DRIVE Net

E-Science Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis

Xiaolei Ma, Yao-Jan Wu, and Yinhai Wang

In past decades, transportation research has been driven by mathematical equations and has relied on scarce data. With increasing amounts of data being collected from intelligent transportation system sensors, data-driven or data-based research is expected to expand soon. Most online systems are designed to handle one type of data, such as from freeway or arterial sensors. Even if transportation data are ubiquitous, data usability is difficult to improve. A framework is proposed for a regionwide web-based transportation decision system that adopts digital roadway maps as the base and provides data layers for integrating multiple data sources (e.g., traffic sensor, incident, accident, and travel time). This system, called the Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net), provides a practical method for facilitating 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 spatially and temporally. Not only can DRIVE Net be used as a practical tool for various transportation analyses, with the use of its online computation engine, DRIVE Net can also help evaluate the benefit of a specific transportation solution. In its current implementation, DRIVE Net demonstrates potential to be used soon as a standard tool to incorporate more data sets from different fields (e.g., health and household data) and offer a platform for real-time decision making.

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

With advances in data-collection technologies and their deployment in intelligent transportation systems (ITS), the availability of transportation data has increased tremendously in recent years. As a future type of traffic management system, IntelliDrive is quickly gaining popularity and is increasingly being deployed. Because IntelliDrive enables frequent vehicle-to-vehicle and vehicle-to-infrastructure communications, the availability of traffic data is expected to explode in coming years. Hence, data-driven or data-based research shall expand and play an increasingly important role, too. Rich data sets will enable the 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 are unsatisfactory for three reasons:

- Various agencies typically manage transportation data independently and store data in different systems, each with a proprietary interface and different data-processing capabilities.
- Extensive communication efforts are required to acquire data from other agencies.
- 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 barriers to data retrieval and exchange have hindered the progress of scientific discovery and of solutions to traffic-related issues.

During the past decade, various web-based archived data user service systems and advanced traveler information systems have been developed in attempts to increase the exchangeability and usability of data. For example, the University of California, Berkeley, has focused 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*). On the basis of the successful experience of using PeMS on freeways, the arterial performance measurement system was developed to estimate travel time on an arterial route using midblock (system) loop detectors (*8*). These two systems are typical advanced traveller information systems with archived data user service support.

Initiated by Bertini et al. (*9*) and Tufte et al. (*10*), the Portland Oregon Regional Transportation Archive Listing system focuses primarily 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 developed by the University of Maryland. The Regional Integrated Transportation Information System is a user-friendly multiple-agency system for sharing, disseminating, and archiving data (*11*). It integrates multiple data sources from different transportation agencies and has a main focus on freeway applications.

X. Ma and Y. Wang, Department of Civil and Environmental Engineering, University of Washington, Box 352700, Seattle, WA 98195-2700. Y.-J. Wu, Department of Civil and Environmental Engineering, University of Virginia, Thornton Hall B228, 351 McCormick Road, Charlottesville, VA 22904. Corresponding author: Y.-J. Wu, yaojan@virginia.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2215, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 37–49.
DOI: 10.3141/2215-04

In addition to the Regional Integrated Transportation Information System, the University of Maryland has developed various platforms for the visualization and analysis of transportation data, each designed for a specific purpose, such as incident analysis (*12–17*). As for other advanced traveler information systems, Seymour and Miller used the Google Maps application programming interface (API) to provide a mapping system for freeway speed, incident information, camera image, and more for the Dallas transportation management center (*18*). Wu and Wang (*19*) and Wu et al. (*20*) implemented a Google maps–based arterial traffic information system for visualizing real-time traffic conditions and performing online data analysis of arterial traffic by using intersection loop and traffic signal timing data from the City of Bellevue, Washington.

Most of these systems are based primarily on a single data source and serve as traditional online data or online traffic information providers. Despite the needs of various transportation-related agencies for online systems to share and analyze transportation-relevant data, few such systems have been developed with the functions of standardized data formats, regional map–based data visualization, and interactive online traffic analysis with consideration for interactions between heterogeneous data. For example, the impacts of freeway incidents on arterials and freeways were not covered in previous research because an explicit architecture to bridge the gaps between heterogeneous data from multiple transportation agencies was lacking.

The goal of this study is to develop an e-science platform for sharing, visualizing, modeling, and analyzing transportation data. Coined in 1999 by John Taylor of the U.K. Office of Science and Technology, e-science refers to computationally intensive science that processes immense data sets by using highly distributed computational resources connected via the Internet (*21*). Although e-science approaches have great potential to solve tough transportation issues, the transportation community has been slow to accept this concept. The new platform, Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net), intends to take advantage of e-science developments for data-driven transportation research and applications.

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

The remainder of this paper is organized as follows. DRIVE Net's system architecture is 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 illustrate data interactions between different analytical modules in DRIVE Net. Then, potential benefits of this research are discussed. The paper concludes 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 current understanding of the platform and future expectations to DRIVE Net. It is intended to be an open-architecture, open-source project so that system design can be improved continuously as system capabilities expand. The current system architecture is primarily composed of three parts: heterogeneous data sources from different agencies, the data warehouse in the Smart Transportation Applications and Research Laboratory (STAR Lab), and web services running on the DRIVE Net system server.

Data Warehousing, Exchange, and Retrieval

The data warehouse is responsible for data archival, with multiple data-retrieval functions supported by the DRIVE Net system. Data retrieval is challenging because every agency has internal policy and security concerns. For each agency to rely on a single uniform method for data retrieval may be infeasible and inapplicable. Moreover, formats for data archival vary across agencies and even within the same agency; data may follow different patterns.

Standardization of data formats would be highly beneficial to transportation agencies and data users. However, few guidelines have been developed for data exchange and standardization. Some, such as the National Transportation Communications for ITS Protocol, focus on standardizing only data communication, not formats for data exchange and storage (*22*). Thus, four data-retrieval methods are proposed and currently used in DRIVE Net; methods and examples follow:

1. Traditional flat file exchange. This method allows data quantity, property, and privacy to be carefully controlled. Although this data-exchange format is most commonly used across agencies, it is inefficient and time-consuming. After being retrieved through physical media (e.g., CD-ROM) or e-mails, these files can be uploaded to the database through the DRIVE Net website or by using the structured query language import function. Two examples of data obtained through flat file exchange are
   – Washington Incident Tracking System (WITS). Most incidents on Washington's state highways and major freeways are logged in the WITS database at the Washington State Department of Transportation (DOT). WITS data sets are disseminated in flat files (Excel) and imported into the DRIVE Net incident database. Detailed incident information (e.g., geospatial location, notification time, and clearance time) is stored in the incident database.
   – Highway Safety Information System (*23*). On a user's request, the Highway Safety Information System provides different types of data for Washington state highways, including accident, roadway inventory, traffic flow rate, curve, grade, and interchange and ramp data. These 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 remote data in a predefined interval via file transfer protocol (FTP), virtual private network, hypertext transfer protocol, or simple object access protocol. This method is considered the most convenient, efficient way to periodically retrieve data from remote servers. Data can be imported into the database following the schema design by

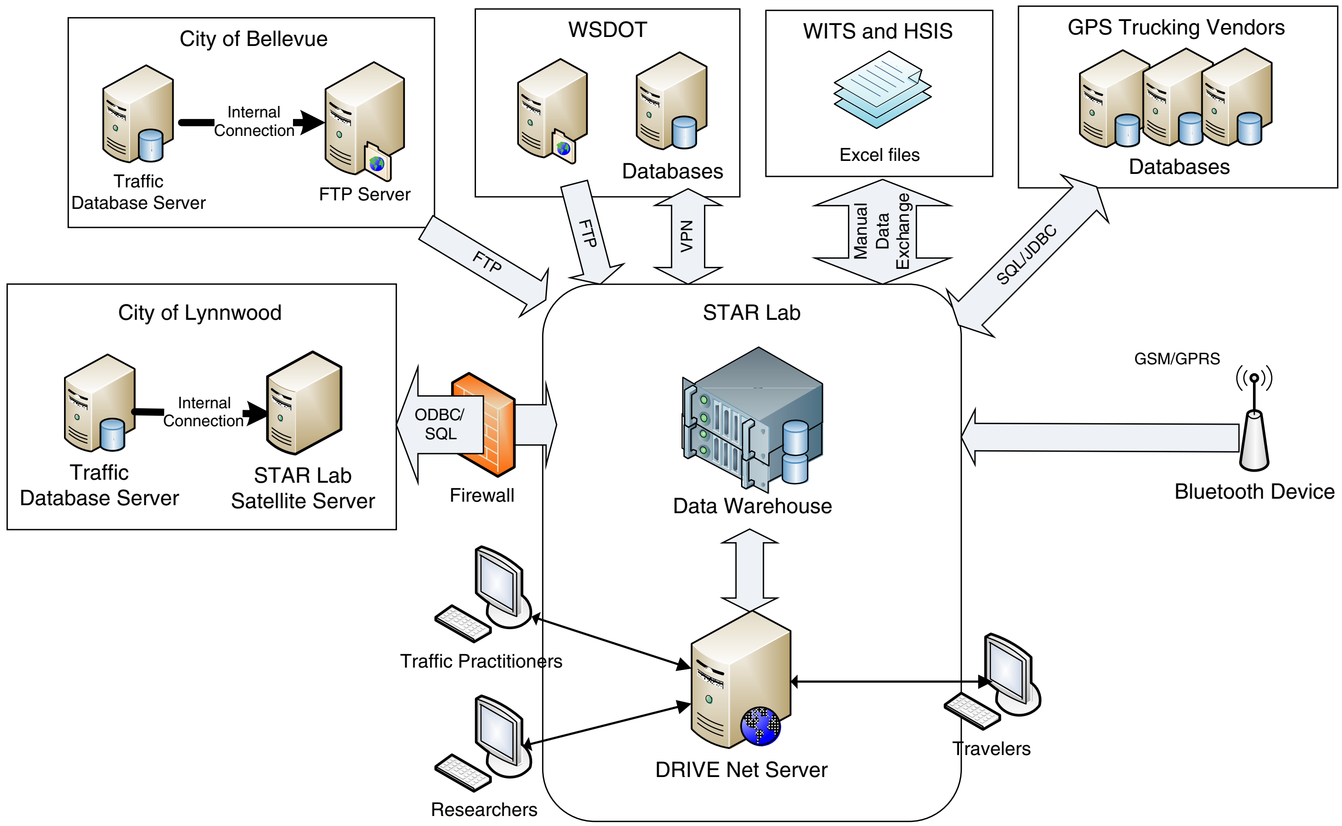


FIGURE 1 DRIVE Net system architecture (WSDOT = Washington State DOT; VPN = virtual private network; JDBC = Java Database Connectivity; ODBC = open database connectivity; SQL = structured query language; HSIS = Highway Safety Information System; GSM = global system for mobile communications; GPRS = general packet radio service).

the research team at the STAR Lab of the University of Washington. Examples include

– Freeway loop sensor data. Washington State DOT operates more than 7,000 inductive loop detectors along freeways in Washington state (*24*); it also shares its 20-s aggregated single loop data via FTP. Data are automatically fetched by the data download module every 20 s and stored in the DRIVE Net freeway database.

– Arterial data. The city of Bellevue, Washington, has more than 500 advance loop detectors at more than 177 signalized intersections (*19, 20*). The controllers at these intersections send cycle-by-cycle real-time traffic data (e.g., flow rate, occupancy, and timing plans) back to an FTP server in the city's traffic monitoring center and stored as comma-separated-value file every minute. DRIVE Net has been fetching comma-separated-value files and importing the data into the DRIVE Net arterial database since 2007.

– Trucking data. Global Positioning System (GPS) tracking systems used by trucking companies are a source of truck probe data for measuring freight performance. Washington State DOT, the University of Washington, and the Washington Trucking Associations have partnered to collect and analyze truck GPS 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 and with frequencies of 1 to 15 min. 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 concerns and limited public access. The STAR Lab provides a satellite server with hardware, software, and data processing tools preinstalled. With a built-in custom service program in the satellite server, the data can be securely pushed back through a firewall to the STAR Lab data warehouse by using open database connectivity. This method is a more expensive but more secure solution to data transmission. One example is intersection detector event data. Second-by-second event data are collected from all video sensors at the 196th Street and SR-99 intersection in Lynnwood, Washington. Data are stored in the STAR Lab satellite server and concurrently pushed through firewalls to the DRIVE Net intersection performance database.

4. Direct data archival. Data can be collected directly from data-collection devices. Data can be sent directly and periodically to the data warehouse from the test site. One example is 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 address at various locations (*26*). This device can transfer data by using general packet radio service and global system for mobile communications communication protocols in real time. Data are sent directly back to the DRIVE Net Bluetooth travel time database every 5 min.

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

DRIVE Net Web Server

The core DRIVE Net system lies in the web server running Apache Tomcat 6.0 on a Windows Server 2008 operating system. This server can render and disseminate data and execute analytical algorithms depending on the user's role. Traffic engineers, researchers, and travelers are 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 by using different techniques of data communication. When necessary, another server can be added to the system. DRIVE Net servers will work jointly like terminals in the grid-computing infrastructure.

SYSTEM DESIGN

The DRIVE Net model is based on a multitier architecture commonly used in software engineering (*28*, *29*). The major merit of the multitier architecture is that 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 contrast to the traditional three-tier client-server model, an additional logic tier handles issues related to data quality. The computational tier is used to control data sharing and execute algorithms. The middleware tier is designed to mitigate the burden in the computational tier usually caused by excessive database access, analytical algorithm calculations, and data quality control (DQC). The client-side presentation tier (web browser) is used to display interfaces, visualize outputs, and receive user input. The overall system flow chart is shown in Figure 2.

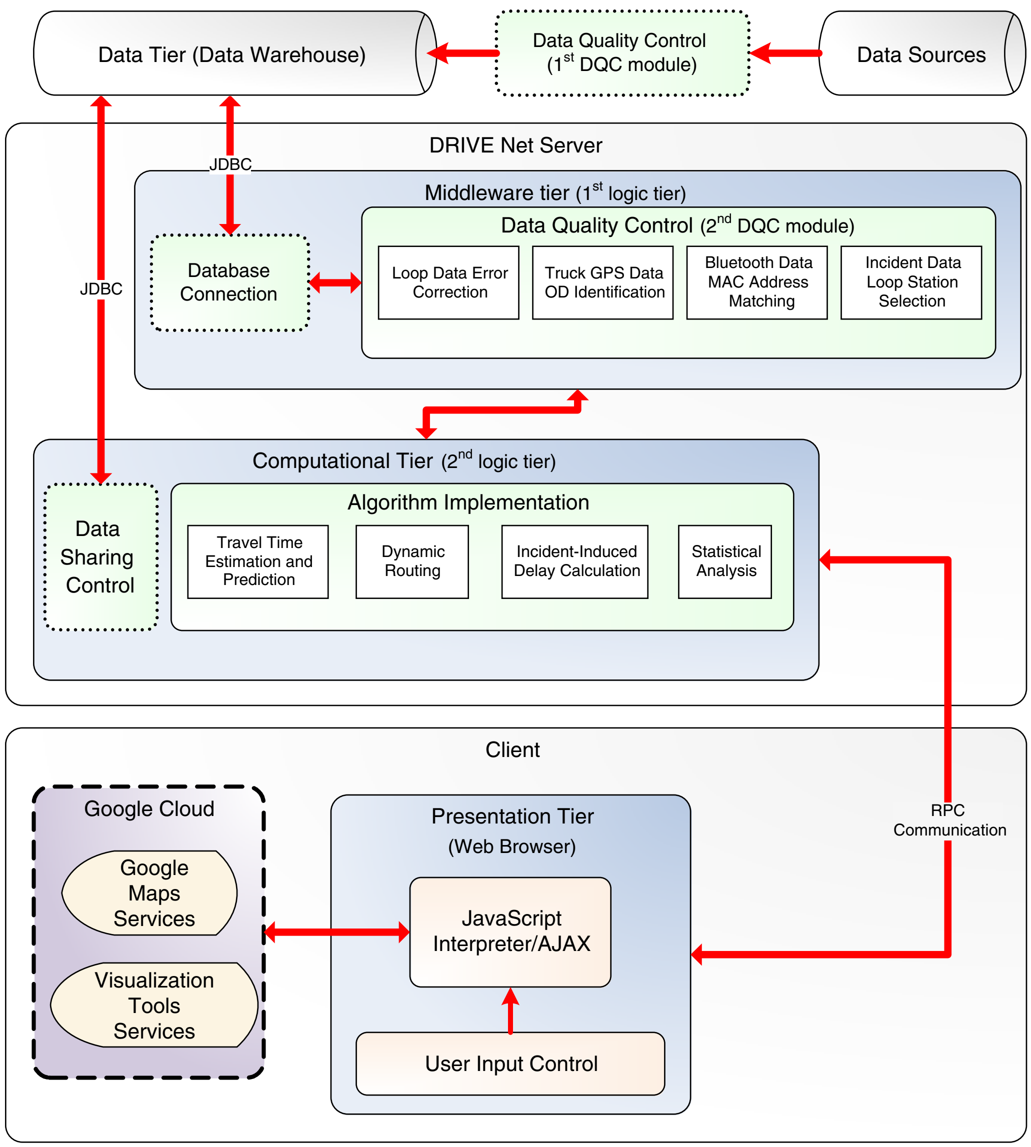


FIGURE 2 DRIVE Net system flow chart (OD = origin and destination; AJAX = asynchronous JavaScript and XML).

Data Quality Control

Data quality is an issue widely recognized by transportation researchers and agencies. An automatic and robust DQC procedure is beneficial for facilitating transportation-related research. To insure data quality, DRIVE Net incorporates a two-step DQC mechanism for cleansing data that detects and removes errors and inconsistencies (*30*). The first step of data cleansing happens during data retrieval from the different data sources, when, for example, erroneous data are flagged or removed. Examples of erroneous data include zero occupancy and negative flow rate in the loop detector data and offset GPS data in the freight database. Additional (i.e., second-step) data cleansing is handled in the DQC module in the middleware tier.

In addition to checking for errors, DQC in the middleware tier 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, which results in undercounting. A probability-based nonlinear model incorporated into the second DQC module corrects the undercounted number of vehicles (*31*). Freeway inductive loop detector data suffer from both misdetections and erroneous occupancy issues as a result of incorrect sensitivity level settings in loop cards (*32*). A software-based algorithm for error detection and correction also is implemented in the middleware tier (*24*).

Another example of DQC is the origin–destination identification algorithm, which incorporates and extracts individual truck origin–destination information for freight performance measurement (*25*). Similarly, the raw Bluetooth media access control addresses collected by the Bluetooth detectors are sent back to DRIVE Net. Redundant data are screened at the first DQC module, and travel time is calculated in the second DQC module of the middleware tier.

Middleware Tier

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

Computational Tier

The computational tier in the DRIVE Net server implements complex algorithms after DQC is complete. This tier also assists in archiving raw data and controlling the data-sharing service. Asynchronous JavaScript and XML technology is implemented to reduce data transfer between the server and the browser and to minimize interference with the display and ongoing activities on the existing page (*33*). This design reduces the server's response time and enhances system performance for displaying dynamic and interactive web pages (*33*).

Multiple algorithms implemented in DRIVE Net use this Asynchronous Java Script and XML technology, including an iterative time-dependent calculation of the shortest (travel time) path (*34*), statistical metrics generation for freight performance measures (*25*), and incident-induced delay (IID) calculation using deterministic queue theory and time-series techniques (*35*).

Presentation Tier

The primary client-side functionality is to provide an interactive graphical user interface. As shown in Figure 2, user input is sent to the computational tier. Computed results are then sent back to the web browser through the remote procedure call. Final results are visualized through Google Maps API (*36*) and Visualization API (*37*), two major third-party components supported by Google's 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 tools services.

Implementation

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

The prototype DRIVE Net system is online (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 eight modules programmed on an objected-oriented basis. Hence, all classes can be "recycled" and "reused" for future development.

CASE STUDY: REGIONWIDE INCIDENT IMPACT ANALYSIS

An analysis of regional incident impact and traffic patterns conducted on the DRIVE Net platform is presented in this section to demonstrate the data interoperability in DRIVE Net.

Incident-Induced Delay

FHWA research indicates that approximately 50% of congestion on freeways is nonrecurrent, and incidents account for 25% of nonrecurrent freeway congestion (*40*). A better understanding of the impact of IID is essential to developing effective countermeasures against nonrecurrent congestion. However, computing IID has been a labor-intensive task that hinders large-scale incident analysis. Wang et al. developed an approach for automatically calculating IID on DRIVE Net based on a modified deterministic queuing diagram (*41*). This approach requires using measured flow rate series and ridge regression at the upstream loop to predict the departure flow rate series at the downstream loop location (*42*).

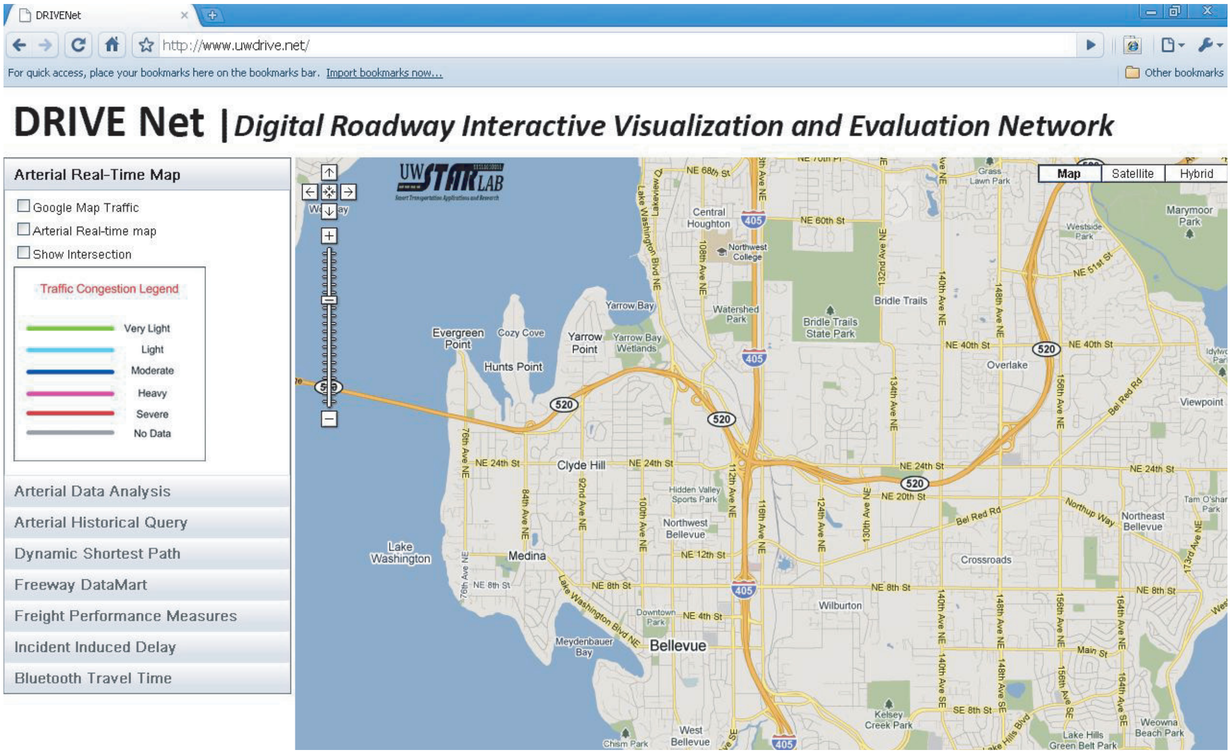


FIGURE 3 DRIVE Net interface (version 1.4).

To apply this approach to IID estimation, the query must 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 IID. The process is enabled by a complicated multilayer graph that properly connects data and roadway components on a Google map. While the WITS data are being imported into the incident database, incident location information and upstream and downstream flow rate data are being retrieved from the freeway database to enhance query efficiency. Flow rate prediction and delay are calculated in the middleware tier; meanwhile, processed results are updated in the incident database. DRIVE Net allows users to specify a time window for the analysis. It can automatically retrieve all incidents that occurred in a selected time window and execute all the complex algorithms simultaneously to produce the result in a timely manner.

To demonstrate system performance, the impact of a noninjury collision incident in the high-occupancy vehicle lane on SR-520 on the DRIVE Net platform is analyzed. This incident, which blocked one lane and caused congestion, started at 9:00 a.m. on January 14, 2009, and was cleared at 9:19 a.m. Detailed incident information (including incident number, notification time, clearance time, and delay) is displayed on the interface, shown in Figure 4.

Incident Impact on Freeways

After the IID is calculated, the duration of incident impact can be easily observed using DRIVE Net. Freeway loop data are retrieved from the closest upstream loop station (es00798) through the DRIVE Net function for query and data analysis. Observed occupancy, flow rate, and calculated speeds indicate that congestion formed before incident notification (9 a.m.; Figure 5), which is reasonable because the incident must happen before the notification time.

Incident Impacts on Arterials

If congestion on a freeway is severe, drivers may divert to arterials (*43*). However, diversion may not be beneficial to travelers if arterials do not have spare capacity. The interaction between arterials and freeways is critical information for integrated corridor management.

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 submodules for arterial performance measures: arterial real-time traffic map, arterial data analysis, arterial data archival, and dynamic routing.

To demonstrate its ability to measure intersection performance, 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 the utilization index (*19*). For this specific incident, Intersection 81 is located next to the off-ramp where travelers are likely to divert from freeway to arterial. A time–domain plot was drawn by specifying incident occurrence time (Figure 6*a*). The high flow rate and occupancy (circled in red) imply a congested period from 9:00 to 9:30 a.m., which is approximately the incident duration. Data collected from January 1 to 31, 2009, were reviewed to evaluate the relationship between flow rate (per cycle) and occupancy; the potential capacity of this approach is about 400 vehicles per hour (Figure 6*b*). When flow rate exceeds this limit, traffic breakdown is likely. Given historical information and statistics, traffic engineers

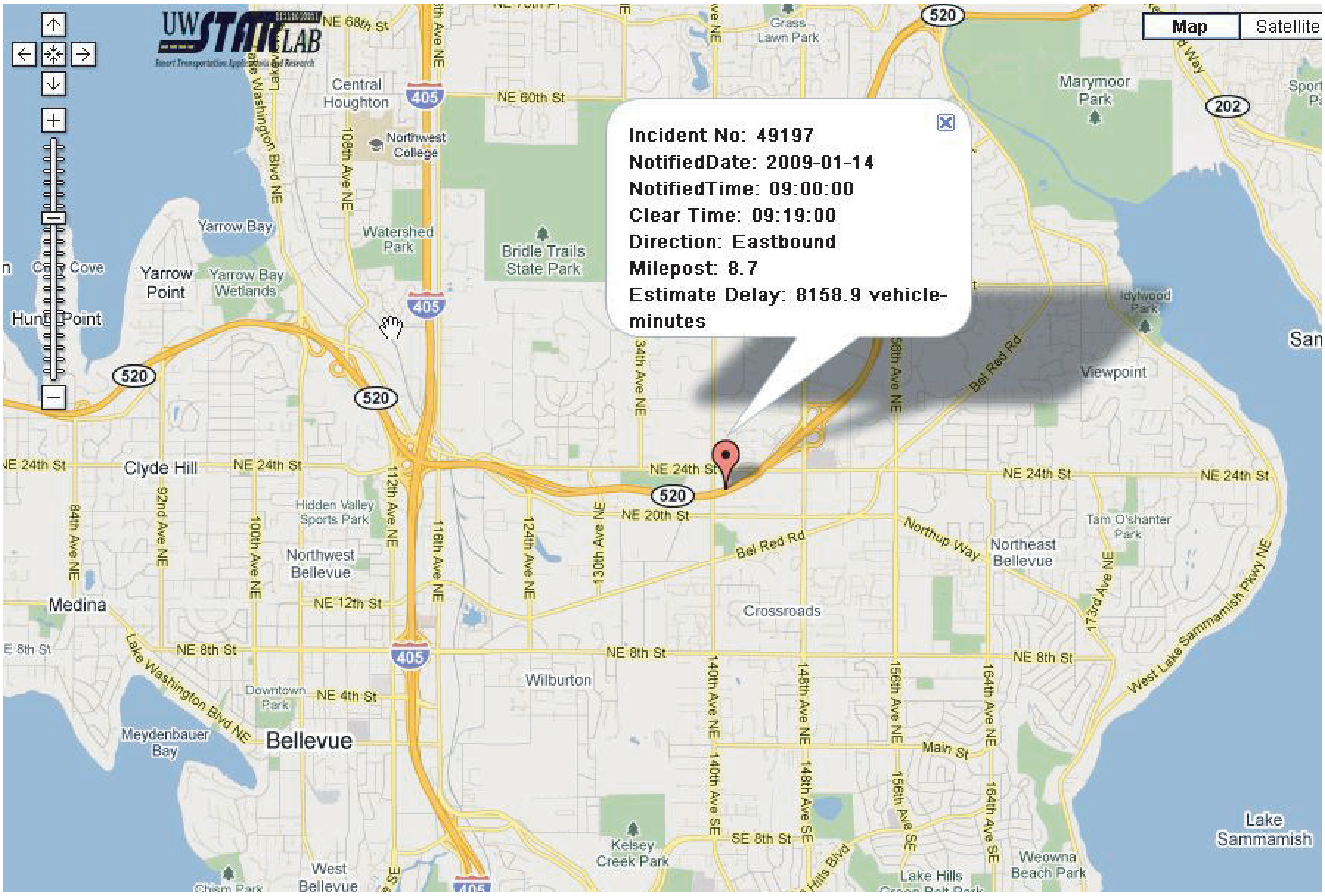


FIGURE 4 Calculation of IID.

could take countermeasures before the traffic breakdown to maximize the throughput.

The impact of this incident on arterials was evaluated with the use of an arterial flow map from 9:12 a.m. on January 14, 2009 (Figure 7*a*). The geographic information system map function in DRIVE Net can display different layers on the Google map, including route and traffic information. With DRIVE Net, a traveler could avoid arterials affected by the incident by visually observing the congestion. The dynamic routing submodule also can be used to guide arterial travelers to the shortest-path route (Figure 7*b*).

This real-time, time-dependent submodule for finding the shortest-path route is based on the A∗ searching algorithm (*44*). Because of dynamic travel time in a 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 the A∗ algorithm iteratively. For each link, the loop spot speed is calculated based on Athol's speed estimation formula (*45*), and the modeled link speed parameters are estimated in advance by using the modified Iowa model (*46*). Then, the travel time predicted for the next step is based on the Kalman filter algorithm (*34*).

All of the above-mentioned algorithms require relational operations to interact with data in the arterial database. Link speed is calculated in the middleware tier in advance to reduce computational complexity. Therefore, system design can effectively reduce computational burden in the browser. Processing time is less than 3 s using a Google Chrome browser on a Pentium D 3.4-gigahertz computer. The travel time for this example route is 2.4 min. The calculation requires real-time database query and computational support in DRIVE Net.

BENEFITS

The case study demonstrates DRIVE Net's ability to enhance connections and interoperability of the incident, freeway, and arterial databases. The application can be easily expanded depending on needs. Additional potential benefits of DRIVE Net that can be foreseen follow.

Interoperable Data Framework

DRIVE Net offers an online platform for traffic-related data sharing, visualization, modeling, and analysis as well as a foundation for combining data sources in a regional map–based system. Unlike traditional advanced traveler information systems, DRIVE Net translates data, standardizes data formats, and centralizes data for multiple agencies. Thus, agencies can better understand each others' data needs, formats, and resources, resulting in significantly reduced costs and improved efficiency.

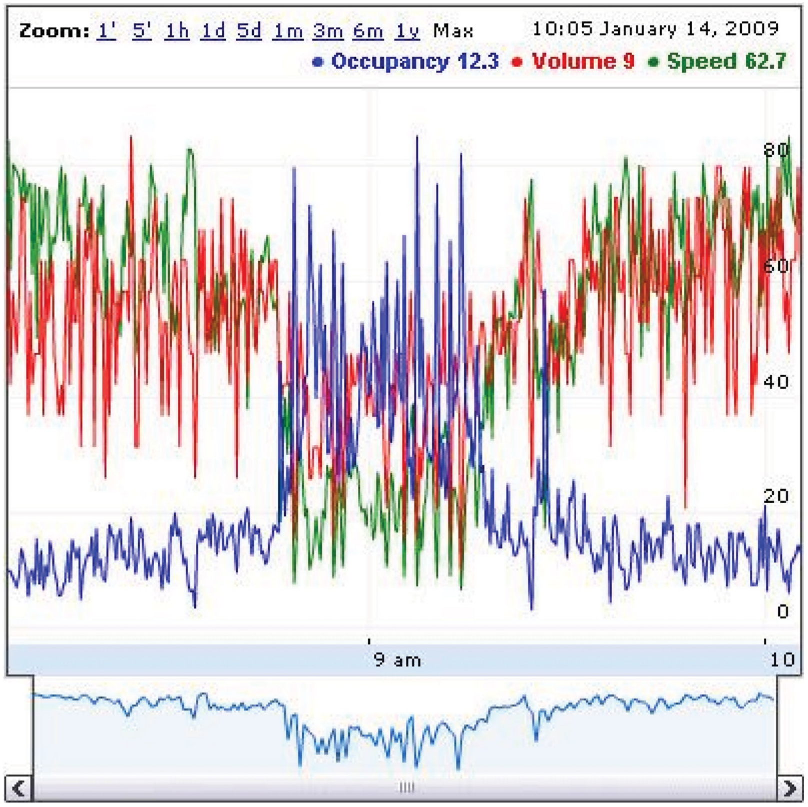


FIGURE 5 Flow rate, occupancy, and speed plots for 20 s at loop es00798 on SR-520 (eastbound, milepost 8.7).

Customized Platform for Knowledge Sharing and Development

DRIVE Net has the potential to become an ideal online tool for 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.

Different user groups may use different resources and functions in DRIVE Net. For example, integration between household data and traffic data could 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 about business and residential locations. Similarly, environmental and public health researchers could investigate relationships among air quality, human disease, and traffic conditions. 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.

E-Science Transportation Prototype

E-science approaches are gaining popularity in areas of engineering and science such as astronomy, physics, and bioengineering (e.g., *48–50*); transportation engineering and planning definitely need to achieve the same goal. DRIVE Net provides a prototype online platform for e-science applications in transportation. It allows data to be easily accessed and broadly visualized and evaluated. With the aid of DRIVE Net, researchers and decision makers can effectively investigate broad interdisciplinary issues, such as whether lung disease is related to geospatial locations of roadways and traffic flow rate and how the reliability of roadway travel time affects commuters' decisions in business and residential locations.

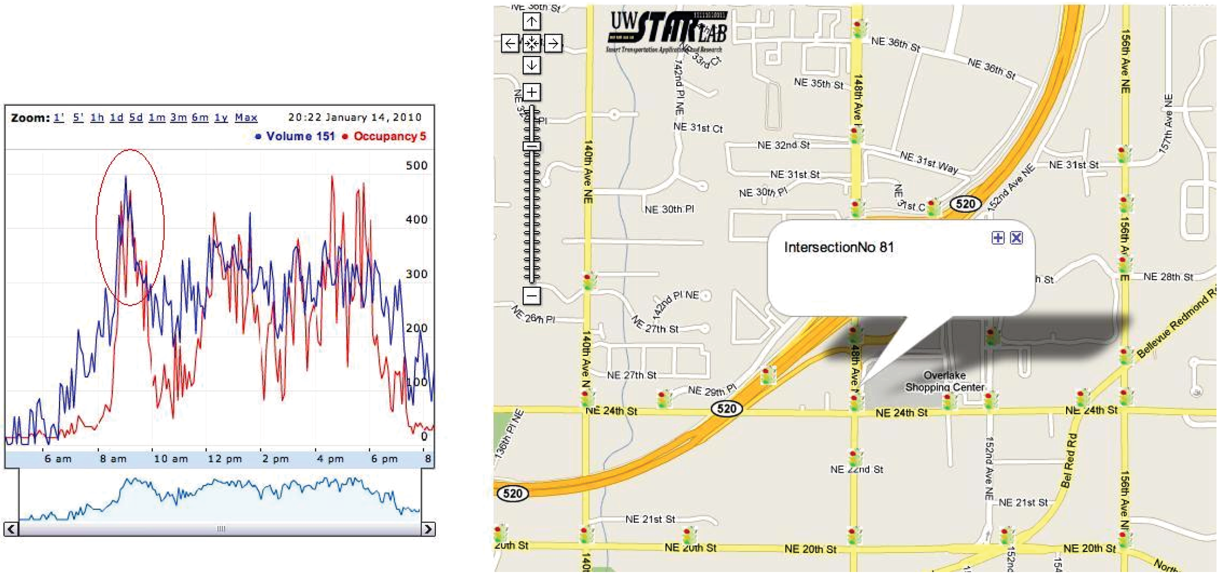
CHALLENGES AND FUTURE VISIONS

Even though the DRIVE Net prototype provides a solid, expandable foundation for road users, researchers, and decision makers, future improvement will be desired and proposed.

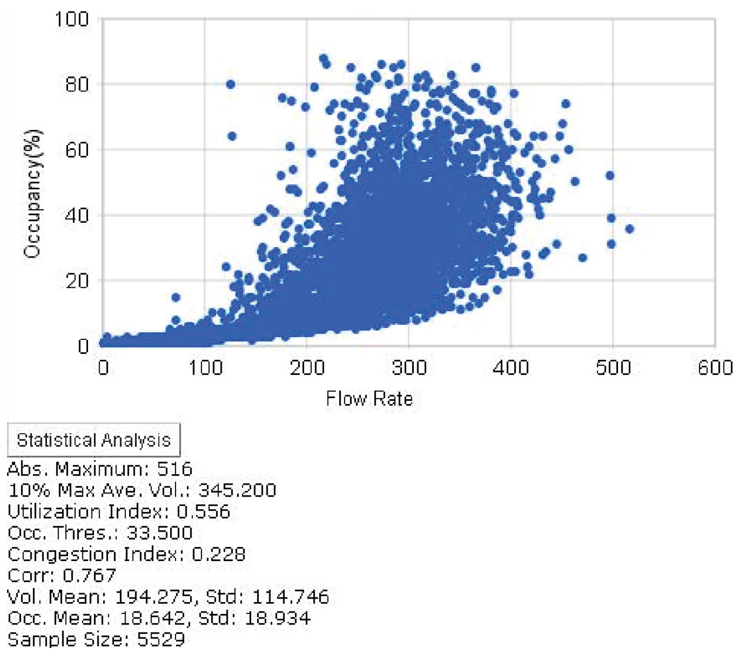
Expanded Data Coverage

Various data sources will be incorporated into DRIVE Net to realize regionwide research, for example,

- Static imagery data from highway surveillance video cameras,
- Real-time really simple syndication feed information for mountain pass conditions and highway traffic information in Washington state,

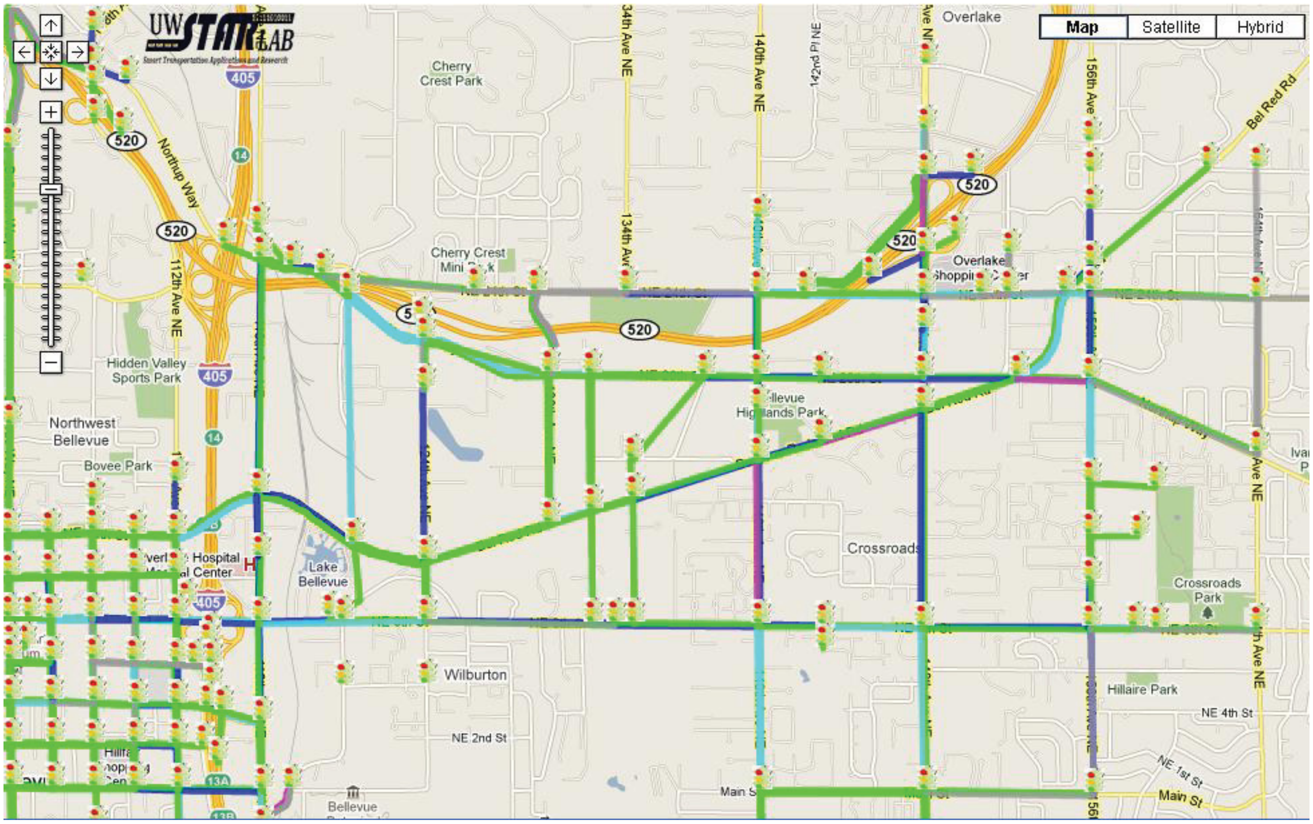


(a)

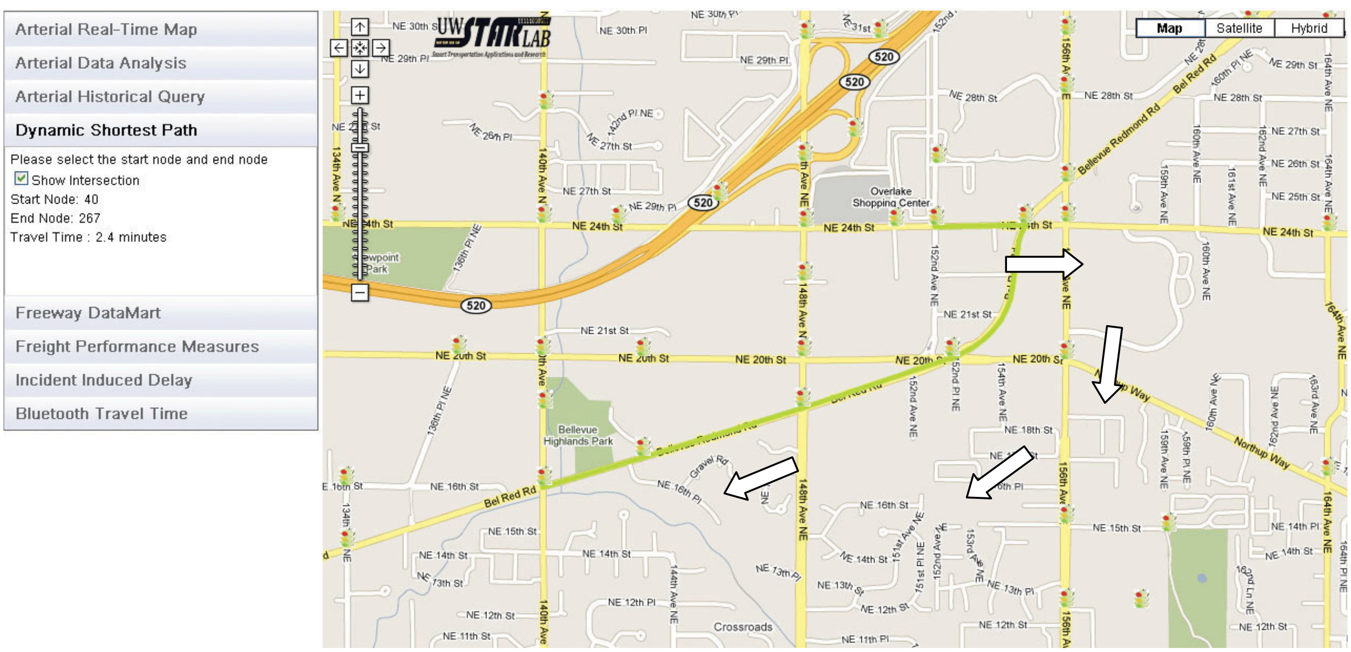


(b)

FIGURE 6 The arterial analysis module in DRIVE Net: (*a*) time–domain plot for intersection 81 on January 14, 2009, and (*b*) occupancy–flow rate scatter plot for January 1–31, 2009.



(a)



(b)

FIGURE 7 Spatial information display: (*a*) historical traffic flow map at 9:12 a.m. on January 14, 2009, and (*b*) dynamic routing.

- City of Seattle and King County Metro transit data, and
- Puget Sound Regional Council traffic count and land use data.

Puget Sound Clean Air Agency and Washington State Department of Ecology operate a monitoring network in this region, and air quality measurement data collected from 23 air quality monitoring sites also will be included in DRIVE Net. In addition, the Washington State Department of Health provides public access to several databases of population-based health outcomes, such as birth certificate data and hospital discharge data, including individual health conditions; with such data, DRIVE Net can investigate the impacts of traffic on human health. However, new data also bring up new challenges. The protection of data privacy will be a crucial issue to be addressed.

Figure 8 shows an illustration of potential future relationships between local and global ontology. Ontology is defined as "formal, explicit specification of a shared conceptualization," which serves a controlled vocabulary cataloging the terms used by all agencies and relates the various terms through synonyms and parent–child relationships (*51*). 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 globally preferred term.

A standardized and centralized data model will be constructed for future data expansion. In Figure 8, existing data sources are colored to show their connections with other data sources. DRIVE Net serves as a system manager between ontologies, integrating them all. With the centralized ontology data model and the proposed DQC mechanism, various agency-specific data sets can be successfully imported and mapped into DRIVE Net.

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 because of intensive data-retrieval and complex algorithm-execution functions. One possible solution is to adopt a cloud computing technique (*52*).

Cloud computing is an Internet-based computing paradigm that fully uses web resources to decentralize workloads at 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 data (transportation-related, health, or environmental). Problems of data integration, cleansing, and aggregation will become more challenging with increasing amounts of nonhomogenous data. For database design, a multidimensional database system could be a possible solution for handling large volumes of complex and interrelated data (*53*). Hybrid online analytical processing—a combination of online analytical processing and multidimensional database system—could be an efficient way to handle the two types of databases with additional costs.

CONCLUSIONS

To facilitate transportation-related research and remove barriers to data acquisition, DRIVE Net—a data-driven platform for transportation analysis and visualization—is proposed. The prototype provides multiple methods for retrieving data from various sources and successfully combining data sets, including arterial, freeway, freight, incident, and Bluetooth travel time. DRIVE Net offers not only a web-based advanced traveller information system with archived data user service support for data sharing and visualization but also an interoperable data framework and a data-rich, regional map–based platform for transportation decision makers and researchers to analyze data and validate models and existing theories.

Currently, DRIVE Net specifically serves transportation information needs. However, given the current system design, DRIVE Net can be easily expanded to a more transparent and accessible online

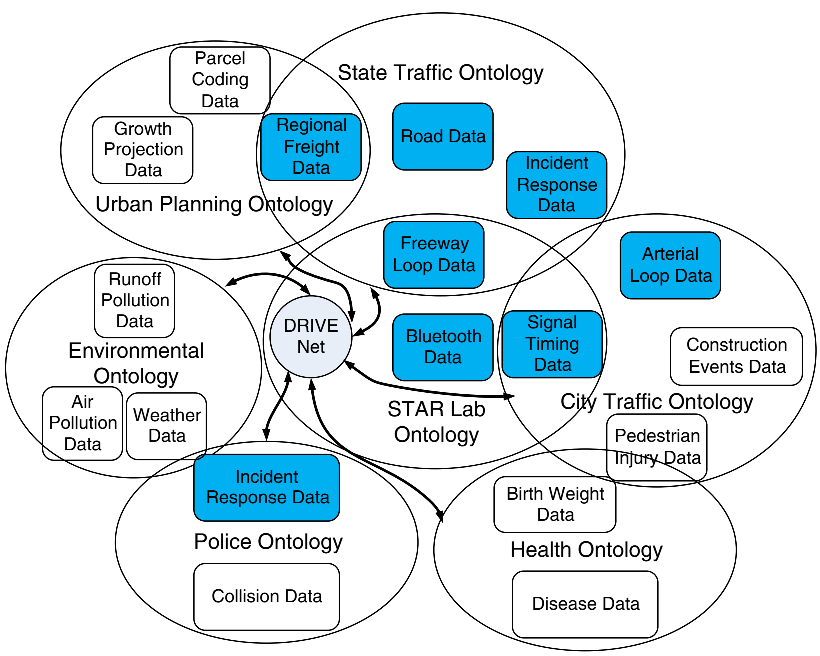


FIGURE 8 Proposed ontology for DRIVE Net.

platform that links sources of other data (e.g., household, pollution, and public health) to facilitate future interdisciplinary studies.

ACKNOWLEDGMENTS

The authors are grateful for data support provided by the Washington State Department of Transportation, the City of Bellevue, the City of Lynnwood, Transportation Northwest (TransNow), and the U.S. Department of Transportation's University Transportation Center for Federal Region 10. The authors thank Jonathan Corey for database maintenance and support.

REFERENCES

1. Chiu, Y. C., and P. B. Mirchandani. Online Behavior–Robust Feedback Information Routing Strategy for Mass Evacuation. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 9, No. 2, 2008, pp. 264–274.
2. Murray-Tuite, P., and H. Mahmassani. *Identification of Vulnerable Transportation Infrastructure and Household Decision Making Under Emergency Evacuation Conditions.* Research Report SWUTC/05/167528-1. Southwest Region University Transportation Center, College Station, Tex., 2005.
3. May, A. D. *Traffic Flow Fundamentals.* Prentice Hall, Englewood Cliffs, N.J., 1990.
4. Pipes, L. A. Car-Following Models and Fundamental Diagram of Road Traffic. *Transportation Research,* Vol. 1, No. 1, 1967, pp. 21–29.
5. Smith, M. C., A. W. Sadek, and S. Huang. Large-Scale Microscopic Simulation: Toward an Increased Resolution of Transportation Models. *Journal of Transportation Engineering,* Vol. 134, No. 7, 2008, pp. 273–281.
6. Haghani, A., Q. Tian, and H. Hu. Simulation Model for Real-Time Emergency Vehicle Dispatching and Routing. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1882,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 176–183.
7. Chen, C. *Freeway Performance Measurement System (PeMS).* Publication UCB-ITS-PRR-2003-22. California Partners for Advanced Transit and Highways (PATH), University of California, Berkeley, Richmond, Calif., 2003.
8. Petty, K., J. Kwon, and A. Skabardonis. APeMS: An Arterial Performance Measurement System. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006. http://www.signalsystems.org.vt.edu/documents/Jan2006AnnualMeeting/SundayWorkshop/Petty_apems_trb.pdf.
9. Bertini, R. L., S. Hansen, A. Byrd, and T. Yin. Experience Implementing a User Service for Archived Intelligent Transportation System Data. In *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. Portal 2.0: Toward a Next-Generation Archived Data User Service. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
11. Pack, M. L., J. R. Bryan, and A. Steffes. Overview and Status of the Regional Integrated Transportation Information System in the National Capital Region. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
12. Wongsuphasawat, K., M. L. Pack, D. Filippova, M. R. VanDaniker, and A. Olea. Visual Analytics for Transportation Incident Data Sets. In *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., P. Weisberg, and S. Bista. Four-Dimensional Interactive Visualization System for Transportation Management and Traveler Information. In *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 M. L. Pack. Mining Multivariate Accident Data. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
15. VanDaniker, M. R., and M. L. Pack. Visualizing Real-Time and Archived Traffic Incident Data. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
16. Lund, A. S., and M. L. Pack. Dynamic Wide-Area Congestion and Incident Monitoring Using Probe Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2174,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 1–9.
17. Pack, M. L., and N. Ivanov. Web-Based, Interactive Temporal-Spatial Traffic and Incident Data Visualization Tool. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
18. Seymour, E. J., and B. Miller. Use of a Google-Based Mapping System for Traffic Management Center Web Navigation. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
19. Wu, Y.-J., and Y. Wang. An Interactive Web-Based System for Urban Traffic Data Analysis. *International Journal of Web Applications,* Vol. 1, No. 4, 2009, pp. 241–252.
20. Wu, Y.-J., Y. Wang, and D. Qian. A Google-Map-Based Arterial Traffic Information System. *Proc., ITSC 2007: IEEE Intelligent Transportation Systems Conference,* Oct. 22, 2007, Seattle, Wash., IEEE, Piscataway, N.J., 2007, pp. 968–973.
21. Hey, T., and A. E. Trefethen. The UK e-Science Core Programme and the Grid. *Future Generation Computer Systems,* Vol. 18, No. 8, 2002, pp. 1017–1031.
22. NTCIP [National Transportation Communications for ITS Protocol] website. http://www.ntcip.org. Accessed July 20, 2010.
23. Council, F. M., and Y. M. Mohamedshah. *Highway Safety Information System: Guidebook for the Washington State Data Files.* FHWA, U.S. Department of Transportation, April 2009. http://www.hsisinfo.org/guidebooks/WA_Guidebook_2009.htm. Accessed July 2, 2010.
24. Wang, Y., J. Corey, Y. Lao, and Y.-J. Wu. *Development of a Statewide Online System for Traffic Data Quality Control and Sharing.* Project 61-6022. Transportation Northwest (TransNow), Seattle, Wash., 2009.
25. Ma, X., E. D. McCormack, and Y. Wang. Processing Commercial GPS Data to Develop a Web-Based Truck Performance Measure Program. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
26. Malinovskiy, Y., Y.-J. Wu, Y. Wang, and U. K. Lee. Field Experiments on Bluetooth-Based Travel Time Data Collection. Presented at 89th Annual Meeting of the 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,* Vol. 13, No. 6, 1970, pp. 377–387.
28. Ran, B., B. P. Chang, and J. Chen. Architecture Development for Web-Based Geographic Information System Applications in Transportation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1660,* TRB, National Research Council, 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,* Vol. 10, No. 1, 1995, pp. 3–20.
30. Rahm, E., and H. H. Do. Data Cleaning: Problems and Current Approaches. *IEEE Bulletin of the Technical Committee on Data Engineering,* Vol. 23, No. 4, 2000, pp. 3–13.
31. Wu, Y.-J., G. Zhang, and Y. Wang. Volume Data Correction for Single-Channel Advance Loop Detectors at Signalized Intersections. In *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. In *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.* Adaptive Path, Feb. 18, 2005. http://www.adaptivepath.com/publications/essays/archives/000385.php. Accessed July 2, 2010.
34. Wu, Y.-J., S. An, X. Ma, and Y. Wang. Development of a Web-Based Analysis System for Real-Time Decision Support on Arterial Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2215,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 24–36.
35. Yu, R., Y. Lao, X. Ma, and Y. Wang. Short-Term Traffic Flow Forecasting for Improved Estimates of Freeway Incident Induced Delays.

Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
36. Google, Inc. *The Google Maps API*. http://code.google.com/apis/maps/index.html. Accessed July 2, 2010.
37. Google, Inc. *Google Chart Tools*. http://code.google.com/p/gwt-google-apis/wiki/VisualizationGettingStarted. Accessed July 2, 2010.
38. Google, Inc. *Google Web Toolkit*. http://code.google.com/webtoolkit. Accessed July 2, 2010.
39. The Eclipse Foundation. Eclipse. Ottawa, Ontario, Canada. http://www.eclipse.org/. Accessed July 2, 2010.
40. *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation*. Prepared for FHWA by Cambridge Systematics, Inc., Cambridge, Mass., 2005. http://www.ops.fhwa.dot.gov/congestion_report/congestion_report_05.pdf. Accessed July 20, 2010.
41. Wang, Y., P. Cheevarunot, and M. Hallenbeck. *Quantifying Incident Induced Delays on Freeway Using Traffic Sensor Data*. Project WA-RD 700.1. Washington State Department of Transportation, Olympia, 2009.
42. Hastie, T., R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York, 2009.
43. Wu, Y.-J., M. E. Hallenbeck, and Y. Wang. Impacts of Freeway Traffic Congestion on En-Route Traveler's Diversion. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 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. In *Highway Research Record 72*, HRB, National Research Council, Washington, D.C., 1965, pp. 58–87.
46. Wu, Y.-J., G. Zhang, and Y. Wang. Link Journey Speed Estimation for Urban Arterial Performance Measurement Using Advance Loop Detector Data Under Congested Conditions. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
47. Van Hee, V., S. Adar, A. Szpiro, R. Barr, D. Bluemke, A. Diez Roux, E. Gill, L. Sheppard, and J. Kaufman. Exposure to Traffic and Left Ventricular Mass and Function: The Multi-Ethnic Study of Atherosclerosis. *American Journal of Respiratory and Critical Care Medicine*, Vol. 179, 2009, pp. 827–834.
48. Szlavecz, K., A. Terzis, E. R. Musăloiu, J. Cogan, S. Small, S. Ozer, E. Burns, J. Gray, and A. S. Szalay. *Life Under Your Feet: An End-to-End Soil Ecology Sensor Network, Database, Web Server, and Analysis Service*. Report MSR-TR-2006-90. Microsoft Corporation, Redmond, Wash., June 2006.
49. Singh, V., J. Gray, A. R. Thakar, A. S. Szalay, J. Raddick, B. Boroski, S. Lebedeva, and B. Yanny. *SkyServer Traffic Report–The First Five Years*. Report MSR-TR-2006-190. Microsoft Corporation, Redmond, Wash., December 2006.
50. Listgarten, J., N. Frahm, C. Kadie, C. Brander, and D. Heckerman. A Statistical Framework for Modeling HLA-Dependent T Cell Response Data. *PLoS Computational Biology*, Vol. 3, No. 10, 2007, pp. 1879–1886.
51. Gruber, T. A. Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, Vol. 5, No. 2, 1993, pp. 199–220.
52. Miller, M. *Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online*. Que, Indianapolis, Ind., 2008.
53. Jarke, M. *Fundamentals of Data Warehouses*. Springer, New York, 2003.

The Information Systems and Technology Committee peer-reviewed this paper.