ctre
Center for Transportation Research and Education

Developing an Architecture to Integrate Safety, Mobility, and Traffic Data

tech transfer summary

This project developed a proof-of-concept architecture to integrate mobility and weather data into available crash data to better understand the role these factors have in crashes and to create a foundation for integrating additional data sets in the future.

June 2021

RESEARCH PROJECT TITLE
Developing an Architecture to Integrate Safety, Mobility, and Traffic Data

SPONSORS
Iowa Department of Transportation (InTrans Project 18-676)

PRINCIPAL INVESTIGATOR
Skylar Knickerbocker, Research Engineer
Center for Transportation Research and Education, Iowa State University
515-294-7916 / sknick@iastate.edu
(orcid.org/0000-0002-0202-5872)

MORE INFORMATION
intrans.iastate.edu

CTRE
Iowa State University
2711 S. Loop Drive, Suite 4700
Ames, IA 50010-8664
515-294-8103

The mission of the Center for Transportation Research and Education (CTRE) at Iowa State University is to develop and implement innovative methods, materials, and technologies for improving transportation efficiency, safety, reliability, and sustainability while improving the learning environment of students, faculty, and staff in transportation-related fields.

IOWA STATE UNIVERSITY
Institute for Transportation

TSF-000-0(446)--92-00

Objective

The objective of this project was to develop a process that allows for easy integration of weather and traffic (i.e., probe) data with available crash data from the Iowa Department of Transportation (DOT) that can support efforts to determine the role these factors have in crashes and ultimately lead to improved decision-making.

Background

The Iowa DOT consumes data from multiple streams that are stored to assist in better decision-making. These include, but are not limited to, a state-of-the-art crash repository, access to detailed weather data through the Iowa Environmental Mesonet (IEM) at Iowa State University, INRIX probe data, Wavetronix sensor data, crowdsourced data, and work zone details. The data archive of all these sources extends back for several years with the cumulative data size for the past five years of data in the range of 20 TB or more.

Problem Statement

Despite access to unprecedented amounts of data, decision-makers are often restricted in their ability to explore multiple data due to factors that include the storage of data in silos and the size and availability of different data sources.

Since most users do not have the time or resources to dig through the available data to find relevant information, pre-canned reports are often serially produced from individual sources of data and circulated to decision-makers without providing a comprehensive picture of the issue.

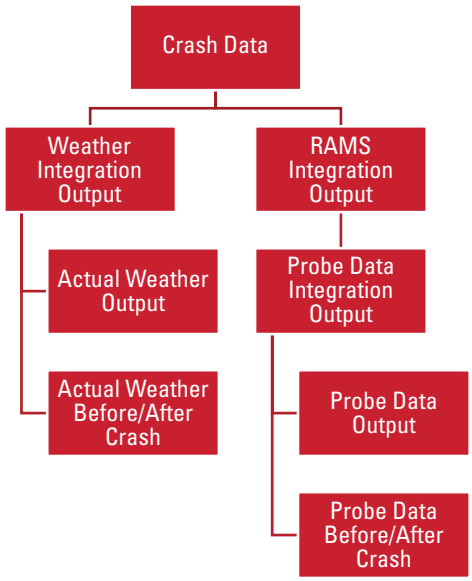
Project Description

Data Sources

The data sources identified and used for the data integration methodology and outputs developed as part of the project include the following:

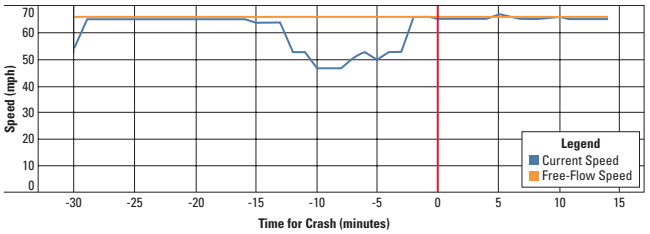
- Iowa DOT crash data
- Weather data provided by the IEM
- INRIX probe data
- Iowa DOT Roadway Asset Management System (RAMS)

The data sources focus on the ability to provide actual conditions of the traffic and weather at the time of the crash.



Overview of the outputs of the data integration process

Weather grid and crash data points



Probe data before and after the crash

Data Integration Outputs

The data integration methodology developed for both the weather and probe data is a multistep process with intermediate outputs created along the way that are needed for later steps in the integration. Seven outputs were created that are all related back to the crash data.

The research team considered any potential output that may be beneficial for future integration efforts and outputted those as separate tables for future use. The primary outputs of the integration process are the weather and probe data at the time of the crash, which allow for the data to be joined directly with the crash data similar to other attributes collected or derived within the crash data. For advanced analysis, a Python script was also developed that allows the probe and weather data to be extracted for a configurable amount of time before and after each crash.

Data Integration Process

Two separate processes were created for data integration including a process to integrate crash and weather data and a process to integrate crash and probe data.

Crash and Weather Data Integration Process

1. Combine crash data with weather shapefile
2. Extract weather JavaScript Object Notation (JSON) files provided by the IEM
3. Join weather and crash data at time of crash files
4. Extract additional weather data before and after the crash

Crash and Probe Data Integration Process

1. Integrate crash data with the RAMS
2. Integrate crash data and probe data
3. Join probe and crash data at time of crash files
4. Extract additional probe data before and after the crash

After developing the integration engine, a prototype online safety and operations evaluation tool was developed to explore the interactions between the crash, weather, and probe data. Methods of analyzing the data were explored to show the potential benefits of using the integration, but it is expected that future research will further explore and understand these relationships.

Key Findings

- The architecture in this report identified the various data sets needed to support crash, weather, and traffic analysis and then developed a process to extract the data. The created outputs from these processes assign the attributes for the related weather and probe data to each crash, which can then be treated like all other crash data within Iowa.

- The considerable size of the data required developing a pipeline that allowed the data to be stored cost-effectively while also reducing the time to query the data sets. To achieve this, the size of the data was reduced through multiple methods including the format was changed to Parquet, the data were partitioned to allow for better grouping of the data, and a big data querying service was utilized in Amazon Athena.
- The crash and probe data integration required additional consideration as compared to the weather data due to the nature of how crashes occur on roadways and how probe data contains data for both directions of travel. Thus, processes were developed to further refine the data to better determine the route and direction of travel.
- For a majority of the data set integration efforts, the RAMS integration developed as part of this project will allow for streamlined integration with other data sets. A simple linear overlay can be used to create the relationship between the data or to extract the relevant information.

Recommendations for Future Research

Future enhancements can be made to the architecture developed in this project to allow for additional summary statistics to be created for each crash as well as the ability to extract additional data for nearby road segments or weather grids. The summary statistics can include information such as the amount of precipitation for a defined amount of time before the crash, whether speeds were trending up or down before the crash, and whether speeds were impacted after the crash. The summary statistics can provide additional attributes but would require additional workflows to extract and summarize these data for other users.

Limitations also exist within the architecture for understanding the impacts on traffic upstream and downstream of the crash. The current process can be modified to support extracting additional probe data for nearby segments.

Finally, the weather data provided by the IEM also includes weather forecasts every six hours based on the same grid and attributes. The forecast data have not been explored fully and may have benefits in some applications.

Future Data Set Integration Opportunities

Additional data sets that can be used to enhance the crash data or used in future research based on input from the Iowa DOT and other relevant stakeholders include the following:

- Advanced Traffic Management System
- Snowplow automatic vehicle location (AVL)
- Winter road conditions
- Traffic and road weather snapshot and videos
- Pavement condition data
- Intersections
- Work zones

Implementation Readiness and Benefits

The proof-of-concept architecture developed provides an initial step in integrating crash, weather, and traffic data for more widespread use within Iowa. The architecture includes the following benefits over the existing system:

- Data from multiple sources are saved in a format that leads to extended ability to query across these data sources. Thus, queries, such as how many crashes occurred during snow and congested conditions, should be easy to perform.
- High-performance computing systems are used to store and manipulate the data, and hence the data processing time will be significantly reduced.
- The data set is available for visual queries. Decision-makers can visually filter and explore the data using a dashboard developed for the project.

The architecture was also designed to support future advanced analysis by developing simple scripts that can be used to dynamically extract the weather or probe data before and after a defined set of crashes. It is expected that future research will utilize these data outputs and processes to improve weather, traffic, and crash related efforts moving forward.