DRIVE Net: An E-Science Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis

Xiaolei Ma
Graduate Research Assistant
Box 352700
Department of Civil and Environmental Engineering, University of Washington
Seattle, WA 98195-2700
Tel: (518) 618-6775
Email: xiaolm@u.washington.edu

Yao-Jan Wu
Graduate Research Assistant
Box 352700
Department of Civil and Environmental Engineering, University of Washington
Seattle, WA 98195-2700
Tel: (206) 685-6817
Email: yaojan@u.washington.edu

Yinhai Wang, Ph.D. (Corresponding author)
Associate Professor
Box 352700
Department of Civil and Environmental Engineering, University of Washington
Seattle, WA 98195-2700
Tel: (206) 616-2696
Fax: (206) 543-1543
Email: yinhai@u.washington.edu

ABSTRACT

Over the past decades, transportation research has been mathematical-equation driven or relied on scarce data. Data-driven or data-based research is expected to expand in the near future with the increasing data collected from intelligent transportation system sensors. So far, most online systems are designed for handling one type of data, such as freeway or arterial sensor data. Even if transportation data are ubiquitous, data usability is difficult to improve. This study proposes a framework for region-wide web-based transportation decision system which adopts digital roadway maps as the base and provides data layers for integrating multiple data sources, including traffic sensor data, incident data, accident data, travel time data, etc. The system is named the Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net). DRIVE Net provides a practical solution to facilitate data retrieval and integration, and enhances data usability. Moreover, DRIVE Net offers a platform for optimizing transportation decisions that also serves as an ideal tool for visualizing historical observations both spatially and temporally. DRIVE Net can be used not only as a practical tool for various transportation analyses but also to help evaluate the benefit of a specific transportation solution using its online computation engine. Based on the current implementation, this system demonstrates its potential to act as a standard tool to incorporate more datasets from different fields, such as health and household data, and offers a real-time decision making platform in the near future.

.

INTRODUCTION

Over the past decades, transportation research has been mathematical-equation driven and relied on scarce data to develop mathematical models and traffic theory (e.g., (*1*), (*2*)). For example, many well-known theoretical models were developed based on a small portion of data but they have been widely used in practice (*3*) (*4*). When the model development is extended to the network level, data availability may reduce further or simply disappear and these theoretical models can often be verified only by simulation data (e.g., (*5*), (*6*)). Options did not seem to exist although researchers knew that the simulation and mathematical models only can capture some of the "facts," many contributors, especially human factors, are not easily reflected in the simulation results.

With the advance of data collection technologies and their deployments in Intelligent Transportation Systems (ITS), transportation data availability has been increasing tremendously over the past years. As a future type of traffic management system, IntelliDrive[TM] (http://www.intellidriveusa.org/) is also quickly gaining in popularity and increasingly deployed. Since IntelliDrive[TM] enables vehicle to vehicle and vehicle to infrastructure communications on a frequent basis, traffic data are expected to explode in the years to come. Therefore, data-driven or data-based research shall expand and play an increasingly important role in the near future. The rich data sets will enable validation of previously developed transportation theories and boost scientific discoveries on transportation planning, system operations, and travel behaviors.

Even though traffic sensor data have been broadly collected and archived. Data accessibility and usability is unsatisfactory for the following reasons: 1) Transportation data are typically managed independently by various agencies and stored in various types of systems. Each system has a proprietary interface and varying capabilities in data processing. 2) Extensive communication efforts are needed to acquire data from other agencies. 3) Data stored in different domains or formats require additional effort and knowledge to standardize the format so that it can be interpreted and used by all data users. Such data retrieval and exchange barriers have hindered the progress of scientific discovery of solutions to traffic-related issues.

In the past decade, various web-based Archived Data User Services (ADUS) systems and Advanced Traveler Information Systems (ATIS) were developed in attempts to increase the data exchangeability and usability. For example, U.C. Berkeley has been focusing on developing the online Freeway Performance Measurement System (PeMS) since 1997. PeMS is capable of analyzing freeway traffic sensor data and providing real-time performance measures, including travel times (*7*). Based on the successful experience of using PeMS on Freeways, Arterial Performance Measurement System (APeMS) was also developed to estimate the travel time on an arterial route using mid-block (system) loop detectors (*8*). These two systems are typical ATIS with support from ADUS. The Portland Oregon Regional Transportation Archive Listing (PORTAL) system was initiated by Bertini et al. (*9, 10*). PORTAL primarily focuses on archiving and analyzing freeway data like PeMS. The most representative system by far found in the literature review is the Regional Integrated Transportation Information System (RITS) developed by the University of Maryland. The RITS is a user-friendly multi-agency system for data sharing,

dissemination and archiving (11). This system integrates multiple data sources from different transportation agencies and has a main focus on freeway applications. In addition to RITS, the University of Maryland also developed a variety of transportation data visualization and analysis platforms (12 - 17), each designed for a specific purpose, such as incident analysis. As for other ATIS, Seymour and Miller (*18*) utilized the Google Map Application Programming Interface (API) to provide a mapping system for freeway speed, incident information, and camera image etc for the Dallas transportation management center (DalTrans). Wu and Wang (*19, 20*) implemented a Google-Map-Based arterial traffic information (GATI) system for real-time traffic condition visualization and online arterial traffic data analysis using the City of Bellevue's intersection loop and traffic signal timing data. However, most of these systems are primarily based on a single data source and serves as a traditional online data or online traffic information provider. Despite of the needs from various transportation-related agencies for online systems to share and analyze transportation relevant data, few such systems were developed with the functions of data format standardization, regional map-based data visualization, and interactive online traffic analysis, with consideration of the interactions between heterogeneous data. For example, the impacts of freeway incidents on arterials and freeways were not covered in previous research due to the lack of an explicit architecture to bridge the gaps between heterogeneous data from multiple transportation agencies.

Therefore, the goal of this study is to develop an e-science of transportation platform for data sharing, visualization, modeling, and analysis. The term "e-Science" was created by Dr. John Taylor of the UK Office of Science and Technology in 1999 (*21*). E-science refers to computationally intensive science that needs to process immense data sets using highly distributed computational resources connected by the Internet. E-science approaches have a great potential to solve some tough transportation issues. However, it has been a slow process for the transportation communities to accept this new concept. The new platform intends to take advantage of e-science developments for data-driven transportation research and applications.

The new platform is named Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net). DRIVE Net is expected to remove data accessibility barriers, allow easy access of real-time regional traffic information, facilitate data sharing and visualization, enable online data analysis for scientific discovery and decision support, and offer opportunities for early stage e-science of transportation investigations. This platform will benefit not only regular road users but also transportation practitioners and researchers. Compared with previous systems, DRIVE Net is not a simple data visualization and archiving system. DRIVE Net not only enables the connections and interoperability among the heterogeneous data sets but also serves as a data-rich visual platform to facilitate scientific discoveries and educational enrichments in the areas of transportation engineering and planning, environmental engineering, and public health science. The design for DRIVE Net clearly considered the need to support future endeavors in e-science of transportation..

The remainder of the paper is organized as follows. DRIVE Net's system architecture is first described, followed by a brief introduction to system design and implementation. Next, a case study of incident impact analysis is presented to demonstrate the utility of the system and to

illustrate data interactions between different analytical modules in DRIVE Net. Then, the potential benefits of this research are discussed. Lastly, the paper will be concluded with future visions and challenges.

SYSTEM ARCHITECTURE

The design of DRIVE Net is critical for its future performance and scalability. The current design shown in Figure 1 reflects our current understanding of the platform and future expectations to DRIVE Net. We intend to make it open architecture and open source so that system design can be continuously improved with expansion of system capabilities. The current system architecture is primarily composed of three parts: heterogeneous data sources from different agencies, data warehouse in the STAR Lab, and web services running on the DRIVE Net system server.

Data Warehousing and Retrieval

The data warehouse is responsible for data archiving with multiple data retrieval functions supported by the DRIVE Net system. Data retrieval is challenging because every agency has its internal policy and security concerns. Relying on one single uniform data retrieval method for each agency may be infeasible and inapplicable. Moreover, data archiving formats in all agencies vary, even within the same agency; data may still follow different patterns. Standardization of data formats is highly beneficial for transportation agencies and data users. However, few guidelines have been developed for data exchange and standardization. Some standards, e.g. National Transportation Communications for ITS Protocol (NTCIP) (*22*), only focus on data communication standardization, not the data exchange and storage formats. Hence, four data retrieval methods are proposed and currently used for data retrieval in DRIVE Net. These methods and data examples are listed as follows:

1) Traditional flat file exchange: Flat files are the most common data exchange format commonly used among all agencies. In this way, data quantity, property and privacy can be carefully controlled. However, this method is also less efficient and more time-consuming. Once these files are retrieved through the physical media, e.g. CD-ROM, or e-mails, these files can be uploaded into the database through the DRIVE Net website or using the Structured Query Language (SQL) import function. Below are two examples of data obtained through flat file exchange*:*
   - *Washington Incident Tracking System (WITS):* Most incidents happening in Washington major freeways and state highways are logged in the WITS database in Washington State Department of Transportation (WSDOT). The WITS datasets are disseminated in flat files (EXCEL) and imported into the DRIVE Net incident database. Detailed incident information, such as incident geospatial location, notification time, clearance time, is stored in the incident database.
   - *Highway Safety Information System (HSIS)* (*23*): Upon users' request, the HSIS provides different types of data for Washington State highways, including the accident,

roadway inventory, traffic volume, curve and grade and interchange/ramp data. The data are stored in the DRIVE Net accident database.

2) Passive data retrieval: DRIVE Net is equipped with customized C# or Java computer programs that are scheduled to fetch the remote data in a predefined interval via File Transfer Protocol (FTP), Hypertext Transfer Protocol (HTTP) or Simple Object Access Protocol (SOAP). This method is considered the most convenient and efficient way to periodically retrieve data from remote servers. The data can be imported into the database following the schema design by the research team at the Smart Transportation Applications and Research Laboratory (STAR Lab) of the University of Washington (UW). Examples:

   - *Freeway loop sensor data:* The Washington State Department of Transportation (WSDOT) operates more than 7000 Inductive Loop Detectors (ILD) along freeways in Washington State (*24*). WSDOT shares its 20-second aggregated single loop data via an FTP website. The data are automatically fetched by the data download module every 20 seconds and stored in the DRIVE Net freeway database.

   - *Arterial data:* The City of Bellevue, WA has more than 500 advance loop detectors at more than 177 signalized intersections (*19, 20*). The controllers at these intersections send the cycle-by-cycle real-time traffic data (e.g. volume, occupancy and timing plans) back to a FTP server in City's Traffic Monitoring Center (TMC) and stored as Comma-Separated-Value (CSV) file every minute. DRIVE Net has been fetching the CSV file and importing the data into the DRIVE Net arterial database since 2007.

   - *Trucking Data:* Global Positioning Systems (GPS) used by trucking companies are a source of truck probe data for freight performance measurement. WSDOT, UW and the Washington Trucking Associations (WTA) have partnered to collect and analyze GPS truck data from commercial in-vehicle fleet management systems used in the central Puget Sound region. (*25*) Data are being collected from three vendors, with various resolutions, ranging from one to 15 minute frequency. DRIVE Net automatically fetches and imports theses data into DRIVE Net truck database via FTP.

3) Active data retrieval: Some agencies may have internet security concern and limited public access. The STAR Lab provides a satellite server with hardware, software, and data processing tools pre-installed. Using a build-in custom service program in the satellite server, the data can be securely "pushed" back through a firewall to the STAR Lab data warehouse using Open Database Connectivity (ODBC). This is more expensive but more secure solution to transmitting the data. Example:

   - *Intersection detector event data:* Second-by-second event data are collected from all video sensors at the intersection at 196th Street and SR99 in the City of Lynnwood, WA. The data are stored in the STAR Lab satellite server and concurrently pushed through firewalls to the DRIVE Net Intersection Performance database.

4) Direct data archiving: The data can be collected directly from the data collection devices. The data can be sent directly and periodically to the data warehouse from the test site.

Example:

- *Route travel time data:* Bluetooth-based travel time detectors developed by the STAR Lab can effectively collect route travel times by matching the unique Median Access Control (MAC) address at various locations (*26*). This device is able to transfer data using General Packet Radio Service (GPRS) and Global System for Mobile Communications (GSM) communication protocols in real time. The data are sent directly back to the DRIVE Net Bluetooth travel time database every five minutes.

For the databases mentioned above, the schemas have been designed in advance to ensure data management and query efficiencies. The relational data model (*27*) is used in the design. All kinds of transportation data can be systematically stored in the DBMS and the relationships between the attributes (columns) can be easily maintained following the designed schema.

DRIVE Net Web Server

The core DRIVE Net system lies in the web server running Apache Tomcat 6.0 in the Windows Server 2008 Operating System (OS) environment. This server can render/disseminate the data and execute analytical algorithms depending on the role of users. Traffic engineers, researchers and travelers are three users groups expected to use DRIVE Net. For example, certain downloading functions are limited to certain user groups. As illustrated in Figure 1, DRIVE Net can be connected to multiple data servers using different data communication techniques. When necessary, another server can be added to the system as well. DRIVE Net servers will work jointly like terminals in the grid computing infrastructure.

SYSTEM DESIGN

The DRIVE Net system is developed based on a multitier architecture model, commonly used in software engineering (*28, 29*). The major merit of the multitier architecture is that the developers can modify or add a specific tier without rewriting the entire application. The model being used consists of a client-side presentation tier (client side web browser), a server-side data tier (data warehouse), and two server-side logic tiers (middleware and computational module). In addition to the traditional three-tier client-server model, an additional logic tier is added to handle data quality issues. The computational tier is used for data sharing control and algorithm execution. The middleware tier is designed to mitigate the burden in the computational tier. The burden is usually caused by excessive access to database, analytical algorithm calculations, and data quality control. The presentation tier is on the client side web browsers, and used for displaying interfaces and visualizing outputs, and receiving inputs from users. The overall system flow chart is shown in Figure 2:

Data Quality Control

Data quality is an issue that is widely recognized by transportation researchers and agencies. Developing an automatic and robust data quality control (DQC) procedure is beneficial to facilitate transportation-related research. To insure the quality data, DRIVE Net incorporates a two-step DQC mechanism handling data cleansing tasks, including error detection, removal and inconsistencies, etc (*30*). The first step of data cleansing service happens at the stage of data retrieval from different data sources. For example, erroneous data are either flagged or removed. Examples of erroneous data include zero occupancy and negative volume in the loop detector data, and offset GPS data in the freight database. Further data cleansing (i.e. second step data cleansing) is handled in the DQC module in the middleware tier. In addition to error checking, DQC in the middleware tier mainly conducts preliminary data analysis and processing to reduce the computational burden in the computational tier. For example, the advance loop detectors at Bellevue's intersections are wired together, resulting in undercount problems. A probability-based nonlinear model developed by (*31*) is incorporated into the second DQC module to correct the undercounted volume. Freeway ILD data suffers from both misdetections and erroneous occupancy issues due to incorrect sensitivity level settings in loop cards (*32*). A software-based error detection and correction algorithm (*24*) is also implemented in the middleware tier. Another example is the Origin-Destination (OD) identification algorithm developed by (*25*), which incorporated and extracted individual truck OD information for freight performance measurement. Similarly, the raw Bluetooth MAC addresses collected by the Bluetooth detectors are sent back to DRIVE Net. The redundant data were screened at the first DQC module and the travel time calculations are also undertaken in the 2nd DQC module of the middleware tier.

Middleware Tier

Middleware is a computer program independently running in the server. As mentioned, the purpose of building a middleware tier is to leverage computational power, manage resources between the server (data and two logic tiers) and the client (presentation tier). In addition to the DQC module mentioned earlier, the data connection module is also developed in the middleware tier. In fact, this module is a program interface to connect with multiple databases using Java Database Connectivity (JDBC) API, allowing the middleware tier to query and receive the results from the data warehouse for further process.

Computational Tier

The computational tier in the DRIVE Net server handles complex algorithm implementation after DQC is complete. In addition, this tier assists in archiving raw data and data sharing service control. The Asynchronous JavaScript and XML (AJAX) (*33*) technology is implemented to

reduce the data transfer between the server and the browser and minimize interference to the display and ongoing activities on the existing page. This design reduces the server's response time and enhances the system performance for displaying dynamic and interactive web pages (*33*).

Multiple algorithms implemented in DRIVE Net use this AJAX technology. These algorithms include a iterative time-dependent A* algorithm performing the shortest travel time routing (*34*), statistical metrics generation for freight performance measures (*25*) and incident induced delay calculation using deterministic queue theory and time series techniques (*35*).

Presentation Tier

The primary functionality of the client side is to provide an interactive Graphic User Interface (GUI). As shown in Figure 2, the users' inputs are sent to the computational tier. The computed results are then sent back to the web browser through the Remote Procedure Call (RPC). The final results are visualized through Google Map API (*36*) and Visualization API (*37*), two major third party components supported by the Google cloud. The Google Maps API allows developers to visualize the results on Google Maps through Google Maps services. The Google Visualization API allows users to interact with the data visualized in the statistical charts, such as histograms and pie charts, through visualization tolls services.

Implementation

A combination of the Google Web Toolkit (GWT) (*38*) and Eclipse (*39*), an open source Integrated development environment (IDE), creates a strong development environment for DRIVE Net. GWT contains Java API libraries, allowing developers to code web applications in Java language and then compile the source code into JavaScript. In this case, development cost and time are significantly reduced compared with traditional web development methods, such as JavaScript and/plus PHP. In addition, debugging in GWT makes traditional JavaScript web development much convenient. A developer is able to access existing widget templates in the GWT library to design web interfaces, and a Java to JavaScript compiler translates and optimizes Java code into JavaScript.

The prototype DRIVE Net system can be accessed online at http://www.uwdrive.net/. The web interface of DRIVE Net (Version 1.4) is shown in Figure 3. All computational functions are located on the left side of the panel, including total eight modules programmed on an objected-oriented basis. Hence, all the classes can be "recycled" and "reused" for future development.

CASE STUDY: REGION WIDE INCIDENT IMPACT ANALYSIS

To demonstrate the data interoperability in DRIVE Net, we present a regional incident impact and traffic pattern analysis conducted on the DRIVE Net platform in this section.

Incident Induced Delay

Federal Highway Administration (FHWA) research reports indicate that approximately 50 percent of congestion on freeways is non-recurrent, and incidents contribute 25 percent of this non-recurrent congestion (*40*). A better understanding of the impact of incident-induced delay is essential to develop effective countermeasures against non-recurrent congestion. However, computing incident-induced delay (IID) has been a labor intensive work that hinders large scale incident analysis. Wang et al. (*41*) developed a modified deterministic queuing diagram based approach for automatically calculating IID on DRIVE Net. The approach requires using upstream loop measured volume series and ridge regression (*42*) to predict the departure volume series at the downstream loop location. To apply this approach in IID estimation, the query needs to be executed in nearly real-time to find upstream and downstream loops, query the upstream loop, predict the departure curve at the downstream location under the incident-free scenario, query the downstream loop, and compute the IID. The entire process enabled by a complicated multi-layer graph that properly connects data and roadway components on the Google Map. When the WITS data are being imported into the incident database, incident location information, the upstream and downstream volume data are retrieved from the freeway database to enhance query efficiency. Volume prediction and delay calculation are executed in the middleware tier; meanwhile, the processed results are being updated in the incident database. DRIVE Net allows users to specify a time window for the analysis. It can automatically retrieve all the incidents occurred in the selected time window and execute all the complex algorithms simultaneously to produce the result in a timely manner.

To demonstrate the system performance, we analyze the impact of a non-injury collision incident occurred in the HOV lane on SR 520 on the DRIVE Net platform as a case study. This incident blocked one lane and caused congestion. It started from 9:00 AM, Jan. 14th, 2009 and was cleared at 9:19 AM. All the detailed incident information including incident ID, notification time, clearance time and delay can be displayed on the interface as shown in Figure 4.

Incident Impact on Freeways

After the incident induced delay calculation is completed, the incident impact duration is of interest and can be easily observed using DRIVE Net. The freeway loop data is retrieved from the closest upstream loop station, es00798, through DRIVE Net query and data analysis function. As illustrated in Figure 5, congestion forms before the incident notification time (9am) based on observed occupancy, volume and calculated speeds. This is reasonable because the incident must happen before the notification time.

Incident Impacts on Arterials

As congestion becomes severe in freeways, drivers may resort to diverting to arterials (*43*).

However, diversion may not be beneficial to travelers who choose arterials as an alternative if there is no spare capacity on arterials. Hence, the interaction between the arterials and freeways is critical information for integrated corridor management (ICM).

As mentioned previously, DRIVE Net can be used as an analytical tool to examine the impact on arterials caused by this particular incident on SR 520. The arterial module in DRIVE Net consists of four sub-modules for arterial performance measures: arterial real-time traffic map, arterial data analysis, arterial data archiving, and dynamic routing.

To demonstrate DRIVE Net's capability of intersection performance measurement, DRIVE Net is used to generate time-domain and flow rate-occupancy scatter plots to investigate intersection performance before and after this incident. Several metrics were developed, such as the congestion index and utilization index (*19*). For this specific incident, Intersection 81 is located next to the off-ramp where diversions from freeway to arterial are likely to take place. A time-domain plot was drawn in Figure 6 (a) by specifying incident occurrence time. The high flow rate and occupancy circled in a red color implies a congestion period starting from 9:00 AM to 9:30 AM, which is approximately the incident duration. Figure 6 (b) depicts the relationship between flow rate (per cycle) and occupancy using data collected from Jan. 1st, 2009 to Jan. 31th, 2009. As observed from this figure, the potential capacity of this approach is about 400 vehicles per hour. When flow rate exceeds this limit, traffic break-down is likely to occur. Based on the historical information and statistics, traffic engineers could take countermeasures before the traffic break down to maximize the maximum throughput.

To evaluate the impact of this incident on arterials, an overview of the arterial flow map at 9:12 AM on Jan. 14th is shown in Figure 7(a). The GIS map function in DRIVE Net can display different layers on Google maps, including route and traffic information. With DRIVE Net, a traveler could avoid going through these incident-impacted arterials by visually observing the congestion. The dynamic routing sub-module can also be used to guide the arterial travelers for the shortest travel time route as displayed in Figure 7(b).

This real-time time dependent shortest path sub-module is based on A* searching algorithm (*44*). Due to dynamic travel time in real-time traffic system, the shortest path based on static travel time is not always the "shortest" route. Therefore, the algorithm calculates time dependent travel time for each link, updates the link cost and performs A* algorithm iteratively. For each link, the loop spot speed is calculated based on Athol's speed estimation formula (*45*), and the link speed parameters modeled are estimated in advance using the modified Iowa model (*46*). Then, the predicted travel time for next step is based on the Kalman Filter algorithm (*34*). and predict next step travel time based on Kalman Filter algorithm (*34*). All the above-mentioned algorithms require the relational operations to interact with the data in the arterial database. The link speeds calculation is conducted in the middleware tier in advance to reduce the computational complexity. Therefore, the system design can effectively reduce the computational burden in the browser. The processing time is less than three seconds using Google Chrome browser on a Pentium D 3.4 GHz computer. The travel time for this example route is 2.4 minutes. The calculation requires real-time database query and computation support in DRIVE Net.

BENEFITS

The case study demonstrates the DRIVE Net's capability of enhancing connections and interoperability of the incident, freeway and arterial databases. The application can be easily expanded depending on the needs. Further potential benefits of DRIVE Net can be foreseen:

Interoperable Data Framework

DRIVE Net offers not only an online platform for traffic related data sharing, visualization, modeling, and analysis, but also a foundation to combine multiple data sources on a regional-map-based system. Unlike traditional ATIS, DRIVE Net conducts data translation, format standardization and data centralization for multiple agencies. Thus, different agencies can understand each others' data needs, formats and resources better, and result in significant cost reduction and efficiency improvement.

Customized Platform for Knowledge Sharing and Development

DRIVE Net has the potential to become an ideal online tool for the various user groups. For example, regular users may create a personal DRIVE Net account with customized travel route information to compute congestion statistics on their commuting routes and explore potential alternative routes. Transportation agencies can use the data sources in DRIVE Net for transportation planning and decision making. For example, integration between household data and traffic data can help researchers investigate important questions such as whether the poor are paying unfair portions of their earnings on new toll roads and how roadway travel time reliability affects decisions on business and residential locations. Similarly, environmental and public health researchers can also investigate the relationship between air quality, human disease and traffic condition. Recent studies have observed associations between traffic pollution and multiple human health issues (47). Similar studies can be easily replicated by joining health data and traffic data in DRIVE Net. Hence, different user groups may use different resources/functions in DRIVE Net.

E-Science Transportation Prototype

With e-science approaches gaining popularity in many engineering and scientific areas, such as astronomy, physics, and bio engineering (e.g. *48, 49, 50*), transportation engineering and planning definitely have the need to achieve the same goal. DRIVE Net provides a prototype online platform for e-science applications in transportation. With this platform, data can be easily accessed and broadly visualized and evaluated. With the aid of DRIVE Net, researchers and decision makers can use DRIVE Net to effectively investigate broader interdisciplinary issues such as if the lung disease is related to the geospatial locations of roadways and volume of traffic and how roadway travel time reliability affects the commuters decisions in business and residential locations.

CHALLENGES AND FUTURE VISIONS

Although the DRIVE prototype provides a solid, expandable foundation for road users, researchers and decision makers, future improvement will be desired and proposed as follows.

Expanding Data Converge

A variety of data sources will be incorporated in DRIVE Net to realize region wide research. For example, static imagery data from highway surveillance video cameras, real-time Really Simple Syndication (RSS) feed information for maintain pass condition and highway traffic alert in Washington state, City of Seattle, King County Metro transit data, PSRC traffic count and land use data, PSCAA together with Washington State Department of Ecology (WADOE) operate a monitoring network in this region comprised of 23 air quality monitoring sites, and air quality measurement data will be included in the DRIVE Net. In addition, the Washington State Department of Health (WSDOH) provides public access to several databases on population based health outcomes, such as birth certificate data, hospital discharge data including individual health conditions. With all these data, DRIVE Net can investigate the impacts of traffic on human health. However, new data also bring up new challenges. Data privacy protections will be a crucial issue to be addressed.

Figure 8 illustrates the future potential relationships between local ontology and global ontology. Ontology is defined as "formal, explicit specification of a shared conceptualisation" (*51*), which serves a controlled vocabulary cataloging the terms used by all agencies, and relates the various terms through synonym and parent-child relationships. Constructing a shared ontology will largely promote the interoperability services, mapped and merged to create a centralized "global" ontology. This shared ontology will translate agency-specific terms to a global "preferred term". A standardized and centralized data model will be constructed for future data expansion. Our existing data sources are colored in the Figure 8 and show their connections with other data sources. DRIVE Net serves as inter-ontology system manager integrating all the ontologies. With the centralized ontology data model and the proposed data quality control mechanism, various agency-specific data sets can be successfully imported and mapped into DRIVE Net system.

Technical Challenges

With more and more data coming into DRIVE Net, insufficient computational power will be a significant bottleneck. High workloads and overheads will aggravate burdens for the DRIVE Net server due to intensive data retrieval and complex algorithm executions. One possible solution is to adopt a cloud computing technique. (*52*) Clouding computing is an internet-based computing paradigm that fully utilizes web resources to decentralize workloads in the server. In the future, the computation tier of DRIVE Net can be placed in the cloud to handle increasing numbers of tasks that must be performed to analyze transportation-related/health/environmental data. Moreover, data integration, cleansing and aggregation problems will become more challenging with the increasing amount of non-homogenous data. For database design, multi-dimensional database system (MDDB) (*53*) could be a possible design solution to handling large volumes of

complex and interrelated data. Hybrid Online analytical processing (OLAP), a combination between OLAP and MDDB, could be an efficient way to handle two types of the databases with additional costs.

CONCLUSIONS

To facilitate transportation-related research and remove data acquisition barriers, a data-driven transportation analytical and visualization platform, namely the Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net) is proposed. The prototype provides multiple methods to retrieve data from various sources and successfully combines different data sets, including arterial, freeway, freight, incident, and Bluetooth travel time data. DRIVE Net offers not only a web-based ATIS with support of ADUS for data sharing and visualization, but also an interoperable data framework and regional-map-based data-rich platform for transportation decision makers and researchers to analyze data and validate models and existing theories. Currently, DRIVE Net specifically serves transportation information needs. Based on the current system design, DRIVE Net can be easily expanded to a more transparent and accessible online platform that links other data sources, such as household data, pollution data, and public health data, to facilitate future interdisciplinary studies.

ACKNOWLEDGEMENTS

REFERENCE

1. Chiu, Y. C., and Mirchandani, P. B., Online Behavior-Robust Feedback Information Routing Strategy for Mass Evacuation. IEEE Transactions on Intelligent Transportation Systems. Vol. 9, No. 2, 264-274. 2008.
2. Murray-Tuite, P and Mahmassani, H. (2005). Identification of Vulnerable Transportation Infrastructure and Household Decision Making Under Emergency Evacuation Conditions. Research Report for Southwest Region University Transportation Center, SWUTC/05/167528-1.
3. May, A., Traffic Flow Fundamentals. Prentice Hall, Englewood Cliffs, NJ, 1990.
4. Pipes, L. A. Car Following Models and Fundamental Diagram of Road Traffic. *Transportation Research,* vol. 1, no. 1, 1967.
5. Smith, MC, Sadek, AW, and Huang, S. Large-Scale Microscopic Simulation: Toward an Increased Resolution of Transportation Models. Journal of Transportation Engineering, Vo. 134, No. 7, 273-281. 2008.
6. Haghani, A., Tian, Q, Hu, H. Simulation Model for Real-Time Emergency Vehicle Dispatching and Routing. Transportation Research Record 1882. 176-183, 2008.
7. Chen, C., Freeway Performance Measurement System (PeMS), California PATH Research Report, UCB-ITS-PRR-2003-22, Jul., 2003.
8. Petty, K., J. Kwon, and A. Skabardonis. Arterial Performance Measurement System—APeMS. In Workshop on Performance Measures, 85th TRB Annual Meeting. CD-ROM. Washington D.C., 2005.
9. Bertini, R.L., Hansen, S., Byrd A. and Yin T., Experience implementing a User Service for Archived Intelligent Transportation System Data, *Transportation Research Record: Journal of the Transportation Research Board,* No. 1917, Transportation Research Board of the National Academies, Washington D.C. 2005, pp. 90-99.
10. Tufte, K. A., Bertini, R.L., Chee, J., Fernández-Moctezuma, R. J., Periasamy, S., Sarkar, S., Singh, P., Whiteneck, J., Portal 2.0: Towards a Next-Generation Archived Data User Service, Preprint CD-ROM for the 89th Annual Meeting of Transportation Research Board, Washington, D.C. 2010.
11. Pack, M.L., Bryan, J., Steffes, A.M., An Overview and Status of the Regional Integrated Transportation Information System (RITIS) in the National Capital Region, Preprint CD-ROM for the 87th Annual Meeting of Transportation Research Board, Washington, D.C. 2008.
12. Wongsuphasawat, K., Pack, M.L., Filippova, D., VanDaniker, M., Olea, A., Visual Analytics for Transportation Incident Data Sets, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2138, Transportation Research Board of the National Academies, Washington D.C. 2009, pp. 135-145.
13. Pack M.L., Weisberg, P., Bista, S., Four-Dimensional Interactive Visualization System for Transportation Management and Traveler Information, *Transportation Research Record:*

Journal of the Transportation Research Board, No. 1937, Transportation Research Board of the National Academies, Washington D.C. 2005, pp. 152-158.

14. Filippova, D. and Pack, M.L., Mining Multivariate Accident Data, Preprint CD-ROM for the 86th Annual Meeting of Transportation Research Board, Washington, D.C. 2007.

15. VanDaniker, M. and Pack, M.L., Visualizing Real-Time and Archived Traffic Incident Data, Preprint CD-ROM for the 86th Annual Meeting of Transportation Research Board, Washington, D.C. 2007.

16. Lund, A.S. and Pack, M.L. Dynatic Wide-area Congestion&Incident Monitoring Using Probe Data. Preprint CD-ROM for the 88th Annual Meeting of Transportation Research Board, Washington, D.C. 2009.

17. Pack, M.L. and Ivanov, N., A Web-based, Interactive Temporal-Spatial Traffic and Incident Data Visualization Tool, Preprint CD-ROM for the 87th Annual Meeting of Transportation Research Board, Washington, D.C. 2008.

18. Seymour, E.J., Miller, V., Google-Based Mapping System for Traffic Management Center Web Navigation, Preprint CD-ROM for the 86th Annual Meeting of Transportation Research Board, Washington, D.C. 2007.

19. Wu, Y.-J., and Wang, Y., An Interactive Web-based System for Urban Traffic Data Analysis, International Journal of Web Applications. Vol 1. No 4. pp. 241-252.

20. Wu, Y.-J., and Wang, Y., and Qian, D., A Google-Map-Based Arterial Traffic Information System, IEEE International Conference on 10th Intelligent Transportation Systems, Seattle, Washington, USA.

21. Hey, T. and A. E. Trefethen. The UK e-Science Core Programme and the Grid. Future Generation Computer Systems, Vol. 18, Issue 8, 1017 – 1031, 2002.

22. National Transportation Communications for ITS Protocol, www.ntcip.org, Accessed by Jul. 20th, 2010.

23. Council, F. M. and Mohamedshah, Y. M., Highway Safety Information System Guidbook for the Washington State Data Files, 2009.
http://www.hsisinfo.org/guidebooks/WA_Guidebook_2009.htm, Accessed July 02, 2010.

24. Wang, Y., Corey, J., Lao Y., and Wu, Y.-J., *Development of a Statewide Online System for Traffic Data Quality Control and Sharing*, Transportation Northwest (TransNow), Project 61-6022, 2009.

25. Ma, X., McCormack, E., Wang, Y., 'Process commercial GPS data to develop a web-based truck performance measures program', Paper 11-1932. Proceedings CD-ROM for the 90th Annual Meeting of Transportation Research Board, Washington, D.C. 2011.

26. Malinovskiy, Y., Wu, Y.-J., Wang, Y., Lee, U., Field Experiments on Bluetooth-based Travel Time Data Collection, Paper 11-3056, Preprint CD-ROM for the 89th Annual Meeting of Transportation Research Board, Washington, D.C. 2010.

27. Codd, E.F.. A Relational Model of Data for Large Shared Data Banks. Communications of the ACM 13 (6): 377–387. 1970.

28. Ran, B., Chang, B. P. and Chen, J., Architecture Development for Web-based Geographic Information System Applications in Transportation, *Transportation Research*

Record: Journal of the Transportation Research Board, No. 1660, Transportation Research Board of the National Academies, Washington D.C. 1999, pp. 114-121.

29. Eckerson, W. W., Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications. Open Information Systems 10, 1995.

30. Rahm, E. and Do, H. H., Data Cleaning: Problems and Current Approaches. IEEE Techn. Bulletin on Data Engineering, Dec. 2000

31. Wu, Y.-J., G. Zhang, and Y. Wang. Volume Data Correction for Single-Channel Advance Loop Detectors at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2160, Transportation Research Board of the National Academies, Washington D.C. 2010, pp. 128-139.

32. Cheevarunothai, P., Y. Wang, and N. L. Nihan. Identification and Correction of Dual-Loop Sensitivity Problems *Transportation Research Record: Journal of the Transportation Research Board*, No. 1945, Transportation Research Board of the National Academies, Washington D.C. 2006, pp. 73-81.

33. Garrett, J.J., Ajax: A New Approach to Web Applications, http://www.adaptivepath.com/publications/essays/archives/000385.php. Accessed July 02, 2010.

34. Wu, Y.-J., An, S., Ma, X. and Wang, Y., Development of a Web-based Arterial Network Analysis System for Real-time Decision Making, *Transportation Research Record: Journal of the Transportation Research Board*. In press. 2011.

35. Yu, R., Lao, Y., Ma, X. and Wang, Y., Short-Term Traffic Flow Forecasting for Improved Estimates of Freeway Incident Induced Delays, Paper 11-3593. Proceedings CD-ROM for the 90th Annual Meeting of Transportation Research Board, Washington, D.C. 2011.

36. Google, Inc. The Google Map API. http://code.google.com/apis/maps/index.html, Accessed July 02, 2010.

37. Google, Inc.,Google Chart Toolshttp://code.google.com/p/gwt-google-apis/wiki/VisualizationGettingStarted. Accessed July 02, 2010.

38. Google, Inc., Google Web Toolkit", http://code.google.com/webtoolkit/. Accessed July 02, 2010

39. Eclipse Foundation, http://www.eclipse.org/, Accessed July 02. 2010.

40. Cambridge Systematics, Inc. Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation. Research Report for Federal Highway Administration. 2005, Available online at http://www.ops.fhwa.dot.gov/congestion_report/congestion_report_05.pdf. (Accessed Jul. 20, 2010)

41. Wang, Y., Cheevarunot, P., Hallenbeck, M. *Quantifying Incident Induced Delays on Freeway Using Traffic Sensor Data*, Washington State Department of Transportation (WSDOT), Project WA-RD 700.1, 2009.

42. Hastie, T., R. Tibshirani, J.H. Friedman, and Springer Link, The elements of statistical learning data mining, inference, and prediction, Springer, New York, 2009.

43. Wu, Y-J, Hallenbeck, M-E and Wang, Y., Impacts of Freeway Traffic Congestion on En-route Traveler's Diversion. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C. Paper Number: 10-3602, 2010

44. Russel, S., and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, N.J., 1995.

45. Athol, P. Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream. *Highway Research Record* 72, 1965, pp. 58–87.

46. Wu, Y.-J., Zhang, G. and Wang, Y., Link Journey Speed Estimation for Urban Arterial Performance Measurement using Advance Loop Detector Data under Congested Conditions, Paper 11-3922. Proceedings CD-ROM for the 90th Annual Meeting of Transportation Research Board, Washington, D.C. 2011.

47. Van Hee, V., Adar, S., Szpiro, A., Barr, R., Bluemke, D., Diez Roux, A., Gill, E., Sheppard, L. and Kaufman J.. Exposure to Traffic and Left Ventricular Mass and Function: The Multi-Ethnic Study of Atherosclerosis. *American journal of respiratory and critical care medicine*, 2009, Vol. 179, 827-834.

48. Szlavecz, K., Terzis, A., Musăloiu-E. R., Cogan, J., Small, S., Ozer, S., Burns, E., Gray, J., Szalay, A.S., Life Under Your Feet: An End-to-End Soil Ecology Sensor Network, Database, Web Server, and Analysis Service", MSR-TR-2006-90, June , 2006.

49. Singh, V., Gray, J., Thakar, A. R., Szalay, A. S., Raddick, J., Boroski, B., Lebedeva, S., Yanny, B., SkyServer Traffic Report – The First Five Years, MSR-TR-2006-190, December 2006.

50. Listgarten,J., Frahm, N., Kadie, C., Brander, C., Heckerman, D., A statistical framework for modeling HLA-dependent T cell response data, PLoS Computational Biology, 3(10): e188, October 2007.

51. Gruber, T. A translation approach to portable ontology specifications. In: *Knowledge Acquisition*. 5: 199-199, 1993.

52. Miller, M., *Cloud Computing: Web-based applications that change the way you work and collaborate on line*, Indianapolis, Ind., 2008.

53. Jarke, M., Fundamentals of data warehouses, Springer, Berlin; New York, 2003.

LIST OF TABELS AND FIGURES

FIGURE 1 DRIVE Net System Architecture

Data tier (Data Warehouse)

Data Quality Control
(1st DQC module)

Data sources

DRIVE Net Server

JDBC

Middleware tier (1st logic tier)

Data Quality Control (2nd DQC module)

| Loop Data Error Correction | Truck GPS Data OD Identification | Bluetooth Data MAC Address Mathing | Incident Data Loop Station Selection |

Database Connection

JDBC

Computational tier (2nd logic tier)

Data Sharing control

Algorithm Implementation

| Travel Time Estimation and Prediction | Dynamic Routing | Incident Induced Delay Calculation | Statistical Analysis |

Client

Google Cloud

Google Maps Serivces

Visualization Tools Services

Presentation Tier
(Web Browser)

JavaScript Interpreter/ AJAX

User Input Control

RPC Communication

FIGURE 2 DRIVE Net System Flow Chart

Ma, Wu, and Wang

22
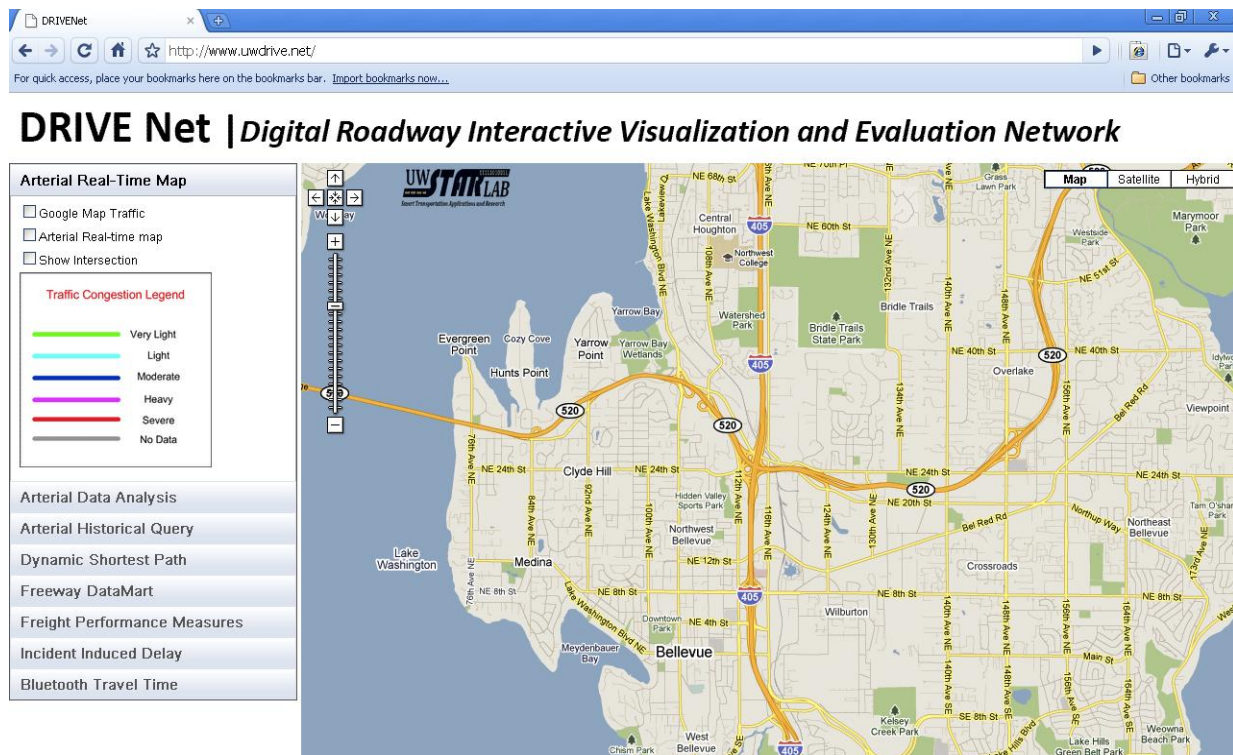
FIGURE 3 DRIVE Net Interface (Ver. 1.4)
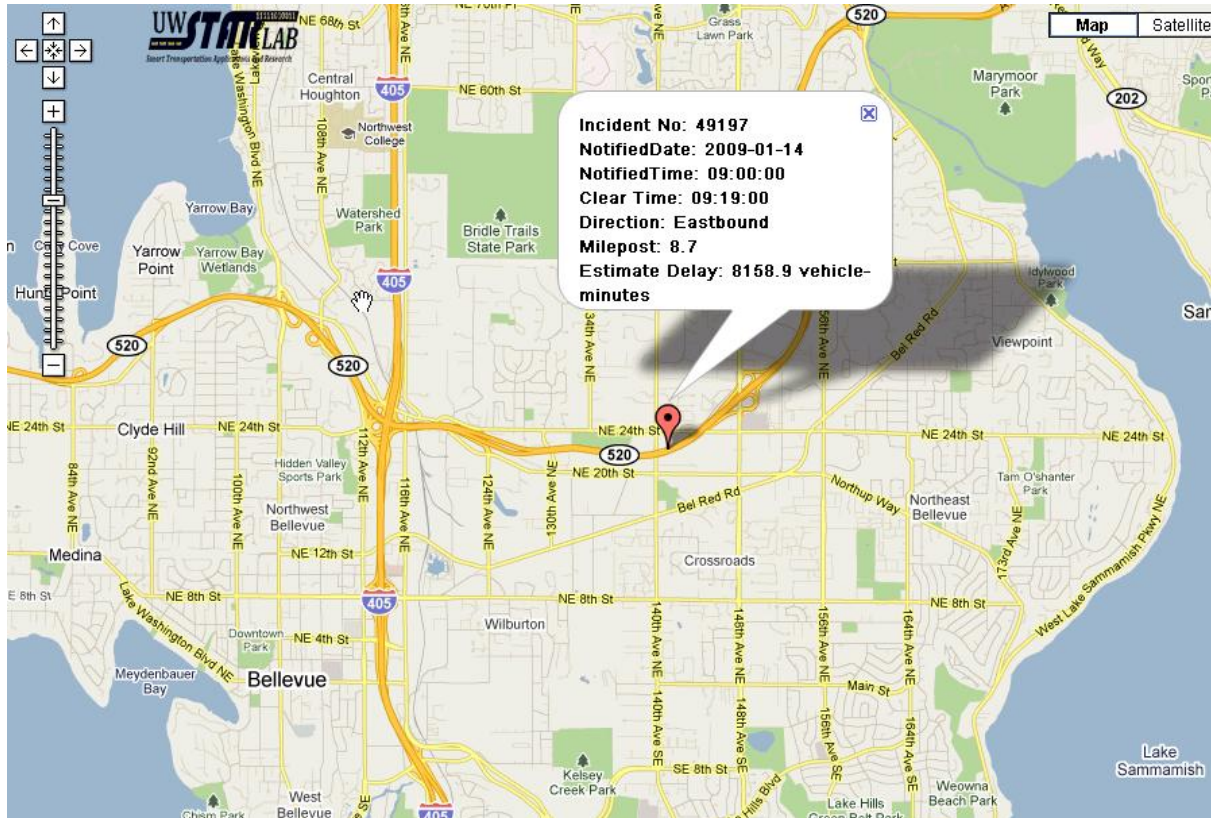
FIGURE 4 Incident Induced Delay Calculation
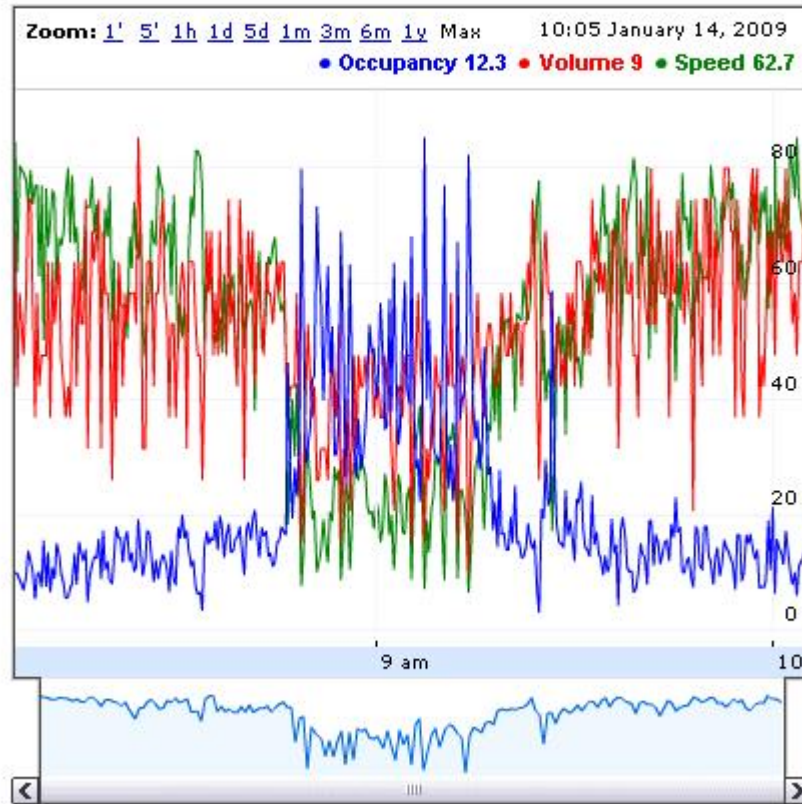
FIGURE 5 20-second volume, occupancy and speed plots at Loop es00798 on SR-520 (Eastbound, Milepost 8.7)

(a)
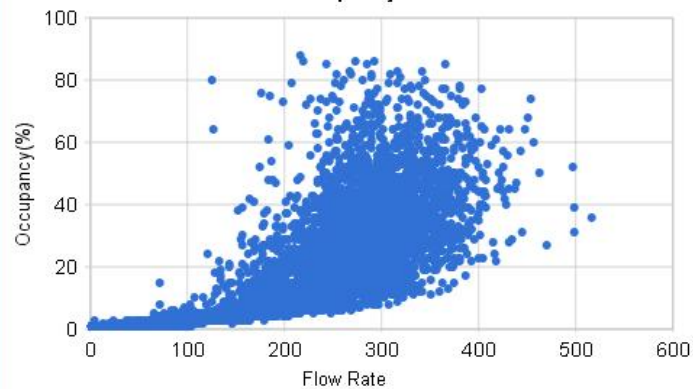
FIGURE 6 The arterial analysis module in DRIVE Net, (a) Time domain plot for Intersection 81 on Jan. 14th, 2009, (b) Occupancy and flow rate scatter plot from Jan. 1st, 2009 to Jan. 7th, 2009
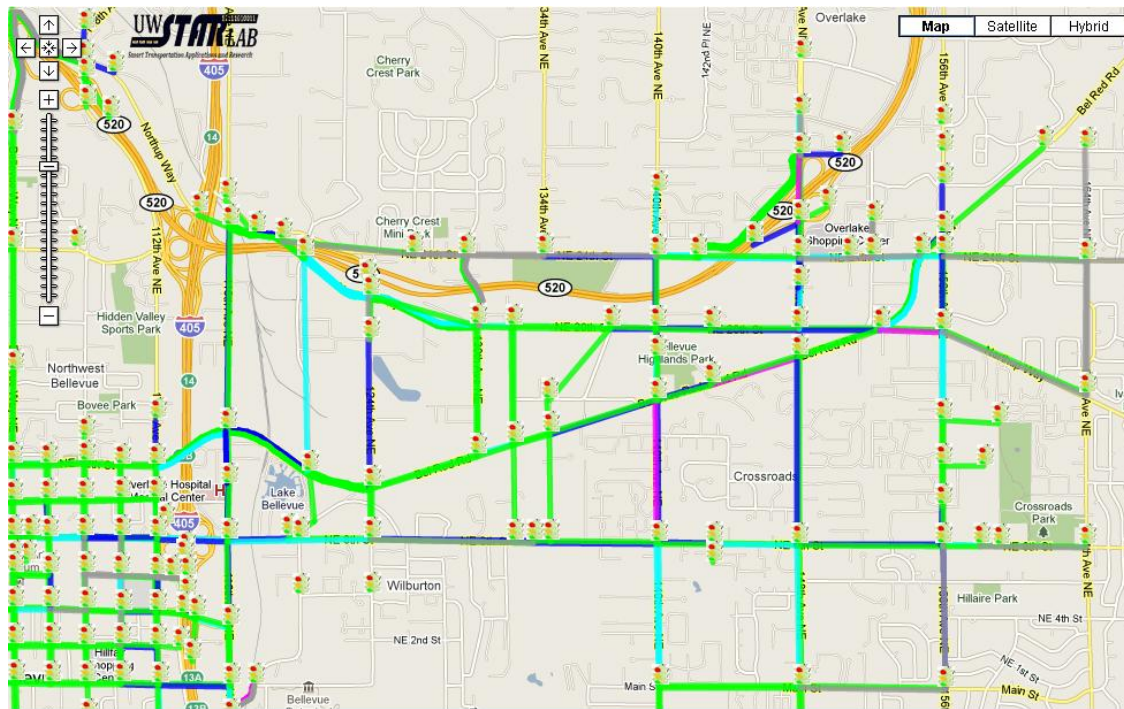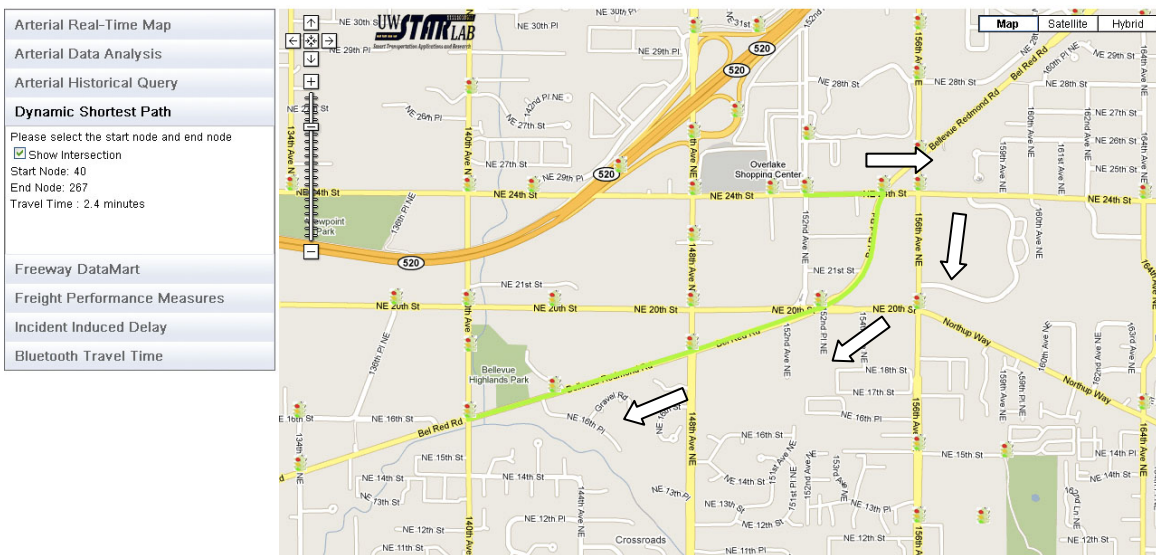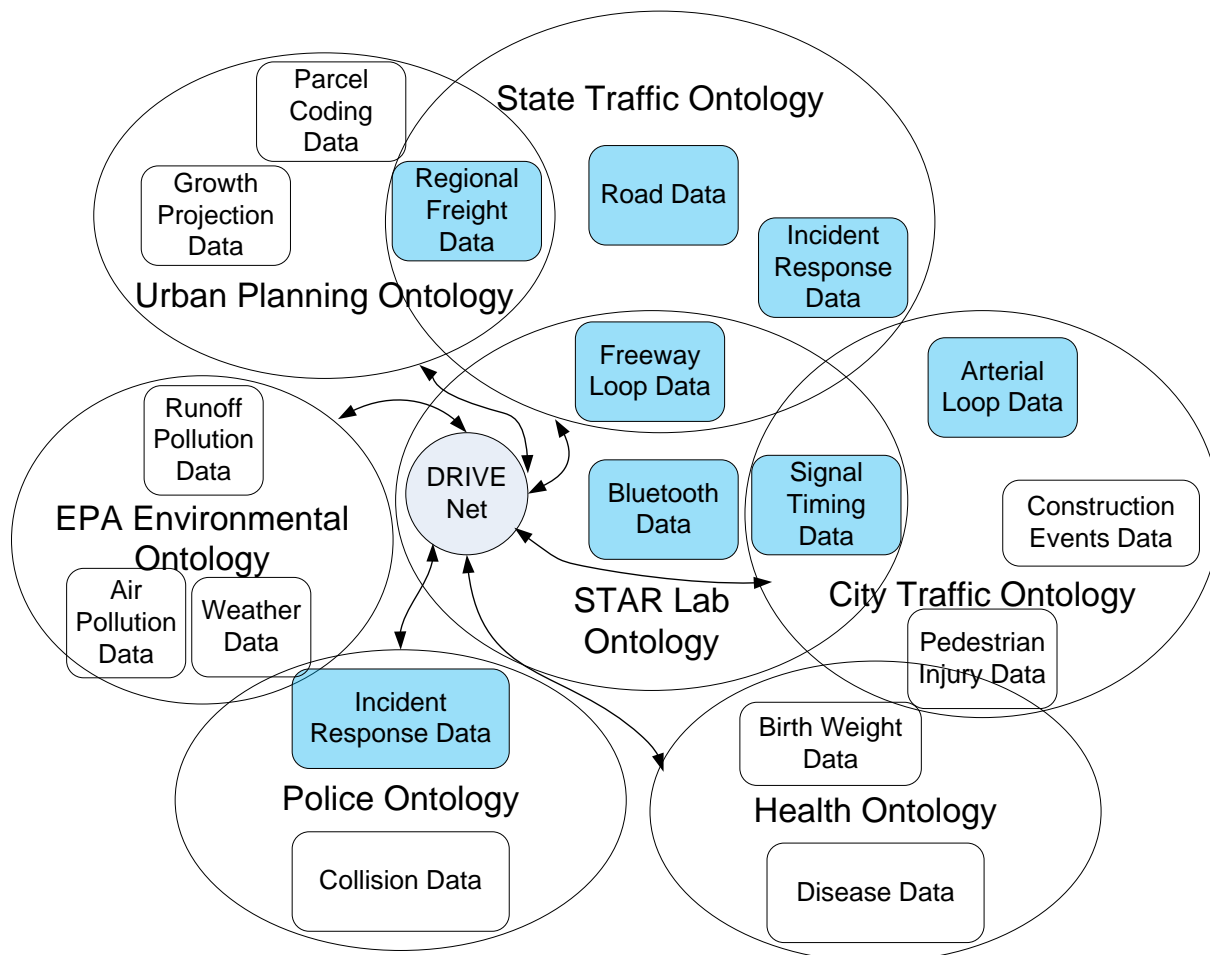
(a)

FIGURE 7 Spatial information display (a) Historical Traffic Flow Map on 9:12 AM, Jan. 14th, 2009, (b) Dynamic Routing

FIGURE 8: Proposed Ontology for DRIVE Net