Regional Safety Data Scoping Study

Developed for the
Metropolitan Washington Council of Governments

By the University of Maryland
Center for Advanced Transportation Technology Laboratory

October 2011

Primary Point of Contact:

Michael L. Pack
PackML@umd.edu
301.405.0722

Executive Summary

This report includes a thorough comparison analysis of Virginia, Maryland, and District of Columbia Police Accident Databases to determine the level of effort in creating a regional data visualization and analysis tool. The remainder of this document describes in greater detail:

1) the existing data analysis tool that was created for the State of Maryland,
2) an analysis of the similarities and differences between the MD, DC, and VA police accident dataset, and
3) options, risks, and a budget for creating regional analysis capabilities.

After completing this analysis, it has been determined that the overall cost of creating a highly intuitive, full-feature, regional safety analysis tool will be approximately $195,000 and take approximately 1-year from receipt of agency data and notice to proceed. There are several risk factors and consideration—technical, legal, and political—that should be considered prior to embarking on this development initiative, but the potential rewards far outweigh the risks.

Background

The condition, safety, and efficiency of the transportation system in the states of Maryland, Virginia, and the District of Columbia affect the quality of life for residents and workers who live in, commute to, or visit these regions. Maintaining and operating the transportation system requires that transportation planners throughout the region are able to understand and respond to it in an intelligent, consistent, and efficient manner.

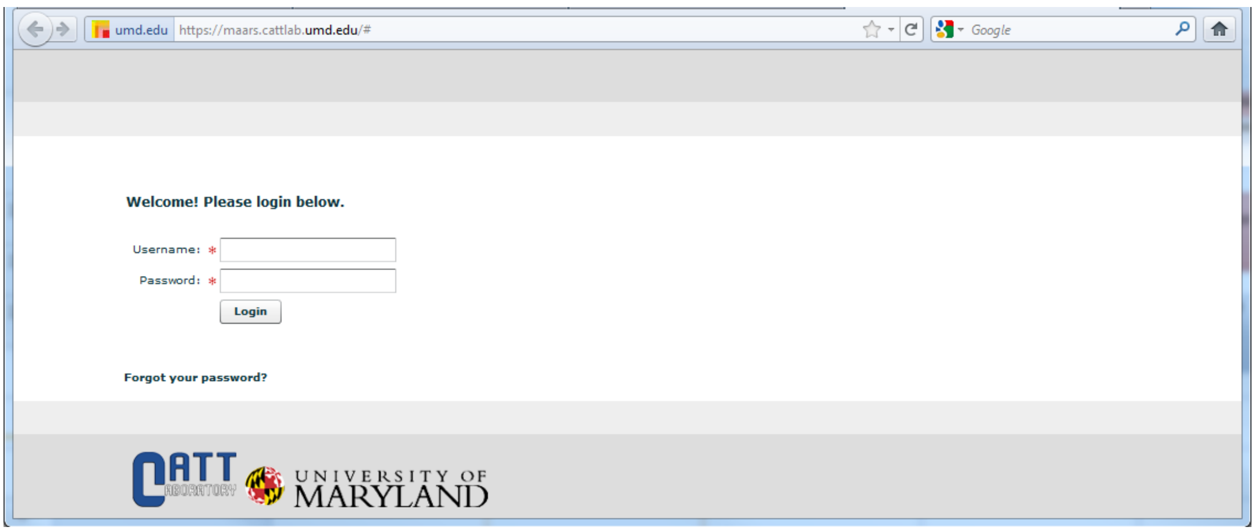
Having access to a highly intuitive, web-based, and, therefore, uniformly accessible crash analysis tool for these three regions would be extremely valuable to state safety analysts, regional planners, and local traffic engineers as it would allow for consistent and efficient analysis using standardized datasets throughout the region. The tool could allow people who do not have in-depth computer training or knowledge of query languages (such as SQL) to build complex queries easily and without worry of error. This opens a wide range of possibilities for policy makers, planners, and traffic engineers, and will enhance their ability to respond to safety concerns from the public. The tool will help analysts to derive accident hot-spots, trends, and other information that was previously unattainable or extremely difficult to glean without significant effort.

A web-based crash analysis, visualization and mining tool for the Maryland State Highway Administration's Office of Traffic and Safety (OOTS) for use by state, regional, and local transportation engineers and analysts is nearly complete. The tool was developed by staff from the Center for Advanced Transportation Technology Laboratory under the direction of Michael L. Pack. The web-based application allows users to dynamically interact with, visualize and query data from the Maryland Automated Accident Reporting System (MAARS) that is managed by the Maryland State Highway Administration. This will help engineers identify locations in need of safety and/or operational improvements.

Safety analysts at the Metropolitan Washington Council of Governments and in the state of Virginia and the District of Columbia are aware of the development of this tool and have expressed an interest in either adapting the tool for the combined region or possibly creating three separate, yet similar tools for each region. Before the University of Maryland team can move forward with recommendations and a proposal to build a regional tool, seed funding is needed to conduct an analysis of Virginia, Maryland, and District of Columbia Police Accident Databases to determine the level of effort of creating a regional data visualization and analysis tool.

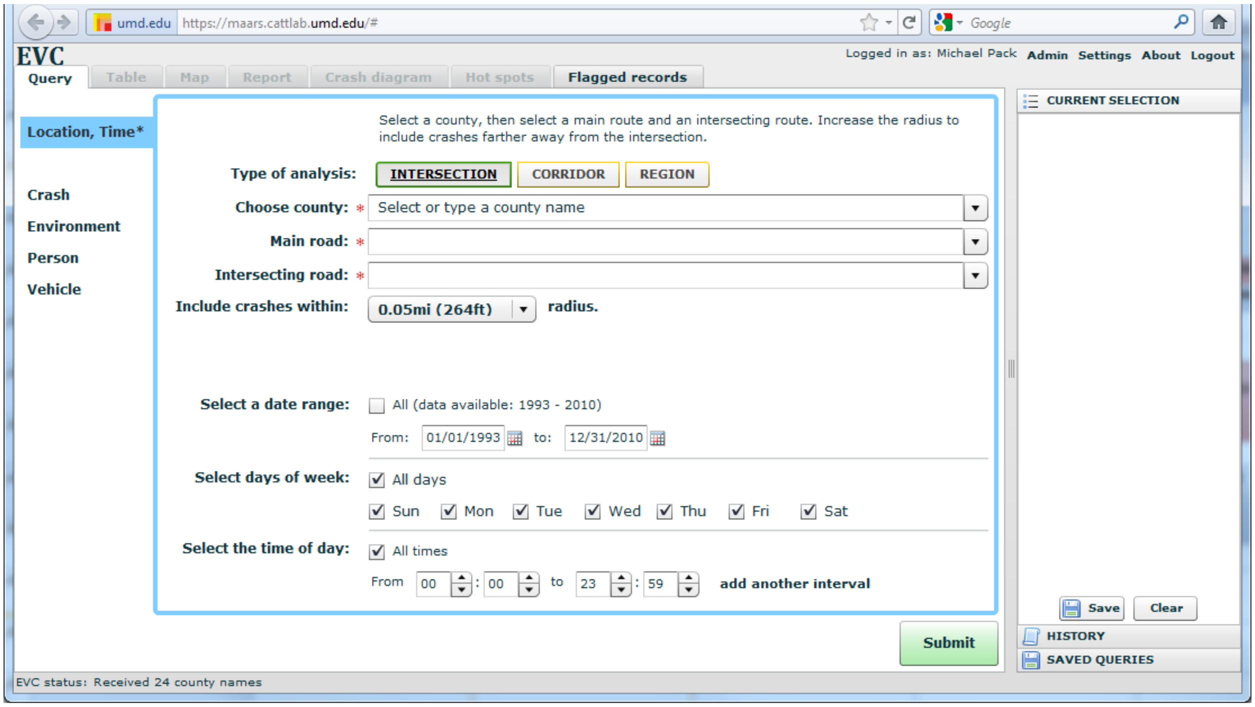
Description of the Existing Tool

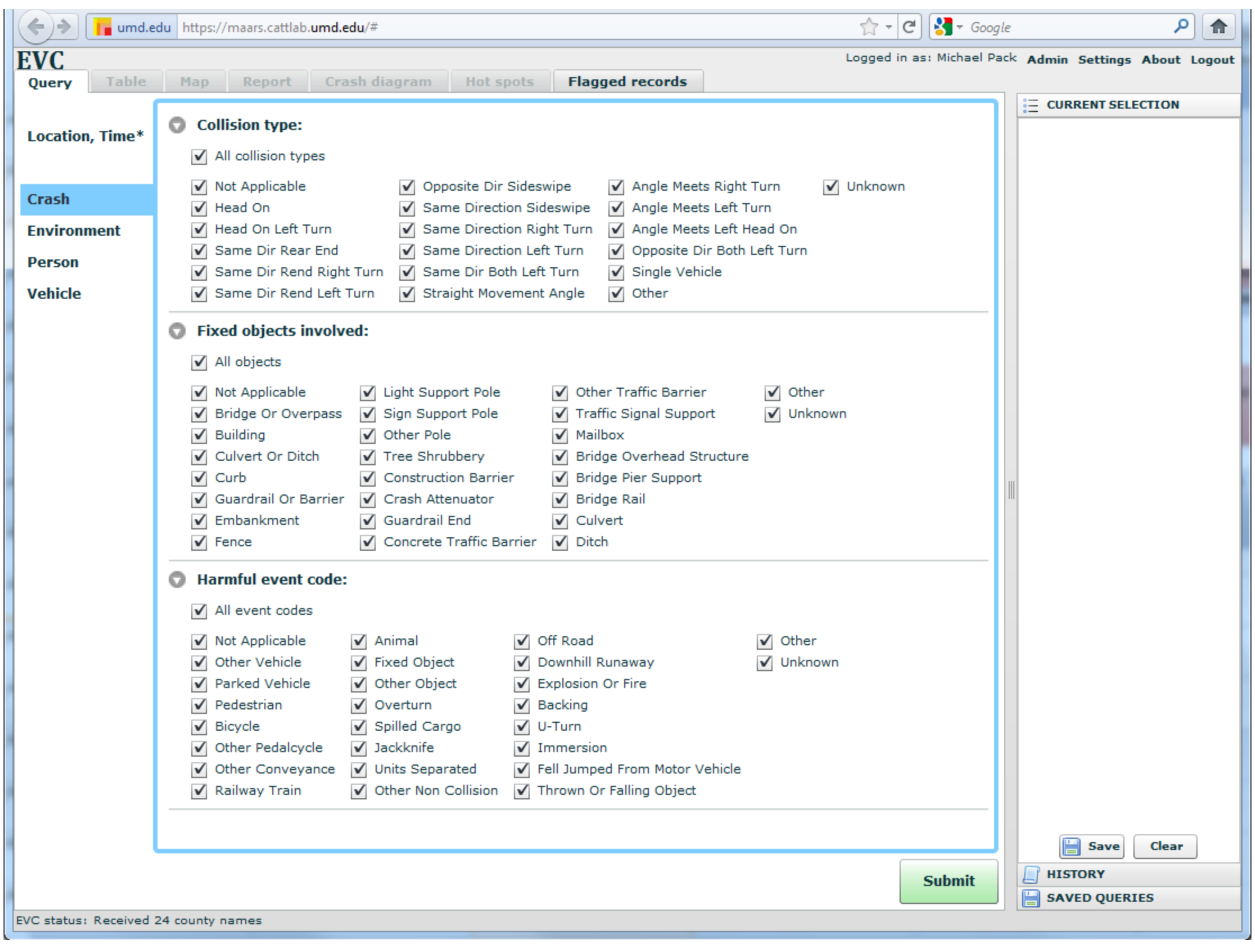
The existing data analysis tool created for the Maryland State Highway Administration and local traffic engineers is a secure, web-accessible data visualization and analysis application that can be accessed via https://maars.cattlab.umd.edu

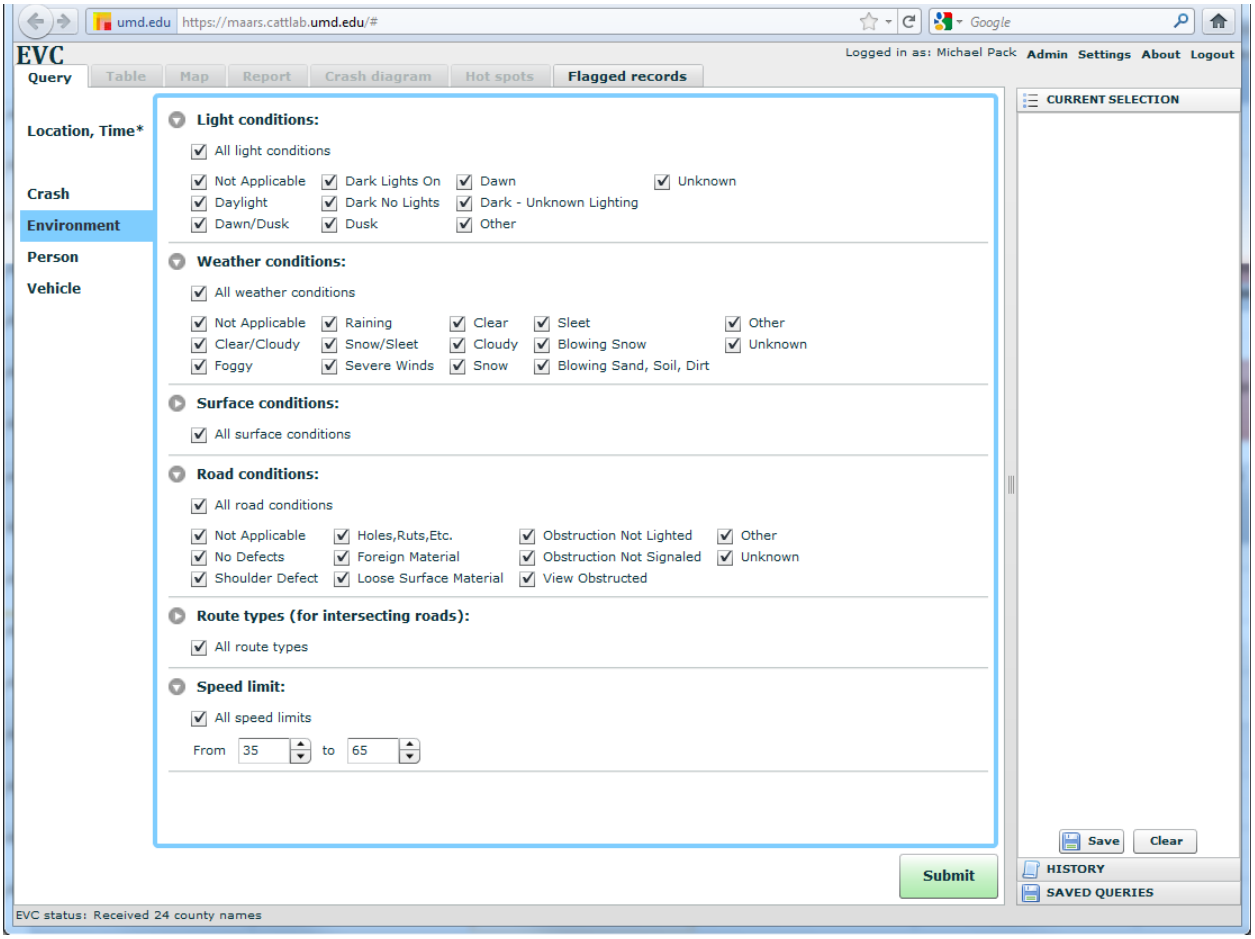


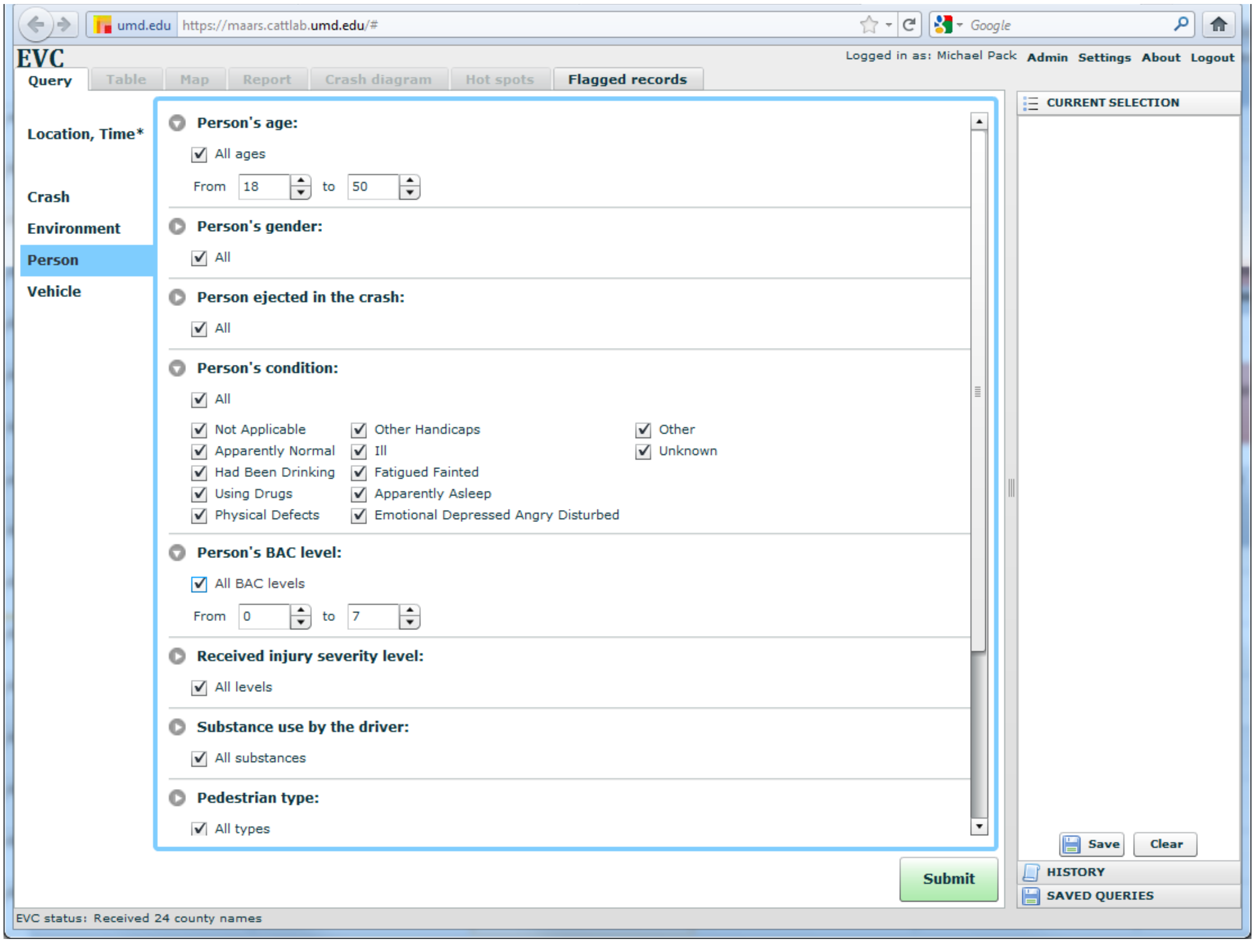
Access to the features of the application can be sectioned off in a variety of ways. Users can be restricted to only certain geographic areas, certain time limits, or other subsets of the data. Users with administrative rights can even have access to edit portions of the data, though this feature is limited only a very small number of agency data administrators.

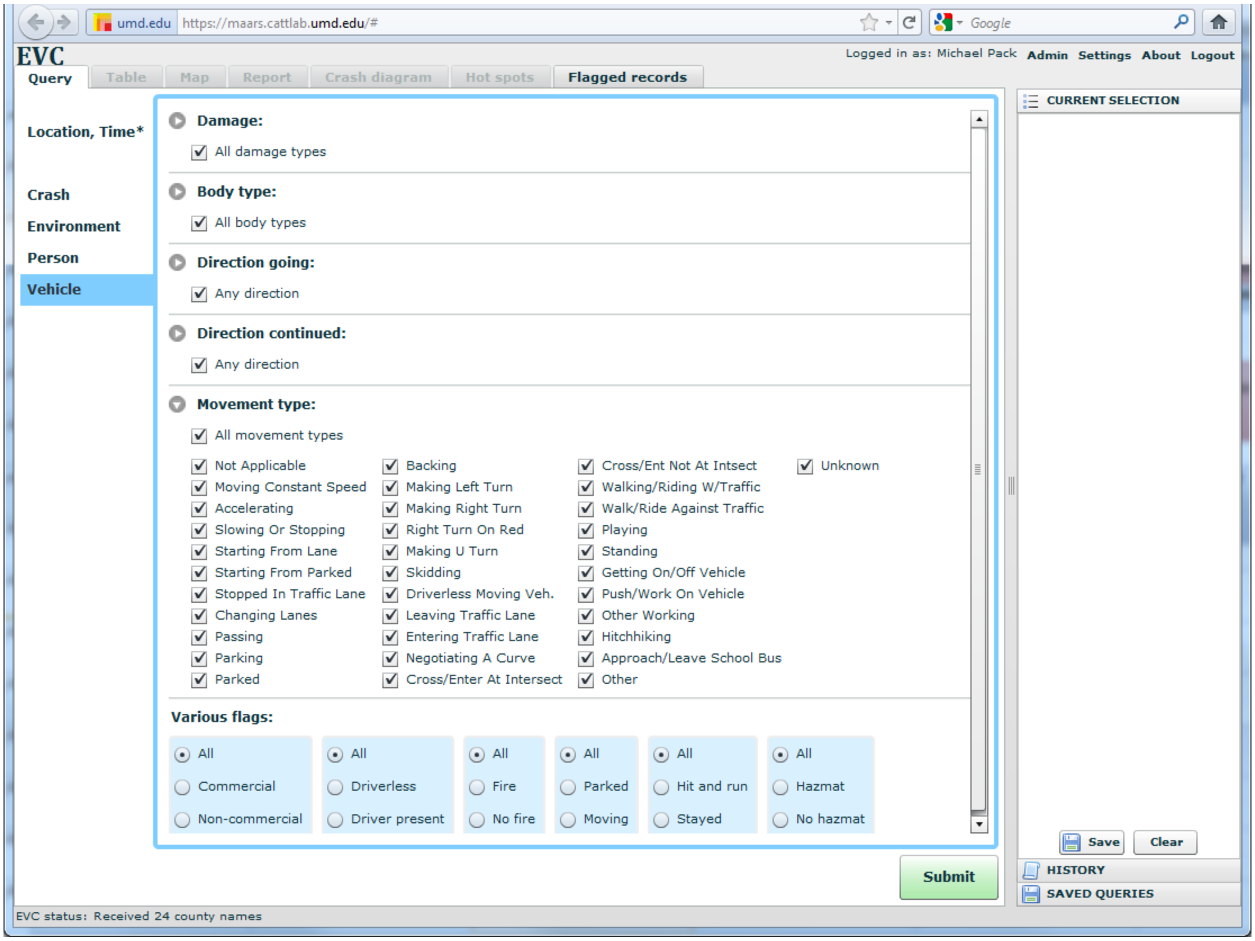
Once logged in, a user can choose to perform an “Intersection” , “Corridor, or “Region” analysis. The user is presented with a series of controls that allow [him, her, they, or the user]to select date ranges, road segments, crash characteristics, environmental variables, person information, vehicle information, etc. All of this is done with commonly used and familiar user interfaces controls. Samples of these controls can be seen in the following five graphics.


umd.edu https://maars.cattlab.umd.edu/#
Google
EVC
Logged in as: Michael Pack Admin Settings About Logout
Query Table Map Report Crash diagram Hot spots Flagged records
Location, Time*
Crash
Environment
Person
Vehicle
Select a county, then select a main route and an intersecting route. Increase the radius to include crashes farther away from the intersection.
Type of analysis: INTERSECTION CORRIDOR REGION
Choose county: * Select or type a county name
Main road: *
Intersecting road: *
Include crashes within: 0.05mi (264ft) radius.
Select a date range: All (data available: 1993 - 2010)
From: 01/01/1993 to: 12/31/2010
Select days of week: All days
Sun Mon Tue Wed Thu Fri Sat
Select the time of day: All times
From 00 : 00 to 23 : 59 add another interval
Submit
CURRENT SELECTION
Save Clear
HISTORY
SAVED QUERIES
EVC status: Received 24 county names


umd.edu https://maars.cattlab.umd.edu/#
Google
EVC
Logged in as: Michael Pack Admin Settings About Logout
Query Table Map Report Crash diagram Hot spots Flagged records
Location, Time*
Crash
Environment
Person
Vehicle
Collision type:
All collision types
Not Applicable
Head On
Head On Left Turn
Same Dir Rear End
Same Dir Rend Right Turn
Same Dir Rend Left Turn
Opposite Dir Sideswipe
Same Direction Sideswipe
Same Direction Right Turn
Same Direction Left Turn
Same Dir Both Left Turn
Straight Movement Angle
Angle Meets Right Turn
Angle Meets Left Turn
Angle Meets Left Head On
Opposite Dir Both Left Turn
Single Vehicle
Other
Unknown
Fixed objects involved:
All objects
Not Applicable
Bridge Or Overpass
Building
Culvert Or Ditch
Curb
Guardrail Or Barrier
Embankment
Fence
Light Support Pole
Sign Support Pole
Other Pole
Tree Shrubbery
Construction Barrier
Crash Attenuator
Guardrail End
Concrete Traffic Barrier
Other Traffic Barrier
Traffic Signal Support
Mailbox
Bridge Overhead Structure
Bridge Pier Support
Bridge Rail
Culvert
Ditch
Other
Unknown
Harmful event code:
All event codes
Not Applicable
Other Vehicle
Parked Vehicle
Pedestrian
Bicycle
Other Pedalcycle
Other Conveyance
Railway Train
Animal
Fixed Object
Other Object
Overturn
Spilled Cargo
Jackknife
Units Separated
Other Non Collision
Off Road
Downhill Runaway
Explosion Or Fire
Backing
U-Turn
Immersion
Fell Jumped From Motor Vehicle
Thrown Or Falling Object
Other
Unknown
Submit
CURRENT SELECTION
Save Clear
HISTORY
SAVED QUERIES
EVC status: Received 24 county names


umd.edu https://maars.cattlab.umd.edu/#
Google
EVC
Logged in as: Michael Pack Admin Settings About Logout
Query Table Map Report Crash diagram Hot spots Flagged records
Location, Time*
Crash
Environment
Person
Vehicle
Light conditions:
All light conditions
Not Applicable
Daylight
Dawn/Dusk
Dark Lights On
Dark No Lights
Dusk
Dawn
Dark - Unknown Lighting
Other
Unknown
Weather conditions:
All weather conditions
Not Applicable
Clear/Cloudy
Foggy
Raining
Snow/Sleet
Severe Winds
Clear
Cloudy
Snow
Sleet
Blowing Snow
Blowing Sand, Soil, Dirt
Other
Unknown
Surface conditions:
All surface conditions
Road conditions:
All road conditions
Not Applicable
No Defects
Shoulder Defect
Holes,Ruts,Etc.
Foreign Material
Loose Surface Material
Obstruction Not Lighted
Obstruction Not Signaled
View Obstructed
Other
Unknown
Route types (for intersecting roads):
All route types
Speed limit:
All speed limits
From 35 to 65
Submit
CURRENT SELECTION
Save Clear
HISTORY
SAVED QUERIES
EVC status: Received 24 county names


umd.edu https://maars.cattlab.umd.edu/#
Google
EVC
Logged in as: Michael Pack Admin Settings About Logout
Query
Table
Map
Report
Crash diagram
Hot spots
Flagged records
Location, Time*
Crash
Environment
Person
Vehicle
Person's age:
All ages
From 18 to 50
Person's gender:
All
Person ejected in the crash:
All
Person's condition:
All
Not Applicable
Apparently Normal
Had Been Drinking
Using Drugs
Physical Defects
Other Handicaps
Ill
Fatigued Fainted
Apparently Asleep
Emotional Depressed Angry Disturbed
Other
Unknown
Person's BAC level:
All BAC levels
From 0 to 7
Received injury severity level:
All levels
Substance use by the driver:
All substances
Pedestrian type:
All types
CURRENT SELECTION
Save
Clear
HISTORY
SAVED QUERIES
Submit
EVC status: Received 24 county names



After selecting the appropriate analysis type (Intersection, Corridor, or Region), and then selecting the other filter criteria, the user hits the "Submit" button and waits for the results to be returned. Depending on the type of search and filters applied the search could finish in a matter of seconds or in several minutes. Results can come back in the form of:

1) Filterable, Sortable, and Exportable Tables
2) Maps
   i. Heat maps
   ii. Cluster maps
   iii. County or State Summary maps
   iv. Other map outputs
3) Summary Reports
4) Crash Diagrams
   i. Intersection Diagrams
   ii. Corridor Diagram
5) Hot Spot Diagrams

All of these visualizations are interactive, allowing the user to explore the results of the data further. Examples of some of visualizations are shown in the images below.

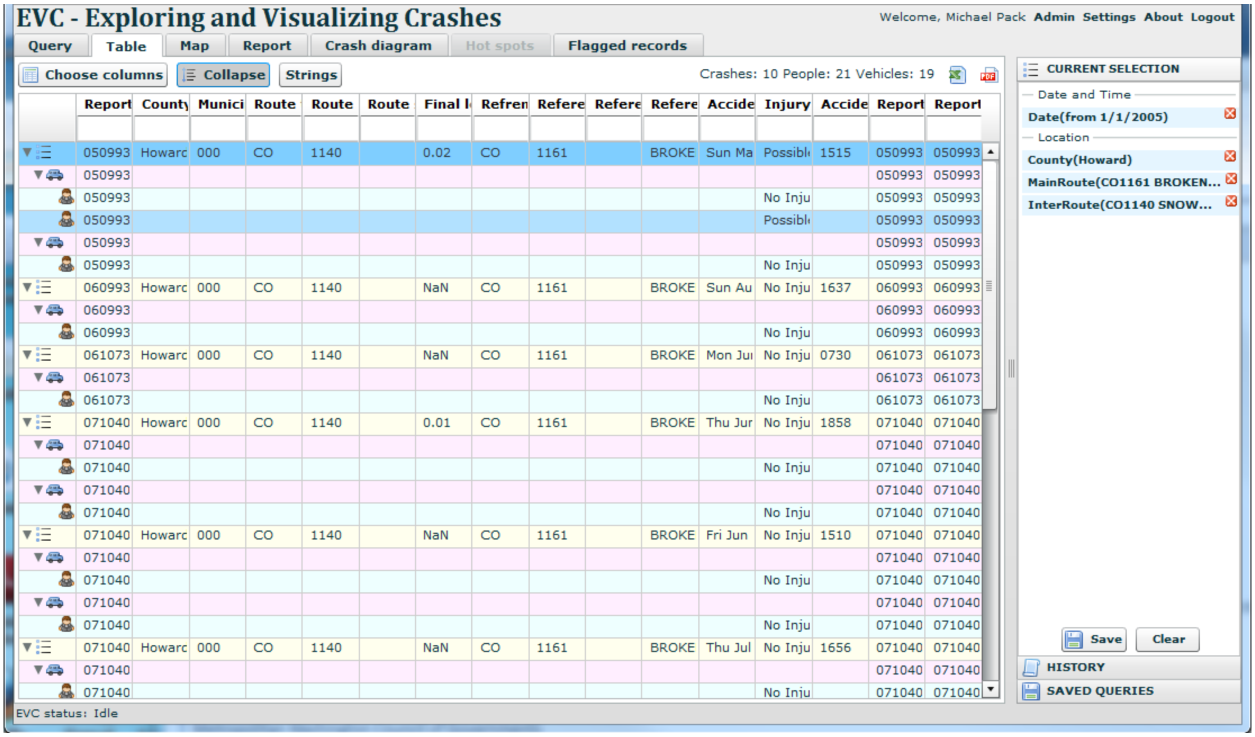


Figure 1: Table output.

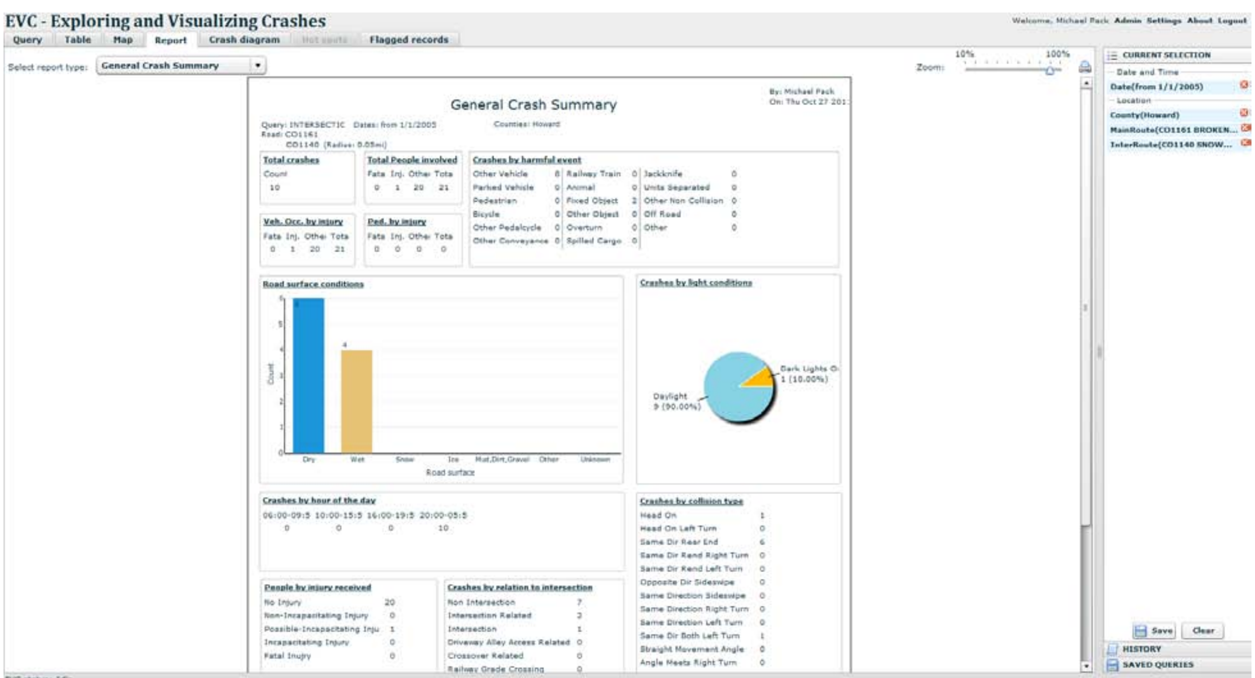


Figure 2: Crash Summary report for an intersection in Maryland.

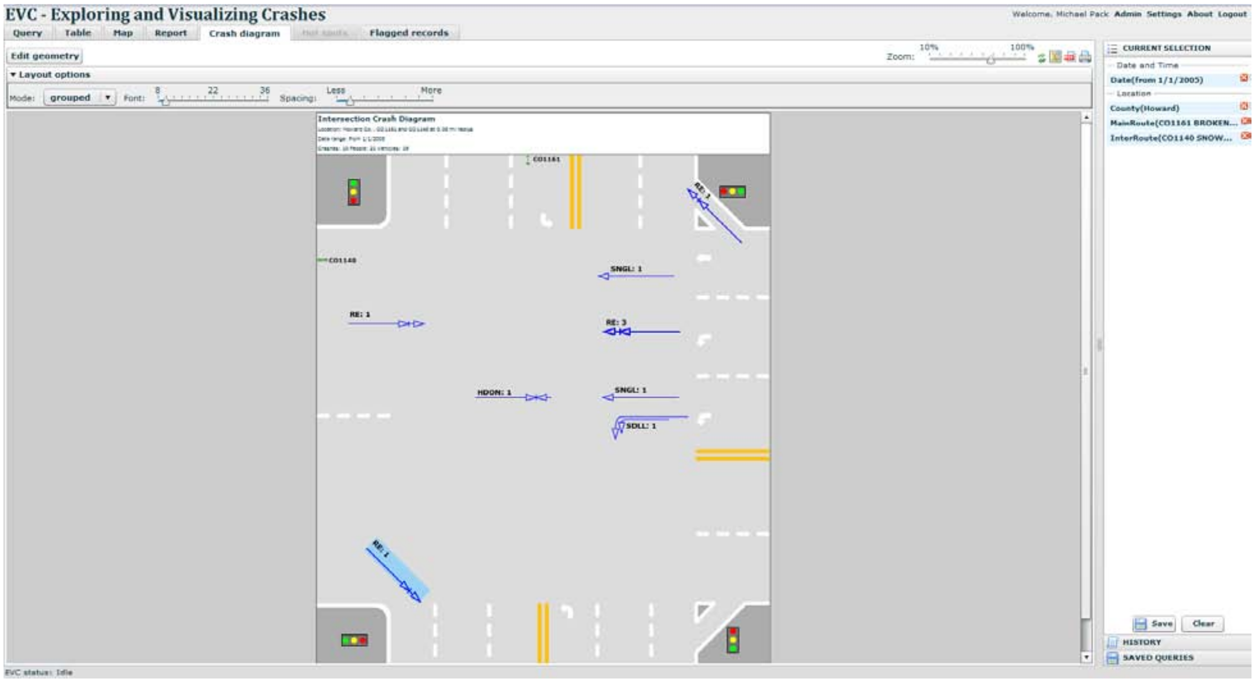


Figure 3: Intersection diagram.

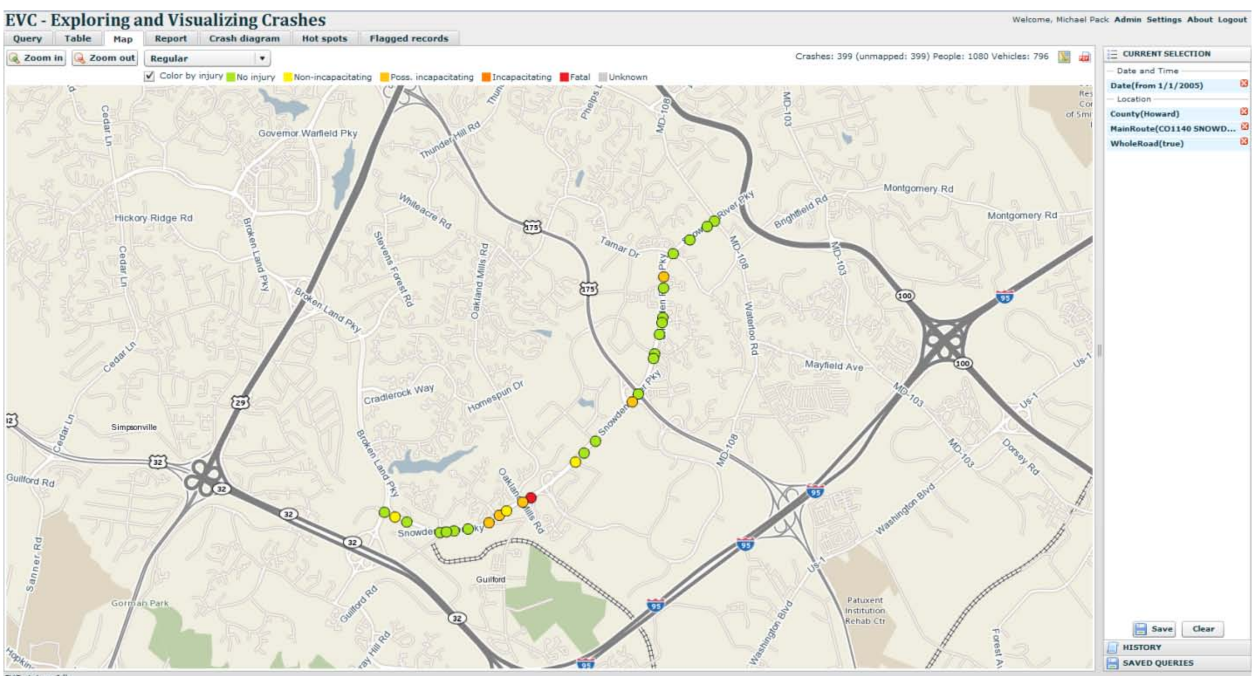


Figure 4: Corridor map showing individual accident locations.

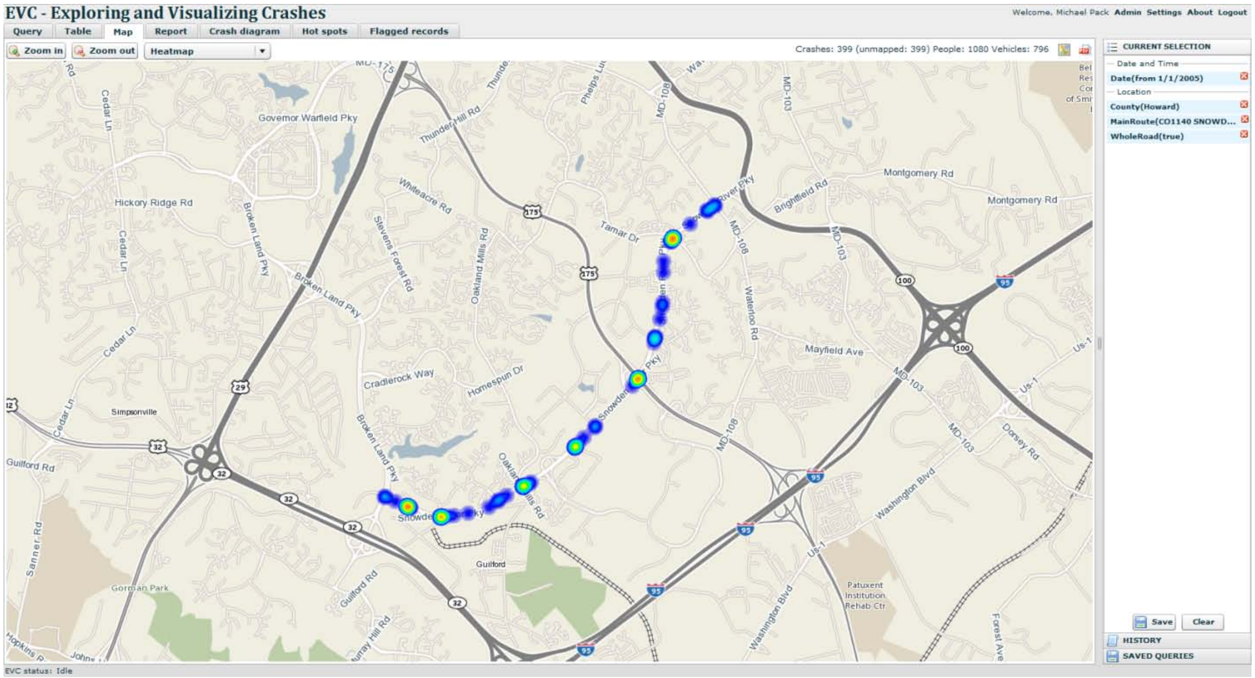


Figure 5: Incident heatmap for a corridor query.

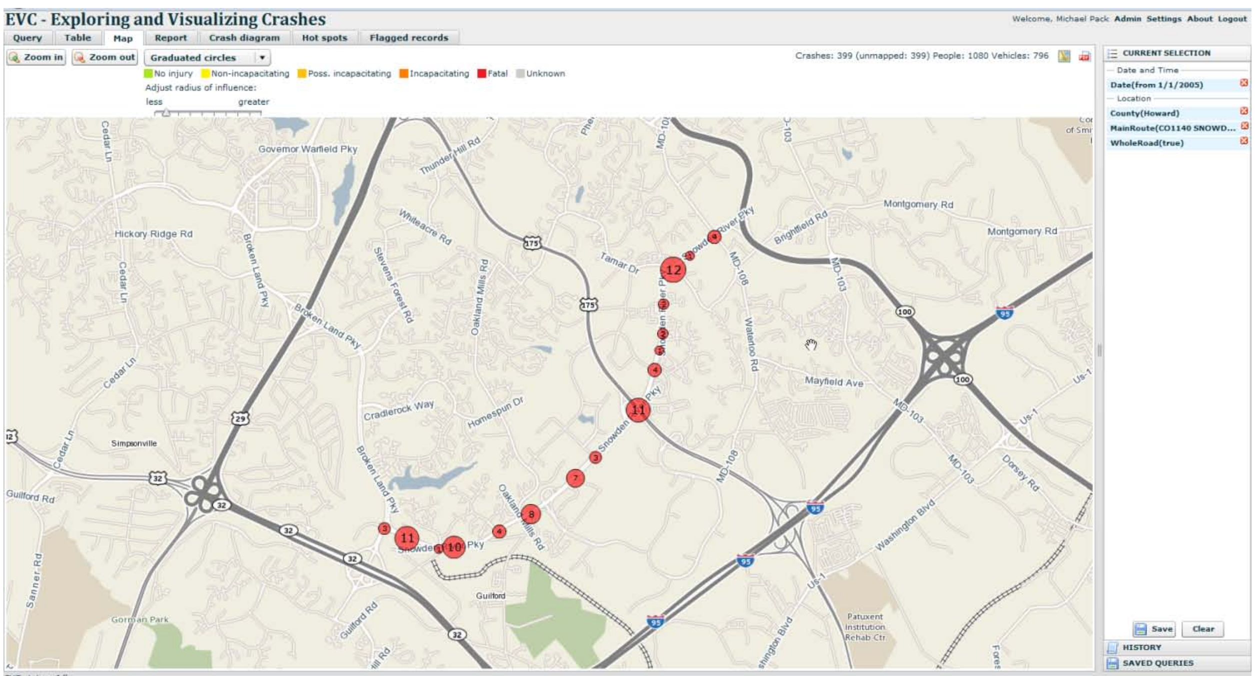


Figure 6: Incident cluster map for the same corridor as above.

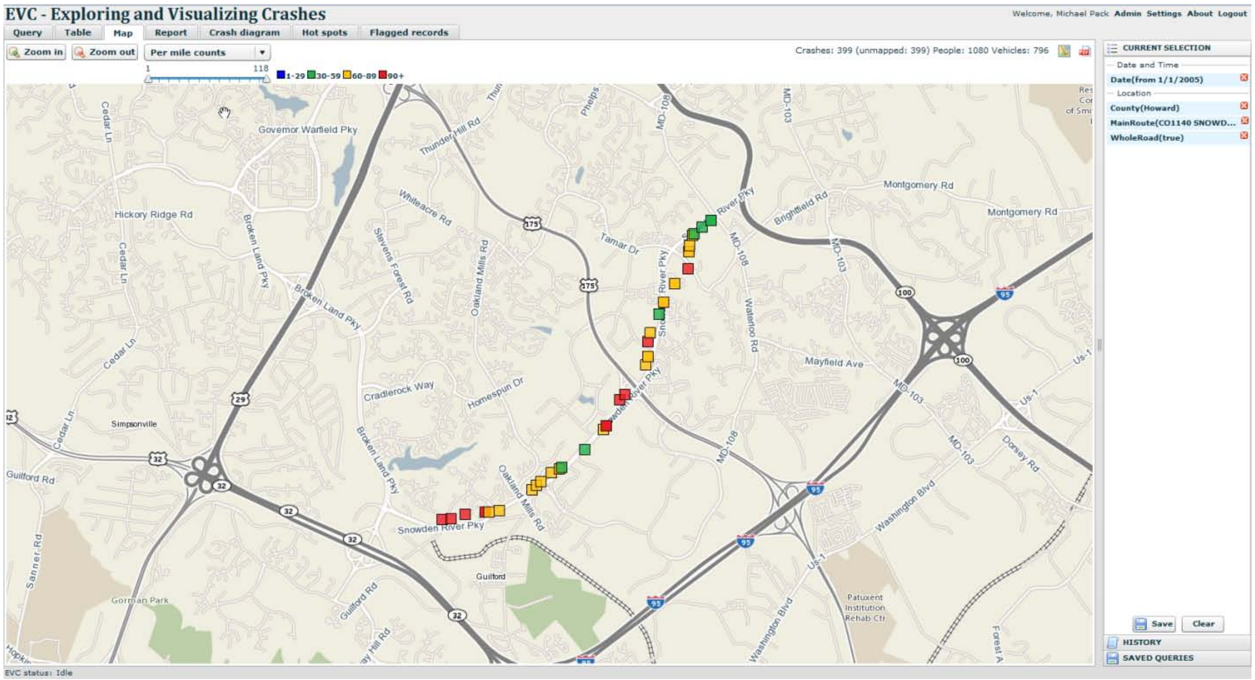


Figure 7: Per mile counts of incidents along a corridor.

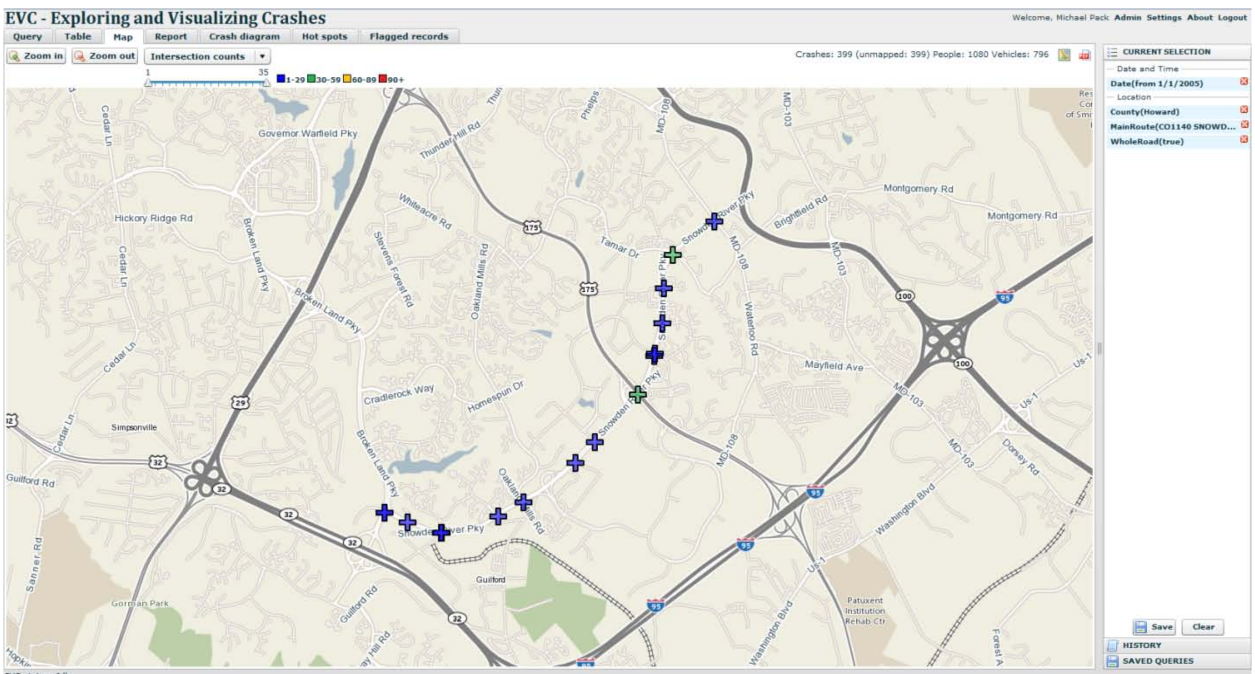


Figure 8: per intersection counts of incident along a corridor.

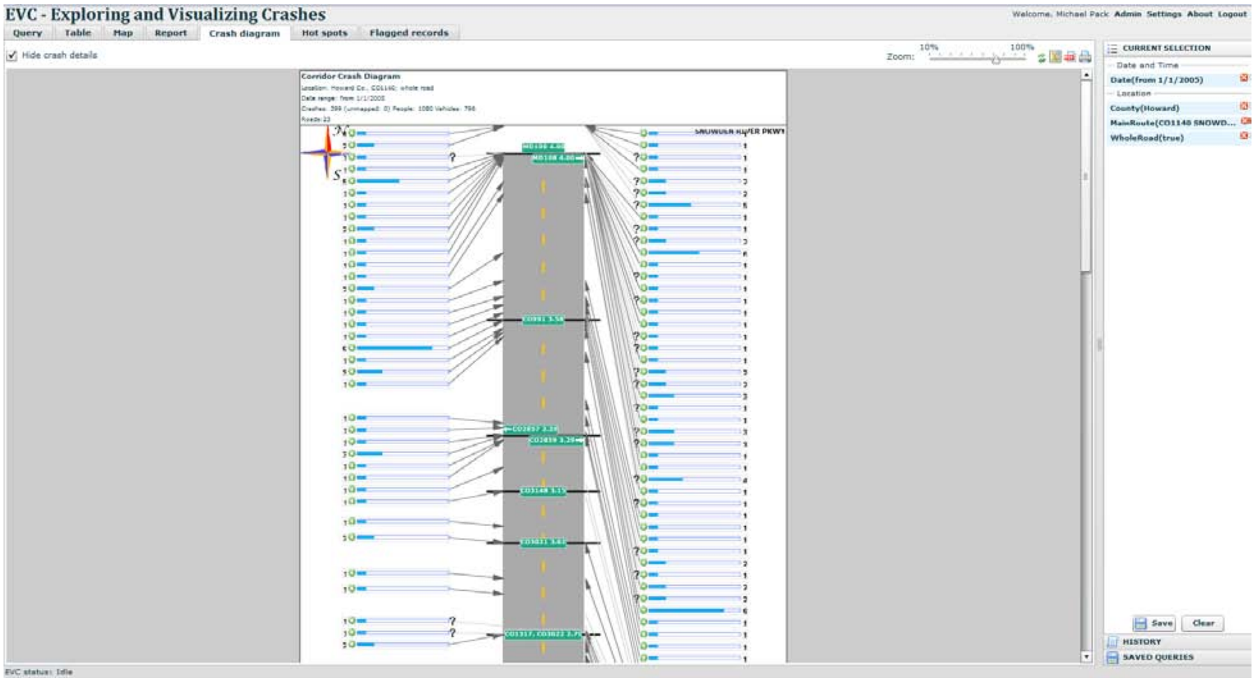


Figure 9: Corridor diagram.

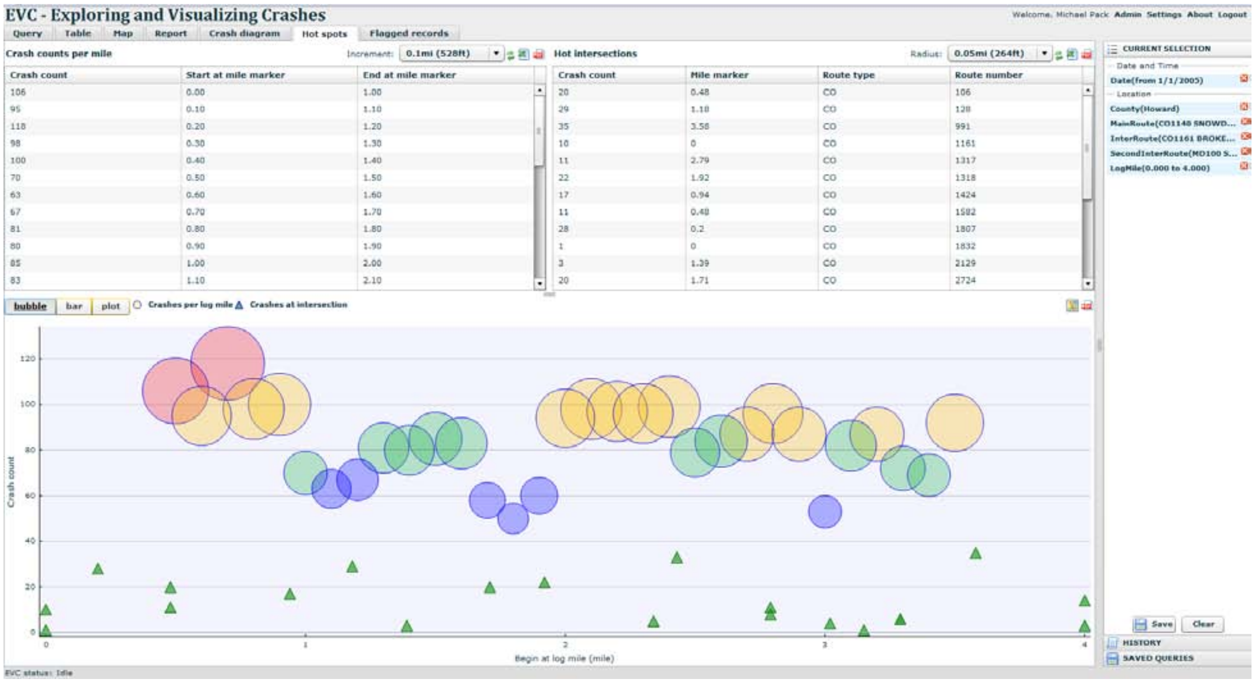


Figure 10: hotspots diagram showing counts per mile and per intersection.

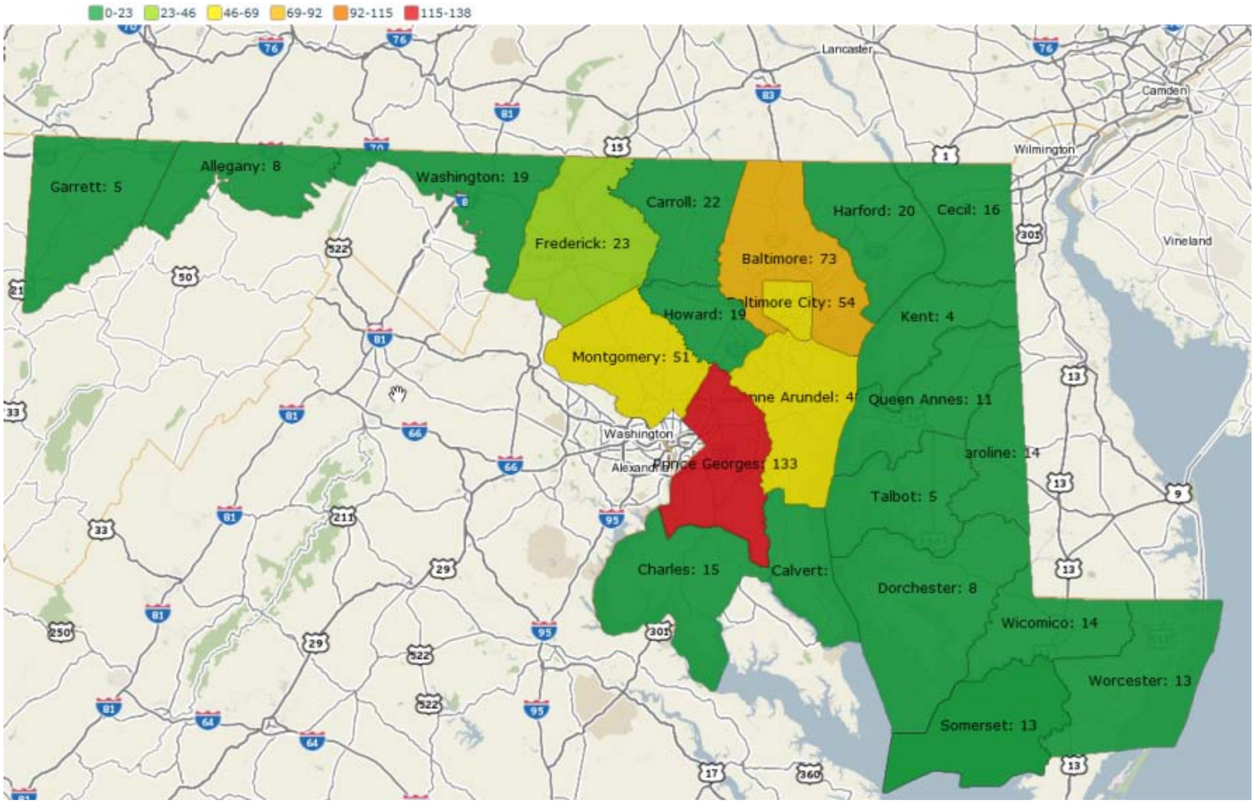


Figure 11: Statewide fatality query.

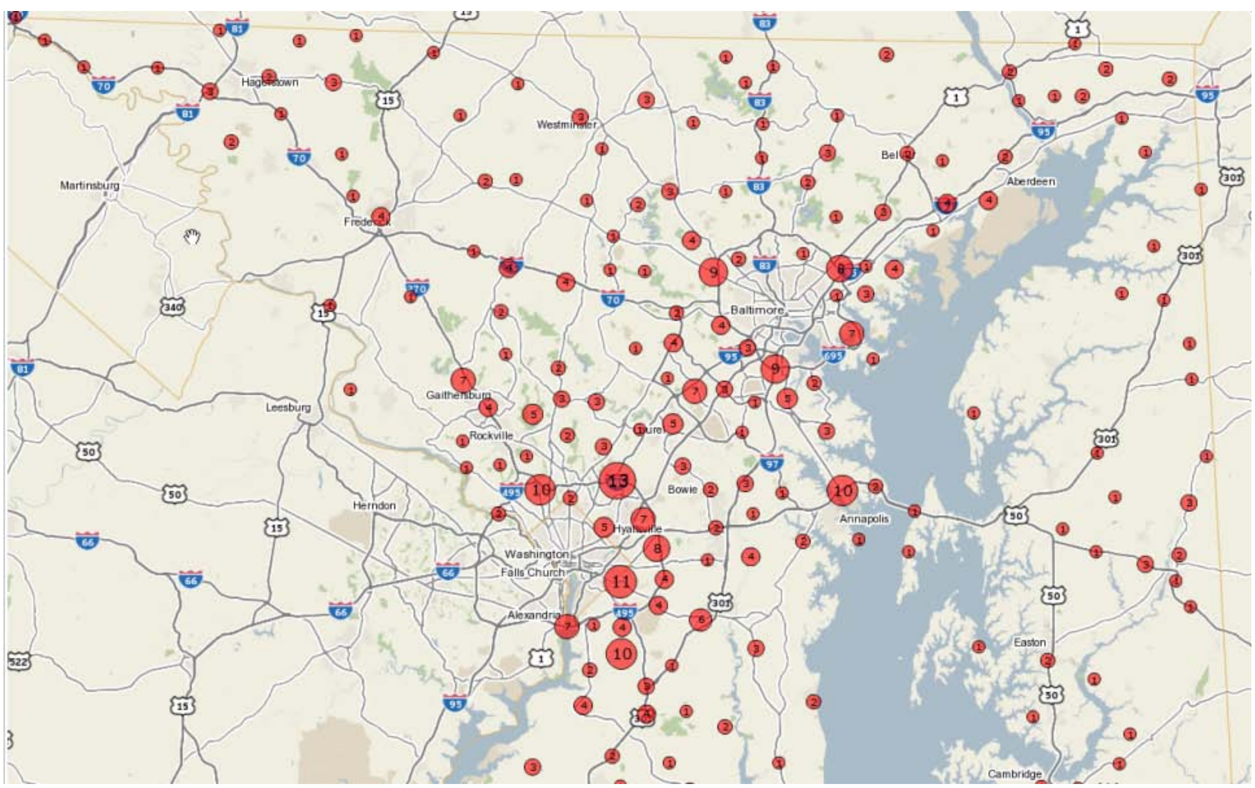


Figure 12: Statewide fatality clustering map.

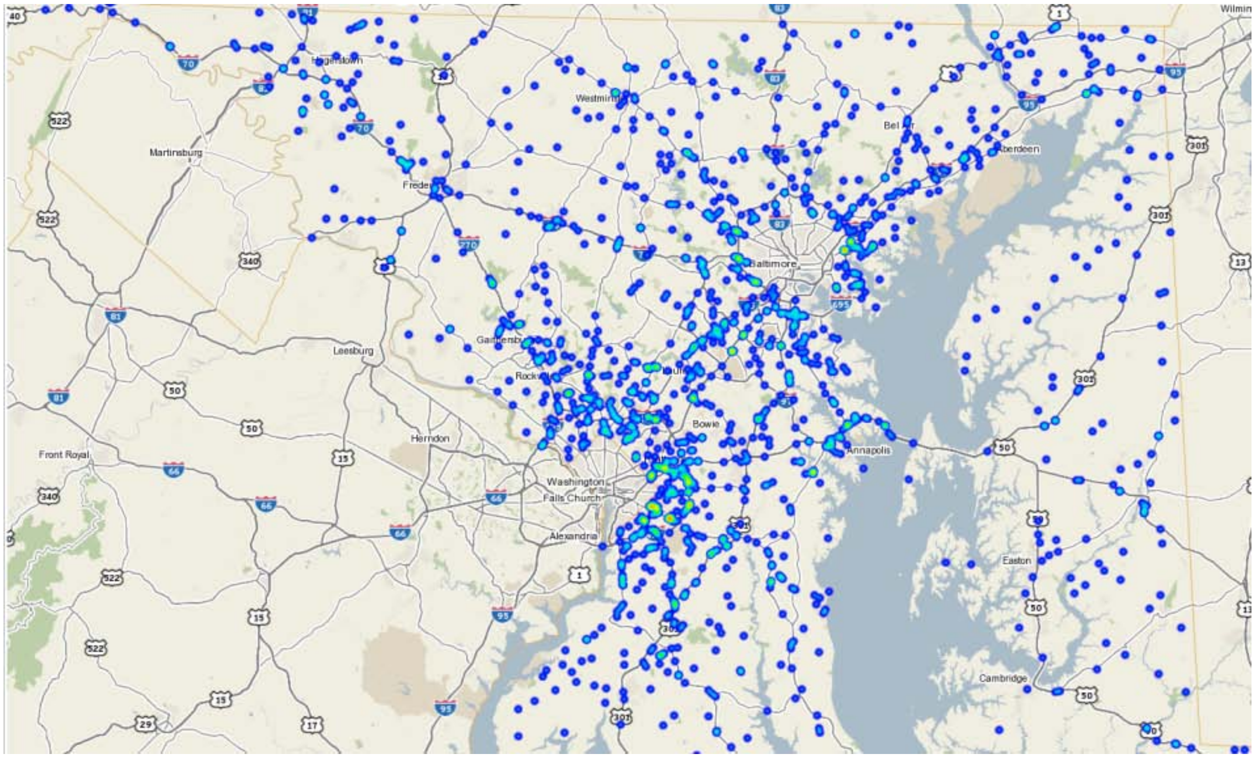


Figure 13: Statewide fatality heatmap.

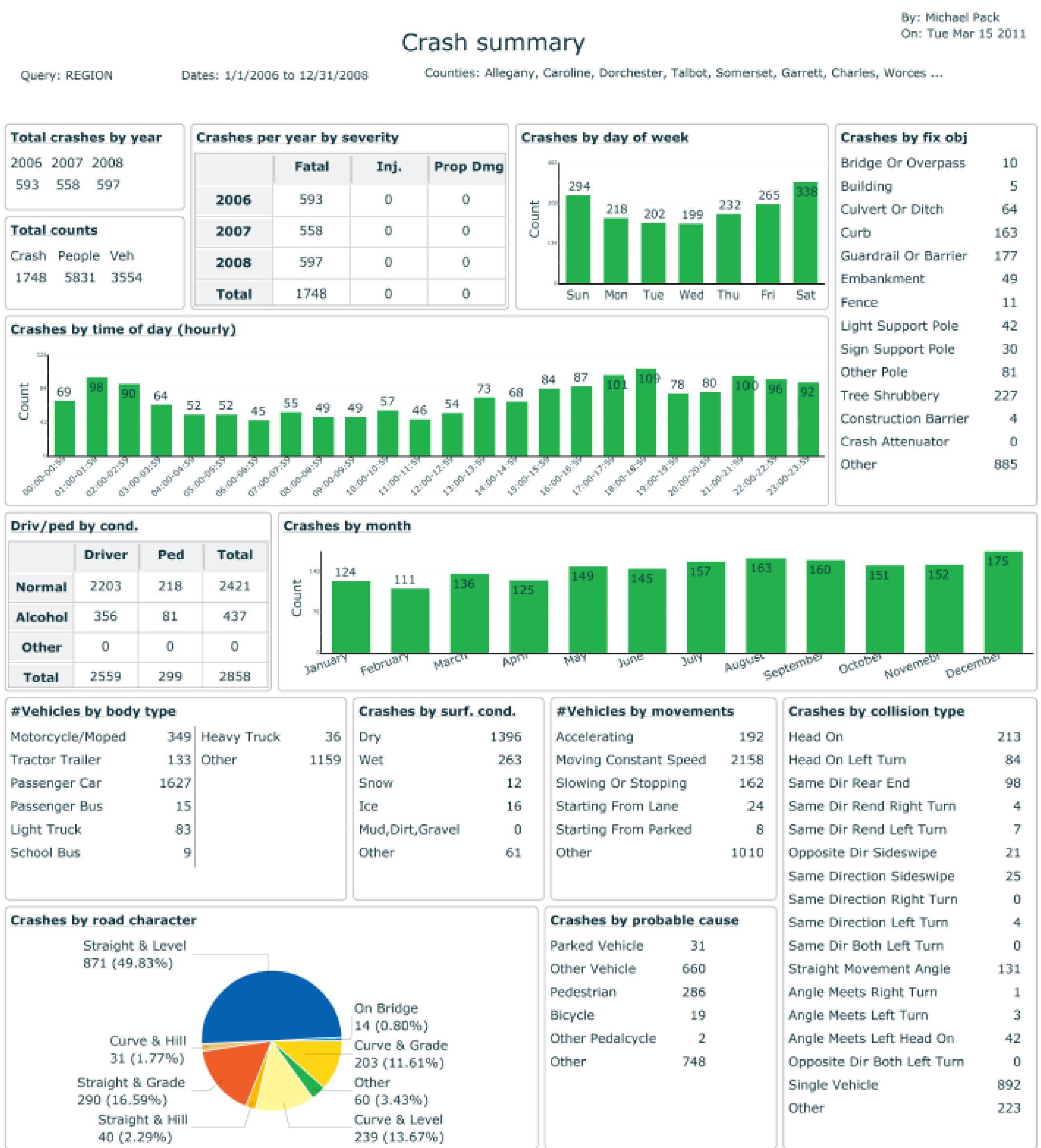


Figure 14: Example of a summary report.

Limitations of the Tool

While the current EVC tool works quite well for localized roadway evaluation, corridors, and specific regional condition reporting (like how many fatal accidents occurred in MD for a given year or two), there are severe limitations of the system when it comes to running larger regional queries that look at many years of data at a single time. This is because EVC returns all query result data directly to the client application, and available memory on the client machine imposes a limit on how many records it can feasibly work with. Furthermore, the limitations of EVC are also determined by the available hardware for the server hosting the database and webserver.

For example, in the current version of EVC, a search for 3 years' worth of accident records for the more populous counties (Baltimore, Montgomery, etc.) requires more than 2 GB of RAM be available to the server. The memory needs of the client machine are related to the number of records being analyzed by the client at any given time. Before running a query, EVC uses ~140 MB of RAM. When working with 10000 records, EVC requires ~270 MB. This number would grow linearly with the number of crashes. Though these figures are modest by modern hardware standards, these numbers are worth bearing in mind as many DOT users have older machines, slower internet connections, etc.

To make EVC more useful for larger queries and regional analysis, the method of querying and returning results to the user may need to be reconsidered. More server side processing and queuing may need to occur. Not all data may need to be returned to the user. In some instances, this could speed up the user experience, but in other cases, this could also slow down the user experience.

Screen resolution is also an issue. While scroll bars make it possible to use EVC at lower resolutions, it is recommended that users have a minimum screen resolution of 1280x720.

Description of the Data

The EVC application is very much tied to the current Maryland Automated Accident Reporting System (MAARS) database. This means that all of the query able fields are part of this database, work off of the Maryland geographic road network, and use specific Maryland log-mile linear referencing schemas. Below are screenshots of Entity-Relationship (ER) diagrams from each agency database schema. Lookup tables have been removed from these screenshots because of space considerations. Note that each schema is quite different. These differences are addressed in the next section. It should also be noted that the D.C. ER diagram is based off of older D.C. data that was provided to the UMD CATT Laboratory nearly two years ago. It is not known if this is still a valid schema that is used by D.C. officials. Unfortunately, the D.C. agency responsible for this data was not able to provide data to the CATT Lab or MWCOG officials in time for this study.

Maryland Data Overview

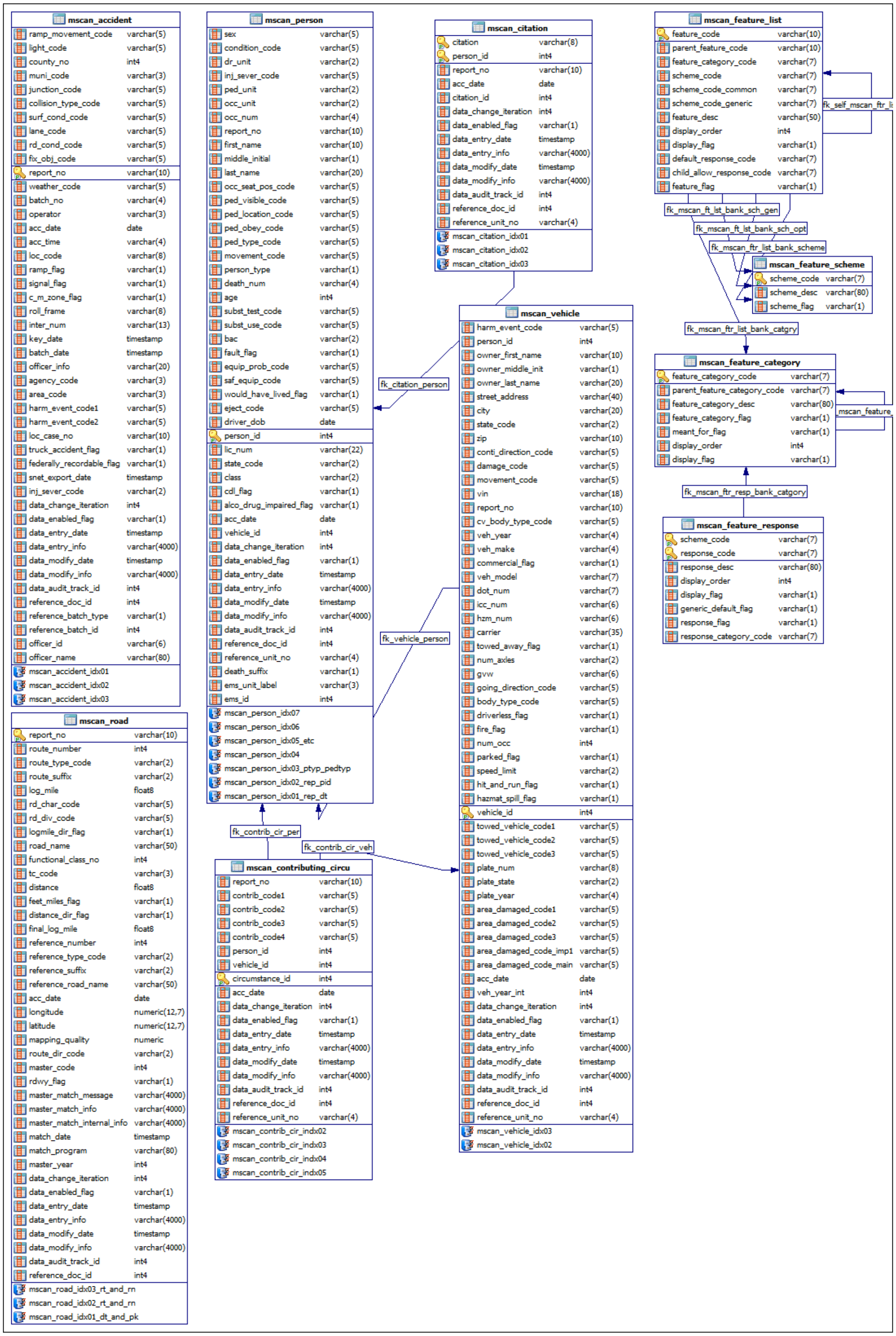
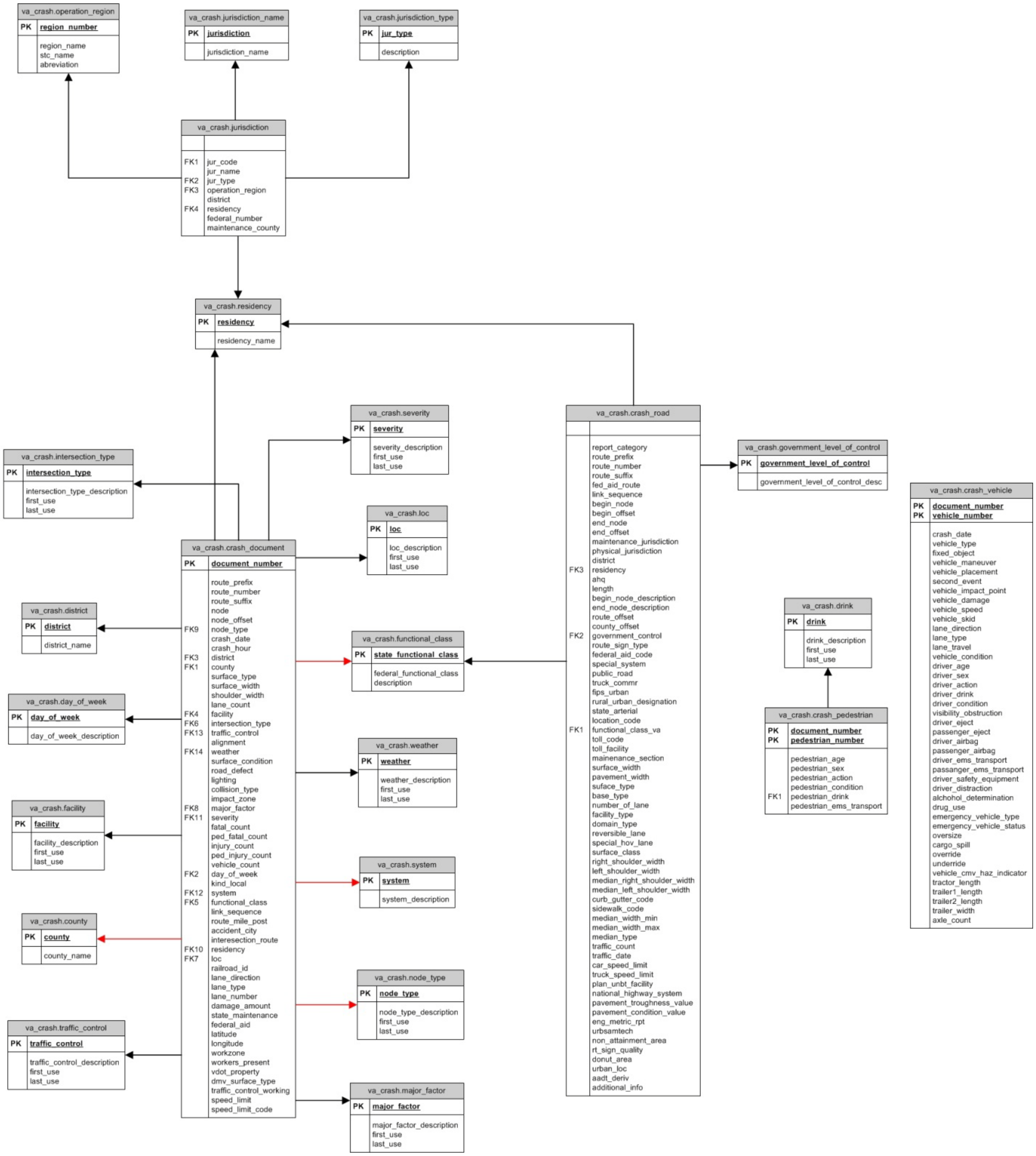
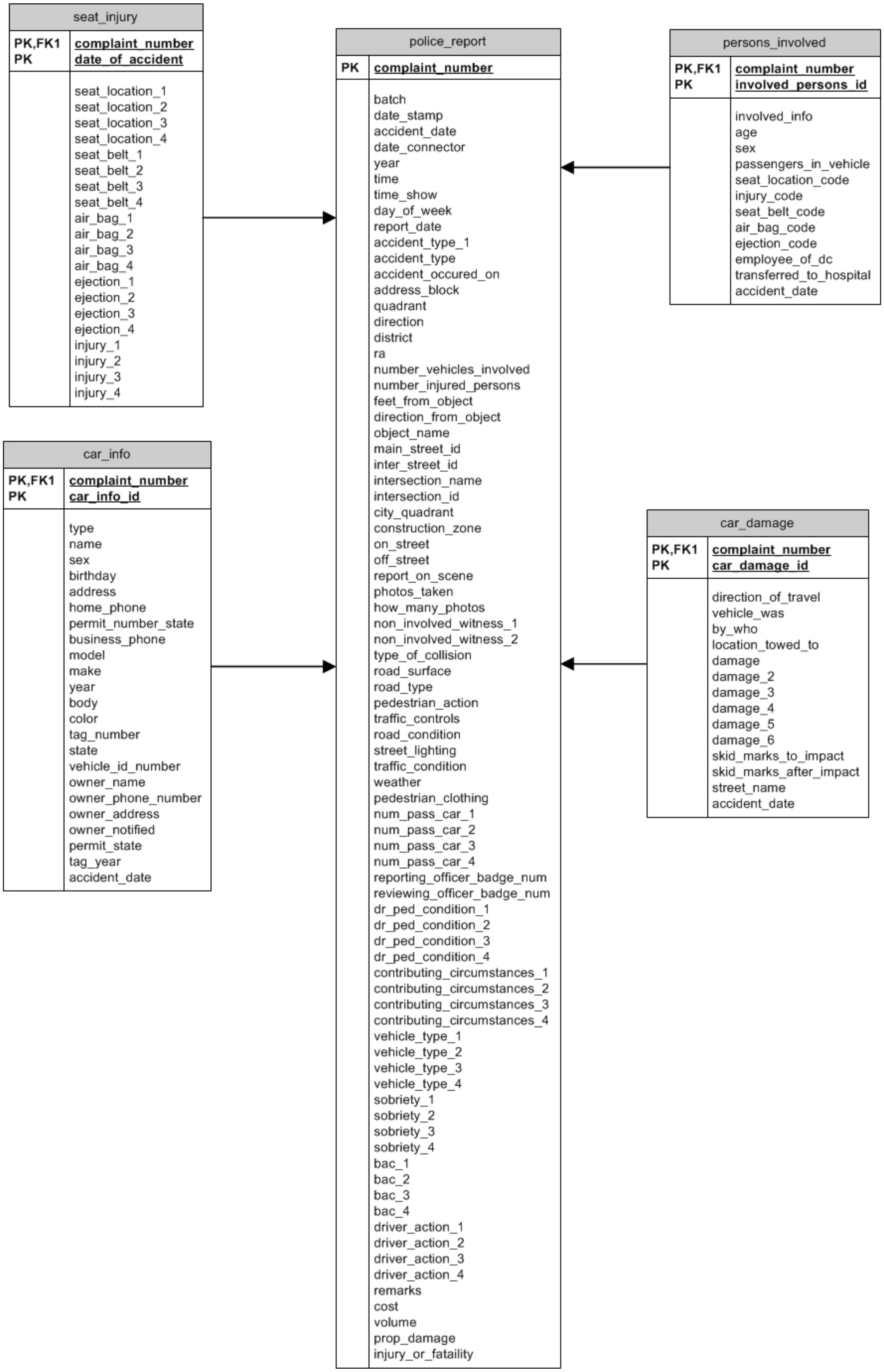


Figure 15: Maryland MAARS Database Entity-Relationship Diagram with lookup tables removed.

Virginia Data Overview



District of Columbia Data Overview



Similarities and Differences in Data

Each agency provides essentially the same types of data; however, the level of detail, terminology, and general structuring of the different components of the overall data set often differs from agency to agency. Below are some examples of the similarities and differences between the data sets for specific data components:

Location Data: All three agencies provide primary road and intersection location information, which serves as the primary means of geolocating the incidents. None of the agencies provide actual latitude and longitude coordinates, although the Virginia dataset contains a space for it in the schema. In order to accurately geolocate each agency's data, we will need to be provided with either explicit latitude/longitude coordinates or a linear referencing system that we can use to geolocate the incidents ourselves. Receiving the coordinates directly from the source agency would be the preferred method, as this would increase the integrity and accuracy of the official dataset.

Date and Time Data: Dates and times are standard across agencies, with the only differences being that some agencies choose to separate the date and time components into their own fields, while others are stored in a single date/time field.

Type of Incident: DC seems to be the only agency that provides direct "accident types" such as hit and run, property damage, etc.; however, all three agencies provide information to derive collision types. Below is a table comparing the various collision types provided by the three agencies. (NOTE: there may be additional collision types for DC not represented due to those types not being present in our data samples)

MD	**DC**	**VA**
	Backing hit parked car	Backed into
	Fixed object	Fixed object in road
		Fixed object off road
Head on	Head on	Head on
Head on left turn		
Same direction left turn	Left turn hit vehicle	
Same direction both left turn		
Opposite direction both left turn		
Other	Other	Miscellaneous or other
	Parked vehicle	
	Ran off roadway	Non-collision, overturned, jackknifed or ran off road
Same direction rear end	Rear end	Rear end
Same direction rear end right turn		
Same direction rear end left turn		

Same direction right turn	Right turn hit vehicle	
Opposite direction sideswipe	Side swiped	Sideswipe - Same direction of travel
Same direction sideswipe		Sideswipe - Opposite direction of travel
	Straight hit pedestrian	Pedestrian
	Right turn hit pedestrian	
Angle meets right turn	Right angle	Angle
Angle meets left turn		
Angle meets left head on		
Straight movement angle		
		Train
		Deer
		Other animal
		Bicyclist
		Motorcyclist
Unknown	<blank>	Not stated
Not applicable		
Single vehicle		

Causality: Causality information differs greatly between the three agencies.

- VA has a single causality field for each collision record, with 9 options to choose from.
- DC allows for 4 contributing circumstances per collision, with approximately 14 options for each (additional options may exist that are not represented in the sample dataset). Although 4 contributing circumstances are allowed per collision, it appears that that these circumstances may be limited to one per vehicle involved, rather than allowing multiple circumstances for a single vehicle.
- The MD dataset stores contributing circumstances in a separate table with links to a specific person or vehicle, allowing for any number of circumstances to be applied to each person and vehicle involved in the collision. There are a total of 83 unique options grouped into 7 categories. This setup allows for very specific causality information to be recorded for a given collision.

Additional Fields: There are significantly more data fields available in each of the three datasets, including vehicle type, color, race, etc. Like the collision types and causality fields discussed above, the level of detail and coverage of these additional data fields varies from agency to agency. It is thought that these fields are less important than the four primary categories of information already discussed, so they will not be covered in detail for this report.

Proposed Path Forward

EVC is very much tied to the MAARS database schema structure. All the filtering options and columns displayed in the UI come directly from the MAARS database. Trying to fold the VA and DC data into the current EVC application would require significant UI reworking in addition to back-end changes, and navigating through all of these changes are estimated to be more work than is worth. Similarly, the EVC tool as it exists is not capable of handling the large amount of data that would result from combining the three agency data sources and doing region-wide queries for the date ranges that we've typically seen Maryland users interested in. Instead, we are recommending creating a new, similarly functioning tool that builds off of the ideas and UI of the EVC application, but uses a more solid foundation, building off of lessons learned from design mistakes within EVC.

Because regional analysis will primarily be looking at combined cities, counties, etc. we could probably leave out the single road/intersection crash and corridor diagrams for the combined tool. Instead, we would invest our time and energy into enhancing map functions, table functions, hot-spot diagrams, and other visualizations more suited for regional analysis. Several CATT Lab developers working exclusively on this task could likely produce a working prototype in less than 6-months.

Assumptions:

- The app would be best built around a standardized data format and would include data from all 3 agencies. This means the CATT Lab will need to create a "common" database schema that includes all three agency data sources, fused together as best as can be accommodated. This also means that some fields that can't possibly be standardized between all three agencies might need to be dropped.
- The application would allow the user to compose a query using all fields and lookup table values available in our standard format, similar to what EVC does now.
- We are also assuming that all agency accident reporting data will be provided to us in a timely manner, and that no major schema changes will have occurred between now and receipt of the data.
- At least one new database server and a web application server will need to be purchased.

Risks

There are several risks to this project that should be addressed early in the development of this application. They include:

- **Data Accessibility**: For the project to be a success, all three agencies will need to be committed to providing their data to MWCOG and the UMD team. Furthermore, the agencies need to agree on when and how to provide updates to the data as new records are reported. The tool will be of little value until all three agencies have updated their data.
- **Application Accessibility**: MWCOG will need to establish an "administrator" for the tool. This person will be responsible for creating accounts for individuals who need access to the tool. The three agencies will need to agree on who should have access to the tool, for what period of time, and for what geographic region. While all three states may be providing data to the tool,

it may be desirable to only give complete regional access to certain individuals. This will be a MWCOG and agency decision.

- Data Retention: The agencies will need to decide how far back in time the data should be available? 3 years, 10 years, or 20 years? This is important in deciding how much storage space is needed. This budget assumes a 15-year period of data will be available.

Estimated Timeline

It is estimated that political challenges, coordination with agency users, and other meetings will likely cause delays within the project. Because of these challenges, the development team envisions a solid year would be needed to deliver a working product to the users.

Estimated Budget

The anticipated funds needed to fully integrate three state-wide data sources and develop regional querying, reporting, and visualization capabilities is $195,000. This includes approximately $25,000 in hardware costs for database servers and web application servers. A detailed budget can be provided upon request. The budget assumes the University of Maryland standard 52.5% overhead rate shall be applied. Special accommodations are sometimes made for agencies that have policies on overhead rates.