

# Transportation Research Informatics Platform (TRIP)

James Pol, P.E., PMP

Turner-Fairbank Highway Research Center (TFHRC)

Federal Highway Administration (FHWA), Office of Safety Research and Development (R&D)

April 28, 2020



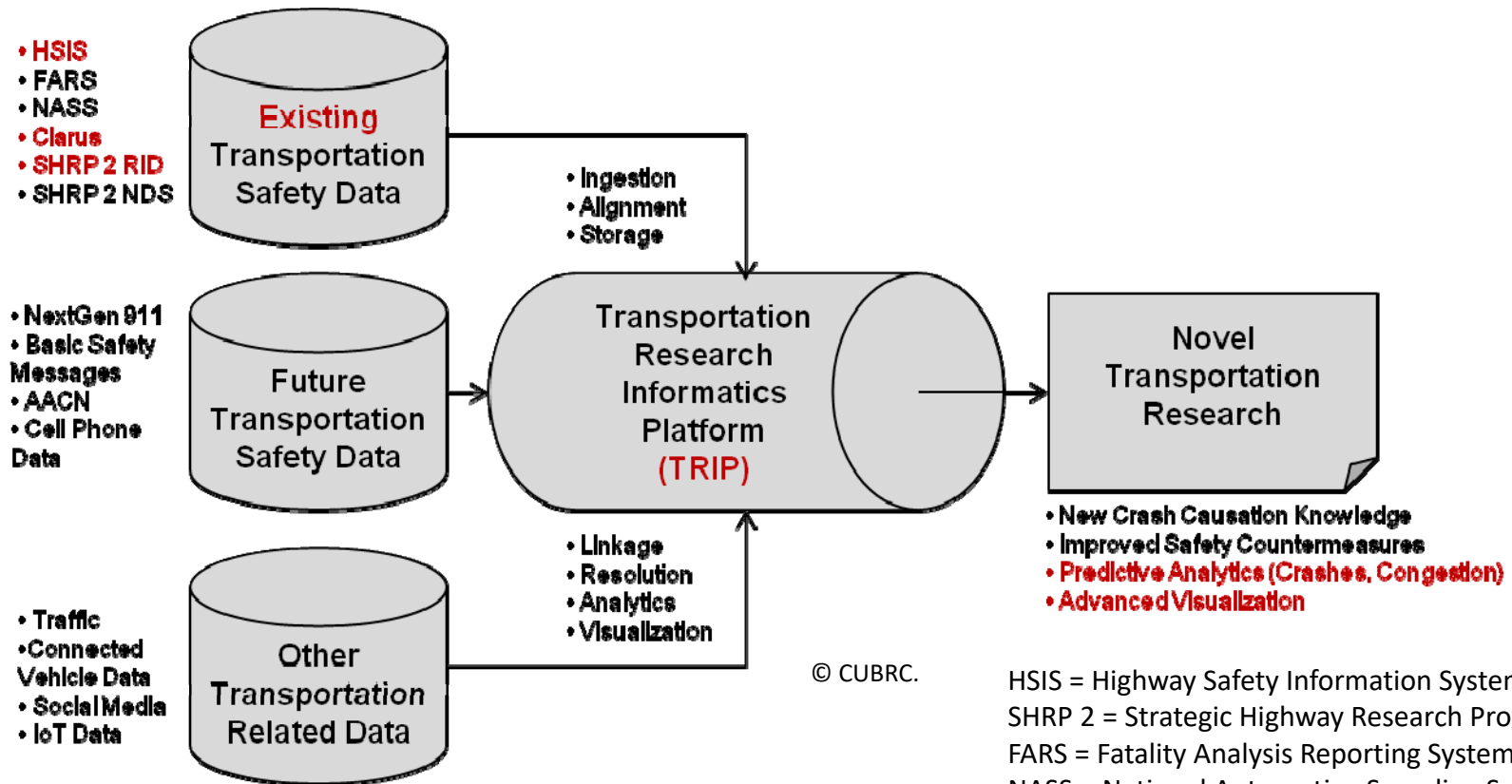
U.S. Department of Transportation  
**Federal Highway Administration**

# TRIP

- Transportation researchers and practitioners have access to **unprecedented amounts of data**, but lack the tools to easily store, manipulate, and analyze this data.
- TRIP is an **end-to-end informatics-based solution** to enable the efficient handling and analysis of 'Big Data' in the transportation domain.
- TRIP **demonstrates the feasibility of utilizing open source**, widely supported state-of-the-art technologies to process massive amounts of data.
- TRIP's main advantages over existing systems include **efficiency, scalability, and customization**.
- Initial design objectives support transportation **safety** analyses, however the platform is capable of supporting **planning** and **operations** activities.
- The goal is to make TRIP readily available to transportation research, planning, and operations agencies.



# High-Level Prototype Diagram



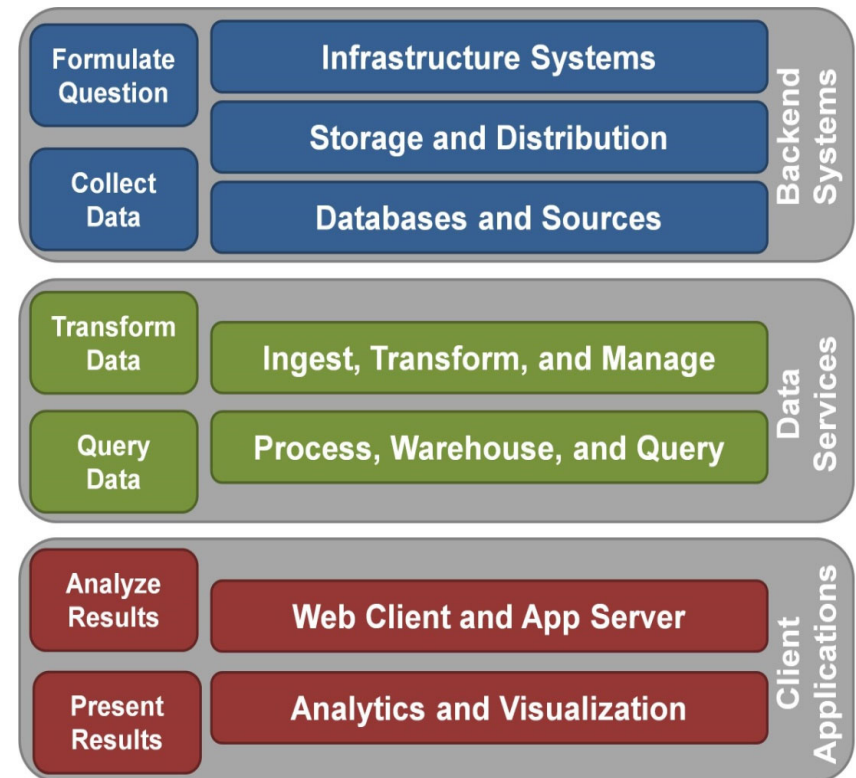
© CUBRC.

HSIS = Highway Safety Information System  
 SHRP 2 = Strategic Highway Research Program  
 FARS = Fatality Analysis Reporting System  
 NASS = National Automotive Sampling System  
 RID = Roadway Information Database  
 NDS = Naturalistic Driving Study  
 IoT = Internet of Things



# Flexible TRIP Architecture

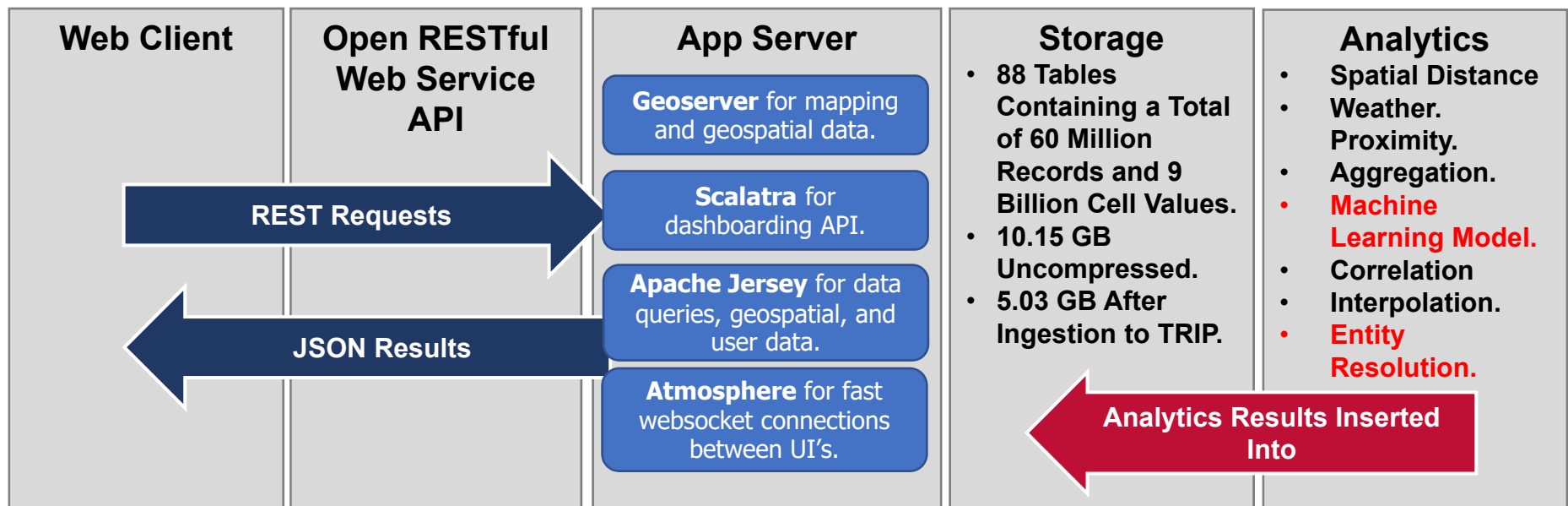
- TRIP uses artificial intelligence and open-source tools to exploit and integrate disparate sources of data, which eases data discovery among existing, unconnected databases.
- TRIP's platform is discipline agnostic and was originally developed for safety research. Database "attachments" only need some common features, such as time or geospatial data.



© CUBRC.



# TRIP Web-based User Interface



© CUBRC.

API = Application Programming Interface  
 UI = User Interface

# Machine Learning Model Research Question Example

<p>What roadway curvature characteristics present an increased risk factor for commercial vehicles?</p>	<p>RID</p>	<p>Stacked box plots, machine learning model</p>	<ol style="list-style-type: none"> <li>1. Determine how to retrieve a crash with a commercial vehicle.</li> <li>2. Retrieve that set of crashes and retrieve the curvature information associated with that crash.</li> <li>3. Plot the roadway curvature characteristics vs. all of the crashes and stack the same plot on top of the roadway condition information.</li> <li>4. Attempt to model the crash types using the curvature characteristics as the source of information.</li> <li>5. Select several different machine learning model types as the basis for evaluation and report performance of those models.</li> </ol>	<p>Complete: A combination of roadway curvature characteristics and event-based crash data can provide many different opportunities to gain insights using a variety of different analytics visualization and training techniques. Three different models were invoked to return results for this question. They include a decision tree, a random forest, and a k-nearest neighbor model.</p>
---	------------	--	---	--

© CUBRC.



# Entity Resolution Research Question Example

<p>Find all senior drivers who struck pedestrians at intersections with crosswalks but without pedestrian crossing lights during twilight hours.</p>	<p>HSIS, RID</p>	<p>Entity Resolution</p>	<ol style="list-style-type: none"> <li>1. Find when twilight was for the time range of interest.</li> <li>2. Grab an oversized sample of drivers (<math>x &gt; 40</math>) that struck pedestrians with the parameters defined.</li> <li>3. Allow user to select age and output a table.</li> </ol> <p>Event Joining Criteria (order of operations/evaluation sequence): Date of Event; Age of Driver; Gender of Driver; Vehicle Year; Vehicle Make;</p> <p>Event Validation Criteria: Time of Event, Vehicle Heading, Speed Limit.</p>	<p>Complete: Tables provide results of the 25 crashes that met the criteria in the HSIS database and also for the 45 crashes that occurred in the RID database.</p>
--	------------------	--------------------------	--	---

© CUBRC.



# Entity Resolution

	RID	HSIS	Match	RID	HSIS	Match	RID	HSIS	Match
Case	NA	201345286	N	NA	2012017191	N	NA	201345286	N
Report	E310559	NA	N	E152536	E152995	N	E162518	E16003	N
Time of Day	819	820	P	1809	1630	N	145	1828	N
Year	1998	1998	Y	2006	2005	P	1995	2010	N
Make	Toyota	Toyt	P	KIA	KIA	Y	Honda	Niss	N
Heading	South	North	N	North	North	Y	West	Northeast	N
Age	44	45	P	50	49	P	19	56	N
Gender	Female	Female	Y	Female	Female	Y	Female	Male	N
Speed Limit	35	35	Y	60	60	Y	55	60	N
Road Type	Straight	Straight	Y	Straight	Straight	Y	Straight	Straight	Y
Road Surface	Dry	Dry	Y	Dry	Dry	Y	Dry	Wet	N
Weather	Clear	Clear	Y	Clear	Clear	Y	Clear	Raining	N
Score	11		Y	-3		P	-11		N

TRIP automates some of the data discovery among disparate data sources. The entity resolution process assigns a “score” of how well records and fields align, enabling more time for analysis.

© CUBRC.





# Query Builder User Interface

The screenshot displays the TRIP Query Builder UI. On the left, a query diagram shows a green triangle labeled 'Returned Results' connected to an orange diamond labeled 'Must Match All'. Below the diamond is a box for 'Filter by Field (timestamp) between Sep 2nd, 2012 03:00 PM and Dec 3rd, 2012 09:00 PM' with a 'Geo Filter: 1 areas selected.' label. To the right of the diamond is a box for 'Filter by Field (alcohol\_test\_result) 48 categories selected.'.

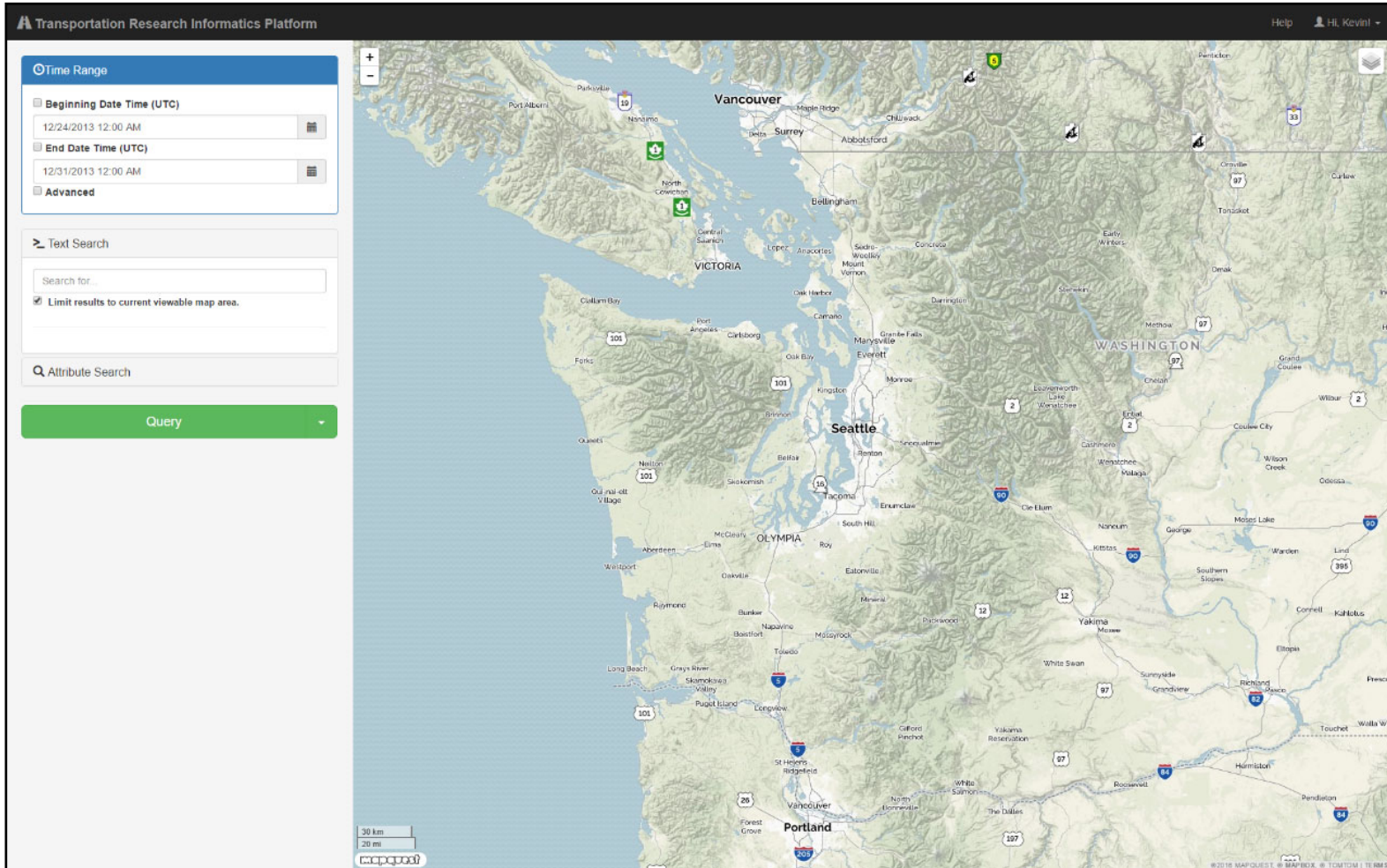
On the right side of the interface, the 'Query Name' is 'Seattle Seahawks Query'. Below this, there are tabs for 'Browse Fields', 'Search Columns', 'Graphlets', and 'Settings / Execute'. A section titled 'Select a field to view browse it's contents.' contains a table:

Field Name	Field Type	View...
age	Numeric Categorical	View...
air_bag_type	Discrete Categorical	View...
alcohol_test_result	Numeric Categorical	View...

Below the table, there are checkboxes for field types:  Identifier (8),  Discrete Categorical (155),  Numeric Categorical (23),  Numeric Continuous (4), and  Standard Columns (3).

At the bottom right, there is a bar chart titled 'Adjust Viewable Fields' and 'Adjust Chart Display'. The chart shows two series: 'crash\_irs' (teal) and 'crash\_other' (purple) across various categories on the x-axis. The y-axis ranges from 0 to 700. The chart is currently set to 'Stacked' display.

# User Interface Intro



© CUBRC. Map: © 2018 MapQuest, © Mapbox, © TomTom.



U.S. Department of Transportation  
Federal Highway Administration

“Transportation Research Informatics Platform (TRIP).” CUBRC. Accessed April 21, 2020.  
<https://www.cubrc.org/index.php/data-science-and-information-fusion/trip>.

# Temporal Queries Example

Transportation Research Informatics Platform

Success! Query Completed successfully. 569 results retrieved in 1493ms.

**Time Range**

- Beginning Date Time (UTC)  
11/01/2012 12:00 AM
- End Date Time (UTC)  
11/30/2012 12:00 AM
- Advanced

Results search to selected days of the week.

Sunday ▾

Adjust how to search with time.

- Time encapsulates entire period
- Time is recurrent between start and end time

Text Search

Search for: \_\_\_\_\_

Limit results to current viewable map area.

Attribute Search

Query

**Crash Locations**

**Crash Attributes**

Data Source	RID
Table Name	crash_irs
Collision Report Number	3127113
Timestamp	2012-11-18 12:57:00
Lat/Long	47.177511, -121.787874
Location	King
Roadway Type	Two Way - Undivided
Weather	Raining
unit_number	1
age	31
person_type	MV Driver
event_1	Ran off the Road

Show Accident Details Show Clarus Details

© CUBRC. Map: © 2018 MapQuest, © Mapbox, © TomTom.



# Spatial Queries Example

The screenshot displays the Transportation Research Informatics Platform (TRIP) interface. On the left, a sidebar contains search filters. The 'Time Range' section includes 'Beginning Date Time (UTC)' (11/01/2012 12:00 AM) and 'End Date Time (UTC)' (11/30/2012 12:00 AM). The 'Advanced' section has a 'Restrict search to selected days of the week' dropdown set to 'Sunday'. The 'Text Search' section has a search box containing 'CenturyLink Field', which is circled in red, and a checked option 'Limit results to current viewable map area'. Below this is a search result snippet for CenturyLink Field. The 'Attribute Search' section is empty. A green 'Query' button is at the bottom of the sidebar. The main map area shows an aerial view of Seattle with several blue location pins. The top of the interface shows the platform name and user information.

© CUBRC. Map: © 2018 MapQuest, © Mapbox.



# Attribute Queries Example

The screenshot displays the 'Data Source Attribute Examiner' window in the Transportation Research Informatics Platform (TRIP). The search term 'alcohol' is entered in the search bar. The results table is as follows:

Database	Field Name	Column Type	Type Of Hit	Matched Value	Add	Show
crash_lrs	alcohol_test_statuses	Discrete Categorical	column name	alcohol_test_statuses		Show
crash_lrs	alcohol_test_result	Discrete Numerical	column name	alcohol_test_result		Show
crash_lrs	contributing_circumstance_1	Discrete Categorical	element in histogram	Under Influence of Alcohol	Add	Show

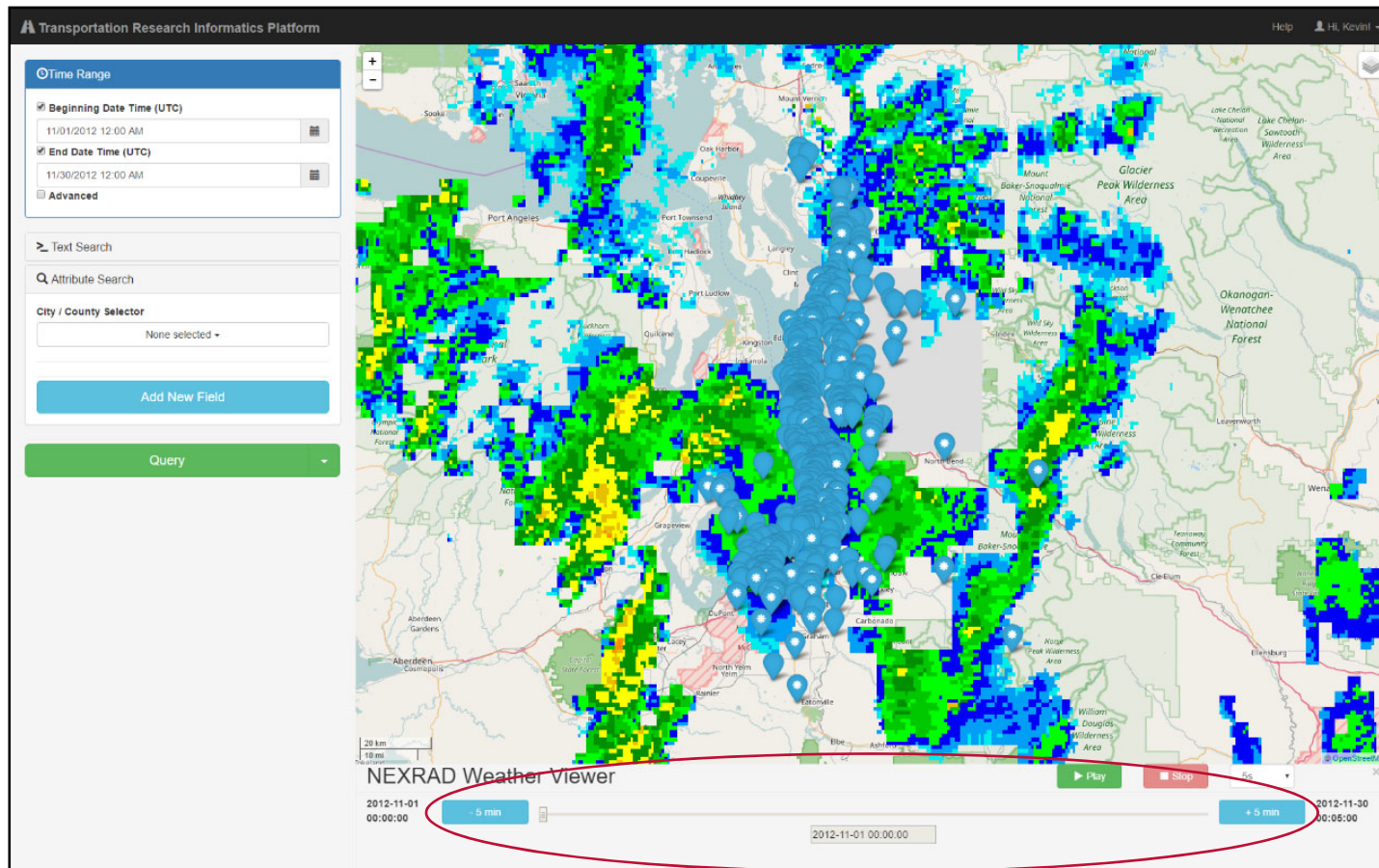
Below the table, it indicates '44 selected' and 'Hide blank attributes from graph' with 'Number of blank or null values: 315156'. A histogram titled 'Histogram Distribution for attribute alcohol\_test\_result' shows a distribution of values from 0 to 0.5, with a peak around 0.15. The histogram is circled in red. The background shows a map with blue location markers.

© CUBRC. Map: © 2018 MapQuest, © Mapbox.





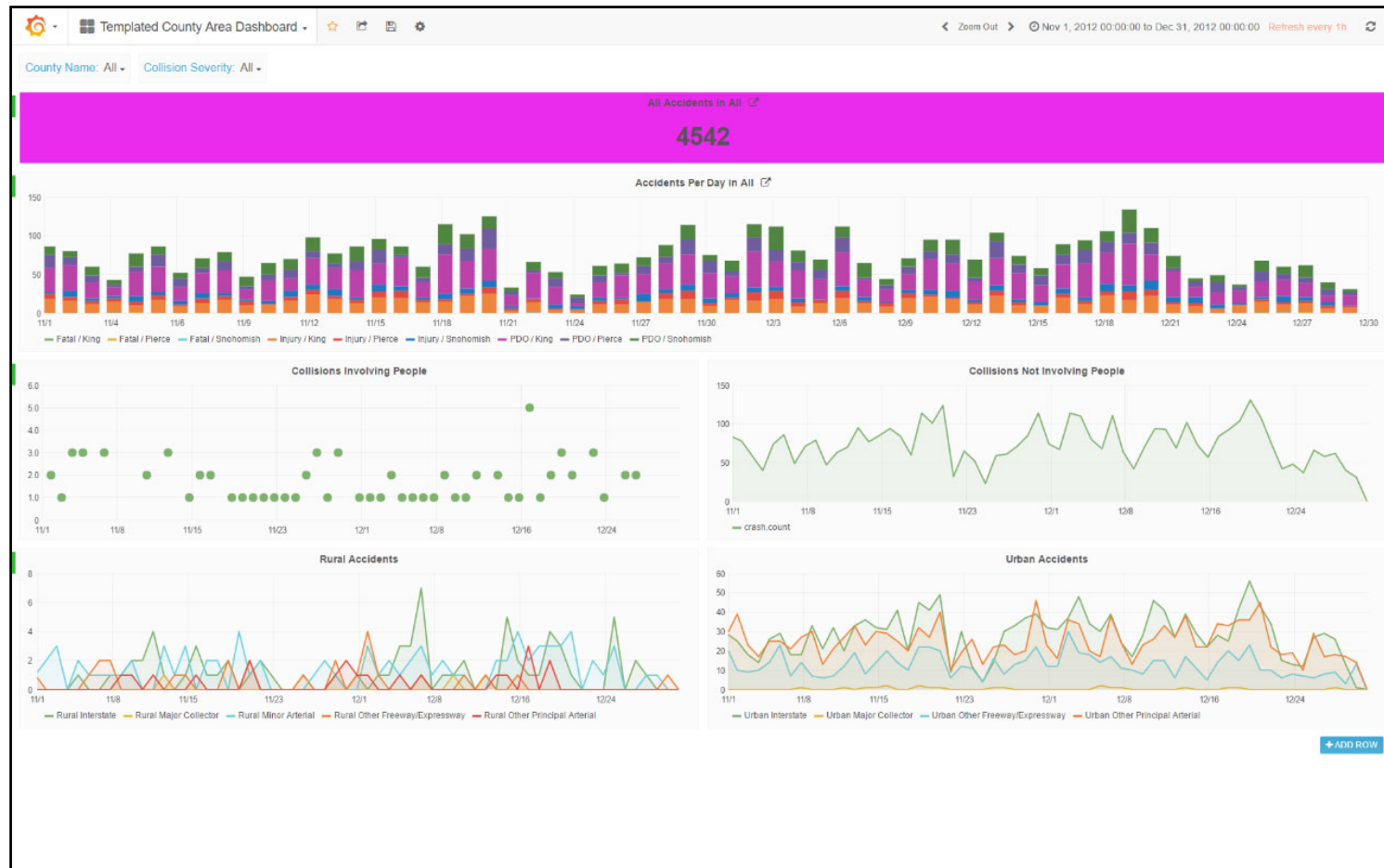
# Query with Weather Radar Imagery



© CUBRC. Map: © OpenStreetMap.

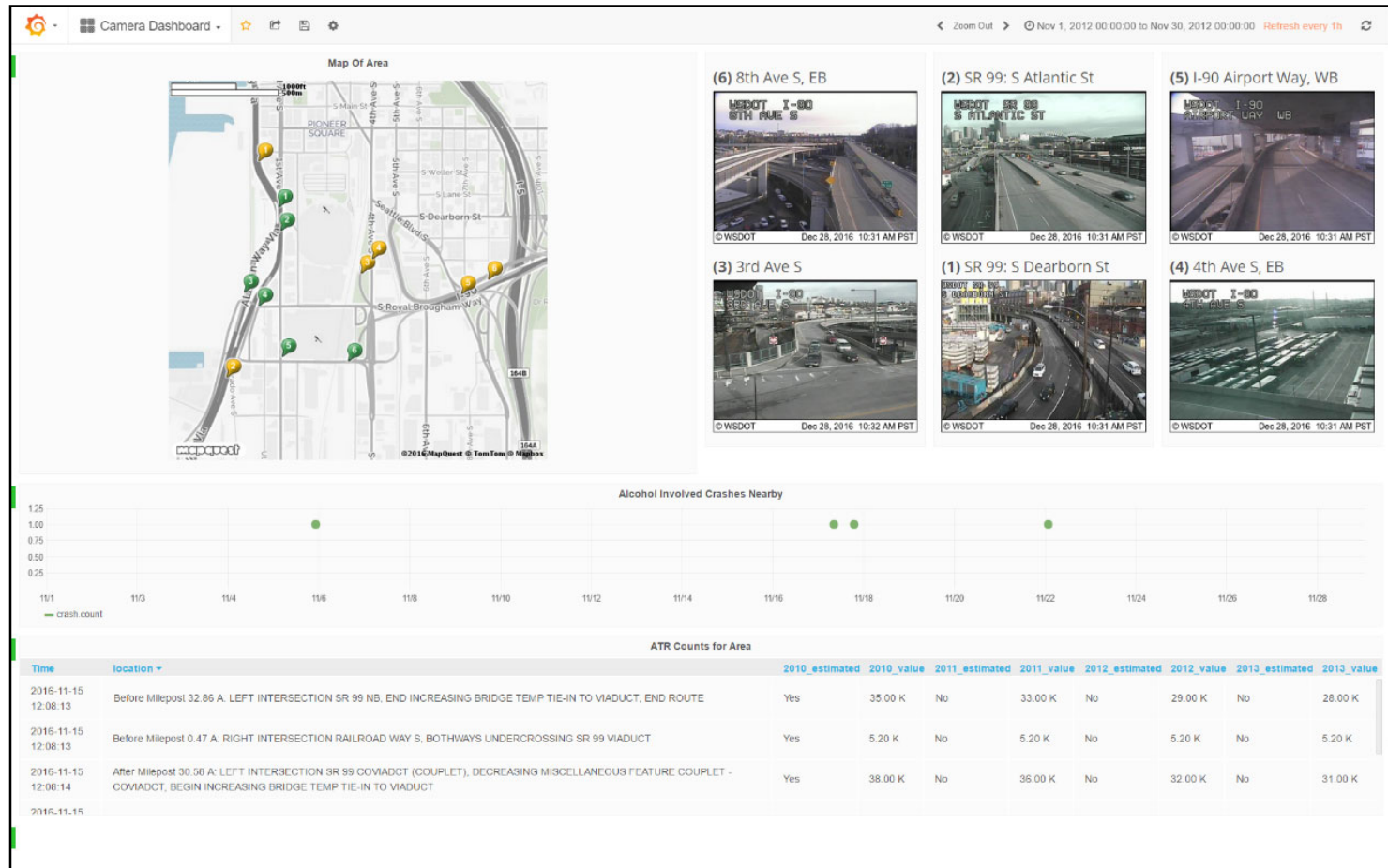


# Dashboarding Example





# Real-time Operations Feeds



© CUBRC. Map: © 2018 MapQuest, © Mapbox, © TomTom.  
Photos: © Wisconsin Department of Transportation.



# The WHY?

No current commercial off-the-shelf system can integrate disparate and unstructured data in a way that provides value for transportation researchers and practitioners.

- The platform accepts structured, semi-structured, and unstructured data in a scalable way; it ingests, aligns, and stores these data and then provides linkages and resolves the data for analysis.
- The platform is capable of handling massive amounts of data and also provides the capability of fast access and analysis times—even on a small server system. The platform easily scales from one to 100 machines with no additional configuration.



# Market...

- All State departments of transportation are data driven organizations. The challenge is that most data stores are defined to suit narrow application purposes.
- Many “connect the dot” opportunities are not suitably explored. For example, a 2009 National Bureau of Economic Research study found that the “introduction of E-ZPass reduced prematurity and low birth weight among mothers within 2 km of a toll plaza.”  
<https://www.nber.org/papers/w15413.pdf>
- The FHWA Resource Center and TFHRC are exploring tools like TRIP to assist with nationally scaled transportation performance measures.



# FHWA Research Objectives

- States that do not have the technical ability to analyze large datasets typically contract this task out.
  - Impedes national performance metrics analysis.
  - Closes off stores of data that may reveal further innovations.
  - Delivers inconsistent analysis of critical issues.
- Safety R&D is considering TRIP to augment HSIS and introducing HSIS-like data to nonparticipating States.



# FHWA Investment to Date

- The Exploratory Advanced Research program developed TRIP to address the topic of working with massive data sets.
- The final report (FHWA-HRT-19-008) features platform set up and use in a variety of environments.
- The Technology Readiness Assessment in March 2017 assigned TRIP a Technology Readiness Level 5, which integrates components demonstrated in a laboratory environment.

<https://fhwa.dot.gov/research/exploratory-advanced-research/research/technology-readiness-assessment-work-ear-program>



# Accelerated Market Readiness

- New activity is launching in 2020 to improve the Technology Readiness Assessment from Level 5 to Level 8.
- More use cases are needed in order to demonstrate to end users how the informatics platform may be used for analysis.
- The paradox is that for TRIP to be attractive to State and local agencies for their analyses, these agencies need to see a variety of analyses conducted to demonstrate the effectiveness of the platform.



# Disclaimer

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this presentation only because they are considered essential to the objective of the presentation. They are included for informational purposes only and are not intended to reflect a preference, approval, or endorsement of any one product or entity.



# James Pol, P.E., PMP

Technical Director

FHWA, Office of Safety R&D

[James.Pol@dot.gov](mailto:James.Pol@dot.gov)

# Thank you!

