Final Research Report
Research Project T2695, Task 32
Statewide Archive

STATEWIDE ARCHIVE

by

Adam Sanderson
Software Engineer

Mark E. Hallenbeck
TRAC Director

Washington State Transportation Center (TRAC)
University of Washington, Box 354802
University District Building
1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

Washington State Department of Transportation
Technical Monitor
Pete Briglia
State ITS Engineer, Advanced Technology Branch

Prepared for

Washington State Transportation Commission
Department of Transportation
and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

September 2005

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 627.1		

4. TITLE AND SUBTITLE	5. REPORT DATE
STATEWIDE ARCHIVE	September 2005
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Adam Sanderson and Mark E. Hallenbeck	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Washington State Transportation Center (TRAC) University of Washington, Box 354802 University District Building; 1107 NE 45th Street, Suite 535 Seattle, Washington 98105-4631	11. CONTRACT OR GRANT NO. Agreement T2695, Task 32

12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Research Office Washington State Department of Transportation Transportation Building, MS 47370 Olympia, Washington 98504-7370 Doug Brodin, Project Manager, 360-705-7972	Final Research Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. ABSTRACT

This report describes an initial effort to develop intelligent transportation system (ITS) data archives that can be linked and accessed through a single, Web accessible, geographic information system (GIS) interface. This project was designed to test where this approach of linking disparate databases can help resolve some of the key issues associated with making transportation system performance data available throughout the Washington State Department of Transportation (WSDOT). These include 1) keeping the basic data structures manageably simple to reduce database cost and complexity, 2) allowing data archive control and primary management to remain at the local level to improve the quality control function, 3) providing easy access to staff throughout the organization, 4) providing an interface that allows staff unfamiliar with the data to easily learn what data are available in each database, and 5) providing a simple way to allow staff to combine disparate datasets that share geographic characteristics.

To test the concepts developed for this project, the project team created three specific databases and linked those databases through the spatial identifiers stored in WSDOT's GIS. Summaries statistics from each of the three databases were developed to be useful to a wide range of WSDOT staff, and that are available through the Internet.

The prototype map interface to the three databases can be found at the following URL: http://trac29.trac.washington.edu/tracmap/mapserver.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Traffic data integration, ITS data archives, ADUS	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None		

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Washington State Transportation Commission, Department of Transportation, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

CONTENTS

FIGURES

EXECUTIVE SUMMARY
STATEWIDE ARCHIVE

The Washington State Department of Transportation (WSDOT) is currently deploying a variety of intelligent transportation systems (ITS). These systems routinely include the deployment of data collection sensors. The data collected are then used in near-real time to make a variety of traffic management decisions. Storage and analysis of these data allow evaluation of the performance of these ITS, as well as improved management decisions based on the analysis of the archived data.

Data collected by ITS deployments can often also be useful for other non-traffic management purposes within both WSDOT and its partner transportation agencies in the state. Consequently, the collection, storage, retrieval, analysis, and reporting of data collected by ITS has considerable potential value to the state. However, just because ITS data are collected and stored does not automatically make those data useful. Two additional actions must occur before the potential benefits of these data can be realized.

First, because the data collected for real-time traffic management are often very fine-grained, they must be aggregated and summarized into more traditional engineering statistics to be useful to most users.

Second, because these data are tied directly to discrete traffic management systems, often only the staff operating the ITS are aware of the data's existence, understand how it was gathered, and have access to them. Thus a mechanism must exist that 1) allows users to easily determine what data are available and then 2) provides those interested users with access to those data at a useful level of detail.

This project was designed to help WSDOT explore the issues involved in storing, analyzing, using, and reporting ITS data related to traffic system performance. It created three specific, Web accessible, ITS-related databases and linked those databases through the spatial identifiers stored in WSDOT's geographic information system (GIS).

The databases created are as follows:

- the CVISN[1] truck tag database, which computes truck travel times between CVISN tag reader locations and can be used to provide measures of roadway system performance
- the Average Car Occupancy database, which makes accessible multiple years of vehicle occupancy counts collected for WSDOT as part of the Department's ongoing evaluation of the performance of HOV lanes in the Puget Sound region
- the Summary FLOW database, which provides access to summary statistics developed from freeway operations data collected by WSDOT's Northwest Region's traffic management center (the FLOW system).

Summary statistics from each of the three databases were developed to be useful to a range of WSDOT staff. All three datasets are accessible through an online, map-based interface, based on WSDOT's maps and linear referencing system, at the following URL: http://trac29.trac.washington.edu/tracmap/mapserver. Details concerning these prototype databases, and the lessons learned while they and their user interfaces were developed, are described in the main body of this report.

[1] CVISN = Commercial Vehicle Information and Safety Network

STATEWIDE ARCHIVE

INTRODUCTION

Intelligent transportation systems (ITS) generate enormous amounts of potentially useful information. Everything from electronic tag readers to loop detectors and the traffic signal controllers to which those loop detectors are connected generate data that may be harnessed to provide valuable performance measures that can, in turn, be used to more effectively operate transportation systems and manage the staff and resources that build, operate, and maintain those transportation systems.

Unfortunately, in many cases these data are not stored, and in many more cases even when they are stored, the data are not in a useful level of detail, or made readily accessible for use in disparate analyses across an agency.

ITS data archives tend to be large in size and complex in their data structures. They also tend to be geographically isolated, as most ITS deployments affect relatively limited miles of roadway in the state. These factors tend to make ITS data, even when they are stored, difficult for many agency staff to use. In many cases few people know that the data exist. Even when staff know that an archive exists, they often are not familiar with how to access the archive, how the archive functions when they do find it, what data exist inside the archive, or how those data can be appropriately used.

One final problem with creating and maintaining useful ITS data archives involves the resources and staffing they require. Because of their size and complexity, these archives do require some staff time and resources for maintenance. Those staff resources are most effectively applied at the regional level, where the field equipment is

located and where a direct understanding of the roadway system performance is most available and desired. This allows staff to identify and resolve problems quickly, and because they are the primary users and benefactors of the archive (it is their management systems that can be made to function better through use of the archive's output), problem identification and resolution tend to happen quickly. However, if too "local" an approach is taken in maintaining an archive, the rest of the agency may not gain the wider benefits that such an archive can provide.

To resolve these issues for the Washington State Department of Transportation (WSDOT), the project team developed an overall schema for a "statewide ITS archive" that really consists of a series of archives that are linked to each other through WSDOT's geographic information systems (GIS) resources, including maps and the linear referencing system. This basic concept is illustrated in Figure 1.

Individual data archives are developed for each ITS. Each archive is designed to efficiently store, quality assure, analyze, filter, and summarize the data collected by that ITS. The data summaries useful to the rest of the Department are then placed in Web accessible tables, given the appropriate spatial linkages to allow the data to be identified through a map interface, and then selected and downloaded by the user.

The result is a collection of databases that can be maintained at the regional/local level, but whose regional data can be readily found by users throughout the Department or even outside of the Department. In addition to serving as an easy means of identifying the location of available data, use of the map interface, combined with Internet accessible data, allows these data to be easily linked to other data maintained by WSDOT.

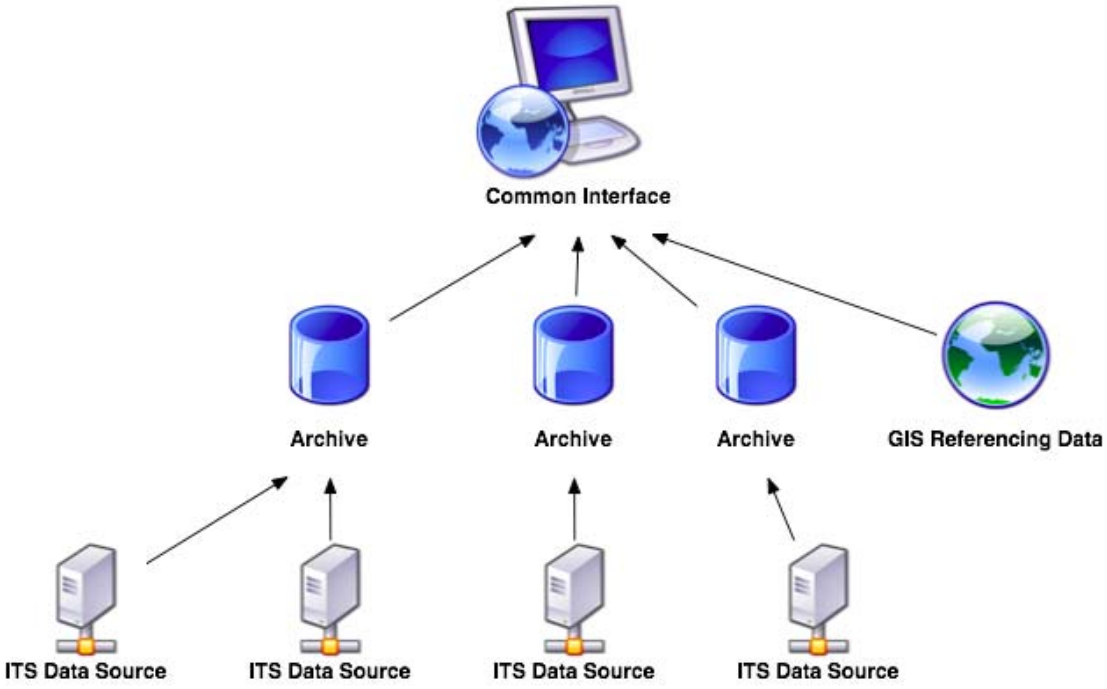

Figure 1: Schematic of ITS Data Archive Relationships

The goal of this project, then, was to demonstrate this concept of ITS archives linked through WSDOT's GIS resources. It was designed to illustrate how data archives from disparate sources can be built and connected in an easily accessible, easily used manner, so that data from those archives can be readily found and used by transportation agency staff around the state.

PROJECT APPROACH

This project developed from the existence of a number of WSDOT ITS deployments for which no data archive existed, combined with the desire to increase WSDOT staff access to the summary performance statistics developed from the Northwest Region's freeway traffic data archive. The intent was to improve access to data resources already collected by WSDOT.

The project focused on two separate aspects of the archive development process. The first was the creation of useful archives, that is, databases that successfully converted large quantities of collected "raw" data into useful summary statistics. The second was the creation of a single user interface that could both inform WSDOT staff of what data were available for their use and direct them to those data.

The output from the project is a series of databases, accessible via the Internet, that are linked through an interface based on WSDOT's GIS. The locations for which data are available through each of the databases and the types of data available at each location can be determined from a map-based user interface. Users interested in obtaining specific data items are transferred from the Web interface to the specific database interface, into which they input specific data queries. Three databases were created:

- the CVISN[2] truck tag database, which computes truck travel times between CVISN tag reader locations and can be used to provide measures of roadway system performance
- the Average Car Occupancy database, which makes accessible multiple years of vehicle occupancy counts collected for WSDOT as part of the Department's ongoing evaluation of the performance of HOV lanes in the Puget Sound region
- the Summary FLOW database, which provides access to summary statistics developed from freeway operations data collected by the Northwest Region's traffic management center (the FLOW system).

[2] CVISN = Commercial Vehicle Information and Safety Network

A fourth database, the Arterial Performance database, is currently being created as part of a separate research project and will be linked to the same user interface when it is completed.

The map interface can be found at the following URL:

http://trac29.trac.washington.edu/tracmap/mapserver

ARCHIVE CREATION

The first task of this project was to create the actual ITS archives. Each of the archives is discussed briefly below. Each data archive in the Statewide Archive will function similarly, collecting, archiving, and presenting its data. Each should be created by a WSDOT group to specifically resolve some key management function, such as improve traffic management capabilities on some stretch of highway. The archive should store the collected data to perform that improved management function, and its primary role should be to help answer questions about whether the management system is operating optimally. However, these same data are quite likely valuable to the rest of WSDOT. The focus of this project was to demonstrate how these data can be stored, used, and shared throughout WSDOT.

As new ITS systems are have been developed and old systems have been expanded, the amount of data being produced within WSDOT has grown dramatically. Storing this rapidly increasing stream of data is not a trivial task. Data must be archived in a manner that is efficient, yet easy to access and analyze. Once the raw data have been converted to “information,” the results need to be placed where users and potential users can easily find, obtain, and understand them.

Currently, a large percentage of ITS do not store the data they collect at all. Others store the data only for very short periods, or merely dump data into a log file.[3] The result is that data which WSDOT has already paid to collect are not available for use to meet other Departmental needs. The Department is then faced with paying for the collection of additional data, or making key management and policy decisions without understanding how the transportation system is actually performing.

Simply storing the raw data collected from the field in an accessible form can be beneficial, but in most cases, data must be processed further to become useful to end users. Raw data are too fine-grained to be used by most people or groups outside of the specific ITS application that collects the data. Thus, to derive meaningful performance statistics, the raw data must generally be aggregated, filtered, and summarized into more widely used performance metrics. Continued storage of performance metrics can then grow over time, allowing the analysis of the archived data to describe

- trends that are occurring in system performance
- the impacts of specific system management actions
- the success/failure of treatments applied to allow the transportation system to function more efficiently.

Therefore, the initial stage of making ITS data accessible and useful at a statewide level is to create an archive that allows it to be easily processed and reported. The goal of every archive should not be simply to store data. The goal should be to support the traffic management decisions that the ITS supplying the data is intended to improve.

[3] A 'log file' is a simple, unstructured text file. Because these files do not contain well defined data fields, it is difficult to efficiently or cost effectively find, organize, and extract data from them.

The CVISN Truck Tag Database

The Commercial Vehicle Information for Safety Network (CVISN) program is a federally sponsored initiative designed to

- decrease the time and effort required by commercial vehicles to comply with motor carrier regulations
- improve the efficiency of enforcement personnel charged with enforcing those regulations
- decrease the delays experienced in transit by vehicles with good safety records and up-to-date regulatory paperwork.

Trucks participating in the CVISN program are equipped with electronic identification tags (transponders) that are read as the trucks approach regulatory enforcement locations. The electronic ID is used to look up that vehicle's regulatory status and safety record. Vehicles in regulatory compliance and with a good safety record are then allowed to by-pass the enforcement location without stopping.

This same basic electronic tag technology has a variety of other functional uses. Tag IDs for participating trucks can be used at ports and intermodal terminals for a variety of purposes, including inventory tracking, theft control, cargo arrival and departure tracking, and various homeland security tasks.

The "raw" data provided by the system are the individual truck tag ID, the location at which the ID was read, the time the tag was read, and any CVISN database information (truck weight, regulatory status) that WSDOT wishes to associate with that tag. From the management perspective of WSDOT's CVISN staff, having an archive of tag observations allows analysis of the level of participation in the system. This in turn

allows determination of the level of benefits being obtained, as well as management review of system performance. As an example, a review of these data showed that as trucks crossed the state, the time required to make that trip could be determined by comparing the time at which a given truck tag was observed at adjacent tag reader locations. These travel times could yield considerable insight into statewide roadway performance.

For this project, the project team synthesized a database from two sources of data to help determine truck travel times. Thanks to cooperation from Transcore (a company operating CVISN readers at a number of international border locations and intermodal terminals as part of ongoing USDOT efforts), the project team was able to collect truck tag time and location data not only from WSDOT but also from a variety of non-WSDOT sites in the state.

After gathering the raw data from Transcore and WSDOT, the archive immediately makes each truck tag ID anonymous by creating a new ID to ensure the privacy of the CVISN participants. Data contained in the database can be shared with any user selected by WSDOT without concern of violating the privacy of truckers participating in the CVISN program because the actual identification number of truck tags is not maintained in the database and cannot be recreated.[4] The new, anonymous tag IDs, the location at which they were observed, and the time and date they were observed are stored in a new relational database.[5]

---

[4] The anonymity function converts a single tag to a single new tag ID at all CVISN tag locations for 24 hours. That is, on August 1st, tag 123ABC is given the new ID 987ZXY each time it is observed. This allows the "matching function" to take place. However, on August 3rd, tag 123ABC will be converted into an entirely different anonymous tag ID.

[5] The details of the CVISN database structure and coding can be found in the appendix to this report, as well as on line at the following URL: http://trac29.trac.washington.edu/wiki/page/show/TruckTagProject

The database also provides the functionality needed to convert these raw data into more useful travel time statistics. To do that, it matches tags observed at multiple tag reader locations and determines the travel times between those locations for that vehicle. These travel times are then "cleaned" to remove inappropriate tag matches. (For example, a truck may pass two tag reader locations more than once per day. A travel time could be computed by using the first pass at the upstream reader and the second pass of the downstream reader, but this value would not represent an actual trip. This "invalid" travel time would be stripped from the database.) The end results are travel times for individual trucks between specific locations in the state, statistics useful for a variety of planning purposes.

The remainder of the CVISN truck tag database consists of the user interface, which allows users to extract data for specific reader locations and travel times for specific "trips" (by time of day, and for selected date ranges) in a variety of output formats.

The CVISN truck tag database can be accessed directly at the following URL: http://trac24.trac.washington.edu/trucks.

The Average Car Occupancy Database

The second database developed under this project contains the vehicle (car and truck) occupancy observations performed by the Washington State Transportation Center (TRAC) for WSDOT to monitor the use of the Seattle metropolitan region's high occupancy vehicle (HOV) lanes. The raw data are collected by TRAC field staff on personal digital assistants (PDAs) and consist of the number of people observed in

individual passing vehicles. Time stamped data are stored for 15-minute intervals, with metadata that describe the specific location at which data are collected, the type of lane (HOV or general purpose), and the date on which the data are collected.

These data are uploaded into a new relational database that allows TRAC supervisory staff to quality check, through a Web application, the collected data and to ensure that the data collection staff went to the appropriate location. The database then computes a variety of summary statistics with these raw data.[6] Commonly computed summary statistics include average car occupancy (ACO) for a location and lane type, percentage of HOV violators, and the percentage of trucks using the HOV lanes. Summary statistics are produced for user selected time periods and can be produced in a variety of formats (e.g., as HTML or as CSV files).

Thus, like the CVISN truck tag database, this archive obtains data (in this case via download from the PDAs), stores the data in a relational database, runs quality assurance checks on the data, and provides user requested access through the Web to the summary statistics that are commonly requested by WSDOT staff.

The vehicle occupancy database can be accessed directly at the following URL: http://trac29.trac.washington.edu/hov.

The FLOW Data Summary Database

The first ITS archive that WSDOT created consists of data collected by WSDOT's Northwest Region freeway management system (FLOW). It consists primarily of 5-minute summaries of vehicle volume and lane occupancy statistics by loop

[6] The details of the Vehicle Occupancy database can be found in the appendix of this report. Programming details are available at the following URL: http://trac29.trac.washington.edu/wiki/page/show/HovProject

for all loop detectors in the system. This database is quite large, representing over 2 GB of data each year in compressed binary file formats.

This archive predates the Internet, and as a result, the archive itself is not currently stored in a Web accessible, modern relational database. However, as a result of a series of research projects that started in the mid 1990s a substantial analytical process has been constructed to convert the 5-minute data archive into a variety of summary statistics that describe the use and performance of the Puget Sound freeway system. The software that computes these summary statistics is called CD Analyst.

The statistics most commonly produced with CD Analyst include the following:

- traffic volume (by time of day, type of lane, and location)
- average vehicle speed (by time of day, type of day, and location)
- average and 95th percentile travel times for 10 specific "commutes" in the metropolitan region.
- the frequency of the occurrence of congestion both at specific locations and for entire freeway corridors (by type of lane).

The quality control functions that exist and must be incorporated as part of the CD Analyst process are modest, and therefore, the resources required to convert the raw 5-minute data into these summary statistics is considerable, even with the CD Analyst software. As a result, even though WSDOT has the ability to compute these statistics whenever they are needed, it makes considerable sense to compute the primary statistics only once and place those summaries in an easily accessible location. This project created a new database structure to allow that access.

These summaries are available through TRACMap, the map interface prototype created to provide a single entry point to all of the ITS data archives described above, as well as any additional databases constructed by WSDOT in the future. This interface is described in the next section. Summaries from the FLOW database can be accessed directly at the following URL: http://trac29.trac.washington.edu/flow/. Details on how these summary data are stored can be obtained from the TRACMap documentation site.[7]

LINKING THE ITS ARCHIVES AND MAKING THE DATA ACCESSIBLE

The next task of the project, given the creation of an archive to produce useful transportation system performance statistics, was to make it easy for staff throughout WSDOT to learn that these data are available, what the data represent, and how the data can be obtained. Factors such as the type of data stored in any given archive, the capabilities of those archives, and the utility of the extracted summaries vary dramatically from one ITS database to another, making this task more difficult. This is not surprising because these data are extracted from ITS that support a highly varied set of management functions, ranging from freeway and arterial traffic management, to snow and ice control, to construction traffic management and traveler information.

Rather than trying to force these widely disparate data into a single data structure, the project team chose to promote the integration of these disparate datasets by simply highlighting the availability of data within each archive and by providing a mechanism for transferring the user directly to any data archive that contained potentially useful data.

The "front end" of each database would need to be designed to help each user quickly learn what data are present and provide an easy mechanism for helping each user

[7] http://trac29.trac.washington.edu/wiki/page/show/FlowProject

extract those data useful for their specific analyses. This approach does require that end users develop their own methods for converting the available data into the formats appropriate for their analysis. (For example, they may need to convert data supplied in a common spreadsheet format – CSV – into a format more suitable for analysis in a statistical package, or they may need to add together data from multiple CSV files.)

Given the variety of potential uses for ITS performance statistics, it was decided that this flexible approach would serve WSDOT needs more effectively and at lower cost than trying to construct more formal data storage and retrieval systems. However, if the usage patterns for specific datasets or combinations of datasets warrant the construction of a more formal database integration between any two archives, this integration can be easily incorporated within the environment developed for this prototype. For example, at some point in the future, it may be possible (and appropriate) to provide summary statistics as GIS data layers, which could be used with other WSDOT GIS resources, in which specially written analytical code can compute routinely desired summary reports that extract data from two or more ITS archives.

For this initial prototype, it was determined that simply locating the available data on WSDOT's digital maps in order to communicate the availability of data and to provide the interface link to the desired ITS archive would provide much more utility and much more analytical flexibility, given the available budget of software development and the evolving nature of data uses within the WSDOT.

The project team decided that linking the ITS archives with WSDOT's linear referencing system was the best option for meeting these functional needs. The ESRI GIS software was selected because 1) it is a software platform supported by WSDOT,

and 2) it supports WSDOT's linear referencing system, allowing data from ITS archives to be linked. ESRI also provides support for Internet access to data stored within its GIS framework.

The Location Referencing System

The GIS framework and map interface allow the visual identification of data within specified geographic boundaries (e.g., "what traffic data exist between milepost 100 and 125 that I can match against my accident data?"), a function important for helping users understand the availability of data. Once the available data are identified, users can create ad hoc analyses that meet their needs, within their time and resource constraints.

The location references are not only useful for locating the data on the TRACMap interface. They are also key data selection variables for the automated data queries that allow users to extract the data they have identified and now desire. The location references also serve as uniform database keys that allow automation of those data extraction or integration tasks that could be performed routinely.

For now, the TRACMap interface assumes that the actual linkage of data extracted from two or more datasets will be performed manually. Manual integration was chosen because the ITS archives are new and it is unclear how these data resources will be used. (That is, it is unclear which datasets or variables from those datasets need to be more tightly integrated in order to routinely perform key analyses.) Consequently, the prototype was designed to allow flexible access to data at the lowest cost for archive development and maintenance. The user (WSDOT data analyst) is then responsible for transforming the data into formats that can be used for specific analytical tasks. This

pushes the cost of the data integration and manipulation to the analysis function and ensures that the value of the integration effort outweighs the cost of performing the required data manipulation. If the value of this analysis task is high, and the task will be performed repeatedly, then WSDOT should consider automating the steps required to extract and format the data and introduce them in a new dataset.

The Prototype Map-Based User Interface

The basic, Web accessible, GIS map interface for the prototype is shown in Figure 2. The interface is divided into four sections: the map itself, the navigation controls in the upper left, the data selection options in the lower left, and the database selection option in the middle at the base of the screen. This last section controls what data are available for selection through the interface. (In Figure 2, the data currently available for selection come from the Vehicle Occupancy database for calendar year 2004. This is shown at the base of the screen image.)

Currently, the availability of data from only one data set (and year) can be viewed at a time. As currently designed, the user selects both the ITS archive (database) and year for which data are desired. When the "update" button at the bottom of the screen is clicked, the map image is revised to show the location of data contained in that database for that year. The map interface can be accessed at the following URL: http://trac29.trac.washington.edu/tracmap/mapserver

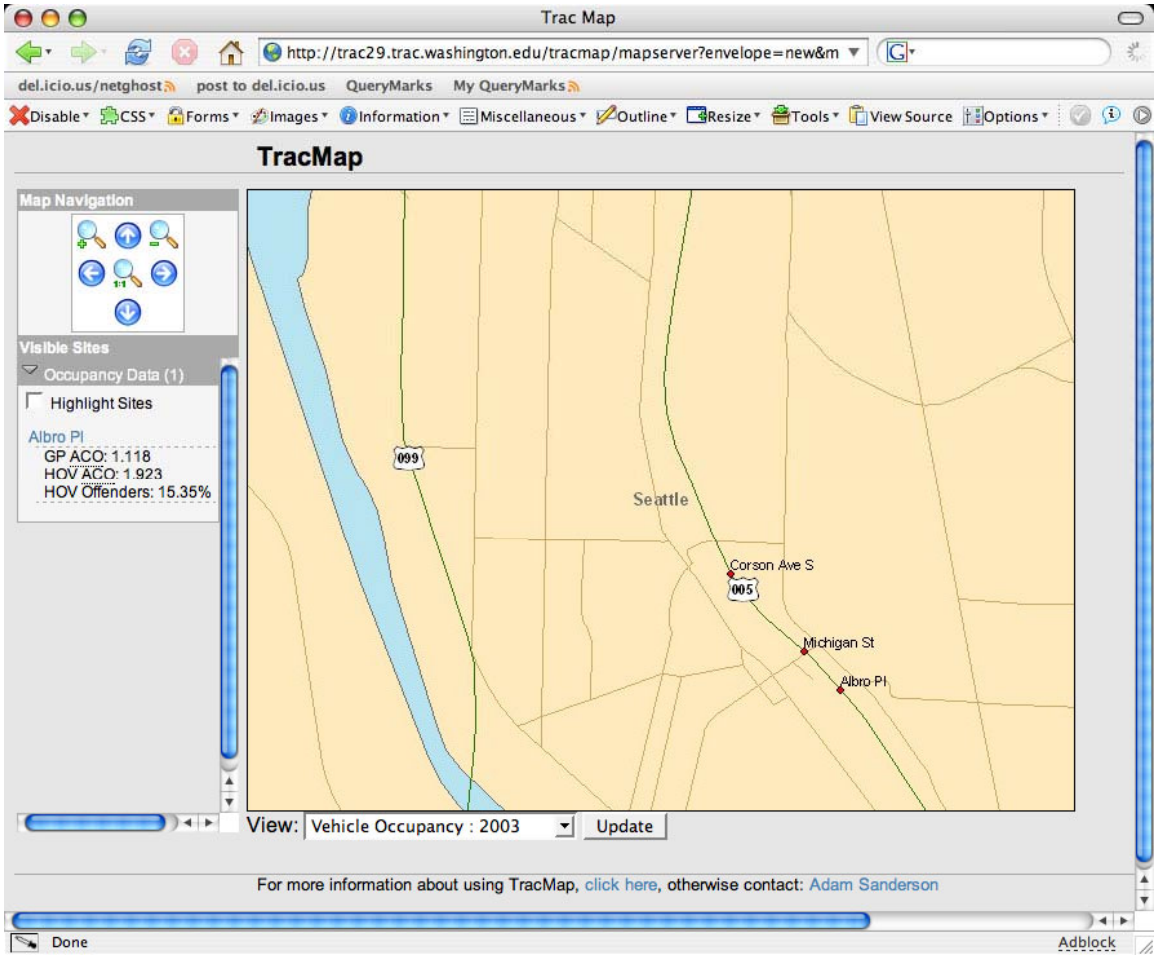

Figure 2: TRACMap Interface

The navigation section in the upper left portion of the screen allows the user to zoom in/out, as well as move the map image directionally (N/S/E/W.) Figure 3 shows a close up of the navigation section of the interface. The four "arrow buttons" allow the user to pan the map. The magnifying glass with a plus symbol zooms the map in (showing a smaller geographic area, but with greater detail), whereas the magnifying glass with a minus symbol zooms out (showing a larger geographic area). The map can be "zoomed out" until the entire State of Washington can be seen. The central button, a magnifying glass labeled "1:1" returns the map to its default scale. The user may also "click and drag" on the map. While dragging, a rectangle will illustrate a selected region,

and when the mouse button is released, the map will zoom in on the geographic region included in that rectangle.

Figure 3: TRACMap Navigation Controls

Only those data points visible on the current map image are visible in the data selection section of the interface. As the map image displayed on the screen changes (by zooming or panning), the data available for "selection" in the box on the lower left side of the screen change. (Note that it may be necessary to scroll through a list of available sites in the data selection area of the screen to see all of the sites for which data are available.)

If users place their mouse pointer over the name of a location listed in the data selection section of the screen, a display box appears over that data point on the map image to help users understand the geographic location of that particular point at which data are available. For example, in Figure 2, it can be seen that vehicle occupancy data are available at Cedar Avenue on I-405. For the two "travel time" archives connected to TRACMap (the CVISN tag-based travel times and the FLOW Data based travel times), the screen shows all of the points at which data are collected for use in computing that trip's travel time. A FLOW trip is illustrated in Figure 4.

Figure 4: Example of Sites Included in a FLOW Travel Time Trip as Displayed on TRACMap

For the vehicle occupancy and FLOW Data datasets, the user can also simply position the mouse pointer over a data collection point, and the name of that location will appear. For these two databases, users can then simply "click" on that data point to obtain data from those locations, or they can "click" on the name of the location in the data selection section on the lower left side of the screen. This functionality does not currently exist for the travel time data sets. At this time, data for the FLOW travel time archives can only be obtained by selecting the trip in the data selection box on the lower left side of the screen.

The data selection box at the left of the screen also has the ability to display summary statistics related to each site for a given database and data summary. Exactly which summary statistics the data selection box displays (if any) is determined by the programmer who constructs the interface between each ITS archive and the TRACMap interface. Currently, if the ACO database is selected, the data selection area displays the average ACO value for the selected year for both general purpose and HOV lanes at each site, as well as the percentage of vehicles using the HOV lane that are considered to be violating the current operating rules. This is illustrated in Figure 5.

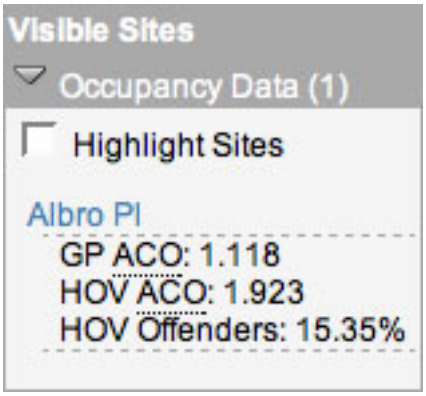

Figure 5: Illustration of the TRACMap Data Selection Section

For the FLOW data archive, the data selection statistics describe the percentage of 5-minute data points that are valid at each location during the selected year. For the CVISN truck tag archive, the number of truck tag reads at that site for the selected calendar year is displayed. The FLOW Trip archive does not currently use this capability.

When users select a location or trip in the data selection section of the user interface, TRACMap transfers users to the selected database and a query interface unique to each archive. (For new archives, this basic functionality would be written into the new archive, not into the TRACMap interface.) The intent of this database-specific data selection screen is to provide users with more specific information about the data available at the selected site for the selected database.

In the prototype interfaces written as part of this project, users are given more detailed site location information (for example, an enlarged map showing the data collection site), as well as a list of all data available at this location from this database. For example, in the FLOW database, users would see how many years of data were present at that site, for each type of lane (HOV/GP) and direction. In the ACO database,

users would see how many vehicle occupancy data collection sessions had been performed by type of lane and direction for each year for which data existed in the database. Other data, such as quality control information, can also be provided in the database specific interface.

Users are then given the ability to download summary datasets related to the selected site and/or allowed to perform a more detailed query of that ITS archive. Other options may also be written into this level of the data archive, depending on the functionality of the individual data archive.

User Interface Design Issues

Because detailed knowledge about the data available in a given ITS archive may be needed to effectively use any archive, the project team recommends that each archive have on its opening screen both contact information (name and e-mail address) and a link to a general description of the data stored within that archive (i.e., metadata[8]). The intent is to help new users understand the nature and limitations of the data to which TRACMap has directed them.

A metadata standard is currently being balloted by the American Society of Tests and Measurements (ASTM) Archived Data User Services committee. The committee's work has been funded by the U.S. Department of Transportation as part of its ongoing ITS standards effort. The intent of the metadata standard is to help transportation agencies understand the types of metadata they should be providing for ITS data archives so that more uses can be made of these data resources and that those uses will be less costly to implement.

[8] "Metadata" is often defined in lay terms as "data about data."

Metadata are particularly important for data sources about whose functioning new users have not received formal training. For major databases (built to Department-wide standards), WSDOT provides training classes to help users understand the database and navigate through the user functions. However, many smaller ITS data archives cannot be supported with that level of commitment by WSDOT even though they may hold valuable information. The primary users of the ITS archives are likely to be small in number, and therefore, the resources needed to develop and provide instruction in the use of the archive are simply not warranted. Similarly, because the external uses of the ITS archives are not always immediately apparent, the project team recommends that the interface to individual archives be kept reasonably simple, with the goal of producing the summary data sets that others in WSDOT are expected to use.

Metadata should be written to explain what these summary datasets represent, how data in them should be interpreted, and any conditions or cautions that a user should be aware of before making use of the data summaries. Contact information for a staff person knowledgeable in the data available in the archive and the use of that archive completes the metadata package by providing an easily used mechanism for answering questions not answered by the metadata.

The current prototype versions of ITS data archives linked to TRACMap lack robust metadata, although recent additions to the ACO archive interface help this database more closely resemble the level of metadata recommended for ITS archives. Further work on metadata for archives should be undertaken if WSDOT decides to fully implement the concepts described in this report.

Adding New Data Archives to the Interface

TRACMap was specifically designed so that additional datasets, whether they are additional years of data in existing ITS archives or completely new archives, can be added to the map interface. Directions for adding data to the interface can be found on the TRACMap documentation site[9]. (See the section on expanding TRACMap.)

To be connected to TRACMap, each ITS archive must include milepost measures that allow association of that data to specific roadway sections or points. The archive must have a Web accessible interface and have the appropriate data retrieval functions. In summary, each ITS archive must

- define queries in SQL to list sites, available years of data, and annotate the results
- write a data delegate, data feature, and an annotator class.

Minor programming changes are required to the TRACMap interface software itself each time a new archive is connected so that the new data sets can be displayed and selected, the appropriate data labels can be shown on the screen, and control is passed to the appropriate URL when the new database is selected. These steps include

- adding a layer to the Arc IMS map
- adding the database connection information and configuration information.

Detailed instructions on performing all of these tasks are included on the website referenced above.

[9] The main TRACMap documentation site can be found on line at the following URL: http://trac29.trac.washington.edu/wiki/page/show/TracMap

CURRENT SOFTWARE STATUS AND SUGGESTED FUTURE WORK

The current Statewide Archive is a promising prototype that is already being used on a daily basis. The prototype TRACMap integration process is currently stable and available to anyone who knows about the site. Two of the three initial ITS archives connected to the site are also currently stable and operational, and the third is likely to be stable and operational in the near future. (The CVISN truck tag software is stable and has functioned for over 18 months, but access to the primary WSDOT CVISN data server that supplies the CVISN tag data was recently disconnected accidentally when WSDOT moved the internal CVISN server without informing the TRACMap development staff. This "software system failure" highlights one of the maintenance issues that will need to be addressed if the TRACMap software is adopted by WSDOT.)

In the opinion of the research team, WSDOT has three options for continuing the work performed in this project. The Department can 1) proceed with user testing, software refinement, and implementation; 2) decide that the approach recommended in this report is significantly flawed and substantial additional work is required to develop statewide access to ITS archives; or 3) decide that the benefits of centralized access to ITS archives do not warrant the cost of providing that access.

The first two of these options are discussed briefly below.

Implementation Requirements

The basic TRACMap interface and two of the ITS archives (the ACO and FLOW archives) have been introduced to interested WSDOT staff who have actively used them. However, formal testing, refinement, and evaluation of the system have not been undertaken. If WSDOT decides that the concepts developed and demonstrated in the

applications discussed in this report are worth adopting, these tasks are the next step toward implementation of these systems.

From the informal reactions we have received to date from WSDOT and outside staff who have used the archives discussed in this report, the research team is of the opinion that more substantial metadata are required to help users understand the nature of the data available from each of these databases.

Other implementation issues that need to be examined, assuming that testing of the system confirms that the benefits of the system outweigh its costs, include determination of the following:

- the cost of ongoing operations and maintenance for the system
- where responsibility for operation, maintenance, and upkeep of the base integration code should sit organizationally within WSDOT (The operation and upkeep of the specific ITS archives should remain with the WSDOT groups that create and use them for their primary purpose.)
- how knowledge of the existence of the primary data availability website can be disseminated effectively throughout the WSDOT
- how an application such as TRACMap fits into WSDOT's Content Management process, including such issues as how changes in ITS applications (such as physical server IP locations) are automatically reflected in changes to the TRACMap software
- how developers of ITS archives can be made aware of the TRACMap application in order to build the necessary functions into their own code in

order to make their summary archive data available to the rest of the Department

- how those developers can determine what summary database statistics are most useful to the rest of WSDOT.

A new research project scheduled for the Fiscal Year ‘06–‘07 biennium will explore these issues.

Possible Future Developments

The initial development of the Statewide Archive focused on gathering data, presenting them, and making them widely available throughout the WSDOT. While this is a huge functional improvement for WSDOT, given its current lack of access to ITS archived data, it is not what many staff expect as an “integrated data system.”

If WSDOT wishes to pursue this more traditional approach to integrating diverse data sets, further work is required. For example, WSDOT might want to consider standardizing an underlying database structure for all ITS operations-related data. It could also adopt a common data transfer protocol to which each archive would adhere. Standardizing the database structure of each archive would ease the coding of automated data queries and integration efforts. The problem is whether the data contained in the diverse ITS archives could actually be restructured into a useful, more standardized data format, or whether the geographic and functional distribution of the ITS archives would limit the cost effectiveness of such an effort.

The databases in the Statewide Archive are currently structured for easy maintenance and flexible queries, at the cost of speed and structural uniformity.

The drawback of this approach is that any program integrating with these diverse databases, such as TRACMap, needs an intimate knowledge of their schemas. To remedy this problem, the ITS archives databases could be restructured to fit a standardized data warehousing schema. This would allow faster, more responsive and scalable queries. A standardized structure would also facilitate queries across multiple databases, making it easier to relate different archives to each other.

Alternatively a thin software layer could be developed to make gathering metadata and generating reports easier. This would allow greater flexibility and expansion. Regional departments would have greater control over the exposure of data and would have the ability to add new datasets in a well defined manner.

Additional discussions with WSDOT staff are needed to determine the appropriate next step for ITS data archives.

APPENDIX A

The ITS archives built and integrated into the Statewide Archive are as follows.

TRUCK TAG DATABASE

Archive interface location: http://trac24.trac.washington.edu/trucks

Data Sources: WSDOT CVISN truck tag database and Transcore CVISN truck tag data stream

Information Provided:

- truck tag read counts
- travel times between tag readers.

AVERAGE CAR OCCUPANCY DATABASE

Archive interface location: http://trac29.trac.washington.edu/hov

Data Sources: Traffic data collected by TRAC temporary staff on PDAs

Information Provided:

- average vehicle occupancies
- vehicle use distributions
- HOV violation rates.

FLOW DATABASE

Archive interface location: http://trac29.trac.washington.edu/flow

Data Sources: CD Analyst output from Northwest Region 5-minute FLOW archive

Information Provided:

- freeway volume, occupancy, and speed data

- travel times for 10 defined trips in the Puget Sound region
- data quality statistics.