EVALUATING CRASH DATA SYSTEMS IN THE WESTERN US

ITE 2020 Joint Western and Mountain District Annual Meeting
Authors: Lia Klein (HNTB) & Emily Johanning (VHB)

INTRODUCTION

The importance of traffic safety data is immeasurable, used to evaluate key areas for safety improvements by government agencies, engineers, policymakers, and the general public. It is essential in identifying deficiencies on public roads and can be an effective tool for evaluating the benefits of infrastructure improvements. However, access to data and methods of reporting varies widely between states. By taking a closer look at the incident data available from regional resources, common differences become apparent, including varying methods of access, formatting, and definitions. To address this problem, we propose opportunities to standardize resources and reporting for a more efficient system, including the development of national standards for definitions and a centralized API for querying incidents. Through this study, we highlight the importance of data availability and user experience. Each state's crash data system has been evaluated quantitatively based on federal guidance and qualitatively based on usability. Comparison of data obtained from each jurisdiction results in discrepancies between format and content. By determining inconsistencies between regions and highlighting effectiveness of system elements, the goal of zero deaths can progress by promoting a more complete and reliable crash database to allow traffic engineers to better identify, analyze, and improve unsafe roadways.

OBJECTIVE

From the many agencies and regions, our analysis focuses on four main jurisdictions: California, Oregon, Washington, and Alaska. These four states represent a huge section of the American population with over 52 million citizens residing along the coast, roughly 15% of the entire nation. The western US is not only populated, but immense in geographic size, with the 4 states totaling almost 1 million square miles of land (over one quarter of the entire US). Crash data can be defined with four main attributes that provide the most basic information related to safety: frequency, rate, type, severity[1]. Quality and consistency of this data is key to determining the substantive safety of the road, the performance of the facility in relation to its contextual characteristics. A roadway's crash history is more relevant in relation to facilities of equivalent traffic volume, location, and terrain. Data from two different jurisdictions with conflicting definitions or inconsistent attributes cannot be properly compared.

QUANTIFYING DATA QUALITY

In an effort to create uniformity at the many state and local level, several federal DOT agencies have been publishing a guideline identifying a minimum set of motor vehicle crash data elements that states should

[1] *Highway Safety Improvement Program Manual,* FHWA, 2010.

include in their data systems: Model Minimum Uniform Crash Criteria (MMUCC)[2]. Created in collaboration with the National Highway Traffic Safety Administration (NHTSA) and the Federal Highway Administration (FHWA), the most recent Fifth Edition (2017) includes a breakdown of 27 unique attributes associated with accurate, uniform, and complete crash data. This study uses these attributes to evaluate and quantify the crash data provided by state and local agencies along the West Coast of the US. These attributes are relevant to safety analysis as they provide the ability to quickly identify potential causes or single out specific factors of each incident. A breakdown of these data attributes used by each state's Department of Transportation's crash data systems can be found in Table 1.

Table 1: MMUCC Attribute Guidelines in Western US Crash Data Systems

		STATE DOT			
ATTRIBUTE		**AK**	**WA**	**OR**	**CA**
C1	Crash Identifier	✓	✓	✓	✓
C2	Crash Classification	✓	✓	✓	✓
C3	Crash Date and Time	✓	✓	✓	✓
C4	Crash County	✓	✓	✓	✓
C5	Crash City/Place (Political Jurisdiction)	✓	✓	✓	✓
C6	Crash Location	✓	✓	✓	✓
C7	First Harmful Event	✓	✓	✓	✓
C8	Location of First Harmful Event Relative to the Trafficway	✓	✓	✓	✓
C9	Manner of Crash/Collision Impact	✓	✗	✓	✓
C10	Source of Information	✓	✗	✓	✗
C11	Weather Conditions	✓	✓	✓	✓
C12	Light Condition	✓	✓	✓	✓
C13	Roadway Surface Condition	✓	✓	✓	✓
C14	Contributing Circumstances – Roadway Environment	✓	✗	✓	✓
C15	Relation to Junction	✓	✓	✓	✓
C16	Type of Intersection	✓	✗	✓	✗
C17	School Bus-Related	✓	✗	✓	✓
C18	Work Zone-Related (Construction/Maintenance/Utility)	✓	✗	✓	✗
C19	Crash Severity	✓	✓	✓	✓
C20	Number of Motor Vehicles Involved	✓	✓	✗	✓
C21	Number of Motorists	✓	✓	✓	✓
C22	Number of Non-Motorists	✓	✗	✓	✓
C23	Number of Non-Fatally Injured Persons	✓	✓	✓	✓
C24	Number of Fatalities	✓	✓	✓	✓
C25	Alcohol Involvement	✓	✗	✓	✓
C26	Drug Involvement	✓	✗	✓	✗
C27	Day of Week	✓	✗	✓	✓
		27/27	**17/27**	**26/27**	**23/27**

[2] *MMUCC Guideline, Fifth Edition (2017),* USDOT/NHTSA, 2017.

This method of evaluation allows us to score each state based on the number of attributes in the raw crash data that match the attributes suggested on a federal level. Results show that Alaska DOT and ODOT most closely followed the MMUCC guidance, with 27 and 26 of the attributes listed respectively. However, it is important to note that some information may not have its own attribute but could be listed within another attribute not identified by the MMUCC. For example, alcohol involvement may be listed under cause of a crash but does not have its own unique attribute. Therefore, a universal adaption of the federal format would create a more transferrable and comparable environment for data analysis.

USER EXPERIENCE

While these attributes may provide a general assessment of data quality, there are other features that represent other aspects of a robust and user-friendly data system. Taking a look at accessibility and ease of use from an analysist's perspective provides a qualitative insight into the data.

Availability

One of the largest discrepancies between state data sets is the availability of data for public or private use. Many users of this data include consultants working on behalf of public agency infrastructure projects or private developments. Collecting and analyzing this data can be a lengthy process, especially when the effort required to obtain raw files requires agency approval and a processing time of up to ten business days. For example, the Washington State Department of Transportation (WSDOT) chooses this more tedious yet secure method for their crash data, whereas the Oregon Department of Transportation (ODOT)reduces the required approval and organization time by providing a web tool that allows the public to freely download crash data from any public roadway instantly. California provides a similar service where the requested data is emailed to the user after they register for a free account linked to the data portal. Since the Alaska crash data system is only available to government employees and no public-facing data request portal exists, the Alaska crash data was not able to be evaluated for many of the following characteristics.

Format

The common format for raw data from state DOTs is comma-separated values (.csv) which can be opened in MS Excel or other common data processing software.

Compatibility

A common tool used for analyzing and representing traffic safety data is ArcGIS, which can transform the raw data into map format using geographic location data, allowing for the creation of a quick visual reference. None of the CSV files available from Washington, Oregon, or California DOTs is compatible with ArcGIS and requires extensive formatting and decoding.

Geographic Coordinate System

One MMUCC attribute is "Crash Location", which is defined as "the exact location in the trafficway to document where the first harmful event of the crash occurred". This information is reported in the form of geographic coordinates. While California and Oregon DOTs use the universal latitude and longitude format, WSDOT reports crash data in Washington State Plane Coordinates. This becomes an issue when

combining data from two different states into the same geographic data network, so standardization of this attribute is recommended.

Comparison of State and City Systems

By creating a data system that is user-friendly, comprehensive, and transferrable, state and local jurisdictions can promote efficiency and thoroughness in traffic analysis. In addition to the state systems, local data systems from the West Coast cities with more independent data sets – the Seattle DOT (SDOT) and the San Francisco Municipal Transportation Agency (SFMTA) – were also evaluated. Each system was evaluated for the general assessment categories previously listed, as well as additional features that may be considered helpful tools for analysts in Figure 1 below.

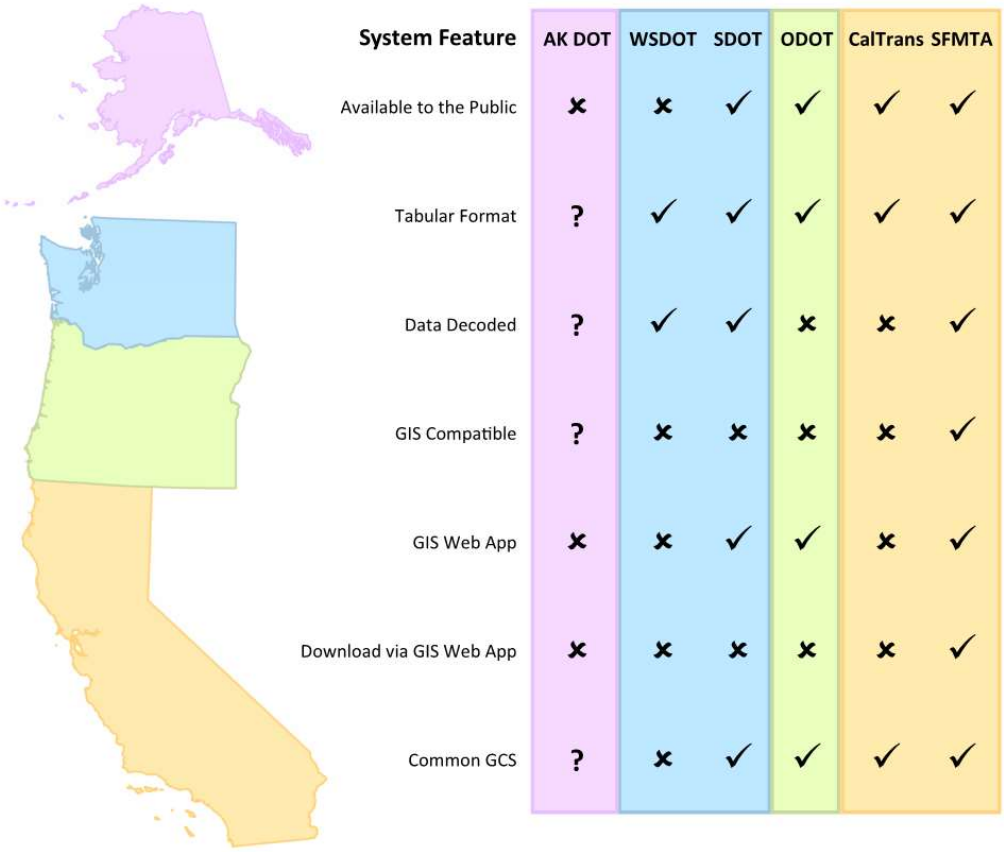

Figure 1. User Experience by State and Local Agencies in the Western US

By these measures, SFMTA is the ideal data system on the West Coast as it provides open, quality data to the public in both a GIS web application and raw format compatible with GIS software.

TAKING A CLOSER LOOK

One outstanding example a future API related to user experience could be based off of is the data system published by SFMTA. Through their TransBASE Dashboard, the user can query the city's crash data by street, neighborhood, intersection, or user-defined boundary and filter crashes by date, time, severity, type, and many other factors. The web application allows for visual and detailed inspection of the data

before downloading, which is available in several forms including pdf, csv, and shapefile (GIS format). Figure 2 shows the user interface of this application.

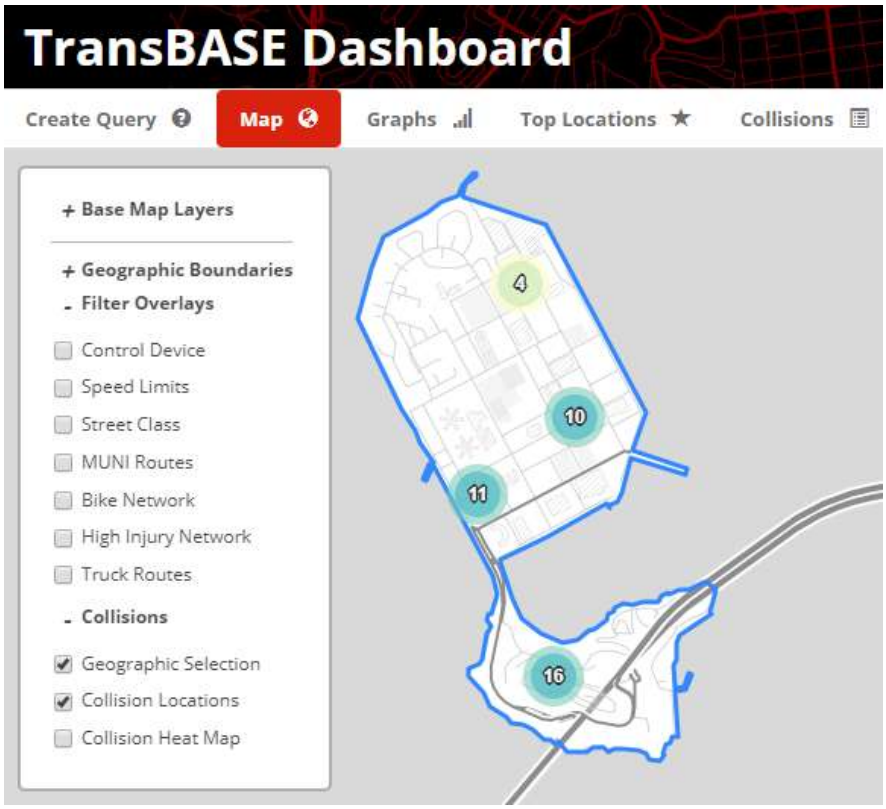

Figure 2. User Interface of the SFMTA TransBASE Crash Data System (Source: TransBASE.sf.gov.org)

The interface not only creates a visual tool for identifying crash hot spots, but also creates graphs, charts, and tables of selected crashes, a quick tool for analysts who could have otherwise spent hours creating these graphics manually. The data is organized, formatted, and decoded, minimizing the effort of the user.

Despite the quality of the user experience, the SFMTA data set matches only 17 of 25 relevant MMUCC attributes (excluding attributes for city and county, as all crashes in this database have the same value). Missing attributes include number of non-motorists involved, school and work-zone relations drug and alcohol involvement, and roadway environment. Although several of these factors can be determined from additional data provided on the TransBASE site or through independent investigation, it is more ideal to have all relevant values included in the main data set.

Conclusion & Recommendations

Several of the data systems published by state and local agencies have stood out as excellent sources of information for traffic engineers and analysts, including the SFMTA's TransBASE and the Alaska DOT data set. However, all systems lacked either standardized attributes or a user-friendly interface. The benefits of quality data can only be fully achieved through an equally quality system for viewing, sorting, and analyzing. If agencies were to adopt the best aspects of both, their systems would be ideally setup to promote the efficient use of safety data critical to the health of our citizens and public infrastructure.