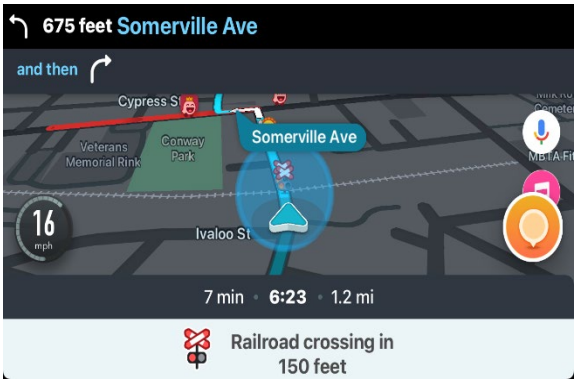

Figure 1. Example of a Waze Grade Crossing Alert

OBJECTIVES

The objective of this project was to research the status and effectiveness of using Waze data to improve safety at highway-rail grade crossings in the U.S. The specific focus was on identifying crossings blocked by standstill rail traffic and crossings with nearby roadway congestion.

METHODS

Ohio had the highest number of reported blocked crossings in the nation over the 10-month period from December 2019 to October 2020 according to the FRA Blocked Crossing

webpage [4]. It represented 27 percent of all reported blocked crossings. Therefore, public at-grade crossings in Ohio were selected for a case study to determine the feasibility of using Waze data to help identify blocked crossings and roadway congestion near crossings.

The U.S. DOT, in partnership with Waze, receives and curates both user-reported and auto-generated data from the Waze Connected Citizens Program [5]. These data sets are stored on the U.S. DOT's Secure Data Commons (SDC) platform, available for approved users [6]. The SDC platform was created to serve as a data warehousing and analysis platform for transportation researchers. The Waze data is available for the entire United States and has been archived since April 2017. The Waze database includes both user-reported and system generated data from Waze users. These data are organized into three tables: alert, jam, and irregularity. For this research, the analysis was focused on alert and jam data.

The alert table contains all user generated incidents reported by Waze users. Drivers using Waze navigation can report accidents, hazards, jams, as well as road closures on the road through the Waze mobile application. By default, an alert is active for 30 minutes, but it can be extended or shortened based on other Waze users confirming or rejecting the presence of these reports [7].

Besides users' reported events, Waze also obtains driving speed, traffic condition and location data by pinging its users' mobile phones every few minutes. Waze takes this real-time update along with its proprietary historical data and blends them together using its algorithm to generate a traffic congestion level as well as delay information. These data are stored as two separate tables: Jam and Jam Point Sequence in the SDC database. The Jam table contains information about average speed, level and duration of the jam, and time. The Jam Point Sequence table contains XY coordinates.

Since streets are represented as lines on a map, and Waze Jam data is collected as a point dataset, a polygon was created around each grade crossing to capture Waze jam data related to that grade crossing. Figure 2 shows a markup of a polygon at Warren Sharron Road grade crossing (Crossing ID 544729H) in Trumbull County, OH. The polygon is 250 feet long on both sides of the crossing with a 50-foot flat-ended buffer on both sides of the road.

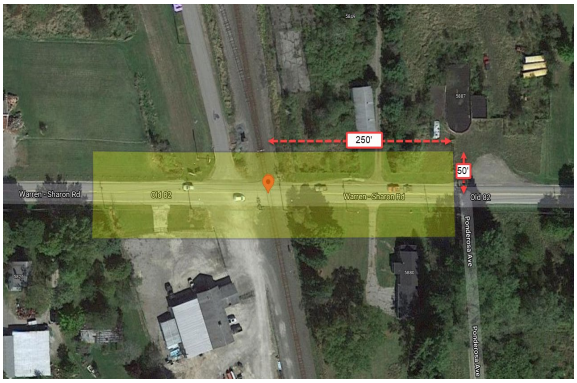

Figure 2. Example of a Buffer around a Grade Crossing

The traffic congestion level data field ("level") in the Jam table describes the traffic condition of a road segment. There are six different levels of traffic jam in the Waze database, from free flow (level = 0) to blocked traffic (level = 5). The thinking is that if a crossing is blocked for extended period of time and there are drivers using Waze navigation stuck at a crossing, the Waze navigation would generate a level 5 traffic congestion level for that road segment.

To find level 5 jams within the crossing polygon, the Volpe team first joined Jam and Jam Point sequence tables using Jam ID. For this pilot project, we selected Jam data reported for the month of January 2019. We then queried for level 5 jams that were reported in the state of OH for the month of January 2019. This dataset was then pulled into a geodataframe (from the Python package geopandas) and performed spatial analysis to determine level 5 jams that occurred within crossing polygons.

In addition to analyzing Waze jam data, we also looked at the alerts that were reported by users

within each crossing polygon between January 2019 and February 2021.

RESULTS

For the month of January 2019, there were a total of 691,880 records of level 5 jams reported in Ohio. As discussed earlier, Waze Jam data are stored in two separate tables (Jam and Jam Point Sequence) in the SDC database. We needed data from both tables to match level 5 jams that occurred within the crossing polygons. Unfortunately, we discovered that Waze reuses jam identification numbers and there was no easy way to ensure that an entry in the jam table that matches our filtering criteria will consistently join to the correct jam point sequence when duplicates exist. For example, a level 5 jam that did not occur near a grade crossing could be joined to a jam point sequence which is near a grade crossing, and our table of results might include the points that fell within the buffer despite being associated with the wrong jam. On the other hand, a level 5 jam near a grade crossing may be left out of the results because it was joined to a jam point sequence which was not near a grade crossing.

We tried some filtering techniques, including making sure the city names matched between the jam table and the crossing polygon data. However, even with these filtering techniques, there were still duplicate records of Jam ID within a city. We also tried matching street names, but since there are many ways street names could be spelled out between Jam table and crossing polygon data (e.g., Street, St, North, N., Avenue, Ave., etc.) there was no easy way to apply this filtering technique without extensive manual processing.

Analysis of the Waze alert data revealed that a total of 26,636 unique alerts were reported within 1,860 crossing polygons over a two-year period from January 2019 to February 2021. The remaining 3,726 public at-grade crossings in Ohio did not have any alerts reported over the study period. Table 1 shows the distribution of 26,636 alerts reported by alert type.

Table 1. Distribution of Alerts at Crossings by Alert Type

Alert Type	Count	% of Total
Accident	574	2.2%
Jam	7,103	26.7%
Weather Hazard	7,227	27.1%
Road Closed	11,732	44.0%
Total	**26,636**	

One of the objectives of this research was to study roadway congestion near grade crossings. Since we were unable to use system generated jams (Waze Jam data), we explored users' reported traffic jams. There are three levels of traffic jams that users can report on -- Moderate, Heavy, or Standstill. The 7,103 unique traffic jams were reported at 1,111 public at-grade crossings. Standstill traffic reports were most common, representing 52 percent of all traffic jams, followed by Heavy traffic at 26 percent, and Moderate traffic at 14 percent. The remaining 8 percent of the traffic jams did not have a subtype associated with them. **Figure 3** shows the top 10 crossings in Ohio with the most reported traffic jams over the study period.

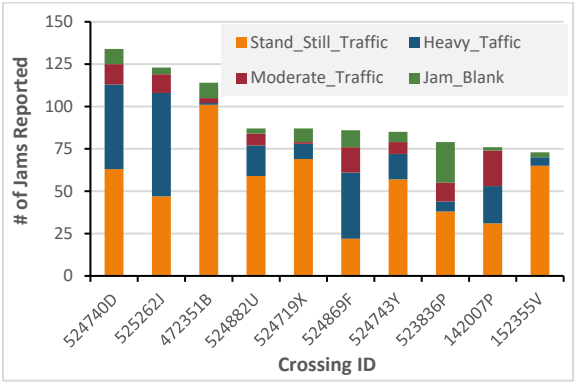

Figure 3. Top 10 Crossings with most Reported Traffic Jams in OH, January 2019 – February 2021

It should be noted that as with any crowdsourced data there can be a variation in where users report traffic jams, independent of when and where traffic jams occur [8].

CONCLUSIONS

Waze jam data are stored in two separate tables in the SDC database. To match level 5 jams (blocked traffic) that occurred within crossing polygons, we needed data from both tables. Unfortunately, it was discovered that there was no easy way to join the two tables due to duplicate records of Jam IDs (related column between two tables). Waze alerts data does contain meaningful information including users reported traffic jams but, due to the nature of crowdsourced data, there are limitations to using this data set, including reliability of the report as well as variation in where incidents occur versus where users report them.

FUTURE ACTION

Crowdsourced traffic data, such as data from Waze, has the potential to aid in the identification and remediation of locations with traffic congestion issues at grade crossings. FRA continues to analyze all available traffic safety data for potential application to rail safety.

REFERENCES

[1] Metropolitan Transit Authority. (June 2018). LIRR Partners With Waze for Railroad Crossing Safety Initiative.

[2] CBS46 Atlanta. (July 2019). Norfolk Southern Partners with Waze to Promote Rrail Safety.

[3] Waze Forum. (October 2021).

[4] FRA. Blocked Crossing Incident Reporter.

[5] Waze Connected Citizens Program.

[6] U.S. DOT. Secure Data Commons Platform.

[7] Goodall, N and Lee,E, (June 2019). Comparision of Waze crash and disabled vehicle records with video ground truth.

[8] Flynn, D., Gilmore, M., and Sudderth, E, (November 2018). Estimating Traffic Crash Counts Using Crowdsourced Data.

ACKNOWLEDGMENTS

This work was performed under interagency agreements between the FRA's Train Control and Communication Division and the Volpe National Transportation Systems Center's Systems Safety and Engineering Division.

CONTACT

Francesco Bedini Jacobini
Program Manager
Federal Railroad Administration
Office of Research, Development and Technology
1200 New Jersey Avenue, SE
Washington, DC 20590
(202) 493-0800
Francesco.Bedini@dot.gov

Tashi Ngamdung
Operations Research Analyst
Volpe Center
Systems Safety and Engineering Division
55 Broadway
Cambridge, MA 02142
(617) 494-2937
tashi.ngamdung@dot.gov

KEYWORDS

Waze, crowdsourced data, railroad safety, grade crossing, blocked crossing

CONTRACT NUMBER

DTFR5317X00026