CHALLENGES AND OPPORTUNITIES FOR ONLINE FREIGHT DATA MAPPING INTEGRATION AND VISUALIZATION

ABSTRACT

This paper presents the issues surrounding the integration and visualization of freight data using internet-based mapping applications. In relation to Internet-based mapping technology in freight data collection and planning we: (a) address implementation issues associated with data integration, (b) present a system architecture to leverage existing publicly available interfaces and web applications to accelerate product development and reduce costs, (c) describe an existing web-based mapping prototype and its capabilities, and (d) state lessons learned and present suggestions to streamline the integration and visualization of freight data. We conclude that despite data integration challenges, Internet-based mapping provides a cost effective and appealing tool to store, access, and communicate freight data as well as enhance our understanding of freight issues. We argue that institutional barriers, not technology, are the most demanding hurdles to widely implementing a freight data web-based mapping application in the near future.

KEYWORDS: Freight Data, Visualization, Integration, Internet Mapping

An internet-based mapping system is utilized to provide geographic context and user friendly information to transportation planning decision makers.

1.　INTRODUCTION

The efficient movement of goods and timely provision of services is critical to the economic and sustainable development of a region. However, the complex economic, spatial, temporal, and operational aspects of logistics activities and freight transportation often hinder effective data collection, communication, and analysis. Public decision makers require a comprehensive picture of freight movements to understand how freight transportation supports economic development, how land use affects freight transportation, and the impacts of transportation infrastructure supply on private sector freight and commercial activity. The need to integrate and coordinate freight data collection efforts is widely accepted and recognized [1]. Freight data is available from many public and private sources. However, the data may significantly vary in terms of collection method, timeframe, format, and quality. The lack of coordination not only prevents the seamless integration of data sources but also limits the scope and quality of transportation studies.

Over the last fifteen years, the ability to collect freight data has significantly expanded through developments in electronics, information and communication technology, and Global Positioning System (GPS) technology. The ability to represent freight data has been greatly enhanced by the development of Geographic Information Systems (GIS) to manipulate and display transportation data at distinct levels of spatial and temporal resolution [2]. More recently, the development of internet-based geographic data visualization platforms, e.g. Google Earth, has dramatically expanded the ability to disseminate and access data [3].

The benefits of sharing maps and spatial data among public agencies are well established [4]. Data and map sharing brings about significant reductions in maintenance activity, increases adoption of GIS technologies, and improves information access and accuracy. Maps and data sharing in GIS transportation (GIS-T) networks require unified and standard semantics, data models, and acquisition methods. For example, a clear semantic hierarchy allows higher order data networks, such as a state highway system, to receive real time updates reflecting any database changes from local government agencies, such as modifications along a local road network [5, 6]. Without common standards and semantics, agencies must devote time and resources to the error prone and expensive process of data conversion [7, 8]. Similar hierarchies and relations are crucial when sharing multi-modal network data, e.g. linking of transit operational data to a road network [9] or the integration of GIS-T applications and transportation asset management [10].

Technological developments in Internet-based mapping tools are creating new challenges and opportunities to collect and communicate freight data. The application of GIS-T to freight has been mostly limited to the display and analysis of truck accident data [11, 12] and truck volumes [13, 14]. The combination of GIS-T and GPS-based data have also been successfully applied to the monitoring of intercity truck movements [15, 16], complement commercial vehicle surveys [17], and the study of commercial vehicle tours in urban areas [18]. The private sector is also swiftly adopting GPS-based technologies, where the monitoring of freight vehicles and containers across the continental US has the potential to reduce cargo damage, control driver behavior, and reduce freight theft [19] [20].

The aim of this research is to highlight the advantages and challenges of using Internet-based spatial data tools and technologies for integrating and visualizing freight transportation data; a working prototype and its system architecture are also presented. Lessons learned from the prototype implementation and recommendations to improve future data collection and visualization efforts are discussed. The paper is organized as follows: Section two describes the project vision and its system architecture. Section three describes data sources. Section four discusses implementation and data integration challenges. Section five describes visualization capabilities of the prototype. Section six presents recommendations for data collection. Section seven ends with conclusions.

2.　PROJECT VISION AND SYSTEM ARCHITECTURE

The system described herein, the Oregon Freight Data-Mart (OFDM), is under continuous development. A primary goal of the OFDM is to provide an online environment to integrate, visualize, and disseminate

freight data in the State of Oregon. It was clear from the outset of the OFDM project that the prototype should handle a diverse set of existing and future data sources and types. A clear vision to make the application flexible and cost effective included the following goals: (a) to provide an intuitive application with minimal user learning, (b) to have powerful visualization and geographical capabilities, (c) to facilitate freight data integration, and (d) to design a system that can leverage existing publicly-available Internet applications in order to accelerate product development and reduce costs.

System Architecture

- Incidents,
- Bottlenecks
- WIM Stations
- Freeway Sensor Data
-Truck Volume
     Locations
- Truck Generator
     Locations

PORTAL Data Archive

Stores textual and numeric data and geo-location information.

Query Results

Web Server Queries database and generates HTML content.

SQL

Web Browser Displays map.

GM API Requests

-Truck Volume
     Graphs
- Truck Generator
     Graphs
- Freight Volume Maps
- Land Use Maps

File System

Stores static images and graphics.

Maps

Google Map Servers Generates map.

System Process

GIS Data → Google Earth

GPS Data

KML/KMZ files

Images

Graphics

File System

OFDM

Google Maps

HTML Browser

User

Excel Files
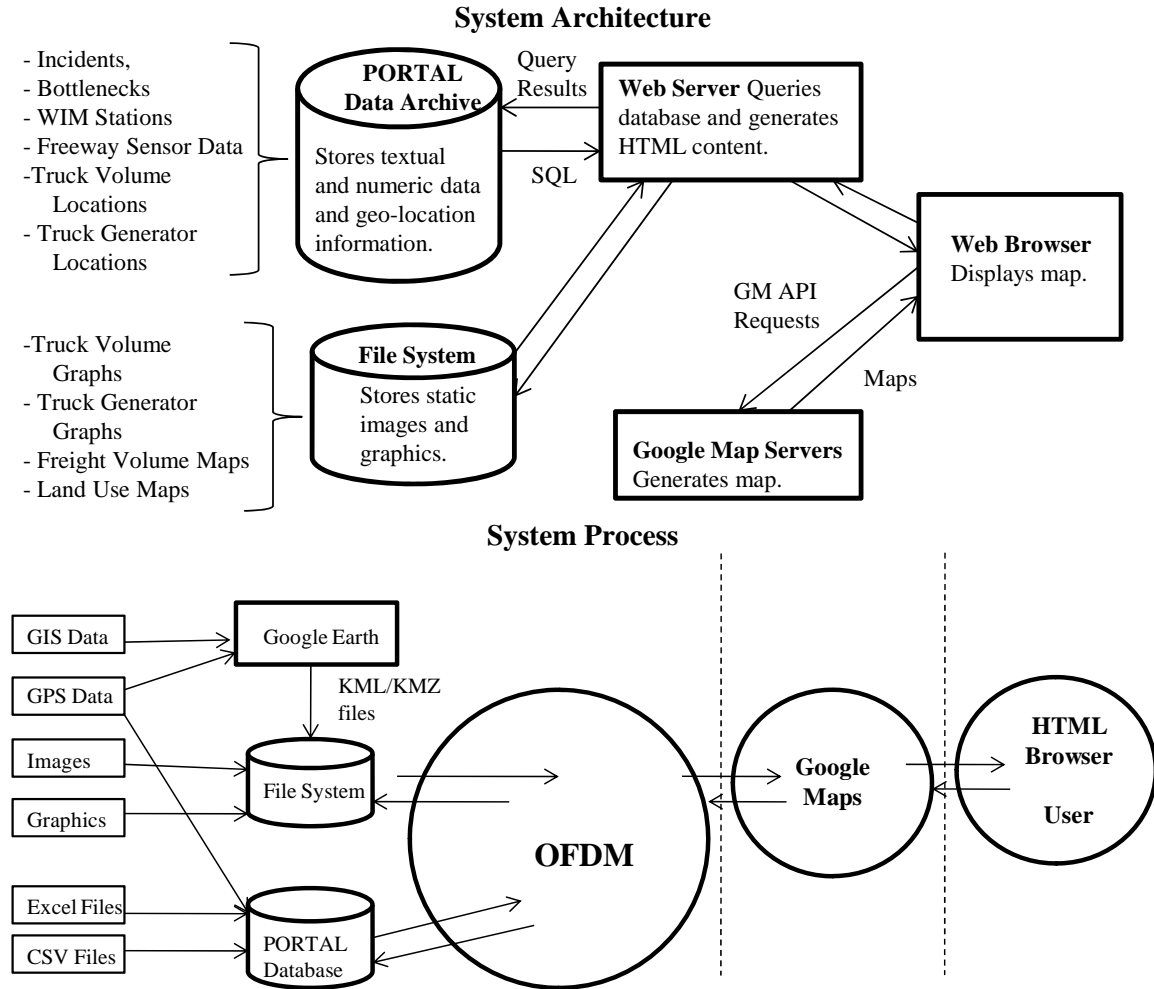
CSV Files

PORTAL Database

Figure 1 The OFDM System

The OFDM is a data visualization tool based on Google Maps® (GM). GM was chosen for visualization because it can be used to combine different types of data and can be accessed by any user from any Internet browser. The intuitive and user-friendly characteristics of GM provide an excellent platform to tailor the display of information. GM fulfills the ease-of-use and visualization requirements of the OFDM, including the ability to integrate images, such as maps, graphics, or digital photos, external links, and HTML content. The user interface also enables integrated visualization of data sources using multiple hierarchical layers and clickable links that can be used to explore and expand details. The OFDM leverages the GM Application Programming Interface (API) which allows a developer to create their own overlays on the basic GM maps. A significant advantage of using GM to display the freight data is the ability to leverage other Google services, such as Google Traffic, Google Street View, and satellite images. The integration of existing freight data with the Google Maps application means that as Google provides more services, the OFDM can take advantage of these services, most often with limited time and monetary

overhead. Finally, Google Earth®[1] can be used as a backbone to develop maps that can be exported to KML/KMZ files, a format that is gaining wide acceptance, which can later be displayed in GM. At a high level, the OFDM system processes and architecture are described in
Figure 1.

A second key component of the OFDM is PORTAL, the official transportation data archive of the Portland metropolitan region [21] . PORTAL consists of a 700GB PostgreSQL database archive and a web site for visualizing that data. The OFDM uses PORTAL for data storage and retrieval. Freight-related data is stored in PORTAL and retrieved for display on the OFDM map interface. Storing data in a database helps support dynamic content by making it easy to select and display only data the user has requested. In addition, as PORTAL expands by adding new data and features, OFDM will automatically be able to leverage that expansion. A current Portland State University (PSU) project involves loading WIM data into PORTAL with the purpose of calculating truck travel times throughout the state of Oregon. Once that project work is completed, the WIM data and associated travel times will be automatically available to the OFDM.

3. CURRENT DATA SOURCES AND CHARACTERISTICS

The OFDM combines a set of diverse data from disparate data sources into a single map-based interface. This interface provides an easy-to-use interface for accessing freight related data while adding geographical context. The OFDM contains data from several sources including the Port of Portland, the Oregon Department of Transportation (ODOT), Portland metropolitan planning organization (Metro), PORTAL transportation data archive, and research analysis and results of several transportation performance related projects at PSU. This section describes the types of data currently contained in the OFDM as well as data sources and original formats. Table 1 summarizes data sources and their characteristics; the disparate data formats include: GIS shape files, Adobe PDF files, Microsoft Word documents, Comma-Separated Value (CSV) files, Microsoft Excel files, and PORTAL data.

Port of Portland

The Port of Portland is one of the major ports in the Pacific Northwest. The Port has an active role in the study of freight movements in the region. Freight data from a recently commissioned data collection study includes truck following studies, truck counts around the Portland metropolitan area, and truck trip generation at major freight facilities (such as a terminal at the Port). This data was collected using counts and surveys and the final deliverables were a series of reports and sets of data in spreadsheets and GIS files.

PORTAL Data Source

As mentioned above, PORTAL archives a wide variety of transportation-related data for the Portland region [21]. PORTAL has been archiving speed, volume and occupancy data from sensors on the Portland-area freeways since July 2004. PORTAL also stores weather, incident, freeway dynamic message signs (DMS), bus movement data from Tri-Met (the local transit agency), and data from twenty-two weigh-in motion (WIM) stations across the state of Oregon. Most of this data is provided to PORTAL by the Oregon Department of Transportation. An initial selection of PORTAL data has been incorporated in the OFDM including the freeway sensor data, which is used to plot highway speeds and reliability, and truck incident data. Data stored in PORTAL is easy to integrate into the OFDM. PORTAL's tabular-style database lends itself to the generation of HTML content and geographical position information compatible with the GM interface. PORTAL data sets, including the freeway sensor data are regularly updated and the retrieval and storage of the freeway sensor data is fully automated. The retrieval and storage of incident and WIM data is semi-automated. When additional data of these types is received and stored in PORTAL, the new data is automatically integrated into the OFDM.

Other Data Sources

The Intelligent Transportation Systems (ITS) Lab at Portland State University leads many transportation-related research projects. Many of those projects produce data which is pertinent to freight transportation. Examples include bottleneck locations produced by recent projects on travel time estimation [22] and

---

[1] Google Maps and Google Earth are trademarked products. For sake of brevity we omit the ® sign henceforward.

automated bottleneck identification [21] as well as truck travel times derived from WIM data. The results of these projects may be in reports or in the PORTAL database. Land use maps are provided by Metro.

Table 1 OFDM Data Sources

Name	Data Provider	Type of Data	Source Instrument	Collection Metadata	Analysis Reports
Truck Incidents	PORTAL/ODOT	Database	Input by ODOT ATMS operators	Database field descriptions	
Truck Volumes	Port of Portland	Survey Data, Excel File	Field Collection, consultants	Report description	
Truck Generators	Port of Portland	Survey Data, Excel File	Field Collection, consultants	Minimal metadata	
Bottlenecks	ODOT/OTREC Project	Text Data	Loop detectors, ground truth GPS data	Detailed description and methodology	Continuous collection and analysis - Reports
Weigh-In-Motion Stations	ODOT	Database	Scales, transponder readers	Detailed description and methodology	Continuous collection
Highway Speed and Reliability	PORTAL/ODOT	Database	Loop detectors, cameras	Detailed description and methodology	Continuous collection and analysis - Reports
Freight Volume Maps	Metro	Maps	Variety of truck counts.	No metadata	
Land Use Maps	Metro	Maps	Norms and regulations	No metadata	

4. INTEGRATION AND IMPLEMENTATION CHALLENGES

Data integration is a process of assimilating data from different sources and formats. The OFDM integrates a wide variety of freight-related data into a single map interface. One of the main challenges of developing the OFDM prototype was the integration of these diverse data sources in an intuitive and useful way and to also add geographic context and the ability to associate and connect data through the use of a map-based interface. In addition, metadata, or data about the data, is poor or non-existing in many data sources (see Table 1). The lack of metadata regarding data collection methods, data semantics, and basic data description greatly complicates the data integration process. This section describes several challenges encountered during the course of the OFDM development.

Geo-location

Geo-location information is provided in a number of different ways; however, the GM API, requires a geographic coordinates, a (latitude, longitude) pair, to display a marker and a list of geographic coordinates to display a polyline or polygon. A subset of the PORTAL incident data is geo-coded with geographic coordinates; typically the major incidents that cause significant traffic delays are also geo-coded. Displaying such geo-located incidents on the OFDM map is relatively straightforward—the incident meta-data (time, description, level, etc.) is retrieved from the database along with the geo-location of the incident

and this data is used to create a GM marker for the incident. The rest of the incident data is typically geo-located by specifying the primary roadway on which the incident occurred and the nearest cross street.

Identifying geographic coordinates information requires identifying the coordinates of roadway intersections, which is typically a manual process requiring some human intervention. The PSU Travel Time study provides bottleneck location in terms of a text description, highway corridor, and approximate milepost; this data was geo-located with the help of GM itself, which can provide coordinates for a point clicked on a GM map. Since the number of bottlenecks was small, this manual method of geo-location was acceptable. The FHWA bottleneck study provides LRS identifiers, which require GIS software to convert to latitude, longitude information. Some of the truck volume and truck generator data was geo-located by hand based on text descriptions of collection locations. Geo-location of WIM stations was done using GM satellite images and approximate highway and milepost information from ODOT.

In the process of testing the OFDM interface, it was observed that geo-location information varies greatly in accuracy. From close observation of the incident data and comparing geo-location with text description, it is clear that the ODOT ATMS incident geo-location information is limited in its accuracy. In contrast, the accuracy of the Highway Speed and Reliability data is quite high as that data was derived from GPS readings taken along the highway.

Raw Data vs. Documents and Images

Raw data formats such as CSV files or Excel files tend to be easier to integrate into the PORTAL database and therefore into the OFDM. In contrast, data which is provided as figures in Microsoft Word or Adobe PDF is more difficult to integrate in a non-trivial fashion. As shown in Section 5, figures can be displayed in the OFDM; however, the data contained in these figures is difficult to integrate into the database and is available only through viewing the images. In contrast, if the data had been provided in a raw format, the data could be loaded into PORTAL and could be queried and displayed dynamically. For example, with the PSU Travel Time bottleneck data, we loaded the bottleneck data into the database and then provided the user the ability to dynamically select which bottlenecks they wanted displayed; such dynamic selection is not possible with data that is provided as an image or document. 'Raw' data lends itself to loading into the database and to dynamic and selective display. Graphs that are provided as images from a document can be stored and displayed, but integration is limited.

Maps

Land Use and Freight Volume data was provided in the form of maps. As discussed above, if the maps are provided as images, the integration is limited to displaying those images to the user. In the case of the Land Use and Freight Volume maps, despite the fact that geo-location data was clearly available at one time, the Land Use and Freight Volume maps has not yet been integrated into the GM display due to the fact that the maps were converted into images. The same problem would occur for maps provided as output from modeling software with limited capabilities for producing latitude-longitude data. However, maps provided in formats such as shape files or KML/KMZ files (the format for Google Earth), can be displayed in a GM interface.

Data Overlap

For many types of information, such as bottleneck locations and freeway speed and travel time, there are several potential data sources. In the case of bottleneck locations, the project had at least three possible sources of data: FHWA bottleneck data [23], the PSU Travel Time Project [22], and the PSU Bottleneck Identification project [24]. All three of these sources provide information about bottlenecks on highways in Oregon, but the scope and type of information provided varied.

The PSU Bottleneck Identification project is investigating automatic bottleneck identification for freeways in the Portland area using the PORTAL data archive. So far, this project has produced a list of possible bottlenecks on the I-5 freeway in Portland [24]; the information provided by this study is bottleneck location in terms of highway, time of the day, and milepost. The PSU travel time project has identified bottlenecks across the freeways in the Portland region; this identification was based on data from the PORTAL data archive and examination of the over 500 ground truth (probe vehicles) travel time runs

collected by that project. The information provided by the travel time study includes bottleneck location (highway id and milepost), activation time, approximate average length of time bottleneck is activated, approximate average extent (in miles) of bottleneck, and a description. Finally, the FHWA provides information on bottlenecks across the state of Oregon (including rural bottlenecks), in contrast to the two previously-described sources which focus only on the Portland metropolitan area. Also in contrast to the two PSU sources, the FHWA provides much greater detail about each bottleneck, including a number of estimated performance metrics, such as AADT, AADTT, Percent Trucks, Annual Truck Hours of Delay, and also classifies bottlenecks. Thus the distribution of the bottleneck locations and the meta-data available about the bottlenecks varies greatly.

At this time, only the bottlenecks from the PSU travel time project have been incorporated into the OFDM. We are in the process of incorporating the additional bottleneck data and several questions have aroused. Should we integrate all three sources, or just one source? If we use multiple sources, should we make separate layers from bottlenecks from separate sources? Making separate layers gives the user flexibility, but may be confusing to a user which does not know how best to select between different sources. If we put multiple sources in one layer, how do we deal with the fact that different bottlenecks have different meta-data? Finally, are there accuracy differences between different sources and if so, how can those differences be communicated to the user?

System Maintenance and Information Overload

Transportation data storage and visualization requires a significant and continuous time commitment and financial support. With any large system, time and resources are required simply to keep the system up and running—hardware and software must be maintained and upgraded; bugs and problems with the system will be discovered and must be fixed and addressed; users need to be trained and their questions must be answered; and finally changes to systems that interact with the transportation archive must be handled—for example, PORTAL receives data directly from the ODOT ATMS, so ATMS upgrades often necessitate PORTAL maintenance.

As new technologies provide an increasing ability to collect large amounts of data, information overload may cloud essential knowledge. With the proliferation of sensor technology, collecting vast amounts of data is inexpensive and relatively easy; the problem then becomes analyzing, filtering and mining that data. In fact, in many cases, people desire to use data collected for uses different from the originally intended use. For example, the ODOT freeway sensors were installed for the purpose of adaptive ramp metering; however, the data from those sensors is now used for travel time estimation, performance metrics and a wide variety of research projects. WIM data which is collected for truck preclearance can now be used to analyze truck volumes and truck travel times. While it is efficient to use already-collected data, in such situations, several issues arise. First, one may simply have much more data than one needs and will need to decide whether to use all of the data or just a subset. In addition, the data quality requirements of the application for which the data was collected may be weaker than the data quality requirements of the new application, so data will need to be cleaned and filtered. Storing and analyzing the vast quantities of freight data which will naturally be collected over the next decade will require careful consideration of system architectures and careful application of emerging technologies to ensure the data is put to its best use.

5.   DATA VISUALIZATION CAPABILITIES

Integrating the data in one central map-based interface greatly enhances a user's ability to relate and correlate data and to understand the context and meaning of the data. The OFDM data is displayed as points and polylines on the map with associated meta-data available via a mouse click or two, or as separate maps or graphs. In this way, all data that is reasonably associated with geographic information is displayed on the map. Figure 2 shows a screenshot of the main page of the OFDM. The primary components of this page are the Data Layers control, which appears on the left side of the screen, and the Map which takes up the majority of the visual field. From this figure, one can see that there are many data layers available to the user in addition to two layers for Google services (Google Traffic and Google Street View). Each layer has a check box and a name. The check box can be used to turn the display of each layer on and off. Most layers also have a small icon between the checkbox and name, which indicates the marker that is used to represent that layer on the map. In addition, passing the mouse over a layer name displays a popup window

with a brief description of the layer and clicking on a layer name takes the viewer to the documentation page which will provide details about that layer. Several of the layers: Bottlenecks, Truck Incidents, and Highway Speed and Reliability have additional options that can be used to further select which data is displayed for those layers. A description of some of the data layers and their features is presented next.
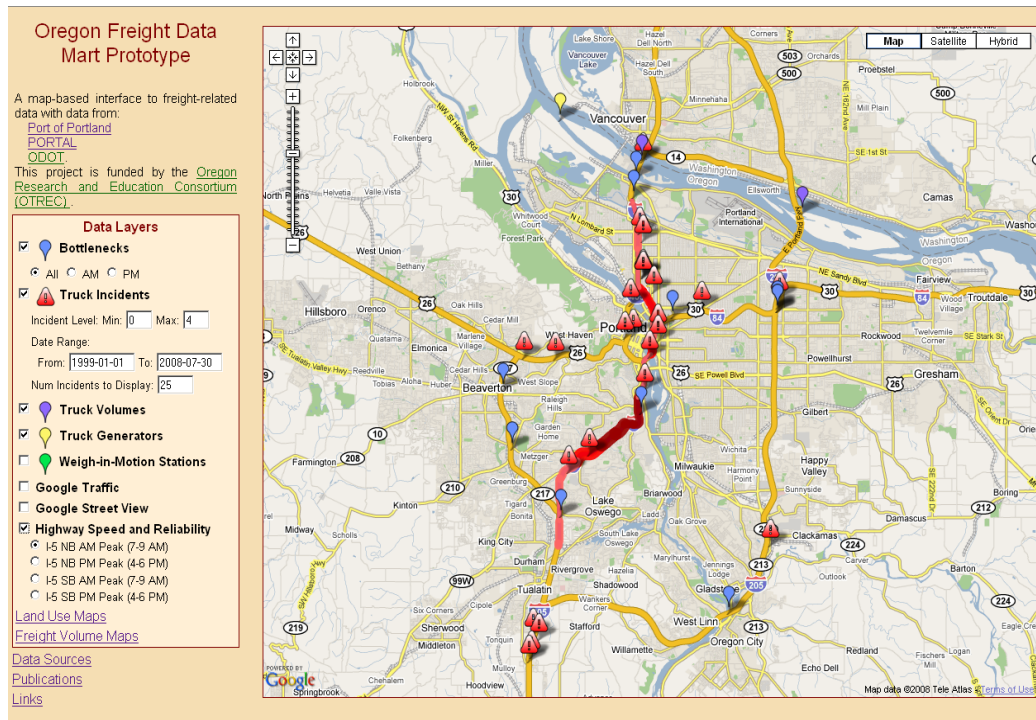

Figure 2 Freight Data Mart Main Screen

Truck Incidents

The PORTAL data archive includes data from the Oregon Department of Transportation on incidents since July 1999. From the incident database, only incidents involving a tractor trailer, a railroad, or a hazardous materials were included in the OFDM (as shown in Figure 3). Each incident is marked on the map with a caution symbol. If the user clicks on the incident location, more detailed data about the incident is obtained. Additional information that is displayed includes incident time, type of incident, the number of trucks involved, railroad cars involved, and the presence of hazardous materials. The user can restrict the date range and level of incidents displayed and also can control the maximum number of incidents to display. If the date range and level produce more incidents than the specified maximum, only the most recent incidents are displayed. This functionality of allowing the user to control which incidents to display directly results from the storage of the incident data in PORTAL.
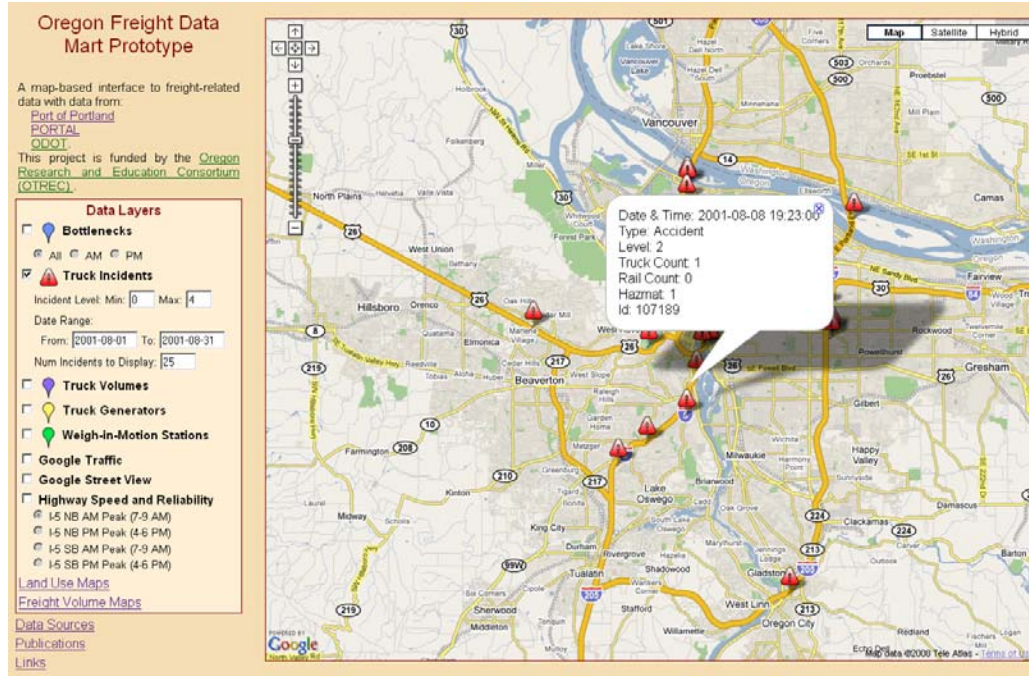
Figure 3 Truck Incident Display

Truck Volume
Truck volume data was collected in a recent freight study [25]. Truckers were surveyed and their responses tabulated to provide information about origin-destinations and trips from key freight generators. Figure 4 shows the map-based display of the truck volume information. Truck volume is available for display in the OFDM only for selected sites and corridors; data for the I-5 corridor is shown in Figure 4. As with the truck incident layer, a user may click on each marker or corridor polyline to retrieve additional information. In this case, a pop up appears with a brief description of the location or corridor and contains a link for 'Additional Information'. Clicking on the link retrieves a web page with a set of graphs, e.g. a graph detailing truck volume by time of day. Figure 5 shows a portion of the figures shown for the I-5 Freeway at Interstate Bridge location.

Highway Speed and Reliability
The OFDM uses freeway sensor data from PORTAL to provide corridor speed and reliability information. Figure 6 shows speed and reliability information for the I-5 NB corridor for the AM peak. In this figure, speed is indicated by the color of the line and reliability indicated by the line width [26]. The key shown indicates how speed and reliability (standard deviation) are translated into colors and line widths. In the future, when one clicks on a link in this segment, one will be automatically directed to plots generated by PORTAL that provide detailed information about that location. Thus data from PORTAL data archive is integrated into OFDM in a visually interactive fashion.
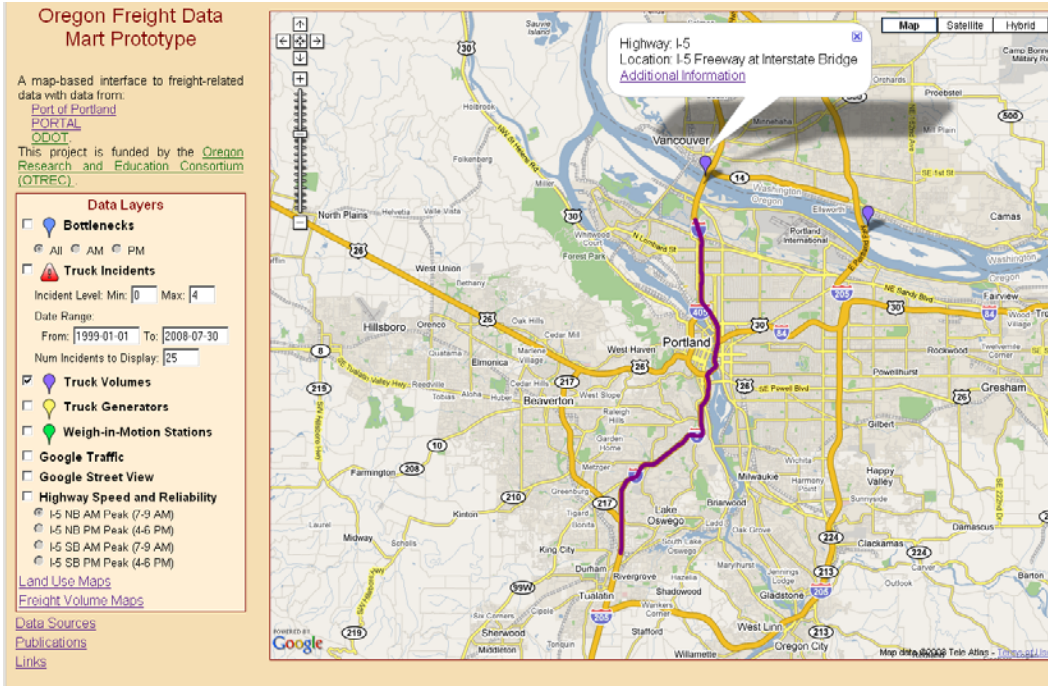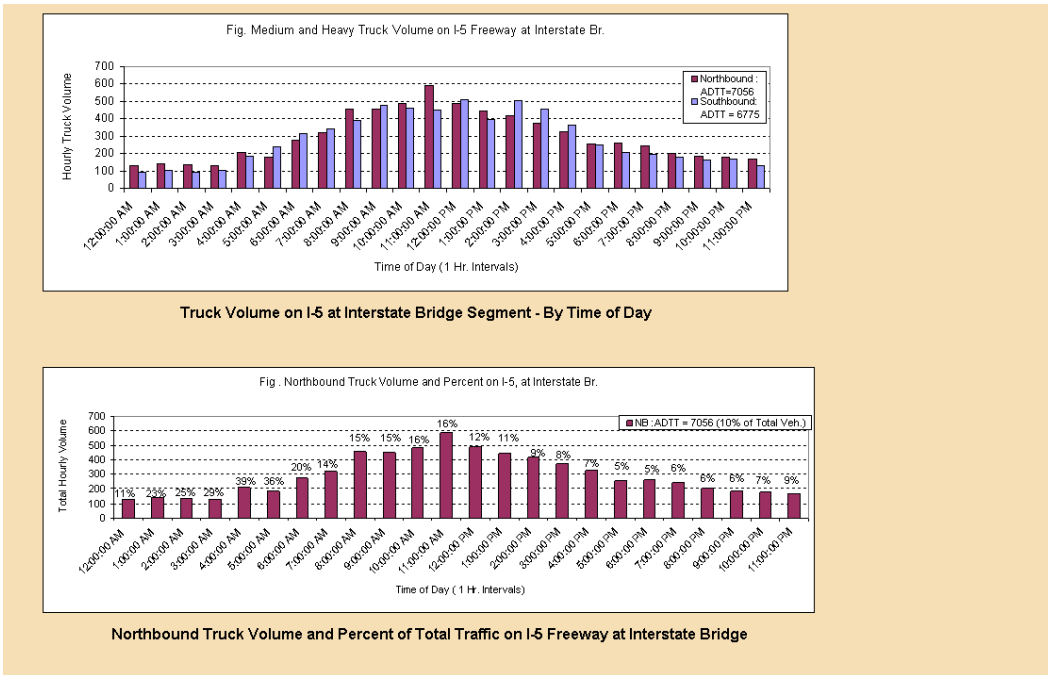
Figure 4 Truck Volume Display – Map View

Figure 5 Truck Volume Display – Detail for I-5 at Interstate Bridge

Figure 6 Highway Speed and Reliability Display

Google Services

A significant advantage of integrating the freight data into a GM interface is the ability to leverage other Google services. At this time, the OFDM incorporates Google Traffic and Google Street View (http://maps.google.com/). Figure 7 shows a geo-located incident and the use of Street View to view the incident location. The ability to view the location of an incident along with information about the incident is quite powerful. Further, the integration of existing freight data with the Google Maps application means that as Google provides more services, the OFDM can automatically take advantages of these services, for example Google Traffic can be used to display real time traffic information.

Figure 7 Google Street View of Incident Location

6.   RECOMMENDATIONS FOR FREIGHT DATA COLLECTION

Powerful lessons for improving freight data collection and communication with minimal cost can be learned from the prototype implementation experience. Integrating diverse freight data in on online mapping system sheds light on intrinsic weaknesses of current data collection methods. We describe weaknesses in data collection discovered during the development of the OFDM and make recommendations for improving the quality of freight data collection. Weaknesses include lack of statistical analysis, documentation, metadata, and geographical information; recommendations include rethinking data collection methodologies, using an Internet mapping mindset, to take full advantage of technology and to improve data quality and the visualization of performance measures.

Metadata is an important but often-neglected part of data collection. Metadata, or "data about data", often includes only basic information such as time and location of collection; however it should be expanded to also include "collection method metadata". For example, in the context of a photograph, the data is the photographic image and metadata typically includes the date and GPS coordinates of the photo; collection and method metadata might include the resolution of the image and information about who or what took the photograph. In the context of a traffic count, the data is the number of vehicles counted, the metadata should include GPS coordinates and the "collection method metadata" would include a picture of the location and installation, model of device used, crew members involved and so on. In addition, semantic data models must also be developed to communicate how different pieces of information relate to each other. For example, links can be created between different data sources that provide similar information,

e.g. traffic counts. Such modeling may provide a new dimension of accuracy if the semantic model providing the network of concepts and the relationships between those concepts is correctly applied [27, 28].

Traditional data collection and reporting methods have not been updated to be Internet mapping friendly. We discovered a frequent lack of documentation of geographic detail and data collection procedures. In certain cases, geo-location information was limited to textual descriptions of collection locations, which is not sufficient for integration into a map-based display. Give that the current cost of GPS logging devices is minor; field data collection endeavors should provide GPS location data. As seen in the previous section, databases or spreadsheets without temporal and geographic location data unnecessarily increase the cost, time, and complexity of data integration. One minute spent in GPS location logging can have a significant payoff throughout the life of the collected data. Similarly, text, or reports, should provide geographic data linking photographic or video recordings to temporal and spatial data. For example, current mapping technology supports transparent access to meta-data; for example, by clicking on the location of an accident a user can immediately access the page of a safety report where photographs of and related information about that accident is contained. Vice versa, in an accident report, there should be a link from the accident photograph to a digital map of the accident location. Public sector agencies should update data collection efforts and procurement practices to standardize the acquisition and access of digital geographic and contextual data using: GPS loggers, photographs, video recordings, traffic cameras footage and other such technologies.

As more data is collected and displayed, data quality issues, including justifiable statements of uncertainty and error, are becoming increasingly important. Data should also be supported by information regarding any sampling or statistical analysis that took place before or after the data collection itself. Users and researchers will benefit from this statistical metadata or "data about data"; in addition, such data can inform the design of future data collection efforts. For example, reports analyzing congestion and bottlenecks have been linked to the existing OFDM prototype and can be accessed via the "Bottleneck" layer. It is strongly recommended that future data collection efforts include metadata indicating the accuracy of measurements in all data sets; currently this metadata is not available in most cases as illustrated by Table 1.

In integrating data into the OFDM, we also observed differences between data from one-time outsourced data collection efforts and continuously-collected data. In any data collection effort, statistical analysis should be performed prior to data collection to estimate required sample sizes; this analysis should also be documented. With the outsourcing of data collection, such analysis is not always performed or documented. In addition, continuously-collected data can be analyzed and checked for data quality with feedback provided to the collecting agencies to improve data collection and communication methods. This feedback loop makes for an improved data set in contrast to outsourced one-time data collection. As technology evolves, a move toward continuous (or at least periodic) and automated collection systems can improve the quality of freight data.

Freight performance measures that take into account data mapping and communication should be developed. Given the premise that most transportation data has a strong spatial component, transportation performance measures should also be expected to have a strong spatial component. The synergy provided by internet mapping allows the visualization and integration of freight-related performance measures and data. For example, the combination of GIS land use data with GIS-GPS truck trip data can provide invaluable insights regarding truck trip demand generation [29] and the regional significance of freight corridors. Similarly, GPS freight data can be effectively combined with Weigh in Motion (WIM) data; although most truck weight and payload information is generated for pavement management purpose it can also be used to estimate the distribution of payloads [30] and analyze the efficiency of urban freight systems [31].

We argue that institutional barriers and a pre-Internet mapping mindset, not technology, are the most demanding hurdles to implementing a freight data web-based mapping application. Leveraging existing applications, we have developed the OFDM prototype so that the software and hardware details are hidden behind standard network appliances and protocols. The OFDM users and information providers are freed from having to know about the details of the low-level technical infrastructure and equipment. Hence, as

the OFDM is expanded with more data and features, the data integration and visualization challenges may be less influenced by technology than by inappropriate data procurement and collection.

7. CONCLUSIONS

This paper presents the core of a digital application for freight data collection, integration, use, and communication. The OFDM system uses an innovative design that combines a searchable database, visualization capabilities provided by a publicly-available online mapping browser (Google Maps) with links to relevant contextual information including reports and documentation. The current implementation focuses on highway performance, safety, truck volumes, bottlenecks, and land use data.

The goals that guided the OFDM application design are to: (a) leverage existing applications to reduce deployment time, (b) employ publicly-available applications and interfaces to reduce development costs, (c) provide an intuitive, user-friendly interface with a minimal learning curve, (d) produce a prototype that can be easily updated, and (e) create a system that can readily incorporate future formats and technologies. This application is expected to provide guidance for freight data collection methods and the design of freight data semantics and protocols.

The largest long-term challenges seem to be rooted in outdated data collection methods and data delivery. Procurement of data must be forward thinking and must incorporate up-to-date technology to record temporal, geographic, and contextual data. In addition, as more data is collected and displayed, data quality issues, including justifiable statements of uncertainty and error, will become increasingly important to avoid information overload and facilitate decision making.

REFERENCES

1. TRB., *Special Report 276: A Concept for a National Freight Data Program.*, in *Committee on Freight Transportation Data, National Research Council, Washington, D.C.* 2003.
2. Fletcher, D.R., *Geographic information systems for transportation: A look forward.* Transportation in the New Millenium: State of the Art and Future Directions, 2000.
3. Butler, D., *Virtual globes: The web-wide world.* Nature, 2006. **439**(7078): p. 776-778.
4. Jacoby, S., et al., *Developing a common spatial data infrastructure between State and Local Government--an Australian case study.* International Journal of Geographical Information Science, 2002. **16**(4): p. 305-322.
5. Dueker, K.J. and J.A. Butler, *GIS-T Enterprise Data Model with Suggested Implementation Choices.* Journal of the Urban and Regional Information Systems Association, 1998. **10**(1): p. 12-36.
6. Dueker, K.J. and J.A. Butler, *A geographic information system framework for transportation data sharing.* Transportation Research Part C, 2000. **8**(1-6): p. 13-36.
7. Abdelmoty, A. and C. Jones, *Towards maintaining consistency of spatial databases*, in *Proceedings of the sixth international conference on Information and knowledge management.* 1997, ACM: Las Vegas, Nevada, United States.
8. Cobb, M.A., et al., *A Rule-based Approach for the Conflation of Attributed Vector Data.* GeoInformatica, 1998. **2**(1): p. 7-35.
9. Trépanier, M. and R. Chapleau, *Linking Transit Operational Data to Road Network with a Transportation Object-Oriented GIS.* URISA Journal, 2001. **13**(2): p. 23-30.
10. Darter, M.T., T.A. Lasky, and B. Ravani, *Transportation Asset Management and Visualization Using Semantic Models and Google Earth.* Transportation Research Record, 2007. **2024**(-1): p. 27-34.
11. Harkey, D.L., *Evaluation of Truck Crashes Using a GIS-Based Crash Referencing and Analysis System.* Transportation Research Record, 1999. **1686**(-1): p. 13-21.
12. Chien, S., G. Li, and J. Daniel, *Geographic Information System-Based Truck Accident Information and Management System for New Jersey Roadways.* Transportation Research Record, 2002. **1818**(-1): p. 134-142.
13. Casavant, K.L., et al., *MODELING WASHINGTON STATE TRUCK FREIGHT FLOWS USING GIS-T: DATA COLLECTION AND DESIGN.* 1995.
14. Alam, M. and E. Fekpe, *Application of Geographic Information Systems Technology in Freight Data Analysis: Case Study of I-90/I-94 Corridor Analysis.* Transportation Research Record, 1998. **1625**(-1): p. 173-183.
15. McCormack, E. and M. Hallenbeck, *ITS Devices Used to Collect Truck Data for Performance Benchmarks.* TRANSPORTATION RESEARCH RECORD, 2006. **1957**: p. 43-50.
16. FHWA, *Freight Performance Measurement: Travel Time in Freight-Significant Corridors.*, in *FEDERAL HIGHWAY ADMINISTRATION, FHWA Report HOP-07-071. Available at http://www.ops.fhwa.dot.gov/freight/freight_analysis/perform_meas/fpmtraveltime. Accessed 08/11/07.* 2006.
17. SURESHAN, S., *ONTARIO COMMERCIAL VEHICLE SURVEY: USE OF GIS FOR DATA COLLECTION, PROCESSING, ANALYSIS, AND DISSEMINATION*, in *Ontario Ministry of Transportation, accessed November 10, 2007, http://www.trb.org/Conferences/2006/FDM/Sureshan.pdf.* 2006.
18. Greaves, S. and M. Figliozzi, *Urban Commercial Vehicle Tour Data Collection Using Passive GPS Technology: Issues And Potential Applications.* Transportation Research Record 2049, 2008: p. 158-166.
19. Elango, V.V., et al., *Visualization of Freight Movement with GT Freight Data Collector and Real-Time Cargo Tracking.* Transportation Research Board Annual Meeting, 2007.
20. Ramachandran, B. and J. Klodzinski, *Internet-Based Mapping Prototype for Integration with an Electronic Freight Theft Management System.* Transportation Research Record, 2007. **1993**(-1): p. 101-108.
21. Bertini, R.L., et al., *PORTAL: Experience Implementing the ITS Archived Data User Service in Portland, Oregon.* Transportation Research Record 1917, 2005: p. 90-99.
22. Tufte, K., et al., *Toward Understanding and Reducing Errors in Real-Time Estimation of Travel Times. Forthcomming Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., 2008.* 2008.

23. White, K. and L. Grenzeback, *Understanding Freight Bottlenecks*, in *Public Roads, March/April, Vol. 70, No. 5*. 2007.

24. Bertini, R., et al. *Using Archived Data to Systematically Identify and Prioritize Freeway Bottlenecks*. in *Applications of Advanced Technologies in Transportation (AATT) 2008 Conference Proceedings*. 2008. Athens, Greece, May 2008.

25. Systematics, C., *Portland Freight Data Collection Project, Phase II*. 2007.

26. Bertini, R., et al., *DERIVING METROPOLITAN PLANNING PERFORMANCE MEASURES FROM A REAL-TIME REGIONAL TRANSPORTATION DATA ARCHIVE.* Proceedings of the 9th International Conference on Computers in Urban Planning and Urban Management Conference (CUPUM), 2005.

27. Shaw, S.L. and X. Xin, *Integrated land use and transportation interaction: a temporal GIS exploratory data analysis approach.* Journal of Transport Geography, 2003. **11**(2): p. 103-115.

28. Goodchild, M.F., *GIS and Transportation: Status and Challenges.* GeoInformatica, 2000. **4**(2): p. 127-139.

29. Fisher, M. and M. Han, *NCHRP SYNTHESIS 298, Truck Trip Generation Data*. A Synthesis of Highway Practice. 2001, WASHINGTON, D.C.: TRANSPORTATION RESEARCH BOARD - NATIONAL ACADEMY PRESS.

30. Figliozzi, M.A., R. Harrison, and J.P. McCray, *Using weigh-in-motion data to calibrate trade-derived estimates of Mexican trade truck volumes in Texas.* Transportation Research Record 1719, 2000(1719): p. 129-135.

31. Figliozzi, M., *Analysis of the Efficiency of Urban Commercial Vehicle Tours: Data Collection, Methodology, and Policy Implications.* Forthcoming Transportation Research - part B, 2007.

LIST OF TABLES

LIST OF FIGURES