2014.12

Mobility Data Analytics Center: the Engine for Building a Regional Traveler Information System for the Power of 32 Region

Traffic21 Institute
Technologies for Safe and Efficient Transportation US DOT University Transportation Center
H. John Heinz III College
Institute for Complex Engineering System

Carnegie Mellon University | 5000 Forbes Avenue, Pittsburgh, PA 15213

Table of Content

Executive summary

The long range vision, consistent with the Power of 32's 2012 Regional Traveler Information System (RTIS) charrette conducted by Carnegie Mellon University's Traffic21Initiative and The Intelligent Transportation Society of America, is to establish a state-of-the-art RTIS system to provide intelligence to multi-modal travelers, freight companies and transportation agencies. The first step to achieve this vision was to develop a Mobility Data Analytics Center at Carnegie Mellon University. This center has begun to collect, organize, and integrate multi-jurisdictional multi-modal transportation data from throughout the Power of 32 region. The data are analyzed to forecast multi-modal transportation system performance, which are provided to individual travelers. Furthermore, the center will also use this data to conduct research and development of data analytical tools for decision making of public agencies, such as transportation models and simulation, assessment of regional planning and design of freight system assets.

The ultimate objectives of the **Mobility Data Analytics Center** are to:

- Provide archived and real-time traffic data of every element of regional multi-modal transportation systems
- Reveal the behavioral information for both passenger transportation and freight movement
- Serve as a key policy instrument for policy makers, regional transportation planners, national researchers and engineers
- Serve as a key information platform for regional travelers and regional freight movement.

We have developed a prototype Power of 32 Regional Traveler Information System that provides travel information most promising and readily available, such as roadway traffic volume and travel time, traffic video, public transit, parking, incidents, weather, and electrical vehicles charging stations. The primary goal of the prototype system is to collect various types of traffic data in the Power of 32 Region, and to demonstrate its effectiveness in facilitating smart decision making among travelers, private firms and public agencies in Power of 32 Region.

Imagine that travelers in this region can access full information through a Regional Traveler Information System (RTIS) that delivers on-demand trip information and recommendations, while public agencies can access the RTIS as a data repository for predictive modeling, ongoing planning, and system management. A regional RTIS would efficiently improve safety, incident management, and emergency management through, for example sharing information between Traffic Management Centers and emergency management partners, or deploying a Road Weather Information System.

RTIS has the potential to build off and leverage many pieces of the existing Intelligent Transportation Systems (ITS) infrastructure and transportation management strategies in the Power of 32 Region.

Establishing a fully functional RTIS can take a few years. The prototype developed in this year one has successfully demonstrated that the Mobility Data Analytics Center, the engine for RTIS, can be a viable tool for integrating massive data and applying analytics to decision-making for travelers, private firms and public agencies. In addition, a Capstone project in Fall 2014 at Heinz College provided possible structural and financial models for the RTIS (see the Capstone project report attached separately). Our Power of 32 Region could develop a RTIS if data can be gathered and consolidated and there is political will. The Power of 32 should work with the state DOTs and local transportation providers to determine how they would like to share their data for a RTIS as a process of planning and operational management. The Power of 32 should also decide whether to move forward with the RTIS.

1. Introduction and background

Transportation has always been critical to quality of life and economic development. It is one of six issue teams formed during the 2010 Power of 32 community-based regional visioning process. This region as well as other post-industrial regions are on the verge economic and society transformation and the state of the transportation system can enhance or impede that process. The increasing complexity of transportation needs from including increasing freight, re-population of urban areas and local energy development leads to more intensified problems in congestion, roadway safety, air quality, accessibility and sustainable mobility. Transportation agencies are turning to smarter ways with a variety of new technologies to improve travel experience and to manage the transportation system more efficiently.

The Power of 32 vision for our region's Transportation and Infrastructure is: Travelers — by foot, bike, car, transit, truck, train, air, or other mode — can plan and execute trips based on real-time, region-wide (i.e. 32 counties) information. The vision is supported by a common goal to create a regional traveler information system to provide real-time information to enable modal and route choices for people and freight.

Over the last decade, new technologies and innovative transportation systems produce **massive data**, which has tremendously enriched ways to monitor and manage our transportation systems. The data, from various sources, altogether provides unprecedented opportunity for the transportation industry to understand travel behavior and to propose efficient management strategies. However, those data sources are usually established by disparate public agencies and the private sector. They rarely communicate with each other and as a result, data is only used and analyzed for a particular piece of transportation system, such as an intersection, a stretch of freeway or bus routes operated by the same agency. With disparate data sources, each part of the system is individually operated and clearly the entire transportation system is far from being socially optimal.

How to integrate data from various sensing systems and how to make the best use of integrated data remain a big challenge. **Integrating** and **learning** the massive data are the keys to success of smarter transportation systems, which consist of the following three components.

1) **Integration of various data sources**. The data sources from various sensing systems include, but are not limited to, inductive loop detectors, monitoring cameras, radars, Bluetooth, GPS tracking, mobile sensors, parking sensors, automatic vehicle identification (AVI), automatic passenger counters (APC), surveys and manual reports. It is crucial to integrate data from federal, state

and local public agencies, and other system operators, as each of them has one or many sensing systems that generate massive data. With the new mobile technology, people have been active in social media. Therefore, human beings are seen as active sensors, who report their daily activities with geo-locations. Massive data from social media, such as Tweets, Flickr, Waze, can make great contribution to the transportation society. Those data sets are complementary. Together they enhance data quality and present comprehensive and reliable information.

2) **Understanding integrated data**. Integrated data convey the information regarding traveler's behavior and system performance. The behavior essentially implies how a traveler makes choices in traveling, (when and whether to make a trip, what mode to choose, where to go, how to drive, etc.). Besides behavioral information, embedded in data is the status of each individual system and how well it is coordinating with other systems. Traditionally, transportation engineers and scholars have relied on hypothetical mathematical models to describe behavior and design transportation systems. The development of these models typically involves assumptions or approximations that are difficult to verify without data. The required data were usually not available, due to either immature sensing technology or a barrier of data sharing. Consequently, solutions developed from theories may not be directly applicable to practice. One of the most challenging issues is that transportation simulation packages, vital to all types of system design, cannot simulate dynamic trips as accurately as researchers wish. Mathematical theories simply cannot resolve this issue due to the complexity of human factors and the nature of complex transportation systems. Consequently, the availability of integrated data fills the gap between theories and engineering practice, which enables a better understanding of systems and human beings.
3) **Optimal decision making for systems management and for individual travelers**. The ultimate goal of understanding big data is to accurately estimate the historical and real-time usage of the transportation infrastructure and to forecast its future performance. To efficiently balance the infrastructure supply and demand, optimal decisions on management strategies, policies and adoption of technologies can be made. In addition, travelers can utilize the information to make better individual trip decisions.

To efficiently manipulate large-scale data, a **mobility data engine** is necessary to accommodate needs of data fusion and analytics. The data engine for this center essentially sets protocols for data exchange from various sources. The data could be stored and managed in distributed servers, but the engine offers **organization, visualization and analytics of transportation data.** The engine can be translated into useful information for people who need it: legislators, transportation planners, engineers, researchers, commuters, privatecompanies. Unlike the traditional single computer stand-alone software or tools for data preparation and system design, the data engine relies on web-based data sharing and browser-based human-computer interaction.

There are quite a few traffic information systems that have been developed over the last decade. Just to name a few, PeMS (pems.dot.ca.gov) in State of California, DriveNet (uwdrive.net) in State of Washington, RITIS (www.ritis,org) in State of Maryland, 511 (www.511pa.com) for State of Pennsylvania. Those information systems have a similar idea of data fusion as the proposed data engine, but they do not work directly with large-scale data learning. In other words, most of the tasks in those systems deal with import/export and visualization of data, but no behavior information is learned, nor is optimal probabilistic decision making revealed. They usually work with state DOT to integrate data from traditional sources, such as inductive loops and cameras. The data coverage, in most cases, is only on the statewide highway or freeway. The city streets are not covered. In addition, most of those systems only work with a single transportation mode, driving, whereas walking, bicycling, public transportation, parking and other types of transportation modes are not considered yet. Google map (maps.google.com) is one of the successful products (other similar products include Bing map, Mapquest, Waze, etc.) that integrate mobile data and traditional traffic data to provide real-time travel information for multi-modal systems. However, due to the nature of the product, Google map does not intend to manage or optimize the entire transportation systems. For this reason, Google map does not provide a platform for data exchange, and it only works with several data sources for the purpose of improving individual traveling experience. For instance, it does not include cameras, APC, freight, and others. At this moment, it does not offer travel time forecast nor system performance monitoring. **Table 1** shows a comparison of those travel information systems.

Table 1 A comparison of several travel information systems.

	Statewide mainlines	**City streets**	**Multi-modal**	**Data exchange**	**Data learning and forecast**	**Transportation management**
PeMS	×			×		×
RITIS	×	×		×		×
DriveNet	×			×		×
511PA	×					×
Google Map	×	×	×			
Our goal	×	×	×	×	×	×

2. Regional Traveler Information System (RTIS)

Our Mobility Data Analytics Center, the data engine for building a regional traveler information system, sets its goal as a comprehensive solution to multi-modal transportation systems in urban, suburban and rural areas. Interacting with the data engine are the three entities, travelers, public agencies, and the private sector. The data with information under the hood, as depicted in **Figure 1**, will help public agencies monitor and manage the system efficiently, help the private sector develop travel-related products or service, and ultimately improve travelers' experience.

The ultimate objective of the Mobility Data Analytics Center is to,

- Provide archived and real-time traffic data of every element of regional multi-modal transportation systems
- Reveal the behavioral information for both passenger transportation and freight movement
- Serve as a key policy instrument for legislators, regional transportation planners, national researchers and engineers
- Serve as a key information platform for regional travelers and regional freight movement.

The Mobility Data Analytics Center supports a Regional Traveler Information System (RTIS) through the following ultimate objectives:

- Work with the Power of 32 leadership to establish a non-profit regional organization (e.g., Move 32) which will use information provided by the Mobility Data Analytics Center.
- Supply regional real-time and predictive information to the traveling public, freight companies and transportation agencies through the web service, online blogs, phone applications, social media, etc..

See **Figure 1** below for diagram of proposed structure and explanation of components.

Imagine that travelers in this region can readily access trip information through a modally integrated Regional Traveler Information System (RTIS) that delivers proactive alerts, on-demand planning information and a data repository for ongoing planning and predictive modeling. For the four state DOTs involved in this information system, a regional RTIS would allow them to improve safety, incident management, and emergency management through, for example sharing information between Traffic Management Centers and emergency management partners, or deploying a Road Weather Information System. RTIS has the potential to build off and leverage many pieces of the existing Intelligent Transportation Systems (ITS) infrastructure and transportation management strategies in the Power of 32 Region.

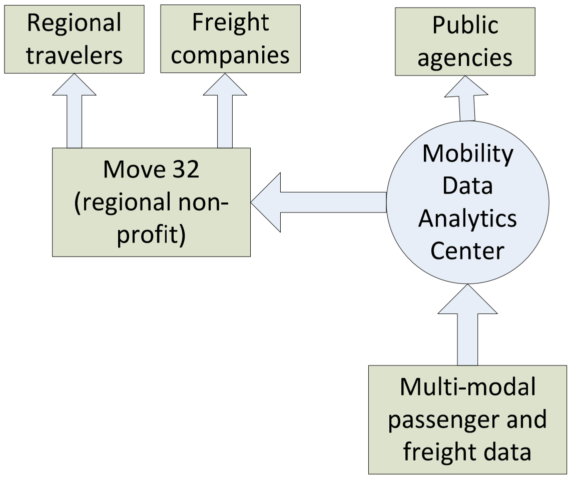

Figure 1 Structure of the mobility data analytics center

The Regional Traveler Information System (RTIS) has the following four advantages for regional development.

2.1. Positively Influence Traveler Behavior

Reliable, real-time data allows travelers to dynamically respond to the transportation system so they can optimize their travel choices. If travelers throughout the region have access to better information, they will be better able to choose routes and transportation alternatives that benefit them. Once users exploit underused options, their behavior will promote better balance across the overall transportation system. This change will therefore benefit both the users of RTIS, and the transportation agencies responsible for systems and infrastructure. Studies have shown that the effects of implementing a comprehensive information system can "improve the performance of transportation networks by maximizing the capacity of existing infrastructure, reducing the need to build additional highway capacity. For example, applying real-time traffic data to U.S. traffic signal lights can improve traffic flow significantly, reducing stops by as much as 40%, and reducing travel time by 25%" (Ezell, 2014)

2.2. Transform Data into Information and decision making

There are a multitude of transportation agencies throughout the Power of 32 region, including over 50 transit agencies, 10 metropolitan planning organizations, 4 state departments of transportation, the federal government, as well as various involvement from municipalities and counties. Many of these organizations collect large amounts of raw transportation data. However, this data does not have much value isolated. These organizations are limited to their geographic distribution and their slice of the transportation system. There is no simple method of examining the Power of 32 region's transportation system as a whole; however, The Mobility Data Analytics Center with the RTIS can fulfill the important role of making the data available to transportation users and travelers to enable improved system management and operations.

The Mobility Data Analytics Center can assemble and integrate transportation data from the many regional sources into a coherent data warehouse. Then, it can apply predictive analytical models to this data, and extract valuable information, and disseminate it back to the user and to the transportation agencies. As transportation systems become increasingly data driven, an RTIS system will be the key to aggregate and process this information in a valuable way.

2.3. Equal Accessibility

As a public good, RTIS should strive to excel in equal accessibility for its users. RTIS will be used by a broad spectrum of users, some with intellectual or physical disabilities, varying levels of income, and those unaccustomed to the latest technological devices. Therefore, Power of 32 should investigate several different user interfaces, rather than relying on a few, like just a smart phone app. A system that connects with users across different interfaces will have a broader reach, and will be able to attract a diverse user base. Using the same data engine and data analytics, Move 32 should also consider implementing a telephone system so it can provide equitable accessibility to uses without smartphone or internet access, which may stem from an economic disadvantage. When website, app, and phone services are running smoothly, RTIS may look for ways to offer its services in multiple language options to users.

2.4. Environmental Benefits

An RTIS would improve the efficiency of the whole regional transportation system by reducing travel times and incentivizing the use of multi-modal alternatives. These results will have positive environmental benefits. Studies estimate that an effective RTIS can reduce travel time by 25%, and therefore cut gas consumption by 10% and emissions by 22%.

3. Prototype Design and Development

We have developed a prototype Power of 32 Regional Traveler Information System that provides the most promising and readily available travel information, such as roadway traffic volume and travel time, traffic video, public transit, parking, incidents, weather, and electrical vehicles charging stations. The primary goal of the prototype system is to collect various types of traffic data in the Power of 32 Region, and to demonstrate its effectiveness in facilitating smart decision making among travelers, private firms and public agencies in Power of 32 Region. We collect data from various sources. For each type of data, we establish standard data format, process them, and furthermore store them in an integrated database. An information system is built upon the integrated database with components of data analytics. We extract historical information from the information system. We also develop models to make prediction of certain traffic information by analyzing the data comprehensively. In addition, we developed a web application (see **Figure 2**) to present information to users, travelers and public agencies. Users can acquire the information they need through various options. Both historical and predicted information can be visualized on maps. More importantly, the information can be used towards smart decision making in the region.

We briefly present the technical structure of the information system prototype. Each data source is further described with future work plan.

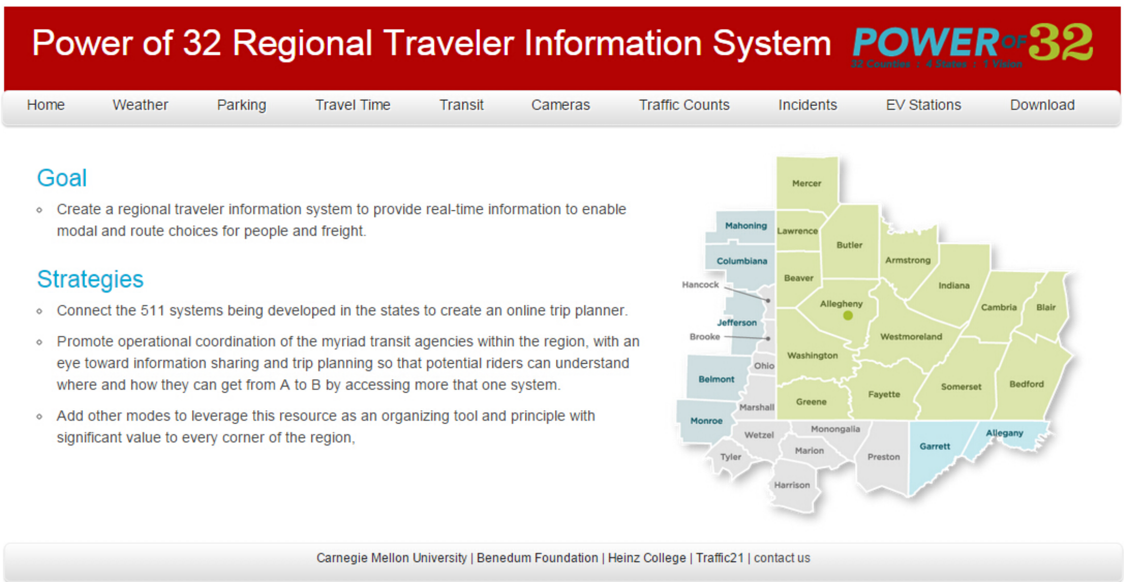

Figure 2 The prototype web application for Power of 32 Regional Traveler Information System

3.1. Structure of the prototype information system

The overall technical structure is shown in **Figure 3**.

MySQL is used as the database. For each type of data, a table is used to store raw data. In addition, tables of processed data are created to support the front end features of the web. MySQL is one of the most popular Relational Database Management Systems. It has plenty of features required by our system. It supports all major platforms including Red Hat Linux used on our server. It works well with the web application developing tool (Python/Django) we adopt. Comparing with other relational databases using SQL, MySQL is open source and thus free of charge. Though MySQL, as a relational database tool, is not as efficient as NoSQL databases, it provides reasonable system performance and efficiency. In addition, MySQL provides more functions such as data integrity check. These functions will be critical for our data maintenance. (Oracle, 2014)

We use the framework Django to develop the web application. Django features Model-Template-View structure. Models describe the back-end database. Each class in the 'Model' file corresponds to a table in the database. 'Templates' are used as the templates of web pages. 'Views' deal with the logics. 'Views' take users' inputs, fetch data from the database through 'Models', and perform logics with the data. Then 'Views' render the relevant 'Templates' to generate web pages, and at last return the web pages to users. Django is a framework suitable for developing light to medium size web applications. It has separate modules responsible for different tasks of building a web application. Therefore, the development process will be very efficient with the help of Django. As a Python based framework, Django also inherits Python's advantages including clean style and sufficient package support. (Behrens, 2012) (Django, 2014)

OpenStreetMap and Leaflet JavaScript package are used for visualization of various data at the front end. OpenStreetMap is a large open source map project. We use it to develop the base maps of our web, namely the contextual raster maps. However, OpenStreetMap also features complete vector road networks with traffic data and analytics tools. (OpenStreetMap, 2014) We will develop more functionalities for our web based on OpenStreetMap in the future, as OpenStreetMap supports both vector and raster data analytics. Leaflet is a light-weighted open source JavaScript library for interactive maps. The map visualization using Leaflet is easy and efficient. In addition, it also has many developers contributing high-quality packages for more functionalities. (Agafonkin, 2014)

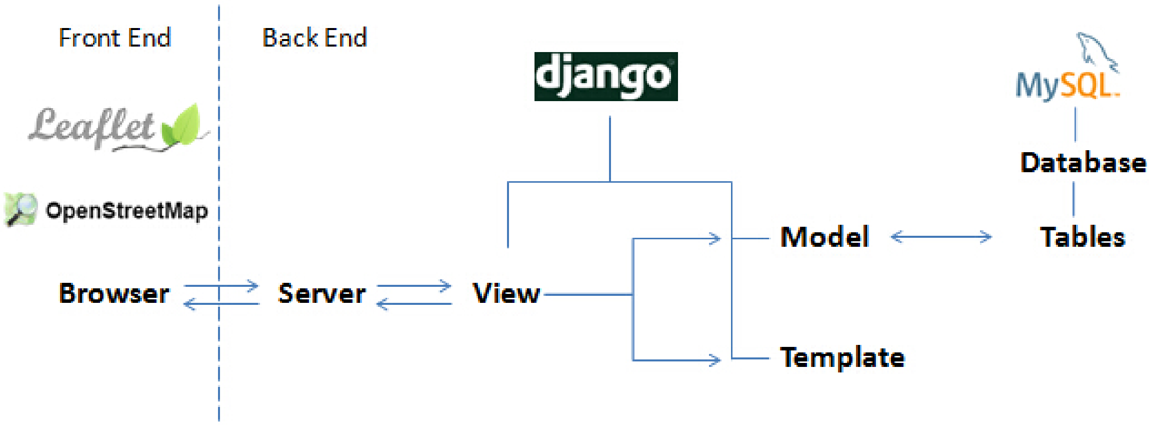

Figure 3 The structure of the prototype information system

3.2. Comparison with other systems

To acquire a better understanding of the advantages and disadvantages of our Power of 32 Regional Traveler Information System (RTIS), we compared our prototype with several other traffic web applications, such as HERE.com, PeMS and RITIS. The results of the comparison and some important features are shown in **Table 2**.

Table 2 Comparison among our web, HERE.com, PeMS and RITIS

	Our Web	HERE.com	PeMS	RITIS
Registration Required?	○[1]	○	●	●
Historical Data	●	○	●	●
Real-time Data	●	●	●	●
Traffic Prediction	●	○	○	○
Data Visualization	●	●	●	●
Data Downloadable?	Soon Available	○	●	●
Data Diversity[2]	●	○	●	●

[1] ○: No ●: Yes

[2] Data Diversity means providing a wide variety of multi-modal information such as parking, the public transit, weather, incidents, cameras, and etc.

The results show that our prototype RTIS web application possesses almost all important features of a traveler information system. For instance, it provides historical data, real-time data, map-based data visualization, and more importantly, it can make predictions of multi-modal traffic information by making use of traffic models, real-time data and historical data. Nevertheless, our RTIS website still requires further development to improve the accuracy of our traffic models, as well as to fulfill the functionalities such as user account management and data downloading.

3.3. Data and Front End

In the following sections, we describe details of each type of data source. We describe each type of data in three aspects: extract, transform and load of data. In addition, web interfaces and future work are also discussed.

3.3.1. Parking

The parking data we have collected are transaction records of parking kiosks in downtown Pittsburgh and near CMU from January to April 2014, and from October 24 to November 6, 2014. From the data, we estimate the occupancy of street parking space over time of day. We display our estimation with different colors on our web pages. Users can acquire historical parking conditions over the time period they select, as well as the time-of-day prediction on any future day. We estimate parking occupancy for cars and trucks separately, so both individual travelers and truck drivers are able to get the specific information they need. We have also collected real occupancy data over the last months. With that data, we can improve our estimation and forecast methods. In the future, we also would like to forecast a specific event's (such as a Steelers' game) impact on parking conditions. With the help of parking information, users can better plan their trips and save time finding parking.

Extract:

① **PPA Data**

Source: Pittsburgh Parking Authority (PPA)
Description: Purchase records of Pittsburgh downtown's parking meters from January 2nd, 2014 to April 22nd 2014.
Content: January 2014 Purchases, February 2014 Purchases, March 2014 Purchases, April 2014 Purchases, Downtown Meter Locations with Latitude and Longitude.
Format: Excel files.

Purchase records files: (Size: 10.2 MB, 247489 records)

Terminal - Terminal ID	Payment type	Transaction Date	Space paid until time	Amount
401512-FORBES0405	Card	2014-04-01 5:11:58 AM	2014-04-01 9:12:37 AM	6.00

Terminal ID: the ID of the meter.
Payment type: Card / Coin. The way the purchase was paid.
Transaction date: the time purchase happened.

Space paid until time: ending time of the purchased parking period.
Amount: dollars. The amount of money charged for the purchase.

Downtown Meter Locations file: (Size: 15.5 KB, 99 meters)

Street Name	Block	Meter ID	Latitude	Longitude
Smithfield St	400	401347-SMTHFD0401	40.43980459193305	-79.9986782670021

Street Name: the name of the street the meter is on
Block: the block number of meter's location
Meter ID: the ID of the meter, correspond to Terminal ID
Latitude, Longitude: the latitude and longitude of the meter

Integrity: Missing records for meter 401347-SMTHFD0401.

② **ParkPGH Data**

Source: ParkPGH.org
Description: Real-time parking information about 24 parking lots in downtown Pittsburgh
Format: The queries are URLs; the responses are in JSON format. Sample response is shown below:
{"id":2,"short_name":"Theater Square","long_name":"Theater Square","address1":"655 Penn Avenue","address2":"","city":"Pittsburgh","state":"PA","zip":15222,"lat":40.443869,"lon":-80.00061,"max_spots":770,"available_spots":613,"reserved_spots":0,"active":true,"keyword":"TSQUARE","keyphrase":"test phrase","passcode":8382,"hours":"24 hours a day, 7 days a week","is_full":false,"spots_last_updated":"2014-12-05T16:36:32-05:00","open_time":null,"close_time":null,"area_id":"1:3:","user_id":null,"created_at":"2013-03-26T10:53:42-04:00","updated_at":"2014-12-05T16:39:07-05:00","raw_number":157,"spots_updated_time":" 9:36 PM","spots_updated":"2014-12-05T16:36:32-05:00"}

Transform:

Description: We processed the PPA data to estimate the occupancies of parking spaces of streets in downtown Pittsburgh. The estimation is made for every 10 minutes from Jan 2nd to April 22nd.

Steps: 1. Group the meters. Since drivers may make a purchase at any meter near their parking locations, we have to group the meters before estimating occupancy. The rule we use is to merge the meters in the same block of the same street into a single group. We estimate parking occupancies in terms of groups, that is, we estimate parking occupancies for each block of the streets.

2. Estimate maximum capacity for each group. We divided the length of each block by the average length needed to park a car (5 meters) to obtain the capacities.

3. Estimate occupancies. Our assumption is that when a driver purchases a parking space, her/his probability to leave increases as the time approaching the ending time of the purchased period. And the driver will always have left by 10 minutes past the ending time. We used a polynomial to model the driver's probability to leave. For each record, denote the purchase time by t_s, ending time of the purchased period by t_e. The probability p of the driver has left at time t ($t_s < t < t_e$) is:

$$p(t) = \frac{t - t_s}{t_e + 10 - t_s}$$

The record's contribution to the occupancies of the group the meter is in at time t when $t_s < t < t_e$ is:

$$c = 1 - p(t)$$

The contribution is 0 otherwise.

For any given time t, we estimate the occupancy of a group by summing up all contributions of records whose period covered time t. Then we divided this estimated occupancy by maximum capacity of the group to get the occupancy ratio of the group at time t.

We estimated all groups' occupancy ratios for every 10 minutes from Jan 2nd to April 22nd, 2014.

Result Sample:

Estimated Occupancy Ratios: (Size: 6.64MB, 5949 records)

Street ID	Date	Occupancy
1STAVE060	2014/1/2	0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0833313846116, 0.163402935868, 0.108833649897, 0.0751556804967, 0.0459473003182,
3RDAVE020	2014/1/2	0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0312499985417, 0.0307557907738, 0.018299323519, 0.0, 0.0, 0.0, 0.0

Street ID, Date: Same as above.
Occupancy: Float number between 0 and 1. The estimated occupancy ratios for every 10 minutes from 0:00 to 23:50 of the day.

Load:

Tables in database:

traffic_streetparking (<u>id</u>, street_id@, date, weekday, occupancy): This table corresponds to the estimated occupancies described above.
id (primary key, int): auto-generated id of each record in the table.

street_id (foreign key from traffic_street, varchar(20)): correspond to the street ID in Estimated Occupancy Ratios.
date (date): correspond to the date in Estimated Occupancy Ratios.
occupancy (longtext): correspond to the occupancy in Estimated Occupancy Ratios.
Index: (street_id, date)

traffic_street (sid, street_name, coordinate): This table stores names and coordinates of street block groups we get at the step 2 of transform.
sid(primary key, varchar(20)): correspond to the street ID in Estimated Occupancy Ratios.
street_name (varchar(20)): name of the street block
coordinate (varchar(256)): two pairs of latitude and longitude, the start and end coordinates of the street block.

traffic_meter (mid, street_name, block, latitude, longitude): This table corresponds to the downtown meter locations file.
mid(primary key, varchar(20)), street_name (varchar(20)), block (varchar(5)), Latitude (double), Longitude (double): all correspond to the columns in downtown meter locations file.

traffic_transaction (tid, type, transaction_time, end_time, amount)
This table stores the raw transaction data.

Front End:

Users can choose a historical time period and a specific weekday to see the estimated average time-of-day street occupancy data of that weekday (Wednesday in the following example) over this period on the map. When "All data" is selected, the average will be taken over all historical data we have (from Jan. 2nd to Apr. 22nd, 2014). See **Figure 4**.

Figure 4 Time-of-day Parking occupancy estimation

A future date can also be selected. In this case, we provide users prediction of time-of-day street parking occupancy by analyzing historical data. See **Figure 5**.

Figure 5 Time-of-day Parking occupancy forecast

For both historical average and prediction, from "time to display" option, users can input a time to visualize the parking occupancy of a specific time. In the following example **Figure 6**, we change "time to display" to 1:00 p.m. from 11:00 a.m.

Figure 6 Time-of-day Parking occupancy forecast with a specific time of day

With the controller on the top right corner of the map, users can switch the content of the map to our estimation/prediction for truck parking occupancy. See **Figure 7**.

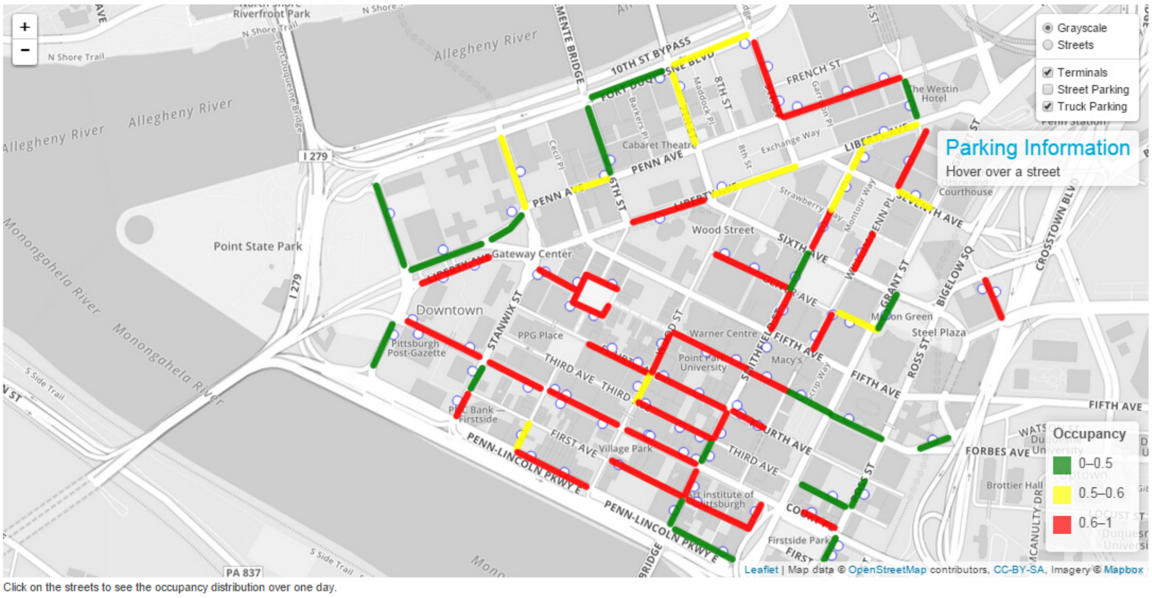

Figure 7 Time-of-day truck parking occupancy estimation

By clicking on a specific street, users can obtain the curve of the occupancy that changes over time of day. See **Figure 8**.

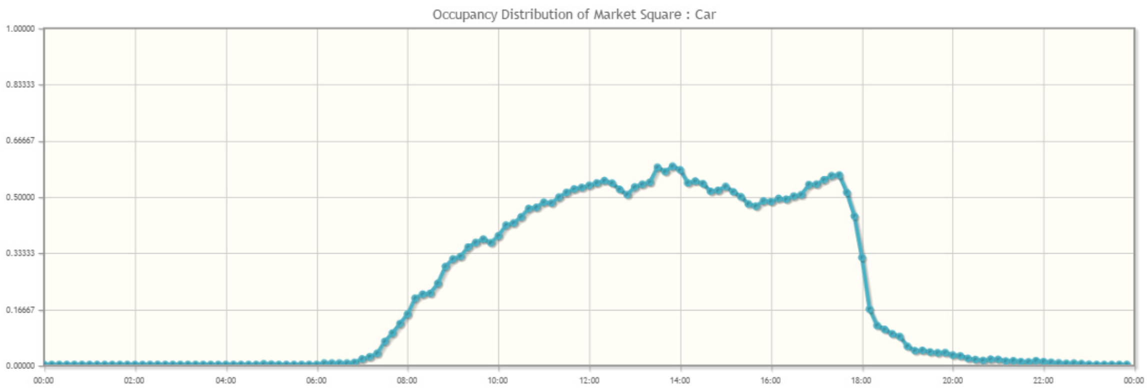

Figure 8 Time-of-day parking occupancy estimation of Market Square

Future Work:

Back End:

Now, we only have 4 months of historical transaction records. The system can be easily extended to estimate or forecast parking occupancy based on real-time data access. In that case, back end needs to be modified.

The estimation/forecast model we use tends to underestimate the occupancy. We will develop a sophisticated model to combine both historical and real-time data to estimate/forecast the occupancy.

Front End:

Besides street parking, we will display the occupancies conditions of the parking lots in downtown Pittsburgh.

3.3.2. Travel Time

Users can obtain the historical travel time of a past time period or our predicted travel time for a future day on selected road segments. In the future, we plan to develop a comprehensive trip planner based on the travel time data. Given a departure time, an origin and a destination, we provide users a recommended route that minimizes travel time/cost. Users can specify the vehicle type (car or truck) for such recommendations. We will also consider real-time traffic conditions and incidents/events in the optimal route calculation.

Extract:

① **INRIX Data**

The INRIX data we have are based on Traffic Message Channels (TMC) that defines geographical locations of road segments. On each TMC, there are travel time and speed data for every 5 minutes in the Year of 2013.

Source: INRIX through RITIS, University of Maryland CATT Lab
Description: TMC-based travel time and speed data for 2013. They are average data for every 5 minutes. The TMCs cover the major highways of Allegheny County.
Content: Each month's travel time and speed data, TMC Identification.
Format: csv files.

Data files: (Size: 1.58 GB)

tmc_code	measurement_tstamp	speed	average_speed	travel_time_minutes
104+04439	2013/4/1 0:00	59.8	60	1.37

tmc_code: 9 characters. The ID of the TMC
mearurement_tstamp: the timestamp of the speed and travel time data
speed: miles per hour. 5 minutes' average speed on the road segment the TMC represents.
average_speed: miles per hour. The historical average speed on the road segment the TMC represents.
travel_time_minutes: minutes. 5 minutes' average travel-time on the road segment the TMC represents.

TMC Identification: information about TMCs of highways in Allegheny County.

Size: 38.5 KB, 355 TMCs.

tmc	road	direction	intersection	state	county	zip	start_latitude	start_longitud	end_latitud	end_longitud	miles	road_order
104+04509	I-279	N	Rosslyn Rd/Exit 2	PA	ALLEGHENY	15106	40.421708	-80.097002	40.420347	-80.0927	0.245826	2
104-04532	I-279	S	Camp Horne Rd/Exit 15	PA	ALLEGHENY	15237	40.584266	-80.099591	40.53015	-80.072424	4.1599	2
104N04532	I-279	S	Camp Horne Rd/Exit 15	PA	ALLEGHENY	15237	40.53015	-80.072424	40.520096	-80.0628652	0.908673	3

tmc: the ID of the TMC, uniquely indentify a segment of road, corresponds to tmc_code in data files.
road: the road name of the segment the TMC represents
direction: the traffic direction of the road segment the TMC represents
intersection: the intersection that define the TMC
start_latitude, start_longitude: the latitude and longitude of the start position of the road segment the TMC represents
end_latitude, end_longitude: the latitude and longitude of the end position of the road segment the TMC represents
miles: the length of the road segment the TMC represents
road_order: the order of the TMC on the road. Order is consistent with the direction.

Integrity: The current data set only covers the highways, and it does not have data and TMC information regarding major arterials owned by the City of Pittsburgh.

Transform:

Description: We separated the measurement_tstamp column in the data files into two columns, date and time.

Load:

Tables in database:

traffic_tmc (<u>tmc</u>, road, direction, intersection, state, county, zip, s_lat, s_lon, e_lat, e_lon, miles, road_order)
This table stores tmc information. The columns correspond to the columns in the TMC Identification file.
Index: road

traffic_tmc_data (<u>id</u>, tmc_id$^{@}$, date, time, speed, avg_speed, travel_time)
This table stores travel time and speed data. The columns correspond to the columns in the data files. The measurement_tstamp column in data files are separated into date and time.
id (primary key, int): auto-generated id of each record in the table.
tmc_id(foreign key, varchar(9)): The record's TMC ID. Foreign key from traffic_tmc table.
date (date): the date of the record.
time (time): the hour-minute time of the record.

speed (double), avg_speed(double), travel_time(double): store data of corresponding columns in the data files.

② SPC CORRIDOR DATA

Source: Southwestern Pennsylvania Commissions (SPC)
Description: Weekday average (typical) travel time, speed, traffic volume and delay data for year 2003-2012. Data contain approximately 100 roadway corridors in the 12 counties of southwestern Pennsylvania.
Content: Node list of each corridor, weekday average travel time, speed, traffic volume and delay data of each corridor in year 2003-2012.
Format: csv files.

Extract:

Data files: (Size: 29.1 MB)

A	to	Z	AM Peak Hour Volume	PM Peak Hour Volume	Speed Limit (mph)	Distance (miles)	Travel Time @ Posted Speed Limit (min)	Weighted Avg. Calc for Avg. Speed @ Posted Speed Limit (mph)	AM Measured Travel Time (min)	PM Measured Travel Time (min)
SR 228 West	to	Sandy Hill Road	850	700	45	2.66553	3.55404	119.94886	3.23	3.22

AM Measured Avg. Speed (mph)	Weighted Avg. Calc for Avg. Speed @ (mph)	AM Delay per Vehicle (min)	PM Measured Avg. Speed (mph)	Weighted Avg. Calc for Avg. Speed @ (mph)	PM Delay per Vehicle (min)	AM Total Delay (veh-hr)	PM Total Delay (veh-hr)
49.38964	131.64958	0.00000	49.79392	132.72720	0.00000	0.00000	0.00000

A, Z: The name of the start and end nodes (all the following measurements are taken on the road segment between the start and end nodes).
AM Peak Hour Volume (vehicles per hour): Weekday average traffic volume of morning peak hour.
PM Peak Hour Volume (vehicles per hour): Weekday average traffic volume of evening peak hour.
Speed Limit (mph): Speed limit on the road segment between the start and end nodes.
Distance (miles): Distance of the road segment between the start and end nodes.
Travel Time @ Posted Speed Limit (minutes): Travel time at speed limit.
AM Measured Travel Time (minutes): Weekday average travel time of morning peak hour.
PM Measured Travel Time (minutes): Weekday average travel time of evening peak hour.
AM Measured Average Speed (mph): Weekday average speed of morning peak.
AM Delay per Vehicle (minutes): Weekday average delay per vehicle of morning peak.
PM Measured Average Speed (mph): Weekday average speed of evening peak hour.

PM Delay per Vehicle (minutes): Weekday average delay per vehicle of evening peak hour.
AM Total Delay (hours): Weekday average total delay of morning peak hour.
PM Total Delay (hours): Weekday average total delay of evening peak hour.

Integrity: The data covers approximately 100 roadway corridors in the counties of southwestern Pennsylvania, including Allegheny, Beaver, Washington, and other 7 counties.

Transform:

Description: We extracted useful columns from the data files and stored them into MySQL database.

Load:

Tables in database:

Corridor#_YEAR_dgs_traveltime(Start_Node,End_Node,AM_Peak_Hour_Volume,PM_Peak_Hour_Volume, Speed_Limit, Distance, Travel_Time_At_Posted_Speed_Limit, AM_Travel_Time, PM_Travel_Time, AM_Avg_Speed, AM_Delay_Per_Vehicle, PM_Avg_Speed, PM_Delay_Per_Vehicle, AM_Total_Delay, PM_Total_Delay, Direction)

These tables store weekday average travel time, speed, traffic volume and delay data of each corridor. In the name of the table, "#" represent the number of corridor, "YEAR" represent which year the data comes from. **The columns in these tables correspond to the columns in the data files.**

Start_Node (primary key, varchar(50)): The name of the start node.
End_Node (primary key, varchar (50)): The name of the end node.
AM_Peak_Hour_Volume (int(11)): **AM Peak Hour Volume**
PM_Peak_Hour_Volume (int(11)): **PM Peak Hour Volume**
Speed_Limit (int(11)): **Speed Limit**
Distance (double): **Distance**
Travel_Time_At_Posted_Speed_Limit (double): **Travel Time @ Posted Speed Limit**
AM_Travel_Time (double): **AM Measured Travel Time**
PM_Travel_Time (double): **PM Measured Travel Time**
AM_Avg_Speed (double): **AM Measured Average Speed**
AM_Delay_Per_Vehicle (double): **AM Delay per Vehicle**
PM_Avg_Speed (double): **PM Measured Average Speed**
PM_Delay_Per_Vehicle (double): **PM Delay per Vehicle**
AM_Total_Delay (double): **AM Total Delay**

PM_Total_Delay (double): **PM Total Delay**
Direction (varchar(20)): The direction from start node to end node.

Front End:

Users can select two nodes on a road, a historical period and a period over a day, and then obtain the average travel time between the two nodes of the time period. See **Figure 9**.

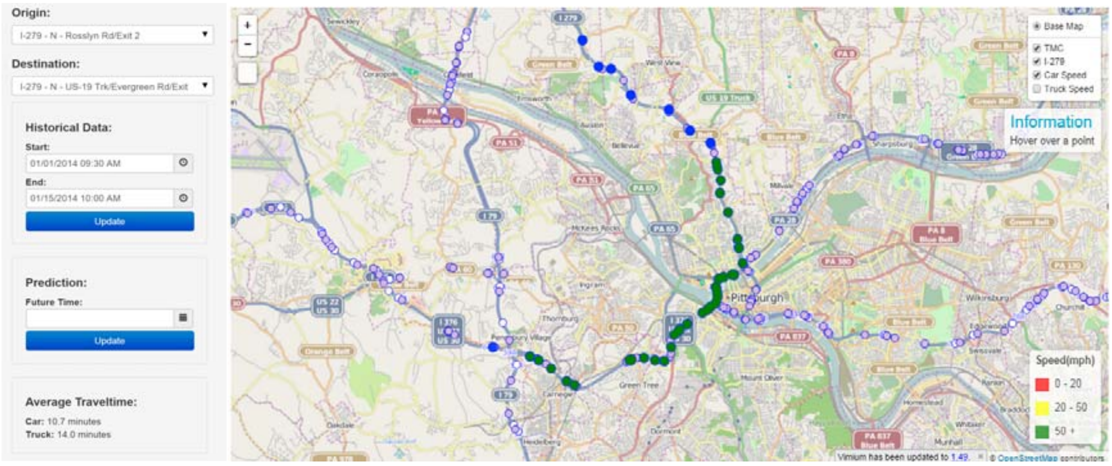

Figure 9 Travel time estimation on highways for cars

With the controller on the top right corner of the map, the map content can be switched to estimate truck travel time. See **Figure 10**.

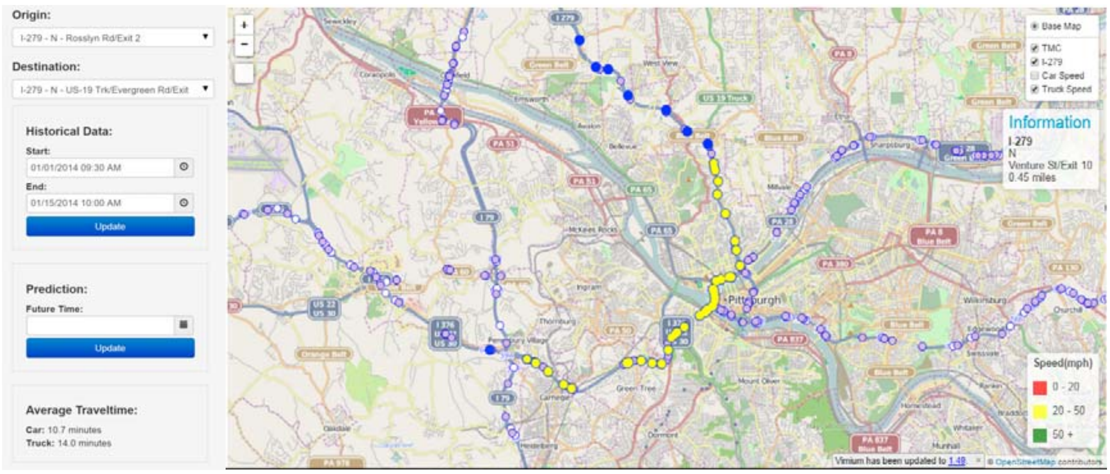

Figure 10 Travel time estimation on highways for trucks

Users can also choose a future date and time of day to obtain the predicted travel time. See **Figure 11**.

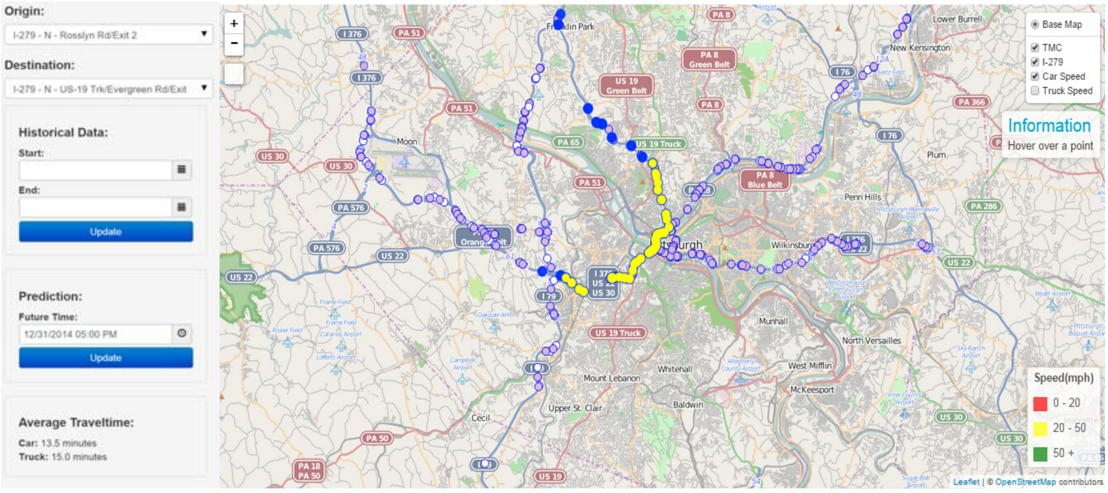

Figure 11 Travel time forecast on highways for cars

Future Work:

The features of the travel time part are quite limited now. In the future, we want to develop a time dependent trip planner to calculate optimal routes for car and truck drivers. The drivers can input arbitrary origin and destination as well as their departure time. Our system will calculate one or several optimal routes for them that minimize travel time/cost. The calculation will take into account the predictive travel time for the entire traveling time period. We also plan to add real-time incident data into the calculation.

To build the optimal trip planner, two prerequisites are required. We plan to calculate the optimal route on a GIS layer. Since the majority of the travel time and speed data are based upon TMCs, a GIS-TMC mapping is necessary. In addition, we need to build a model to predict future travel time and speed with historical data and real-time data feed. We have made some progress on both of the two tasks.

3.3.3. Transit

For the transit section, we have Pittsburgh's general transit feed specification (GTFS) and transit APC-AVL data (Automatic Passenger Counts and Automatic Vehicle Location) from Port Authority in 2013. With GTFS, we are able to show users the geographic information and stops of transit routes. We are building models to calculate metrics of transit routes, including travel time, on-time performance, crowding, incidents and waiting time from the transit data. Public transit passengers are provided with those metrics of a route over a historical time period or a future time period. In addition, real-time public transit vehicle positions will also be displayed on our website with the help of the BusTime API.

Extract:

General Transit Feed Specification (GTFS)

Source: Pittsburgh Port Authority, Google
Description: GTFS is a common data format that describes schedules and associated geographic information of public transit. The GTFS we have describes public transit information of Pittsburgh. (Google, 2012)
Content: agency.txt, stops.txt, stop_times.txt, routes.txt, trips.txt, calendar.txt, calendar_dates.txt, shapes.txt, transfers.txt
Format: txt files. Detailed descriptions of the GTFS format can be found at https://developers.google.com/transit/gtfs/reference

The following are brief descriptions and samples.
agency.txt: Information about the agency who provide the GTFS

stops.txt: Information about stops/stations: (Size: 469 KB, 6932 stops)

stop_id	stop_code	stop_name	stop_desc	stop_lat	stop_lon	zone_id
E00175	2593	26TH ST AT PENN AVE FS (SPRING WAY)		40.454635	-79.97817	1
E00180	2161	28TH ST AT LIBERTY AVE		40.456288	-79.97465	1

stop_times.txt: times that a transit vehicle arrives and leaves a stop: (Size: 44.2 MB, 721248 records)

trip_id	arrival_time	departure_time	stop_id	stop_sequence	pickup_type	drop_off_type
7898615-1406-Collier-Weekday-28	4:59:00	4:59:00	W07587	1	0	0
7898615-1406-Collier-Weekday-28	5:01:00	5:01:00	W06630	2	0	0

routes.txt: transit routes; each route has a collection of trips: (Size: 3.86KB, 103 routes)

route_id	agency_id	route_short_name	route_long_name	route_des	route_type	route_url
001-152	PAAC	1	Freeport Road		3	
002-152	PAAC	2	Mount Royal		3	

trip.txt: trips for each route: (Size: 1.54 MB, 13372 trips)

route_id	service_id	trip_id	trip_headsign	direction_id	block_id	shape_id
020-152	1406-Collier-Weekday-28	7898615-1406-Collier-Weekday-28	Inbound-KENNEDY TO DOWNTOWN	1	1199520	200033
020-152	1406-Collier-Weekday-28	7898616-1406-Collier-Weekday-28	Inbound-KENNEDY TO DOWNTOWN	1	1199523	200033

shapes.txt: geographic information of the trips: (Size: 8.26 MB, 234846 shapes)

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence
10060	40.600608	-79.757111	10001
10060	40.600583	-79.757112	10002
10060	40.600435	-79.757089	10003

Transit Automatic Passenger Counts and Automatic Vehicle Location data (APC-AVL)

Source: Pittsburgh Port Authority
Description: Passenger boarding and alight counts and bus location data for every single bus trip served by Port Authority in 2013

Format:

DOW	dir	ROUTE	TRIPA	BLOCKA	VEHNOA	daymoyr	STOPA	QSTOPA	ANAME	HR	MIN	SEC	DHR
1	0	1	0615	001005	5210	02-Jan-2013	999	00009999	Not Identified - Cal	6	44	50	6

DMIN	DSEC	ON	OFF	LOAD	DLMILES	DLMIN	DLPMLS	DWTIME	DELTA	SCHTIM	SCHDEV	SRTIME	ARTIME
44	51	0	0	0	0.00	0.0	0.000	0.02	1,500	9,999	99.00	99.90	99.90

Port Authority Bus Time API

Source: Pittsburgh Port Authority
Description: APIs that return various real-time data of Pittsburgh's public transit.

Transform:

Description: We processed GTFS data in order to show users public transit routes at the front end. For every route in routes.txt, we collected the stops it passes on both inbound and outbound directions from stops.txt. We also collected the geographic information about direction of every route from trips.txt and shapes.txt. We formed the stops and geographic information in geoJSON format. The geoJSON format is basically JSON, it is used to store geographic related information. It can be used at the front end by Leaflet.js package. The details about geoJSON format can be found at http://geojson.org/

Load:

Tables in database:

traffic_route(<u>route_id</u>, agency_id, short_name, long_name, route_type, inbound_stops_geoJSON, outbound_stops_geoJSON, inbound_geoJSON, outbound_geoJSON)
This table stores information about routes.

route_id (primary key, varchar(10)): uniquely identify a route.
agency_id, short_name (UNIQUE), long_name, route_type: correspond to the columns in routes.txt
inbound_stops_geoJSON, outbound_stops_geoJSON (longtext): geoJSON format data about the stops a route passes on its inbound/outbound direction
inbound_geoJSON, outbound_geoJSON (longtext): the geographic information (shape) about the route on its inbound/outbound direction
index: short_name

traffic_trip (<u>trip_id</u>, route_id$^{@}$, service_id, headsign, direction_id, block_id, shape_id)
This table stores data about trips. It corresponds to trips.txt
route_id(foreign key, varchar(10)): reference to traffic_route.

traffic_stop (<u>stop_id</u>, code, name, lat, lon, zone_id, geoJson)

This table stores data about stops. It corresponds to stops.txt
geoJson(longtext): stores information about a stop in geoJSON format for the convenience of front end.

traffic_stop_time (id, trip_id@, stop_id@, arrival_time, departure_time, stop_sequence, pickup_type, drop_off_type)
This table stores data about trips' schedules at each stop. It corresponds to stop_time.txt
trip_id (foreign key, varchar(50)): reference to traffic_trip.
stop_id(foreign key, varchar(10)): reference to traffic_stop.

traffic_transit_shape (id, shape_id, lat, lon, sequence)
This table stores data about geographic information. It corresponds to shapes.txt.
index: shape_id

traffic_transit_data (id, dir, route, tripa, blocka, vehnoa, daymoyr, stopa, qstopa, aname, hr, min, sec, dhr, dmin, dsec, on, off, load, dlmiles, dlmin, dlpmls, dwtime, delta, schtim, schdev, srtime, artime)
This table stores 2013 public transit data of Pittsburgh.

Front End:

Users can obtain the geographic information and stops of a route. **Figure 12** is 61A inbound:

Figure 12 Geographic information of Bus route 61A

Now we select 71B outbound in **Figure 13**:

Figure 13 Geographic information of Bus route 71B

Users can select two stops of a route and a time of day, and get 3 metrics (on-time performance, crowding and waiting time) for that segment and time. See **Figure 14**.

Figure 14 Three metrics (on-time performance, crowding and waiting time) are computed for the time periods users select on Bus Route 61A, between two stops on Forbes Ave.

Now we choose two other stops of 61A inbound and another time period in **Figure 15**:

Figure 15 Three metrics (on-time performance, crowding and waiting time) are computed for the time periods users select on Bus Route 61A, between a stop on Forbes Ave. and a stop on 5th Ave

Future Work:

Back End:

We have 7.23 GB of public transit data and are likely to have more in the future. In order to calculate the metrics efficiently, we need to pre-process the data and set up some intermediate tables. We also plan to collect data with Bus Time API.

We are also developing sophisticated algorithms to compute various metrics to indicate the reliability of bus transit systems, such as bus and car travel time ratio, on-time performance, crowding, waiting time and incidents. Those reliability metrics can be used to adjust scheduling of the public transit system, as well as to improve passengers' traveling experiences.

Front End:

With Bus Time API, we are able to get the real-time data of public transit. We will display the real time information for users.

3.3.4. Traffic Counts

The traffic counts data cover interstate highways and major roads of Pennsylvania. The data are annual average daily traffic (AADT) counts. We match the counts data with their geographic information and visualize them on our web.

Extract:

AADT (Annual Average Daily Traffic) Count

Source: PA Department of Transportation

Description: Daily traffic counts data on segments of main roads of the Power of 32 region in PA. The data are collected with a 3-year cycle for interstates and a 5-year cycle for other roads. The interstate AADT counts we have are in the cycle of 2010-2013, other roads' AADT counts are in the cycle of 2008-2013.
Content: All Routes AADT Information, Interstate AADT Information
Format: Excel files.

All Routes AADT Information: the AADT count data of all major routes of Pittsburgh and nearby area (Power of 32, in PA): (Size: 1.34 MB, 9136 records)

State Route	County	District	Segment Begin	Offset Begin	Segment End	Offset End	Segment Length	AADT	Count Date
8	2	11	10	0	10	920	920	8792	07/07/2011
8	2	11	10	920	30	0	1519	8873	08/13/2013

State Route: the ID that identifies a route in PA.
County: the ID of the county a route segment is in.
District: the ID of the district a route segment is in.
Segment_Begin: the ID of beginning segment.
Offset_Begin: feet. The offset from the start of the beginning segment. Indicate the starting point of a road segment on which the count is collected.
Segment_End: the ID of ending segment.
Offset_End: feet. The offset from the start of the ending segment. Indicate the ending point of a road segment on which the count is collected.
Segment_Length: feet. The length of the road segment on which the count is collected.
AADT: annual average daily traffic count.
Count Date: the date the count was collected.

Interstates AADT Information: the AADT count data of highways of Pittsburgh and nearby area (Power of 32, PA part): (Size:512 KB, 242 records)

State Route	County	District	Segment Begin	Offset Begin	Segment End	Offset End	Segment Length	AADT	Count Date
0079	02	11	0496	0	0520	1560	14,156	9,136	10/18/2012
0079	02	11	0497	0	0521	1725	14,364	8,761	10/18/2012

Format is the same as All Routes AADT Information.

PaTraffic

Source: Pennsylvania Spatial Data Access (PASDA)
Description: The GIS map with the same segments of the above AADT files.

Transform:

Description: We processed data above and combined them into a geoJSON file.
Steps:

1. For each record in All Routes AADT Information and Interstate AADT Information, find the geographic information (latitudes and longitudes) of the segment in PaTraffic's GIS file.
2. Combine the count data and geographic data into a record of geoJSON format.

Result Sample:

Size: All Routes: 18.2 MB; Interstates: 577 KB

{"type":"Feature","properties":{"aadt_car":7634,"aadt_truck":1158,"aadt":8792},"geometry":{"type":"LineString","coordinates":[[-79.86854624660636, 40.43436520259753], [-79.86866819726758, 40.43442699048575], [-79.86909153778792, 40.434618503483345], [-79.86955059086685, 40.434828449914676], [-79.87002906763982, 40.43504866373033], [-79.87092007033, 40.43546094024394], [-79.87142971178257, 40.43570092679435]]}}

AADT counts data are in "properties"; geographic information is in "geometry".

Load:

Tables in database:

traffic_all_route_count (id, state route, county, district, segment begin, offset begin, segment end, offset end, segment length, aadt, count date)
This table stores raw data in the All Routes AADT Information.

traffic_interstate_count (id, state route, county, district, segment begin, offset begin, segment end, offset end, segment length, aadt, count date)
This table stores raw data in the Interstates AADT Information.

Other files:

all_route_aadt_geometry.js, interstate_aadt_geometry.js:

These two files store data processed in the transform step and can be directly used by the front end.

Front End:

With the controller on the top right of the map, users can switch between AADT counts of interstate highways (see **Figure 16**) and all roads of Power of 32 in Pennsylvania (see **Figure 17 Traffic counts of all roads**).

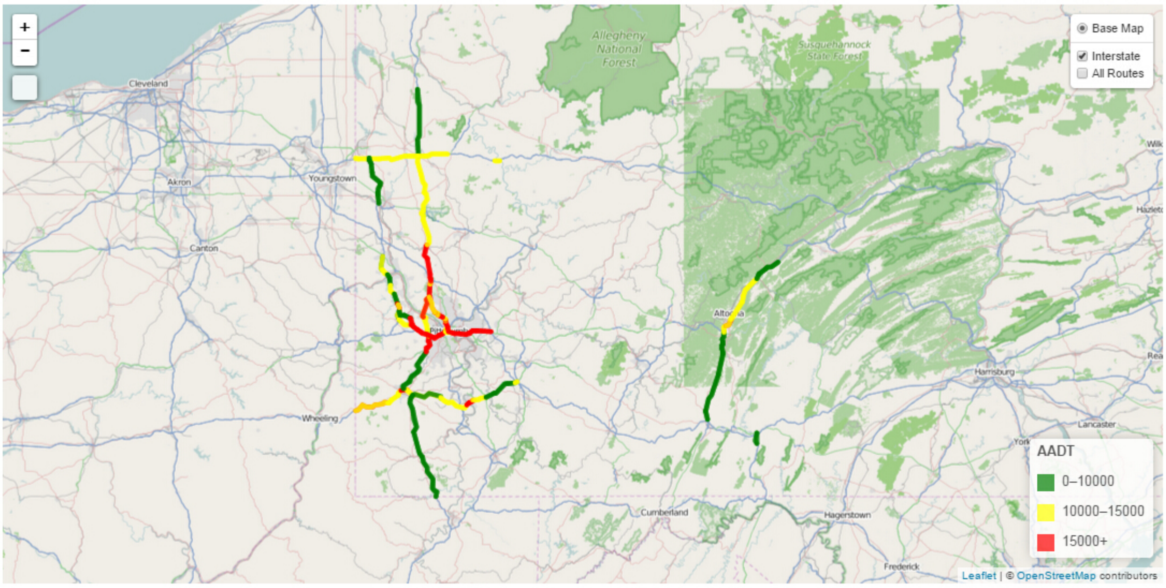

Figure 16 Annual average daily traffic (AADT) counts of interstate highways

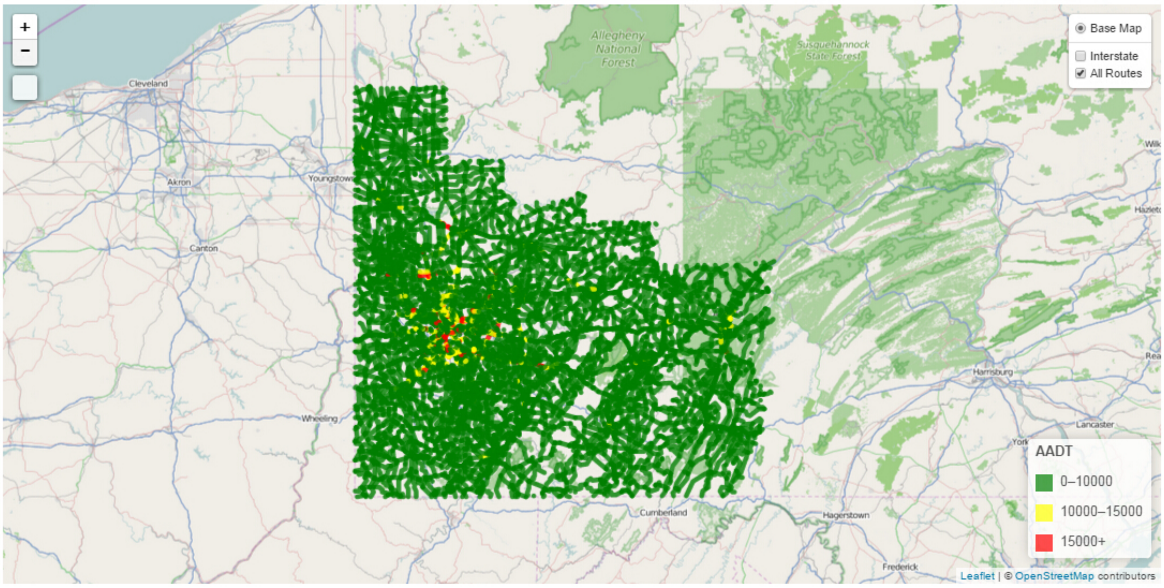

Figure 17 Traffic counts of all roads

Future Work:

We now have data regarding AADT counts. We will try to obtain counts for car and truck separately. We also want to take into account the congestion level indicated by the counts for computing optimal routes.

In addition, we are requesting time-varying traffic counts in the Pittsburgh area from SPC, which can be used to estimate time-varying origin-destination traffic demand and forecast traffic conditions in the dynamic networks for decision-making.

3.3.5. Weather

We collect the weather data of the 32 counties from Wunderground. On the web, users can obtain county level weather conditions. They can also view the weather changes of every county. In the future, we plan to add weather conditions into the trip planner we mentioned in the travel time section.

Extract:

Wunderground Weather API

Source: Wunderground, http://www.wunderground.com/
Description: Wunderground API provides a variety of weather information including real-time weather conditions and weather forecast.
Format: json or xml
Sample Response:
"FCTTIME": {

"hour": "15","hour_padded": "15","min": "00","min_unpadded": "0","sec": "0","year": "2014","mon": "12","mon_padded": "12","mon_abbrev": "Dec","mday": "6","mday_padded": "06","yday": "339","isdst": "0","epoch": "1417896000","pretty": "3:00 PM EST on December 06, 2014","civil": "3:00 PM","month_name": "December","month_name_abbrev": "Dec","weekday_name": "Saturday","weekday_name_night": "Saturday Night","weekday_name_abbrev": "Sat","weekday_name_unlang": "Saturday","weekday_name_night_unlang": "Saturday Night","ampm": "PM","tz": "","age": "","UTCDATE": ""

},

"temp": {"english": "41", "metric": "5"},

"dewpoint": {"english": "37", "metric": "3"},

"condition": "Rain",

"icon": "rain",

"icon_url":"http://icons.wxug.com/i/c/k/rain.gif",

"fctcode": "13",

"sky": "98",

"wspd": {"english": "12", "metric": "19"},

"wdir": {"dir": "N", "degrees": "2"},

"wx": "Rain",

"uvi": "0",

"humidity": "83",

"windchill": {"english": "34", "metric": "1"},

"heatindex": {"english": "-9999", "metric": "-9999"},

"feelslike": {"english": "34", "metric": "1"},

"qpf": {"english": "0.02", "metric": "1"},

"snow": {"english": "0.0", "metric": "0"},

"pop": "72",

"mslp": {"english": "30.17", "metric": "1022"}

}

US County GIS File

Source: Topologically Integrated Geographic Encoding and Referencing (TIGER), U.S. Census Bureau
Description: GIS files containing geographic information about boundaries of counties in the U.S.
Format: GIS

Transform

Description: We extracted geographic information of the Power of 32 counties from US County GIS file and formed them into geoJSON format.

Result Sample:

2.14 MB, 32 counties.

{"type":"Feature","properties":{"county":"Allegheny","api":"Allegheny County","lat":40.4697574,"lon":-79.9804515,"state":"PA","weather":{}},"geometry": {"type": "Polygon", "coordinates": an array of longitude and latitude pairs here }

The real time weather data from Wunderground API can be combined into the geoJSON records at the "weather" attribute.

Load:

Tables in database:

traffic_weather (<u>county</u>, state, api, update_time, geoJson, weather)
This table stores the information about the counties as well as the weather we get for the counties.
county (primary key, varchar(20)): the name of a county.
state (varchar(2)): the state a county is in.
api (varchar(30)): the suffix of the Wunderground api to get a county's weather.
update_time (time): the update time of a record.
geoJson(longtext): the geographic information of a county. Real time weather data can be combined with this data for front-end usage.
weather (longtext): the weather data from the Wunderground api of a county.

Front End:

Users can check temperatures, wind speeds and humidity of 32 counties. We visualize them on map with different colors. See **Figure 18**.

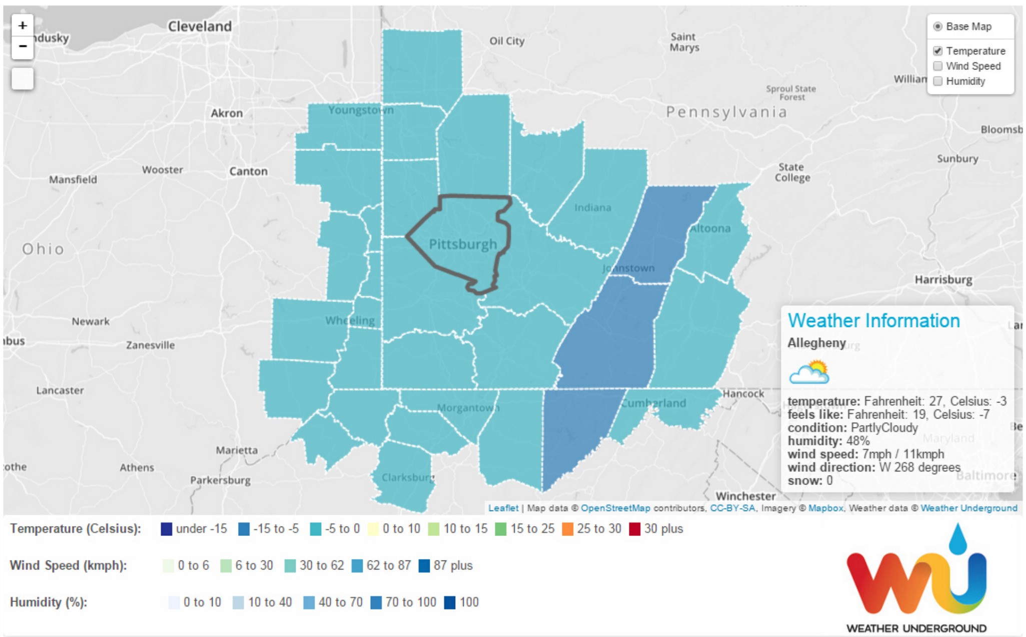

Figure 18 Visualized temperatures, wind speeds and humidity of 32 counties

We also provide a chart to show time-of-day weather forecast for the next 10 hours of any selected county. To see this, users can click on a county in the map. For example, we click the Allegheny County and the following chart of temperature over the next 10 hours is demonstrated in **Figure 19**.

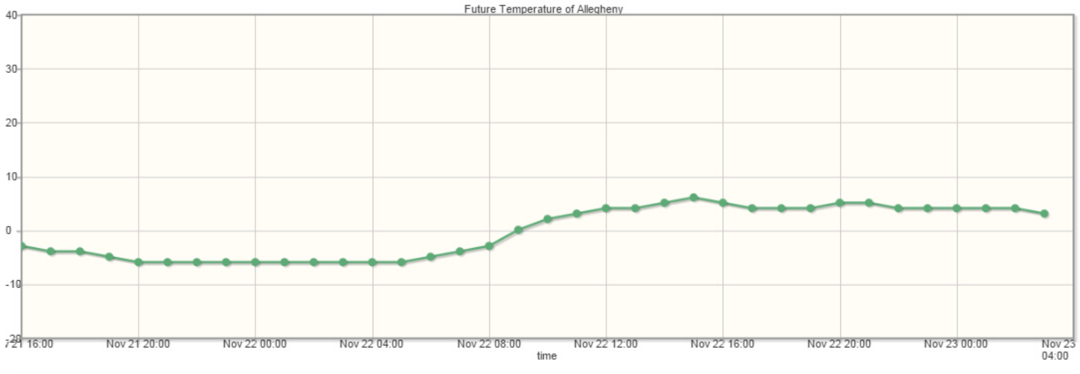

Figure 19 The weather trend of next 10 hours of a county

Future Work:

We plan to establish relations between travel time and weather (temperature, humidity, etc.) so that optimal routes are computed based on real-time weather information.

3.3.6. Cameras

Description: The data contains geographic information and URLs of CCTVs on some segments of Pennsylvania highways. On our web, we display the locations of the CCTVs using highlighting marks. Users can obtain the video images of road conditions by clicking on those marks.

Extract:

PennDOT CCTV List

Source: PA Department of Transportation
Description: The URL and locations of CCTV on the highways of PA. The photo of each CCTV is refreshed every 5 seconds.
Content: CCTV List
Format: Excel files.

Purchase records files: (Size: 112 KB, 702 CCTVs)
DISTRICT: the district a CCTV is in.
Statewide_ID: the ID that identifies a CCTV in PA.
District_ID: the ID that identifies a CCTV in the district.
Descriptive Location: the location of a CCTV.
MILEMARKER: the mile marker of a CCTV's location on a route.

COUNTY: the county a CCTV is in.
STATE ROUTE: The ID of the state route a CCTV is on.
LATITUDE, LONGITUDE: the latitude and longitude of a CCTV's location.
URL: the URL to get the photo of a CCTV.
511 Description: the description of a CCTV in 511 project.
511 Location: the city a CCTV is in.
511 Route: the route name a CCTV is on in 511 project.

Integrity: Missing URL for 33 CCTVs. Data show "URL AWAITING VIDEO SHARING SOLUTION" for those 33 CCTVs.

Transform:

Description: We convert each record in CCTV List into geoJSON format for front-end usage.
Result Sample:
Size: 328 KB, 702 CCTVs.
{"type":"Feature","properties":{"DISTRICT":"2","Statewide_ID":"CAM-02-001","District_ID":"CCTV 2","Descriptive Location":"SR 0322 ByPass Westbound","MILEMARKER":"999.9","COUNTY":"CENTRE","STATE ROUTE":"0322","URL": url here,"511 Description":"US-322 w/o I-99","511 Location":"State College, PA","511 Route":"US 322","SortValue":"50",},"geometry":{"type":"Point","coordinates": coordinates here}}

Load

Tables in database:

traffic_CCTV (statewide_id, district, district_id, descriptive location, milemarker, county, state route, latitude, longitude, url, 511 description, 511 location, 511 route, sortvalue)
This table stores data of CCTV List file.

Other files:
CCTV_geoJSON.js:
This file stores CCTV information processed into geoJSON format and can be directly used by the front end.

Front End

We use markers to indicate the locations of all CCTVs. Users can click on a marker to obtain the video images (traffic condition) of any selected CCTV as seen in **Figure 20**.

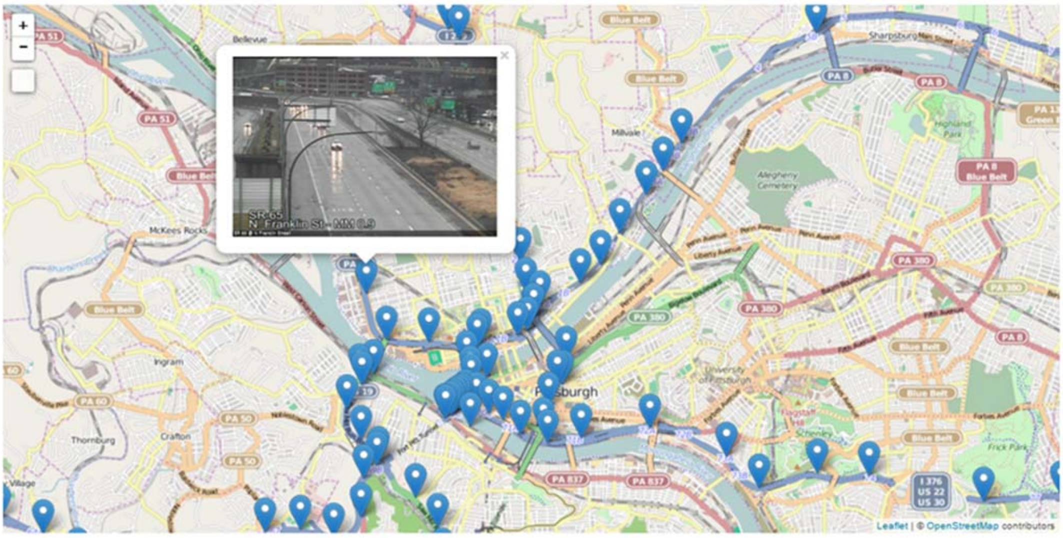

Figure 20 Markers indicating all the locations of all CCTVs

Future Work:

We will provide HD video streams of the CCTVs in the future that are updated by PennDOT. Through the video streams, we will be able to extract traffic flow characteristics data such as traffic counts, speeds, etc.

3.3.7. Electric Vehicle Charging Stations

The electric vehicle (EV) stations data contain names and addresses of all existing and planned EV charging stations in Pittsburgh area. We geocoded the addresses to obtain the geographic information of the stations and display them on our web. In the future, we plan to recommend optimal routes for EVs with limited battery traveling ranges.

Extract:

Electric Vehicle (EV) Charging Stations

Source: Pittsburgh Region Clean Cities, http://www.pgh-cleancities.org/
Description: The name and address of EV charging stations in Pittsburgh area, including public, private and planned stations.
Content: Electric-Vehicle-Charging-Stationsupdated-102520131.
Format: docx files.
Size: 447 KB, 76 records.

Planned Station Names	Address
City of Latrobe Parking Garage(1)	Weldon & Spring St, Latrobe, PA 15650
Private Station Names	**Address**
Montose Nissan Service Center	1455 N. Hermitage Rd, Hermitage, PA 16148
Public Station Names	**Address**
Montose Nissan	1455 N. Hermitage Rd, Hermitage, PA 16148

Transform:

Description: We used the Google Geocoding API (Google, 2014) to get the geographic information (latitude, longitude) of each station and then converted each record into geoJSON format. The details about the Google Geocoding APT can be found at https://developers.google.com/maps/documentation/geocoding/

Result Sample:

Geocoding:

Station Names	Address	type	number	lat	lng	comments
Greensburg Museums	206 Main St, Greensburg, PA 15601	1	2	40.3053839	-79.5458240	
Mark Arbuckle Nissan Service Center	1080 Philadelphia St, Indiana, PA 15701	2	1	40.6230287	-79.1601883	

type: the type of a station; 1-planned; 2-private; 3-public.
number: the number of chargers at a station.
lat,lng: the latitude and longitude of a station.
comments: additional information about a station
GeoJSON:
{"type":"Feature","properties": {"station":"City of Latrobe Parking Garage","address":"Weldon & Spring St, Latrobe, PA 15650","number":1,"type":1},"geometry": {"type": "Point", "coordinates": [-79.3848263 ,40.3177825]}},

Load:

Tables in database:

traffic_EVstation (id, name, address, type, number, lat, lng, comment)
This table stores the data after geocoding.

Other files:
EVstation_geojson.js:
This file stores EV station data processed into geoJSON format and can be directly used by the front end.

Front End:

We use circle markers to show the locations of all EV charging stations. Users can click on markers to get detailed information. See **Figure 21**.

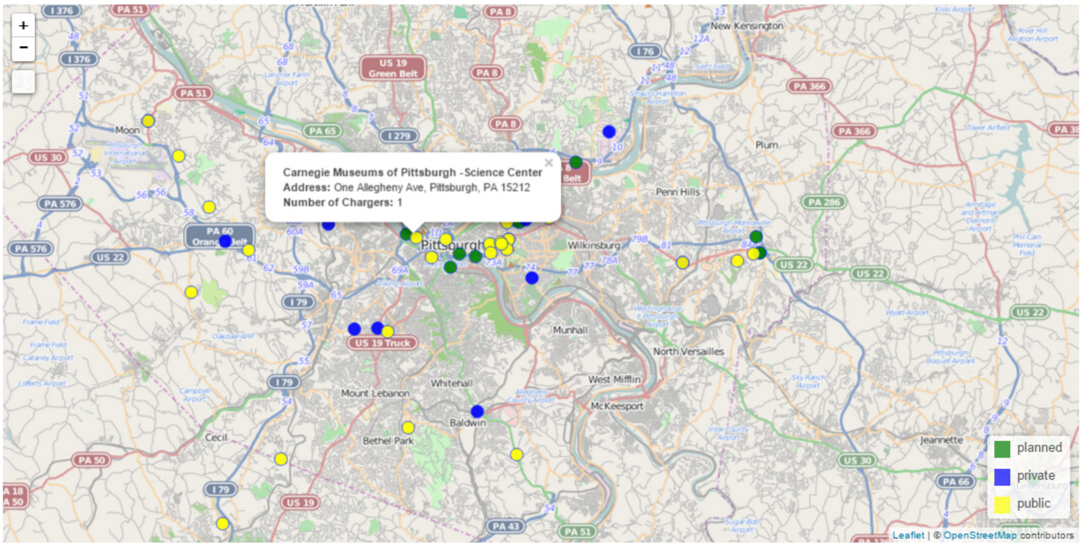

Figure 21 Circle markers to show the locations of EV stations

Future Work:

In the future, we plan to recommend optimal routes for EVs with limited battery traveling ranges. We also plan to acquire data of other fuel stations like natural gas stations, biofuel stations, hydrogen charging stations, etc. This data can help route special vehicles more efficiently.

3.3.8. Incidents

We have data of incidents from several data sources, such as PennDOT Crash, PennDOT 511 RCRS, and Allegheny County 911 Traffic Incident Data. Our purpose is to combine data from those three data sources and build an integrated incident database for the Power of 32 Region.

Extract:

Source: PennDOT, County 911
Description: As mentioned above, our data is from 3 data sources, which are PennDOT Crash, PennDOT 511 RCRS, and Allegheny County 911 Traffic Incident Data. And the data formats from these sources vary substantially.
Format: csv files, access database, excel files.
Data files:
PennDOT Crash:

Data from PennDOT is mainly about crashes on the highways, and it contains several attributes to describe a single crash incident, such as date, time, the district and county where crash took place, the coordinate of the incident, and so on.

CRN	DISTRICT	COUNTY	MUNICIPALI'	POLICE_AGC'	CRASH_YEAR	CRASH_MONTI	DATE_OF_MOI	DAY_OF_WEEI	TIME_OF_DA'	HOUR_OF_DA'	ILLUMINATI(	WEATHER	RO
2012000005	06	09	09101	09101	2012	01	01	1	0050	00	3	1	0

Size: 2.6GB, 369401 records

PennDOT 511 RCRS:

The attributes used to describe an incident are district, county, coordinate, date, time, incident cause and so on.

EVENTID	ST_RT_NO	SR	JURIS	DIRECTION	DISTRICT_	CTY_CODE	COUNTY_I	DISTRICT_	CTY_CODE	COUNTY_I	CAUSE	STATUS	DATETIME	DATE_TIM	DATE_TIM	ACT_DATE	ACT_DATE	RECCREAT	RECCREAT	LASTEDITE	LASTEDITE	INTERSEC	LC
29115	611	0611 (PA -	1	BOTH	6	9	BUCKS (09	6	9	BUCKS (09	ACCIDENT	CLOSED	20100101_	01/01/201	2.01E+11			01/01/201	2.01E+11	01/01/201	2.01E+11	9	40

Size: 71.3MB, 250726 records

Traffic Incident Data from 911:

An incident is described by attributes such as type, address, dispatch date, dispatch time of police, and so on.

CallNumber	Zone	CallType	Call_Final	Priority	Disposition	Unit	Address	DispatchDate	DispatchTime	Month	Day_of_Wee
130000240	Zone 5	TRAFFIC - WITH INJURIES (UNK # PATIENTS)	Traffic Incident	2	Unable to Locate	C3529	BUTLER ST/WASHINGTON BLVD, PGH	1-1月-13	1:12:59 上午	January	Tuesday
130000294	Zone 2	TRAFFIC-HIGH MECHANISM (AUTO-PEDESTRIAN)	Traffic Incident	1	No Report Needed	C3224	OLIVER AVE/GRANT ST, PGH	1-1月-13	1:28:30 上午	January	Tuesday

Size: 721KB, 8763 records

Transform:

The main task of integrating data from all three data sources is to check whether there are any duplicate records reporting the same incidents among three data sources. Since all those data sources record incidents by date, time, and coordinate (or address), we can check the proximity of these attributes.

Pre-process:

Because the representations of coordinates (or addresses) in all three data sources are different, we have to pre-process the coordinates (or addresses) data to uniform them.

For instance, in PennDOT Crash database, the coordinate is represented in six decimal, such as 40 10:03.325, 74 50:50.311; in RCRS, the coordinate is represented in decimal, such as 40.5517341975725,-75.1818159867978; Traffic Incident Data from 911 uses addresses instead of coordinates, which is represented as the cross of two streets, OLIVER AVE/GRANT ST, PGH, or the house number, 1311 SPRING GARDEN AVE, PGH.

For simplicity, we decide to transform coordinates or addresses into decimals, in other words, we transform the location data in PennDOT and 911 to the formats of location data in RCRS.

We use the Google geocode API to convert the address data in 911 into decimal coordinates.

Sensitivity Analysis:

Now we have data from all three data sources, which are PennDOT Crash, PennDOT 511 RCRS, and County 911. What we need to do next is to compare data from different sources and check whether they are representing an incident reported by other database. To do so, we concern four attributes: Date, Time, Type(if mentioned), and location.

We define the following: for record A from data source I, and record B for data source II, A and B are describing the same incident if and only if their dates are the same, the discrepancy of their time of day is less than one hour, and the L2 norm of their geographic coordinates is less than 0.004. Those parameters are set using the following methods.

Parameter selection:

Time of day:

Incidents data in PennDOT may be derived from sensors, and the incidents data from 911 is derived from the phone calls to 911. Thus, for the same incident, the record time may vary substantially among different sources. In this case, we set the time discrepancy to be no more than one hour.

Spatial difference:

Because the data in PennDOT Crash and RCRS are both described by district and county, we consider these two attributes when checking duplicates.

First we extract data in PennDOT Crash and RCRS that contain the same district, county, date, and the difference of time is less than one hour, and we call these constrains 'Constrains *' below.

Take data from 2012 in PennDOT and RCRS for instance. In 2012, there are 123,671 records in PennDot, and 21,564 records in RCRS. Among these data, there are 512 data which satisfy Constrains *.

1	year	month	day	time_rcrs	time_penn	district	county	lat_rcrs	lng_rcrs	lat_penn	lng_penn	l2dis
2	2012	12	27	337	315	10	16	41.19845	80.1539	41.19583	79.5042	0.64971
3	2012	7	26	2200	2230	12	64	40.26556	79.12784	40.63189	79.64933	0.637301
4	2012	10	29	2000	2026	12	64	40.39655	79.15067	40.59178	79.73239	0.613607
5	2012	11	7	655	558	10	10	41.17106	79.71969	40.69886	80.10118	0.607046

The amount of data against spatial difference is shown below in **Figure 22**.

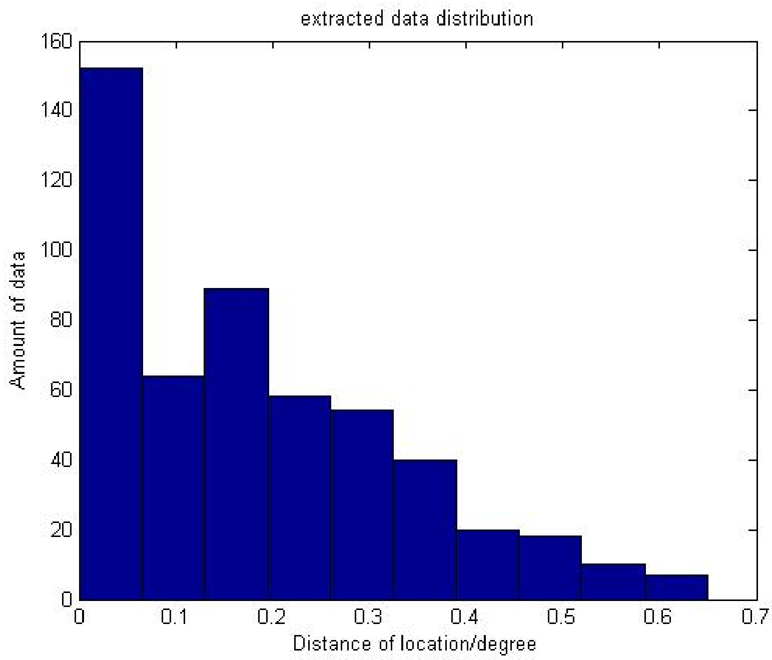

Figure 22 Incident data's distribution

As we know, the radius of earth is 6378.1km. Thus, 1 degree represents a 111.32km-distance on earth. Suppose that we accept a 5km-error for the location record, which means for the data satisfy Constrains *, if their location is less than 5km, we regard them as the same incident. Convert it to degree, we accept a 0.05-degree-error for l2norm of coordinates in RCRS and PennDOT. These data are in the first rectangle from left in figure above. The amount of them is 134.

Then we apply the same constraints to data from 911 and from the combination of RCRS and PennDOT, and obtain the integrated incidents database.

Load:

Tables in database:

incident (ID, Date, Time, District, County, IncidentType, Latitude, Longitude, DataSource, RecordNumber)

Front End:

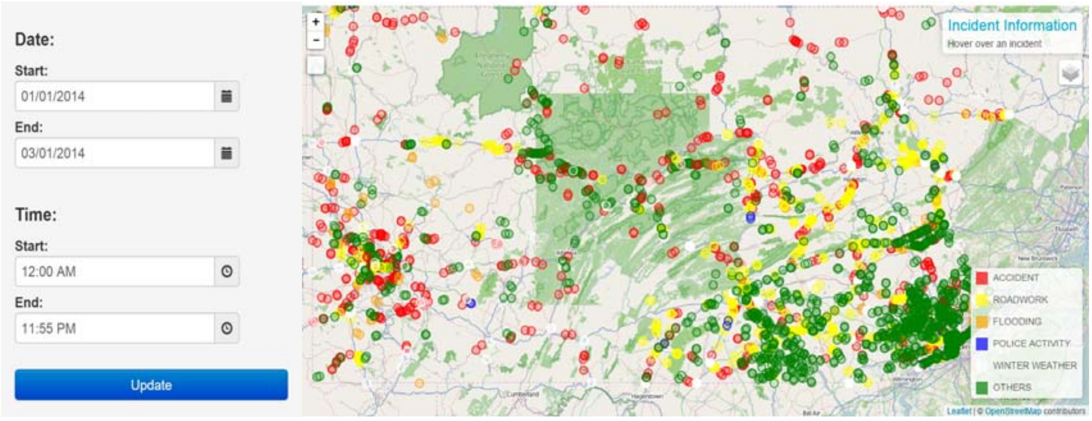

Figure 23 Front end of Incident function

We plot the incidents on the map, and use different colors to distinguish different incident types (see **Figure 23**). On the left of screen, the user can choose a period of time of their concern and show incidents in this period on right side.

Future Work:

Database: We may have more data in the future, thus we need to keep updating our database. We will improve the method of matching incident coordinates to GIS maps to make the geographic visualization more accurate.

In addition, the probability of incidents on all roads can be computed to obtain incidents hotspots for safety management. Real-time incident data access to RCRS also allows us to route individual travelers efficiently, as well as to make effective traffic management strategies for agencies.

4. Conclusions and future work

We have developed a prototype Power of 32 Regional Traveler Information System (RTIS) that provides travel information which is the most promising and most readily available, such as: roadway traffic volume and travel time, traffic video, public transit, parking, incidents, weather, and electrical vehicles charging stations. The primary goal of the prototype system is to collect various types of traffic data in the Power of 32 Region and to demonstrate its effectiveness in facilitating smart decision making among travelers, private firms and public agencies in the Power of 32 Region.

Imagine that travelers in this region can access full information through a Regional Traveler Information System (RTIS) that delivers on-demand trip information and recommendations, while public agencies can access the RTIS as a data repository for predictive modeling, ongoing planning, and system management. A regional RTIS would efficiently improve safety, incident management, and emergency management through, for example sharing information between Traffic Management Centers and emergency management partners, or deploying a Road Weather Information System. RTIS has the potential to build off and leverage many pieces of the existing Intelligent Transportation Systems (ITS) infrastructure and transportation management strategies in the Power of 32 Region.

Establishing a fully functional RTIS can take a few years. In this year one, we have developed a prototype, which successfully demonstrated that the Mobility Data Analytics Center, the engine for RTIS, can be a viable tool for integrating massive data and applying analytics to decision-making for travelers, private firms and public agencies. In addition, a Capstone project in Fall 2014 at Heinz College provided possible structural and financial models for the RTIS (see the Capstone project report attached separately). Our Power of 32 Region could develop a RTIS if data can be gathered and consolidated and there is political will. The Power of 32 should work with the state DOTs and local transportation providers to determine how they would like to share their data for a RTIS as a process of planning and operational management. The Power of 32 should also decide whether to move forward with the RTIS.

Future work: a trip planner for travelers

Ultimately, the RTIS should be able to understand customer needs and desires, and to have the capability to personalize the information provided to travelers. The best way for RTIS to understand what routes and modes are of interest to customers is to incorporate a trip planner into the service by taking into account multi-modal, multi-jurisdictional data.

With a trip planner (Google Maps is a prominent example), the traveler specifies from-location and to-location, and sometimes also specifies the travel mode to be used, and the system informs the user as to the best route(s) to take to minimize travel time. The trip-planner could also be programmed to be able to minimize travel cost at the user's request, or be optimized to minimize parking costs, etc. Furthermore, the trip planner can be integrated with the Integrated Corridor Management system recommendation of this study.

By having the origin-destination information, the RTIS will be able to inform the customer of travel time and delay, as well as to advise the customer of alternate routes and modes. This traveler information can be integrated with additional information, such as the availability and cost of parking at the chosen, traffic impact of real-time incidents, etc. The trip planner can also be integrated with the toll payment system and the transit's fare payment system.

In this manner, the RTIS will be able to provide the customized information needed by the traveler, and at the same time, the system will have a better understanding of travel patterns in the 32-county region.

A capstone project in Fall 2014 at Heinz College has conducted a survey to interview travelers. The survey results indicate the following are recommended as means through which to disseminate traffic information to the users of the system.

(1) Web interface:

The Internet can be considered as perhaps the most flexible means of providing information to end-users. Such an interface will be able to provide all data including weather, traffic, routes, costs, incidents, camera feed, history, parking information etc. The web interface provides excellent visual input for users, when the user is able to access such visual cues (such as while not driving). It will also provide the extensible features that would be replicated on the mobile app.

The customized information afforded by the trip planner described above can be displayed via the web interface.

(2) Mobile app

The mobile app, which should be available on as many mobile platforms as possible, will provide a convenient and quick reference for end-users. In transit usage (in-compliance to NHTSA regulations) for weather, traffic, routes, costs, incidents, parking information etc. While perhaps not as comprehensive as the web interface due to limitations of screen size, the mobile app will be very extensive in its own right. The mobile phone's GPS information provides locational information for the trip planner. This is especially helpful to travelers on foot, bicycle, the access and egress portion of the transit or airplane travel, to supplement the locational capability of the vehicles on which the customer will be traveling for all or a portion of their trip.

In addition, when it comes to regional traveler information systems, rural areas are sometimes overlooked while the entire focus remains on urban centers. Many agencies did not recognize until recently the need or potential benefits to providing traveler information in rural areas. Providing traveler information in rural locations has proven to be very valuable in terms of reduced user delay and safety benefits. The economic importance of moving goods and services to serve farming in rural areas is further enhanced by the tremendous expansion of natural gas extraction from fracking; there is need for trucking to serve this burgeoning industry, both to provide materials used in the process and to handle produced water and other byproducts of the process. The need for such information to the public is also important in order for travelers to avoid construction congestion during summer months and to travel safely during the winter months.

Future work: a trip planner for freight transportation firms

Analogous to the trip planner for travelers, truck drivers specifies from-location and to-location, and their desired time of arrival (delivery time), and the system informs the users the best route(s) that is feasible to trucks and minimizes travel time and cost. The trip planner should take into account 1) transportation constraints (such as height, width, weight restrictions, parking) information and 2) traffic information for heavy-duty trucks, both are not of concern for individual travelers. It is critical to communicate with regional trucking companies to better understand their demand and needs.

Future work to facilitate decision making for public agencies

In order to serve decision making for agencies, the Mobility Data Analytics Center will build a sophisticated transportation network model for the region that describes individual travel activities on roadway systems, transit systems and parking systems. Individual trips can be modeled and simulated in the network with full details, when, where and how to travel. The multi-modal network model is the key to systematic planning and operations of transportation infrastructure. Operational strategies and transportation policies can be fully examined in the network model in terms of system delay, reliability, vehicle-miles traveled (VMT), fuel consumption and emissions. Ultimately the RTIS provides not only system performance information to agencies (as demonstrated in the prototype), but more importantly, qualitative and quantitative assessment of transportation management strategies.

References

Agafonkin, V., 2014. *Leaflet.* [Online]
Available at: http://leafletjs.com/
[Accessed 27 November 2014].

Anon., 2006. *511 Deployment Coalition.* [Online]
Available at:
http://www.itsknowledgeresources.its.dot.gov/ITS/benecost.nsf/ID/F2A8B11B82686B44852577620069ED77?OpenDocument&Query=LLFocusArea
[Accessed 21 November 2014].

Anon., 2010. *I-95 Corridor Coalition.* [Online]
Available at:
http://www.i95coalition.org/i95/Portals/0/Public_Files/uploaded/Vehicle-Probe/VP%20Project%20benefits%20General%2012%20August%202010_FIN.pdf
[Accessed 24 November 2014].

Behrens, M., 2012. *The Django Book.* [Online]
Available at: http://django-book.readthedocs.org/en/latest/index.html
[Accessed 25 November 2014].

Django, 2014. *Django overview.* [Online]
Available at: https://www.djangoproject.com/start/overview/
[Accessed 25 November 2014].

Ezell, S., 2014. *Bringing U.S. Roads into the 21st Century..* [Online]
Available at: http://issues.org/26-4/ezell/
[Accessed 7 December 2014].

Google, 2012. *What is GTFS?.* [Online]
Available at: https://developers.google.com/transit/gtfs/
[Accessed 16 December 2014].

Google, 2014. *The Google Geocoding API.* [Online]
Available at: https://developers.google.com/maps/documentation/geocoding/
[Accessed 27 November 2014].

J. Tusuf, C. W. S. K., 2006. *Evaluation of the Eastern Kentucky Rural Highway Information Project 511 Tourism Service.* [Online]
Available at: http://ntl.bts.gov/lib/jpodocs/repts_te/14307_files/14307.pdf
[Accessed 23 November 2014].

OpenStreetMap, 2014. *OpenStreetMap.* [Online]
Available at: https://www.openstreetmap.org/about
[Accessed 15 December 2014].

Oracle, 2014. *Why MySQL?.* [Online]
Available at: http://www.mysql.com/why-mysql/

[Accessed 15 December 2014].

USDOT, 2001. *US DOT Federal Highway Administration..* [Online]
Available at: http://ntl.bts.gov/lib/jpodocs/repts_te/13477.html Accessed
[Accessed 14 November 2014].