ILLINOIS CENTER FOR TRANSPORTATION

CIVIL ENGINEERING STUDIES
Illinois Center for Transportation Series No. 09-050
UILU-ENG-2009-2024
ISSN: 0197-9191

REGIONAL DATA ARCHIVING AND MANAGEMENT FOR NORTHEAST ILLINOIS

Prepared By
Sharan Dhanaraju
Yanfeng Ouyang
Umit Deniz Tursun
University of Illinois at Urbana-Champaign

Research Report ICT-09-050

A report of the findings of

ICT-R27-22
Regional Data Archiving and Management for Northeast Illinois
Illinois Center for Transportation

June 2009

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-ICT-09-050		

4. Title and Subtitle Regional Data Archiving and Management for Northeast Illinois	**5. Report Date** June 2009
	6. Performing Organization Code
7. Author(s) Sharan Dhanaraju, Yanfeng Ouyang, Umit Deniz Tursun	**8. Performing Organization Report N o.** ICT-09-050 UILU-ENG-2009-2024
9. Performing Organization Name and Address Illinois Center for Transportation University of Illinois at Urbana-Champaign Department of Civil and Environmental Engineering 205 N. Mathews Ave, MC 250 Urbana, IL 61807	**10. Work Unit (TRAIS)** **11. Contract or Grant No.** ICT-R27-22 **13. Type of Report and Period Covered**
12. Sponsoring Agency Name and Address Illinois Department of Transportation Bureau of Materials and Physical Research 126 East Ash Street Springfield, IL 62704	**14. Sponsoring Agency Code**

15. Supplementary Notes

16. Abstract

This project studies the feasibility and implementation options for establishing a regional data archiving system to help monitor and manage traffic operations and planning for the northeastern Illinois region. It aims to provide a clear guidance to the regional transportation agencies, from both technical and business perspectives, about building such a comprehensive transportation information system. Several implementation alternatives are identified and analyzed. This research is carried out in three phases.

In the first phase, existing documents related to ITS deployments in the broader Chicago area are summarized, and a thorough review is conducted of similar systems across the country. Various stakeholders are interviewed to collect information on all data elements that they store, including the format, system, and granularity. Their perception of a data archive system, such as potential benefits and costs, is also surveyed. In the second phase, a conceptual design of the database is developed. This conceptual design includes system architecture, functional modules, user interfaces, and examples of usage. In the last phase, the possible business models for the archive system to sustain itself are reviewed. We estimate initial capital and recurring operational/maintenance costs for the system based on realistic information on the hardware, software, labor, and resource requirements. We also identify possible revenue opportunities.

A few implementation options for the archive system are summarized in this report; namely:

1. System hosted by a partnering agency
2. System contracted to a university
3. System contracted to a national laboratory
4. System outsourced to a service provider

The costs, advantages and disadvantages for each of these recommended options are also provided.

17. Key Words	18. Distribution Statement
	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	99	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

ACKNOWLEDGMENTS

This report is based on the results of research project ICT-R27-22, *Regional Data Archiving and Management for Northeast Illinois*. This project was conducted in cooperation with the Illinois Center for Transportation; the Illinois Department of Transportation, the Illinois State Toll Highway Authority, and the Federal Highway Administration. The research team thanks the Technical Review Panel (TRP) for the guidance, direction, and support. The members of the TRP and others who provided valuable support, include (alphabetically):

John Benda (TRP chair)
Clair Bozic
Jeff Galas
Craig Heither
Joseph Hill
Jeffery Hochmuth
Jerry Hron
Anthony Khawaja
Patti Killinger
Mark Kinkade
Scott Lee
Duana Love
Abby Malloy
Tom Murtha
Jonathan Nelson
Justin Potts
James Powell
Greg Sanders
Charles Sikaras
Micheal Szubryt
Gerwin Tumbali
Aaron Weatherholt
Mark Yow
David Zavettero

Particular thanks go to Jeffery Hochmuth and James Powell for their tremendous contribution throughout the project. Diego Klabjan and Paul Pei contributed to the earlier set-up of this project.

DISCLAIMER

EXECUTIVE SUMMARY

This project studies the feasibility and implementation options of establishing a regional data archiving system to help monitor and manage traffic operations and planning for the northeastern Illinois region. This project aims to provide clear guidance to the regional transportation agencies, from both technical and business perspectives, about building such a comprehensive transportation information system. Several implementation alternatives are identified and analyzed. This research is carried out in three phases.

In the first phase, existing documents related to ITS deployments in the broader Chicago area are summarized, and a thorough review is conducted of similar systems across the country. Various stakeholders are interviewed to collect information on all data elements that they store, including the format, system, and granularity. Their perception of a data archive system, such as potential benefits and costs, is also surveyed. In the second phase, a conceptual design of the database is developed. This conceptual design includes system architecture, functional modules, user interfaces, and examples of usage. In the last phase, possible business models for the archive system to sustain itself are reviewed. The initial capital and recurring operational/maintenance costs are estimated for the system based on realistic information on the hardware, software, labor, and resource requirements. We also identify possible revenue opportunities.

A few implementation options for the archive system are summarized in this report; namely:

1. System hosted by a partnering agency
2. System contracted to a university
3. System contracted to a national laboratory
4. System outsourced to a service provider

The costs, advantages and disadvantages for each of these recommended options are provided. The primary function of this archive is to assist the participating agencies in their operations, planning, and performance evaluation. Therefore:

- The archive should aim to generate revenue from the beginning, though the actual return is likely to be minimal for at least the short term.
- Because the data archive must be suitable for operations (in addition to planning), it must have a quick and easy interface and be able to generate results quickly.
- The decision on which option to pursue should be dictated by the partnering agencies.
- The business option chosen must also be matched to the technical requirements.
- Regardless of which business plan option is pursued, the development of the northeast Illinois data archive must be considered in step with short- and long-range needs of the GCM Gateway Traveler Information System.
- Also regardless of which business plan option is pursued, the development of the northeast Illinois data archive must be consistent and compatible with the existing Northeastern Illinois ITS Architecture and other legacy systems.
- Similarly, pending full development of CMAP's data warehouse, there needs to be a link between the data archive and the data warehouse to allow the free flow of information.

TABLE OF CONTENTS

CHAPTER 1 BACKGROUND

1.1 INTRODUCTION

In the transportation industry, multiple jurisdictions operate and maintain various facilities that combined make up the transportation network that we all use. Each has different responsibilities, capabilities, and operational procedures. Consequently, each has developed different levels of data storage and usage. The capabilities of these local systems are typically limited. Most of them permit basic operations such as backups, but, for example, reporting functions are limited, or non-existent.

Such decentralized information systems hinder global decision making since their original purpose and motivation were to meet the requirements of the specific agency and not broader community needs. To enable centralized and global decision making, these systems need to be integrated to provide a coherent information system that can serve the broader transportation community.

The northeast corner of Illinois is heavily populated, and as a result traffic management is extremely complex. To monitor and handle real-time operations, several transportation agencies have built their own information system(s), which are mostly used internally to analyze real-time operations and related analysis. To name a few, the IDOT Traffic Systems Center records five minute speeds and counts; PACE keeps track of bus schedule on-time performance, incident locations and times, and many other records; and the Illinois State Toll Highway Authority (ISTHA) Traffic and Incident Management System (TIMS) records five minute I-PASS travel time data, incident locations and times, and other key data.

Most of these systems provide real-time data at different levels of granularity. As a result, simply ‘pulling’ all these data together is not meaningful, nor it is straightforward. One extreme is to combine all of the data into a single database by first identifying overlapping fields. This might require a large database with expensive data storage requirements and maintenance necessity. Some of the data in this combined database might never be used and therefore redundancy will occur.

The Illinois Department of Transportation (IDOT), ISTHA, and other regional agencies agree that by having a single system, more advanced modeling, decision making, and analysis could be performed. This project evaluates potential benefits of a centralized system, assesses the underlying cost, and studies integration options.

IDOT, ISTHA and other operating agencies covering the transportation system throughout the northeast corner of the state realize that it is suboptimal to have several standalone transportation information systems not communicating to each other. For real time operations, the GCM Gateway was created to help solve this problem. Greater efficiencies could be achieved by having a single system with various decision support systems built on top of it, including a data archive.

The expense of creating and maintaining an archive system covering all agencies can be significant. Thus for the functions that require archived data such as planning and research, no one agency has assumed this regional responsibility. As a result, the region’s agencies are considering putting together a single warehouse for data archiving and mining. Combining approximately a dozen information systems into a single one will not be an easy undertaking. Such a complex task requires meticulous planning and a thorough analysis of all possible options and alternatives. In addition, benefits and costs must be identified.

A single data warehouse would provide better decision making based on a single repository encompassing the entire region. For example, travel demand patterns could be determined across the entire region as opposed to limited regions that interact with each other.

(e.g., travel time) and various factors (e.g., agencies operational decisions or incidents) could be analyzed in a more cohesive manner across the entire region, yielding a better understanding of underlying causes and effects. The ready availability of extensive data may help operators in real time make more informed and thus better control and traveler guidance decisions. With aggregated data from several agencies, correlations among events stored in several existing systems can be studied and used to improve operations. Benchmarking among agencies provides yet another benefit of an integrated database. The regional travel demand model might be more easily validated with additional correlated data. Such a system can be used not only as an operational tool, but also as a planning tool and thus providing a wide variety of new functionalities and a great new opportunity.

1.2 RESEARCH OBJECTIVE AND APPROACH

The main objective of this project is to provide clear guidance to the regional agencies about building a comprehensive transportation data archive. Several implementation alternatives are identified and each one analyzed thoroughly.

The researchers started the project by first reviewing data archiving principles and existing archives around the country, including related deployments in northeastern Illinois. Regional agencies interested in or in some way already engaged in the task were contacted, and the results are presented. Another component is the review of the existing Regional ITS Architecture.

Next, the needs and scopes for the proposed data archive are indentified. This includes identifying major data elements including the format, system, and granularity. This is based on extensive interviews and discussions with the stakeholders at the interested agencies. While interviewing the stakeholders, each was asked to identify their goals, objectives, and requirements of a regional data archive via a survey. The requirements and specification of the regional data archive rely on the data available at the agencies.

Based on the identified opportunities, needs, and requests, the research team designed a specification of the database, including conceptual schematics, software architecture, functional modules, user interfaces, and examples of usage.

Finally, the research team reviewed possible business models for the proposed archive system to sustain itself. The operational costs for the system were estimated based on realistic information on the hardware, software, labor, and resource requirements. Possible revenue opportunities were identified. A few implementation options for the archive system are summarized and recommendations are provided.

The remainder of this report is organized as follows. Chapter 2 reviews the literature on transportation data archiving practice, and similar systems in peer states. Chapter 3 summarizes the interview and surveys with the multiple stakeholders. Chapter 4 analyzes possible centralized and decentralized data system architectures and presents a conceptual design of the system. Chapter 5 discusses costs, revenue, operations of the proposed archive system, and makes recommendations on the business model.

CHAPTER 2 LITERATURE REVIEW

2.1 INTRODUCTION

Transportation data archiving is defined as the systematic retention of transportation data so that the collected data can be used to fulfill the needs for short term transportation operations / management and for long term planning / design. Data warehousing aims to archive transportation system characteristics, operation details, and performance measures to improve overall management of transportation systems.

The main types of data elements that are usually archived, according to *Guidelines for Developing ITS Data Archiving Systems* (Turner 2001), are as follows:

- **Traffic monitoring and detection data:** Traffic volume, speed, travel time, vehicle classification, vehicle weight, video imaging
- **Traveler information data:** Travel time, traffic flow speed, level of congestion, incident data, construction and maintenance activities, lane closures
- **Traffic control parameters/data:** Display and parameters for ramp meters, traffic signal, lane control sign boards, message sign boards, dynamic message signs
- **Incident and emergency management data:** Time sequence of incidents (detection, notification, response, clearance), location, extent, cause, participants and response activities
- **Public transit data:** Ridership, origin-destination, travel time, route information
- **Construction and work zone data:** Location, time, date, extent of blockage or closure

A data archiving activity is the integration of operational transportation data with more static information covering road system geometrics and characteristics. Effective use often requires higher level compilation and analysis of these different kinds of data to yield meaningful insight into system performance. Proactive traffic operations management requires historic data in order to forecast the likelihood of problematic sections and assist in the preparation of strategies to mitigate anticipated problems. A collective data pool, furthermore, increases the cost effectiveness of data retention and storage activities for transportation planning, design, operations and research needs.

Archived Intelligent Transportation System (ITS) data are distinct from traditional data sources in three ways:

(1) ITS-generated data are temporally intensive (i.e., collected in very short intervals).
(2) ITS-generated data fill some major data gaps that could not be filled in the past due to resource (storage) limitations.
(3) Nowadays, nearly all ITS-generated data are available on electronic media, thereby expediting data analysis and information dissemination.

As a result, ITS-generated data can both replace and supplement data collected through traditional methods, such as paper-based documentation.

2.2 PRINCIPLES OF DATA ARCHIVING

2.2.1 Overview

Collecting transportation data to determine the performance of a transportation network requires a logistical plan that addresses the following: detailed preparation for the type of data to be collected, frequency, locations, analysis techniques and responsibilities, database management and performance analysis reporting (U.S. DOT 2008). Freeway Management Systems (FMS) of transportation agencies generate data about the transportation networks and their status under varying traffic conditions. The next step is to compile the collected data into useful summary information. Ideally, most of this effort is completed automatically, but usually the Traffic Management Center (TMC) personnel need to be directly involved.

Part of the data archiving effort is the transformation of raw data into performance measures that help guide operations, management, planning and design. In general, raw data are created in the FMS and processed at the TMC before utilization. Although data should also ideally be gathered for all surface streets and transit modes, surface street data archiving has received relatively little attention to date.

According to the *Freeway Management and Operations Handbook* (U.S. DOT 2008), the primary reasons for archiving the generated data are to:

- Provide more and better information for managing and operating the system.
- Maximize cost-effectiveness of data collection infrastructure.
- Reduce costs compared to manual data collection.
- Conform to established and proven business practices in other industries.

For northeastern Illinois, all four of these reasons apply.

The data can be useful for any time period including the immediate, mid-term and long-term applications, as well as for various travel modes. The major applications of the compiled summaries fall into four main areas:

- Highway operations and maintenance
- Safety
- Multi-modal transportation planning, including freight systems
- Transit management

The key characteristics of successful data archiving practices identified in the literature review are summarized as follows:

- Leading entities (not necessarily the stakeholders), who are primarily responsible for operating and maintaining the data archive, have to be appointed by the stakeholders.
- A prototype using a few well-defined types of data should be established first. As the system becomes operational, the responsible entities should work aggressively to add new sources and types of data. History has shown that starting with too complex a system often leads to start-up failure, and archiving systems that do not actively expand may never achieve the level that they were intended to.
- Quality control is a crucial part of data archiving, to find and remove erroneous data. To the extent possible, missing data replacement should be part of the procedure. The archiving entities should work with those agencies that provide consistently erroneous data to find solutions, or eliminate them as a data source.
- Data compilations have to be available at a level of aggregation that is consistent with agency needs, and must consider analysis methods and statistical variability of the compiled data. Aggregation level is defined as the time interval at which raw data are summarized, since raw data may be fine-grained and voluminous. Flexibility should be built into the archiving system for unanticipated needs, and the range of aggregation solutions can be from a single fixed aggregation level to complex multiple aggregation levels. In advanced systems, a good choice is to have several aggregation levels (e.g., multiple fixed aggregation levels, or dynamic aggregation levels).
- The original raw field data should be stored offline in its entirety.

- Diagrams and guidelines that show how data are collected, processed and presented should be provided so that users can better understand the origin of data, their content and the presentation style.
- The data archive has to be user friendly so that an ordinary user with a typical desktop computer and internet access should be able to access and analyze the data without special training. Distribution of stored data can be done by portable storage devices or over a web-based medium.

2.2.2 Issues and Costs

Implementation issues can be categorized into five areas:

1. Institutional impediments

These impediments generally center around agency cooperation and proprietary data concerns. Agency cooperation is usually an issue related to financial issues. There may be political issues that usually have nothing to do with the data archive itself. The issue of proprietary data can be over ownership, value, or a range of other issues.

2. Privacy and liability

Most public agencies take the issue of data privacy very seriously. They are entrusted with the public good on a variety of responsibilities, and there is always some data the organization creates that they protect as part of their responsibilities. The issue of liability for making private data available – even if it is anonymous – is always present with the public agencies.

3. Data quality and integrity

Any field infrastructure requires a certain level of maintenance in order to maintain a working system. The level of commitment by an agency to maintenance of these systems varies considerably. Generally, data quality and integrity are a function of the amount of importance an agency places on maintenance. If an agency cannot produce reliable data over the long-term, then it should not be archived with the other systems.

4. Data standards and compatibility

While attempts have been made at creating a national standard for data formatting (e.g., TMDD), there is still considerable variability in how data is stored and by what standards. Especially troublesome is the location referencing used by various agencies. A regional data archiving system will have to accommodate such variability at least in the short term.

5. Other technological issues

Besides the obvious issues with getting two (or more) independently developed software systems to share data, there are issues with the integration of non-electronic data. Additionally, there are issues relative to what format the final combined data should use.

To address these issues, development time and dollars must be spent to:

- Articulate and communicate the needs for archived data and clearly state benefits
- Forge mutually beneficial partnerships among data-producing agencies and data users
- Identify the lead entities responsible for data archiving
- Address data quality/integrity issues while defining appropriate aggregation levels
- Review privacy concerns and liability issues, and develop policies to address these

- Define data saving methods and hardware
- Reconcile data incompatibilities among different data sources
- Revamp existing software to accommodate archived data
- Integrate the archived data with non-electronic data to meet broader data needs
- Reformat archived data into a user-friendly format
- Make archived data accessible in a timely manner
- Provide complete training on use of archived data in engineering and planning practice

In the course of doing the literature review, many instances of very detailed design requirements were identified. Some of the more detailed design issues are as follows:

- Some database formats may be inexpensive initially but may require high level programming support on an on-going basis that is expensive.
- Firewall and security issues need to be addressed as early as possible.
- Whenever possible, commercially available or open source software should be used instead of proprietary or custom-designed systems. Early systems (e.g., Washington State TRAC and Minnesota TMC) have a history of sharing software code.
- A configuration management plan for system software and hardware is required.
- Proper documentation of software is fundamental.
- Mapping tools, most likely based on GIS, should be included.

To help identify benefits, it may be appropriate to conduct targeted field tests. Data archiving tools for incident management will be complex because they require the compilation and assimilation of data from several sources, but the pay-off in improved incident management may be great. The larger safety application area poses challenges relating to privacy and liability, in addition to the complexity of correlating geometric, weather and operational data, but likewise may yield large benefits by reducing accidents. As an added benefit while the full data archive develops and comes into active use, a natural constituency may develop that demands and supports more active monitoring and maintenance of ITS detectors and related equipment.

2.2.3 Archiving Opportunities

Many of the benefits of data archiving were addressed in the initial application for funding. There are a wide variety of groups within transportation that may rely on archived data on a recurring basis. Table 2-1, adopted from *Cross-Cutting Studies and State-of-the Practice Reviews: Archive and Use of ITS-Generated Data* (FHWA 2008), summarizes the archiving opportunities by application area.

Table 2-1. Archiving Opportunities by Application Area

Operations & Maintenance
1. Assessment of Security Vulnerability of Highway Network 2. Performance Monitoring 3. Management of Traffic Delay 4. Planning and Managing Special Events 5. Planning for Management of Unplanned Events 6. 511 Information Validation 7. Adaptive Signal Timing Strategies 8. Asset Monitoring and Maintenance
Safety
1. Intersection & Segment Safety 2. Work Zone Safety 3. Evaluation of Alternative Strategies for Speed Management 4. Pedestrian and Bicycle Safety 5. Safety Public Awareness and Training
Multi-Modal Transportation Planning & Freight Systems
1. Urban Travel Demand Forecasting 2. Data Reporting 3. Spatial Movements of Travel Demand 4. Performance Monitoring 5. Freight Management and Movement
Transit Management
1. Route Planning and Fleet Use Evaluation 2. Transit System Operations and Maintenance 3. Paratransit Operations for Special-Needs Groups 4. Emergency Management and Preparedness of Transit Systems 5. Assessment of Security Vulnerability of Transit Systems

2.3 ARCHITECTURE AND STANDARDS

Data archiving is best accomplished when there are regional standards that help to promote interagency sharing. Additionally, regional ITS Architectures help to illustrate the wide variety of sources an archived system should consider.

The National ITS Architecture is a common framework for planning, defining, and integrating intelligent transportation systems. A policy that requires all ITS projects to conform to the Architecture was issued by FHWA (2001). The National ITS Architecture Version 6.0 and its contents can be found at http://www.iteris.com/itsarch/.

All proposed system architectures and data have to be compatible with adopted, applicable standards. Although FHWA has not yet formally adopted any standards, numerous standards have been promulgated by three main Standard Development Organizations (SDOs), and may eventually be adopted by FHWA. See http://www.standards.its.dot.gov/ and http://ops.fhwa.dot.gov/int_its_deployment/standards_imp/standards.htm for information and status on standards.

The Chicago region has several ITS Architectures in place, including a statewide, regional (MPO) and a three state Gary-Chicago-Milwaukee (GCM) ITS corridor architecture. (Note: it is expected that the GCM ITS corridor architecture will eventually transition into the new Lake Michigan Interstate Gateway Alliance ITS corridor architecture.) Systems in the region mainly provide real-time information to various agencies, and they highlight available ITS resources for developing a data archiving system in northeastern Illinois. More detail about these systems is summarized in Section 2.4.

The National ITS Architecture has an Archived Data User Service (ADUS) that defines a general framework for data archiving. Functions of the ADUS are to:

- Collect, archive, manage and distribute data from ITS sources.
- Provide proper formatting, quality control and metadata (descriptive data about detailed data).
- Provide data fusion between data elements from numerous data sources.
- Present processed data for federal, state and local agencies.

The National ITS Architecture also includes three archiving Market Packages: ITS Data Mart (AD1) – the focused archive of a single agency; ITS Data Warehouse (AD2) – pulls together ITS Data Marts of several agencies into a single repository, and; ITS Virtual Data Warehouse (AD3) – same functionality as AD2, but rather than a single repository, provides the tools to *virtually* link the individual agency archives.

During the initial development of the National ITS Architecture, user requirements for ITS data archiving were defined and documented. As a starting point for local initiatives, the documentation at http://www.iteris.com/itsarch/ is useful. The documentation constitutes a comprehensive resource listing the different components of ITS and the many connections that unite them into an overall framework for deployment.

2.4 CASE STUDIES

This section briefly reviews some existing transportation data archives at federal and state levels. The FHWA's effort to establish a national level transportation data network is first presented, followed by practices in various leading states. The developments of these systems are at different stages and the progress has not been uniform. Every system has its unique features, and collectively they may set a roadmap for the regional data archiving and management for the northeastern Illinois project.

In this section, ITS efforts related to transportation data collection and real time operations are also presented, mainly because these activities are closely related to data archiving efforts. It is not uncommon in the literature that a few peer states (e.g., Florida and Virginia) first developed data collection and compilation systems before building data archiving systems. Their situations somewhat resemble the case for Illinois, where a real-time information system is already in place before the development of an archiving system.

2.4.1 FHWA Mobility Monitoring Program (MMP)

In 2000, FHWA initiated a program in which archived freeway detector data from several cities (Atlanta, Cincinnati, Detroit, Hampton Roads, Houston, Los Angeles, Minneapolis-St. Paul, Seattle, and San Antonio) are collected in order to develop a performance monitoring program. This program mainly aims to monitor the mobility and reliability of each system at the city and national level and plans to add more cities into the system over time. The study is unique because it is the first step in a national level transportation data archiving effort.

Although the collected data were originally in different formats, all are summarized into 5-minute and lane-by-lane format for processing and analysis. Currently, only freeway detector data are gathered and saved offline on CD, although summary statistics are available online. Quality control measures are applied to the data to flag and remove erroneous data. The data record, data collection system, and data processing are documented for each city.

The MMP calculates system performance metrics based on data archived at traffic management centers (TMCs). These data are highly detailed measurements from roadway surveillance equipment installed for operational purposes. Data from spot locations (volumes and speeds) are used as well as travel time estimates from probe vehicles where available. For each participating city, the MMP develops congestion metrics at both the corridor and area levels. Early work from this project has provided a basis for measuring travel time reliability. Beginning from 2002, traffic incident data has been collected from TMCs where these data exist. Also, continuous traffic data from signalized highways is being explored as a potential source for system performance monitoring. A major product of the MMP is the monthly Urban Congestion Report; this report characterizes the emerging congestion trends at the city and national level. As of late 2007, these reports are being prepared for 23 cities around the nation.

2.4.2 Texas - Austin, Texas (Turner 2001)

Freeway detector data are archived by the operations department of the Texas Department of Transportation (TxDOT). The data are provided on CD when requested. The data are aggregated at the hourly level, creating 365 x 24=8,760 comma-separated ASCII-text files every year. TxDOT is considering some improvements to this relatively old system. Some improvement recommendations are as follows:

- Original 1-minute level data should be stored offline on CD.
- Summary statistics on spreadsheets should be provided at the 5- minute, 15-minute and hourly level.
- Data quality control measures should be exercised and historical data should be reviewed for erroneous data as well.
- Distribution of data over the internet should be provided.

Data Flows and Data Consumers in the Austin, Texas case are shown in Figure 2-1 below. The data archiving effort is conducted in parallel with the provision of real-time information.

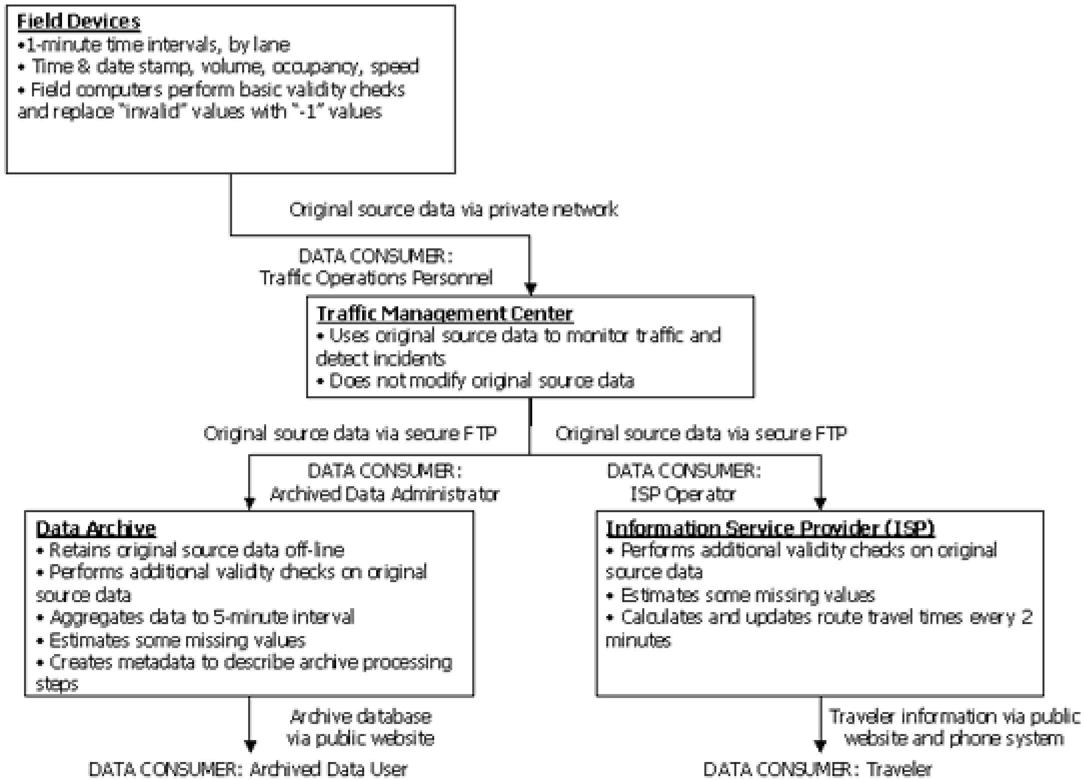


Figure 2-1. Data flows and data consumers in Austin case study.

2.4.3 California PeMS (Varaiya 2006, 2008)

The Freeway **Pe**rformance **M**easurement **S**ystem (PeMS) is developed and operated by the University of California - Berkeley, with the cooperation of the California Department of Transportation (Caltrans), California Partners for Advanced Transit and Highways, and Berkeley Transportation Systems (BTS).

The intent of PeMS is to collect historical and real-time freeway data from freeways in the State of California in order to compute freeway performance measures. Raw real-time freeway detector data (e.g., every 30 seconds) from Caltrans districts and TMCs are collected and summarized into 5-minute time intervals and loaded into the PeMS data warehouse. PeMS also obtains and stores California Highway Patrol CHP-published incident data. Data from PeMS are available through the http://PeMS.eecs.berkeley.edu website, which receives 4000-5000 web hits per day. The hardware was originally housed at UC Berkeley; a functional backup copy has been housed at Caltrans since 2006. The maintenance and upgrades are contracted to BTS.

PeMS contains 2 Terabytes (TB) of historical data online and adds 2 GB of data every day. Access to the data site requires a username and password. The website of archived data is accessible to the general public (public, academia, private firms and public institutions) after a registration process; access to real-time data feeds by value-adding resellers require extra approval process. The system provides several tools that present data summaries and reports. The main goal is to monitor freeway performance by using indices such as: travel flow speed, estimated travel time, and vehicle counts. PeMS is a unique project because it is one of the few statewide transportation archiving efforts and has received positive reviews. Success in part is due to the fact that Caltrans has extensive detector coverage on its freeways, for which PeMS presents a valuable tool with extensive capabilities to calculate system performance indices. Future improvements to the system include the addition of better prediction and data generation modules, which are generally traffic simulation tools based on macroscopic traffic flow models (such as cell-transmission).

PeMS performance in the coming years should provide excellent insight into whether or not a centralized statewide data archive can be useful to local agencies. PeMS underlying software architecture can be described as follows:

- The bottom layer is database administration: disk management, crash recovery, table configuration.
- The middle layer comprises software that works in real time:
  - –Aggregates 30-second flow and occupancy values into lane-by-lane, 5-minute values.
  - –Calculates the g-factor (a parameter to characterize average vehicle length and loop characteristics) for each loop at each time interval, and then the speed for each lane. (Note: *speed = g-factor × flow / occupancy*).
  - –Aggregates lane-by-lane values of flow, occupancy, and speed across all lanes at each detector station.
  - –Computes basic performance measures including vehicle-miles-traveled, vehicle-hours-traveled, travel times, and congestion delay.
- The top layer comprises additional built-in applications and analysis modules that can be accessed via the PeMS website. PeMS allows users to pull out the trip travel estimates and shortest routes for specific O-D pairs. For any selected O-D pair, fifteen shortest paths on the district freeway map can be displayed along with travel time estimates. In addition, it is possible to predict travel time for the future, using a traffic projection algorithm that utilizes historical and real-time data. Another feature of PeMS is ‘plots across space’ that display bottleneck locations in a particular section of freeway.

The communications architecture of different components of PeMS is shown in Figure 2-2.

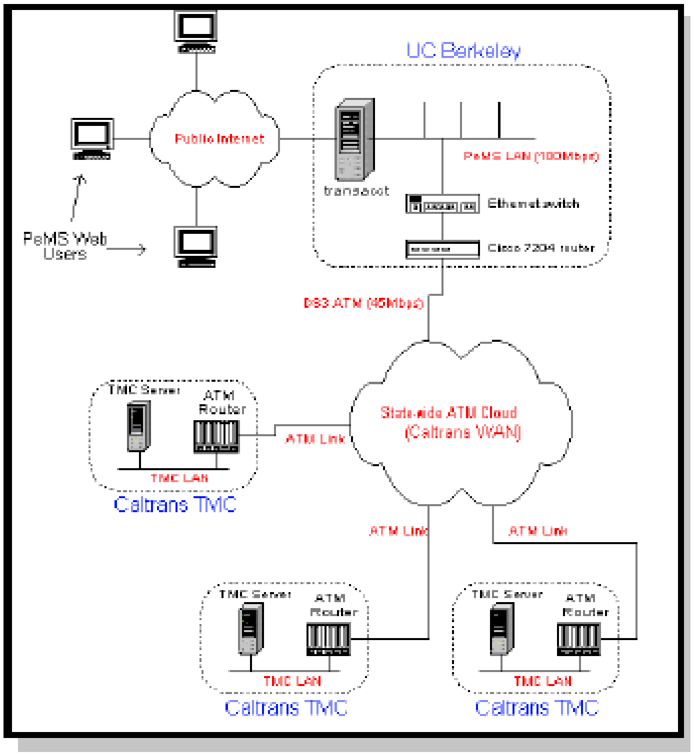


Figure 2-2. High-level communication architecture of PeMS.

2.4.4 Washington – Seattle (Washington State DOT, 2006)

The Washington State DOT is an agency that has collected and stored freeway data since 1981, although there were some technological restrictions at the early stages of the effort. The agency has a history of institutionalized data archiving efforts, and this has been an evolutionary effort. The archive is the product of collaboration between the University of Washington, Washington State Transportation Center (TRAC), and the Washington State DOT.

Seattle freeway detector data are collected every 20 seconds from field detectors and are summarized into 5-minute aggregations and distributed by CD upon request. The data are converted into estimates of vehicle speed and travel time. Quality control measures are applied to the collected data. Distributed CDs contain data extraction and summary tools besides the 5-minute aggregated data. The data provided are used for many operational purposes such as evaluating ramp metering strategies, analyzing HOV lane effectiveness, monitoring freeway performance, quantifying freight transportation efficiency, and pavement design.

2.4.5 Illinois and Neighboring States

2.4.5.1 Gary-Chicago-Milwaukee (GCM) Priority Corridor (Turner, 2001)

The Gary-Chicago-Milwaukee (GCM) Priority Corridor is a joint effort among the Illinois Department of Transportation (IDOT), the Indiana Department of Transportation (INDOT), and the Wisconsin Department of Transportation (WisDOT). The GCM ITS Priority Corridor was one of four multi-agency ITS coalitions formed as a result of the Intermodal Surface Transportation and Efficiency Act of 1991 (ISTEA). Officially started in 1993, the GCM Corridor was made up of all of the major transportation agencies in the 16-county area connecting Gary, Indiana through Chicago, Illinois to Milwaukee, Wisconsin, and is home to about 11 million people.

The Gateway System serves as an information clearinghouse for the information coming from a variety of transportation sources. The information is collected in each state's traffic management center, forwarded to the Gateway, and sent back to the traffic management centers. Each state manages its own system, but due to the Gateway, has better information on what traffic conditions are on the facilities adjoining its system. The Gateway currently collects, validates, fuses, and distributes real-time traffic information for expressways and tollways. The GCM website provides travelers in the corridor area with information to assist them in making transportation choices. The website provides:

- A real-time map of road congestion and construction data
- Construction and closure reports
- Congestion and travel time reports
- Historical travel time reporting
- Traditional vehicle detection data
- Dynamic Message Sign reporting
- Incident Reports
- Links to video images available in the corridor
- Links to transportation-related sites in the corridor

Data is stored for roughly 30 days on the Gateway, but the data are not archived. A companion system at www.gcmtravelstats.com stores and provides travel time data at a fairly high aggregation level for comparison purposes, by route, day of the week, and date. Current and average travel times along with differences are presented in a table. A graphical representation showing the average travel time by time of day together with the travel time for the current day is shown below the table. The graph also shows a shaded area representing one standard deviation from the average and lines indicating the travel times under different congestion scenarios (low, medium and high). Custom queries allow a range of dates to be selected for each of two routes for comparison purposes, and also allows for filtering the data based on the day of the week. The custom query also provides the average, minimum average, and maximum average travel times for specific segments.

2.4.5.2 Wisconsin Transportation Data Hub – WisTransPortal (Wisconsin Traffic Operations and Safety Laboratory, 2007)

The Wisconsin Department of Transportation (WisDOT) collects a significant amount of transportation operations related data to support long-range planning, improvement projects, and the metropolitan Milwaukee and Madison freeway management systems. In order to provide tools that effectively manage, aggregate, archive and analyze multiple data sources, the Wisconsin Traffic Operations and Safety (TOPS) Laboratory (based at the University of Wisconsin - Madison) has started the development of a data management system (WisTransPortal) to collect and archive ITS data in Wisconsin.

The goal of the WisTransPortal project is to develop capabilities for a statewide ITS data hub to support multiple applications in traffic operations and safety. Those capabilities include integration, management, analysis, and dissemination of real-time and historical ITS / traffic operations data through a centralized database and communications infrastructure. The data archiving component of WisTransPortal will consist of automated services that connect to various WisDOT data sources and prepare the data for archiving. The system includes web-based query tools and GIS mapping tools. Some of the web applications available for the users include the following: (i) interfaces to query and download data from crash and weather databases; (ii) ability to analyze the traffic detector data; (iii) map-based tools to retrieve information. In addition, other applications (e.g., for better data retrieval and integration of lane closure information) are in the development and testing phase.

2.4.6 Florida

2.4.6.1 SunGuide Florida (2008)

SunGuide is the statewide name for the Florida Department of Transportation (FDOT)'s Intelligent Transportation Systems. The SunGuide system does not provide data archiving, but it is a successful real-time operational data management system that is similar to the Gary-Chicago-Milwaukee (GCM) Priority Corridor.

A real-time map on the Smart SunGuide website shows traffic-impacting events providing travelers the information they need to make informed decisions on what route to take according to incident and congestion location information. The map also includes dynamic message sign and closed-circuit television cameras (CCTV) locations and status. Video images from the cameras and sign messages can be viewed by "mousing-over" map icons.

TMC operators coordinate with the Road Rangers, Florida Highway Patrol, FDOT District 6, Florida’s Turnpike Enterprise, and the Southeast Florida 511 Advanced Traveler Information System (ATIS) to determine incident and congestion locations. The information collected is then used for a better response and management of incidents and to disseminate appropriate messages to travelers using dynamic message signs, the 511 system, and traveler information websites. Once all the information from various sources regarding incidents is verified, the details of the event are entered into the SunGuide database system. These details are analyzed to determine the degree of disruption on the roadway and the potential impact on traffic. Response plans are generated and communicated to all stakeholders by emergency agencies.

2.4.6.2 Central Florida Data Warehouse (University of Florida Transportation Research Center, 2008)

The Central Florida Data Warehouse (CFDW) prototype system is on its way toward implementation. In March 2002 a report “*An evaluation plan for the conceptual design of the Florida Transportation Warehouse-Phase 1*,” was compiled to draw the road map for the implementation stage as phase 1. As of April 2008 the implementation is in Phase 2, which is studying the proof of concept of the project. This project has been fully funded by the Florida Department of Transportation Research Office.

The principal function of the CFDW is to collect and store statewide information from many data sources and make it available for use by different applications. The archived data are used to prepare performance reports at different levels (such as detector, system and operational levels), including the following information:

- Maximum Flow Rates
- Vehicle Miles of Travel (VMT)
- Vehicle Hours of Travel Time (VHTT)
- Average speed
- Delay
- Level of Service (LOS)

CFDW will interact with other traffic control centers, transportation management centers, emergency management centers, weather centers, disaster management centers, and motor carrier control centers. A prototype website has been created to access the data from the system. For future development, more of SunGuide's TMCs are to be brought into the archiving process.

2.4.7 Virginia

2.4.7.1 Smart Travel Virginia (University of Virginia, 2008)

Smart Travel Virginia, like SunGuide Florida, is quite similar to the GCM system in terms of providing real time transportation information. This system is a collaborative effort between the University of Virginia (UVA) and the Virginia Department of Transportation (VDOT). It has two major components: freeway management and arterial management.

The Freeway Management System includes:

- Closed Circuit Television (CCTV) for surveillance
- Vehicle Detection Stations (VDS) for measuring volume, speed, and lane occupancy
- Variable Message Signs (VMS) for specific en-route traveler information
- Highway Advisory Radio (HAR) for broadcast en-route traveler information
- Communications Subsystem (COMM) consisting of fiber optics, leased phone lines, microwave and wireless
- Servers and software for logging incidents, generating response plans, and posting messages.

The Arterial Management System includes:

- Traffic signal controllers for controlling intersections
- Communications Subsystem (COMM) consisting of fiber optics, leased phone lines and wireless
- Servers and software for gathering vehicle data, uploading/downloading timing plans, and monitoring intersections.

All of the data and information are collected and disseminated by the Smart Traffic Center (STC). Operators monitor the system, verify incidents and respond to them. Smart Travel also provides interstate traffic cameras that are available for viewing on TrafficLand.com so motorists can plan their routes prior to departure. VDOT also has a color-coded map displaying real-time traffic conditions, available to the public via the NOVACommute.com website.

2.4.7.2 ADMS Virginia (University of Virginia, 2008)

The Archived Data Management System (ADMS) Virginia enables the use of archived traffic operations-related data for transportation applications such as planning, mobility performance measurement, improving operational effectiveness, and supporting decisions. FHWA and VDOT originally funded ADMS Virginia as an operational test. VDOT continues to sponsor this project and is leading the effort with team members from the Virginia Transportation Research Council (VTRC) and UVA's Center for Transportation Studies (CTS). UVA subcontracted the software development part of the project to Open Roads Consulting, Inc. (ORCI). The equipment necessary for the project was originally hosted at the Smart Travel

Laboratory (STL), a jointly-owned facility of VDOT and UVA. The production version was turned over to VDOT's information technology center in Richmond in 2006.

DMS Virginia provides query capability to users. Authorized users may log onto the system through a standard web interface and query the database for historic traffic, weather, and incident information for specific routes, segments, or detector stations on user specific dates, at specific times, and at specific levels of aggregation.

The system has been developed in incremental efforts, implemented at different locations progressively with each incremental step increasing the functionality and the coverage of the system. According to the information available from the project's webpage, the system has been built in four 'phases' or 'builds'. The traffic control and monitoring operations of the system are controlled from various traffic management centers (TMC) across the state; they are Northern Virginia Smart Traffic Center, Hampton Roads Smart Traffic Center and Smart Travel Lab at University of Virginia and most of the data in the system corresponds to the freeway detector data. Annual costs of operation and maintenance are estimated at $300,000 and funded by VDOT. Future upgrades in the system include improvements to the user interface, in order to make it easier for the public to use the system.

Additional information on ADMS Virginia is presented in Section 5.2.2 in the context of a business model for the northeastern Illinois data archive.

2.5 SUMMARY

The most important points from this review can be summarized as follows:

- FHWA aims to have a national transportation data archiving network; it will be advantageous for the northeast Illinois regional data archive system to make preparations for integration into the national system (e.g., by following similar ITS standards and technical requirements).
- Data archiving and real-time transportation operation systems have similar characteristics. Many of the reviewed cases are joint efforts among several governmental agencies and academic institutions that deal with a great deal of transportation data.
- There are not many efforts that can be used as working examples for regional data archiving and management for northeast Illinois. Some of the applications tend to consist of raw storage of freeway only data that can be reached by requesting CD based data. But there are a few examples such as the California PeMS, and developing efforts in Florida, Wisconsin and Virginia that may serve as good starting points.
- Typically the original raw data are fed into one system by different entities. Centralized data storage is a common practice.
- Current technological advancements allow processing and storing a large amount of data. For example, PeMS contains 2 TB of historical data online and adds 2 GB of data each day.
- A sound systems engineering approach to system design is critical to ensure the developed system meets the needs of the intended users. Additionally, software development principles such as scalability, modularity, and portability must be followed.
- Using an incremental build approach is beneficial from the standpoint of getting feedback from the stakeholders early in the process, when system design modifications are still feasible. In other words, rapid prototyping should be advantageous since it allows the stakeholders and developers alike to try and test the system during development, before the final release, in an iterative process.

- Stakeholder involvement is identified as an important aspect through providing ideas, reviewing documents/demos, testing prototypes, and reporting any abnormal findings.
- Focusing on providing user-friendly access to the data is crucial. Domain knowledge of the data and information is an important part of the process, and software can only do what it is programmed to do. If the underlying data is not understood, the services are unlikely to provide the desired results.

CHAPTER 3 INTERVIEWS WITH STAKEHOLDERS

3.1 INTRODUCTION

As part of the efforts to create a data archive system in northeastern Illinois, a series of interviews were conducted over the period of June - July 2007 covering a range of potential stakeholders and participating agencies: Chicago Office of Emergency Management and Communications (OEMC), Illinois State Toll Highway Authority (Tollway), Chicago Metropolitan Agency for Planning (CMAP), Regional Transportation Authority (RTA), Lake County, Will County, Illinois Department of Transportation (IDOT) District 1, and IDOT Central Operations. Each interviewee was provided a written list of questions in advance, to help focus discussion. This section provides the results of the interviews, in summary form, with details in an Appendix. Data archive goals and objectives, followed by high-level requirements, are presented as well.

The proposed data warehouse will aim to accumulate as much data as is available concerning traffic and transportation within northeastern Illinois, and make it available to interested users or participating agencies. The data warehouse is not expected to exchange real time information between systems, since the GCM Gateway currently accomplishes that task.

3.2 INTERVIEW SUMMARY

The following section presents a summary analysis of the interview results. The detailed questions and agency responses are presented in Appendix A.

The transportation agencies in northeastern Illinois have a variety of data that are potentially available for archiving and further analysis. An initial review of data types (surveillance, surface condition, events, etc.), sources, and users is presented in this chapter. Generally, the interviewed agencies feel that a successful regional data archive will significantly improve current practices in transportation planning, project programming, demand forecasting, traffic simulation, network operations, and advanced management and reporting. Regarding the format of data storage, most agencies prefer SQL-compliant databases, which is also the standard IT specification for IDOT. Some agencies would like GIS compatibility to be included for enhanced spatial analysis capabilities.

A certain format conversion functionality would be necessary in the archive system, i.e., it is expected that most electronic data feeds will require some form of conversion or manipulation at the data archive. In terms of data transfer, most agencies prefer the XML format for its compactness and convenience. Data in XML format can be systematically generated from databases. The preferred data aggregation level ranges from 5 minutes to weekly. Since the majority of available data feeds (e.g., traffic surveillance) are available on a finer scale (e.g., 20 seconds), the archive system needs to be equipped with temporal aggregation functionality.

Most agencies prefer web-based access to the data archive over the internet in which access security is directly addressed. A few agencies also suggest automated periodical data dumps or “offline” CD-ROM deliveries. Due to the nature of a data archive, it is acceptable that the lead time for some data collection, processing and arrangement be as much as a few months. Agencies prefer a much more timely access to data, but recognize that some items may take longer (e.g., crash reports often are not available until several weeks or longer after the occurrence). For data retrieval efficiency, most agencies suggest that, depending on the type of data, the most recent 2-4 years of data should be immediately available, while 10-15 years of retrievable “offline” data is acceptable.

It is also worth noting that only a few agencies currently do any data archiving, and even fewer agencies would be able to host the data archive system. This suggests that establishing a new centralized data warehouse is the best approach.

On the business side, most agencies will not be able to guarantee financial investment in this effort, especially long-term commitment for system maintenance and upgrades. Therefore, the data archive needs to have a sustainable financial model. One possibility is to charge users for data usage and to compensate providers for data provision. Most of the agencies say that they will be able to find ways to absorb any potential "profit" generated by the data archive, so long as it is not in the form of a direct cash payment to the agency. Data quality control and an equitable, fair pricing scheme are desirable for the system. Also, each data provider will need to identify the types of data that cannot be released due to legal restrictions, privacy, or security concerns.

3.3 CURRENT DATA EXCHANGE

Figure 3-1 summarizes the current exchange of data in the region at a high level, including both automated data exchanges and those requiring human intervention of some kind, based entirely on the interviews.

To further establish a framework for the data archive, both agency websites and the *Northeastern Illinois Regional ITS Architecture* were reviewed (via the on-line source). The purpose was to identify data exchanges that were listed as currently available, but apparently not shared, based on the fact that they were not identified in the agency interviews. This expanded list of potential data exchanges that might be implemented in the data archive is presented in Appendix B.

IDOT

- **SYSTEM TO SYSTEM**

Safety Datamart
Surveillance Data

- **SYSTEM TO HUMAN**

Construction / Maintenance Data
Incident Data
Bridge Restriction Data
Weather Data gathered by weather sensors

IDOT DISTRICTS (District 1)

- **SYSTEM TO SYSTEM**

Surveillance Data
Loop Detector Status

- **SYSTEM TO HUMAN**

Incident Data
Weather Data
Construction / Maintenance Data
Lane Closure Data

- **UNSHARED DATA**

Com Center Reports
ETP Reports

TOLLWAY

- **SYSTEM TO SYSTEM**

Surveillance Data

- **SYSTEM TO HUMAN**

Weather Data
Incident Data
Traffic and Revenue Forecast
Construction / Maintenance Data
Lane Closure Data

NOTE

SYSTEM TO SYSTEM: Automated electronic exchange

SYSTEM TO HUMAN: Human interaction required; may include email, hard copy, fax etc.

GCM

- **SYSTEM TO SYSTEM**

Congestion Data
Dector Reports
Incident Reports
Dynamic Message Sign Reports
Construction / Maintenance Data
Lane Closure Data
Travel time Reports

WILL COUNTY

- **SYSTEM TO SYSTEM**

Construction / Maintenance Data

- **SYSTEM TO HUMAN**

Weather Data
Signal Inventory

LAKE COUNTY

- **SYSTEM TO SYSTEM**

Video and loop detector data
Surveillance Data

- **SYSTEM TO HUMAN**

Weather Data
Incident Data
Construction / Maintenance Data

CMAP

- **SYSTEM TO HUMAN**

Land Use Data
Transit Data
Surveillance Data
Incident Data
Parking Data
Construction / Maintenance Data

RTA

- **SYSTEM TO HUMAN**

Environment Data
Transit Data
Performance Data
Weather Data
Incident Data
Parking Data
Construction / Maintenance Data
Asset Inventory

TRANSIT AGENCIES (CTA, PACE etc.)

- **SYSTEM TO HUMAN**

Transit Data
Parking Data

CITY OF CHICAGO & OEMC (including 311 Center and 911 CAD)

- **SYSTEM TO SYSTEM**

Traffic Signal Data
Video Surveillance Data

- **SYSTEM TO HUMAN**

Incident Data
Construction/Maintenance Data
Land Use Data
Incident Reports (not geocoded)

AIRPORT WEATHER STATIONS & PRIVATE WEATHER INFORMATION PROVIDERS

- **SYSTEM TO HUMAN**

Transportation Weather Information

Figure 3-1. Current Data Exchanges – High Level (see next page for descriptions of the data).

Following are the details of available data identified by the agencies in Figure 3-1:

AGENCY	Surveillance data (real time traffic/conditions)	Construction / Maintenance / Condition data	Event-based data (incidents, causalities, etc.)	Other (parking, public transit, environmental and weather, passenger information)
OEMC	▪ Real-time Signal (3000 signals inventory) ▪ ADT (1989-2006, 1200 location,15 min intervals, GIS based) ▪ Video detector (15 min aggregate level) ▪ Redlight enforcement information	▪ SunTrack (construction, city operations information)	▪ Crash report (monthly from Chicago police department) ▪ 911 (real time incident)	▪ Plow truck information ▪ Contracted weather service (highway advisory radio) ▪ Dynamic message signs (40 portable signs controlled by TMC-Traffic Management Center)
TOLLWAY	▪ Traffic (speed, volume, occupancy, OD)	▪ Lane closures and estimated impact	▪ Incident/crash report	▪ Traffic and revenue forecast 17 weather sensors in pavement
IDOT District I	▪ Traffic counts, volume, occupancy, speed, travel times (20-sec, 5-min, 1-hour data for up to 48 hrs, daily data for up to one month) ▪ Detector status ▪ Ramp	▪ Lane closures (7500 records per half year)		▪ DMS message (hourly updates) ▪ Weather data (snow operations) ▪ Metering rates (fixed rates)
IDOT	▪ Datamart ▪ ATMS, ISP	▪ Daily statewide road construction map ▪ Bridge restrictions	▪ Incident report	

AGENCY	Surveillance data (real time traffic/conditions)	Construction / Maintenance / Condition data	Event-based data (incidents, causalities, etc.)	Other (parking, public transit, environmental and weather, passenger information)
RTA	▪ Transit info (headway, travel time, ridership, level of service) ▪ Performance data by route (ridership for 10-12 years, with 6 months of lag) ▪ AVL reports (per 2 minutes, highly noisy) ▪ Signal priority (signal timing, travel time) ▪		▪ Incident data	▪ Asset inventory
WILL COUNTY	▪ Surveillance data ▪	▪ Construction data		▪ Signal Inventory
CMAP				▪ Schematic network model of regional highway and transit systems ▪ Specific inventory data, including a regional highway traffic signal inventory. ▪ Limited traffic data that is mainly for the production of traffic demand model ▪ Land use data, including a land development database
LAKE COUNTY	▪ Travel time estimate from video and loop detector data ▪ Traffic volume, speed, occupancy		▪ Police reports for incident data	▪ Weather data available from airports

3.4 DATA ARCHIVE GOALS AND OBJECTIVES

The mission for a regional data archive was addressed in the application for this project: to provide a single source of transportation related information to facilitate not just additional planning and research, but also improved operations. To achieve this need and meet the requirements of the systems engineering process, goals and objectives must be identified. Goals are a high level statement of some result that when accomplished, achieves all or part of the mission. Objectives are a set of statements that include specific measurable accomplishments that are directly related to the opportunity, need, or problem addressed by a Goal. The Goals are aims that support fulfillment of the mission and the Objectives are the steps/actions necessary to accomplish the goals. Table 3-1 identifies the various goals and objectives for the regional data archive system.

Table 3-1. Data Archive Goals and Objectives

Goals	Objectives
1-To provide better transportation data for all agencies in northeastern Illinois	a. Collect both transit and highway data from every transportation agency in the region above township level b. Collect data from the larger municipal transportation agencies. c. Allow open access to the data for all public uses.
2-To be a sustainable resource to the public agencies throughout northeastern Illinois	a. Maintain data for an extended period of time (up to 20 years) b. To increase the type of data provided by agencies over time. c. To address funding for the archive from the beginning, including potential costs for services. d. To ensure the integrity and longevity of all data collected, so that it is secure and can be migrated to new formats
3-To provide service back to those that contribute	a. Incoming data is to be checked for errors, and host agencies informed if any potential problems are detected. b. Participating agencies should have a voice in the management of the data archive c. The data archive should produce recurring reports summarizing the information available and highlighting any anomalies identified. d. The data archive should provide assistance, such as spatial reconciliation, to assist participating agencies in providing data

3.5 HIGH-LEVEL REQUIREMENTS FOR THE DATA ARCHIVE

Following the Systems Engineering process, the requirements of a system flow from the goals and objectives. The following requirements are matched to their respective objectives identified in Table 3-1. Based on interview responses and agency discussions, the following high-level requirements for the data archive have been identified, along with identifiers of the associated Table 3-1 Objectives in brackets:

- The data archiving system should be managed with input and oversight from the participating agencies [2a, 3b]

- Depending on the business model, the data archiving system may be managed by a third party [2c, 3b]
- The data archiving system should have at least 10 years of data (eventually), with recent data on-line and older data easily retrieved [2a]
- Data content should expand or contract over time in response to user requirements [2b]
- Users should have remote access to the databases [1c]
- The user interface should be friendly and appropriate [1c]
- The data should be as complete and current as possible [1a, 1b]
- The system should perform validation checks on incoming data [3a]
- The archive should have data aggregation capabilities, and produce and distribute reports [3c]
- There should be some payment for use of the system [2c]
- Data must be secure from tampering and loss [2d]
- As much data as possible should be included in the warehouse, and it should not be limited by storage constraints [2a]
- The warehouse must resolve spatial and time differences, and should have inherent GIS capabilities [3d]

These requirements address the design at a high level. In Chapter 4, more detailed technical requirements for the regional data archive will be developed, and specific types and quantities of data for inclusion will be identified.

CHAPTER 4: CONCEPTUAL DESIGN

4.1. CENTRALIZED vs. DECENTRALIZED ARCHITECTURE

Residency of data is a major issue in the physical model of a data archive. There are intermediate options between the extreme cases where the data either resides in a single, centralized repository, or resides in disparate locations on a network. A decentralized archive system mainly is defined by different physical data storage locations; logically each one of them works as a single data processor that is connected by a virtual network to the others. Decentralized systems appear as a non-distributed database system to the user, although each one of the individual databases is managed by the local authority that carries that particular part of the system.

Both centralized and decentralized archives have their advantages and disadvantages; the details are summarized in Table 4-1.

Table 4-1: Comparison between Centralized and Decentralized Archives

Features	Centralized Archive	Decentralized Archive
Resource requirement (e.g., storage devices and communication infrastructure)	Lower	Higher
Autonomy and control	Centralized control with little local autonomy, once the data is provided	Much local autonomy at data storage sites
Performance	Occasionally, less efficient data query and retrieval because there is more data to sort through.	More efficient only in case of one or two databases needing to be queried.
	Much more often, likely faster performance because the system will be designed from the start for data storage/ access	In more typical case of retrieving information from multiple agencies, performance depends on weakest link – if one agency does not have the resources, their system could slow the entire query and retrieval down
Reliability/availability	A single point of failure is vulnerable to system/site breakdown, so redundancy must be built in	Improved reliability against total site breakdown, but with increased points of failure, higher likelihood of performance issues
Economics	Economy of scale for initial build	Dis-economy of scale for initial build

Features	Centralized Archive	Decentralized Archive
Query and reports	Some flexibility in reporting to different formats	Requires strict adherence to rules and standards
Physical complexity	Lower	High (vulnerable to bugs in software)
Communications requirements	Less in that once data is loaded to central repository, only one fast, secure linkage needed for data query and retrieval	More in that query and retrieval typically requires connections to several different sites
Development and maintenance	Simpler and less costly when compared to the sum of the other improvements that are required in the decentralized architecture	Much more complex and costly though costs are shared among participating agencies/entities
Security concerns	Easier to control access	It is open to more points of entry, so rules and checks must be in place to protect databases
Ability to Modify	Simpler since it is only one system	Requires high level of coordination/ synchronization between all partners

During the interviews with multiple transportation agencies in northeastern Illinois (see the interview summary), the majority requested to have an independent database unit that manages and maintains the data archive. They mentioned that their agencies do not have enough financial and personnel sources to support the regional data archiving system. Often there is a lack of resources to support any basic archiving, let alone actively managing a portion of a distributed system. With active and coordinated support and participation from various agencies, it is likely that a centralized system is much more suitable to fulfill the functional requirements.

Decentralized systems have advantages such as storing data closer to the source, and increased overall reliability by not putting all the information into a single storage site. But they are more complex and more costly for development of the computer software to link the different storage sites together. Decentralized systems require a high investment in maintenance, synchronization, and upgrades. Generally, security of the decentralized systems is more at risk, since there would be more than one database administrator agency and more potential entry points to "hack in" to the database. Distribution of control also has potential problems such as difficulty acting in a coordinated manner in case of a change and improvement, and configuration management is more difficult. Additionally, each agency has to continually work in a coordinated fashion to create a homogeneous system, which is a challenge for any regional project. Finally, institutional issues at various agencies may generate resistance.

In reviewing the basic architecture of the archive, considerations are the amount of data to access, the frequency of access, and the consequences of not being able to access that data. On balance and in light of the needs and capabilities in northeastern Illinois, a well planned and executed centralized system is recommended.

4.2 CONCEPTUAL SCHEMATIC

4.2.1. Overview

The schematic shown in Figure 4-1 illustrates a conceptualization of the proposed regional data archive system for northeastern Illinois. The components in the schematic, as numbered for reference, describe the system architecture. This section explains these components, their major functionalities, and flows of information.

The system would consist of three layers/stages, as follows:

i. Input layer (external to the data archive system)
ii. System layer (includes multiple stages and components of the data archive system)
iii. Output layer (external to the data archive system)

4.2.1.1 Input layer (Component 1)

The input layer includes all the data sources such as loop detectors, incident reports, and meteorological stations. Such data will be automatically fed into the system as shown in Component 1a. The agencies that may provide the initial data to the system include: Chicago Metropolitan Agency for Planning, IDOT Central Operations, IDOT District 1, Illinois State Toll Highway Authority, Lake County, City of Chicago Office of Emergency Management & Communications, Regional Transportation Authority and Will County, plus potentially large cities such as Aurora, Joliet, and Naperville.

Certain information (such as roadway topology, and device installation and removal) may occasionally change over time. Whenever such change happens, the data archive maintainer should be able to update the databases accordingly through an interface represented by Component 1b in the schematic.

4.2.1.2 System layer (Components 2-14)

The system layer includes all key system components and the various stages that the data go through inside the system. This layer will be further discussed in Section 4.2.2.

4.2.1.3 Output layer (Components 15)

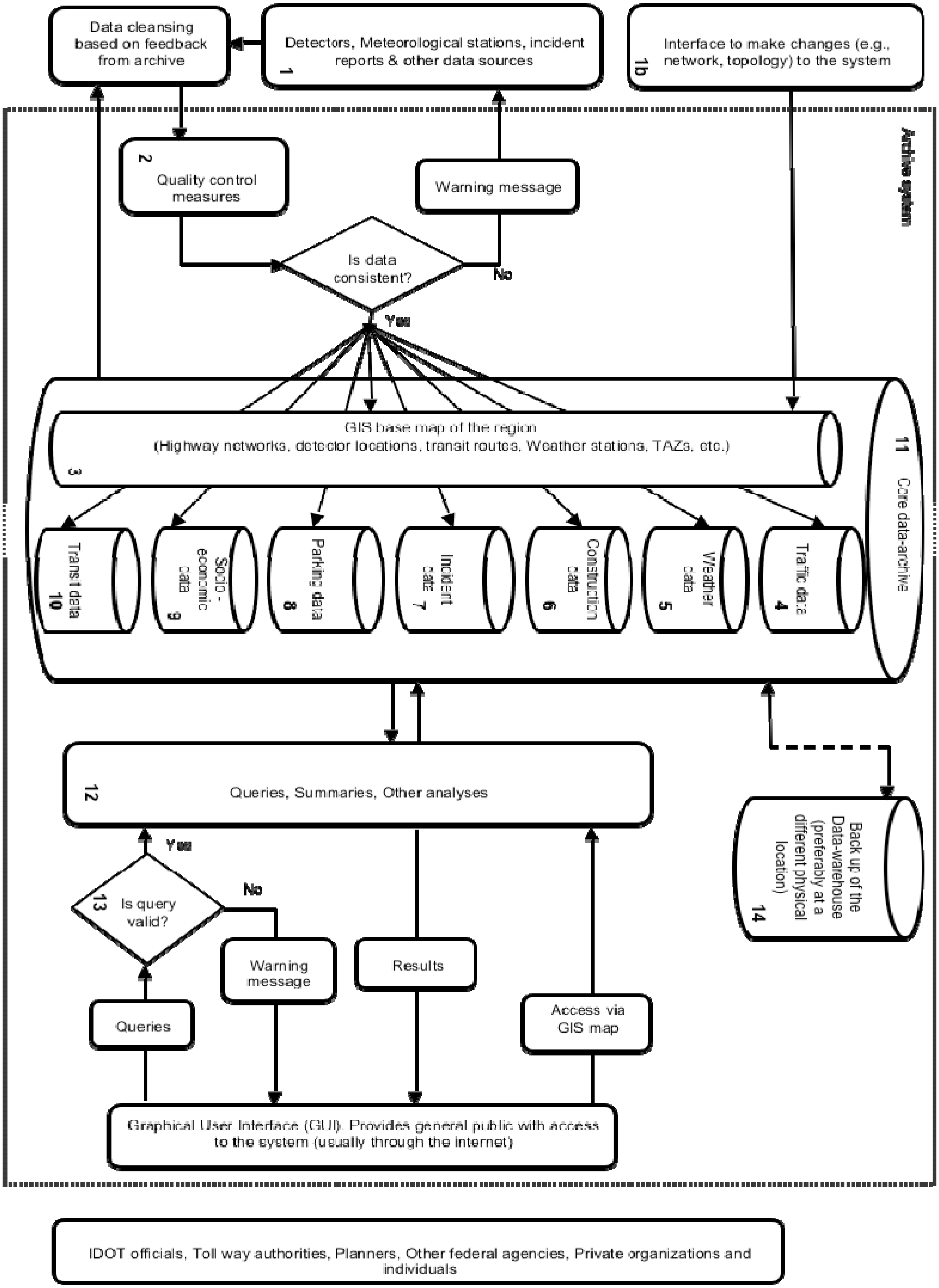


Figure 4-1. Conceptual schematic of the regional data archive system.

This layer refers to the outside users who will interact with the system through a suitable interface (most likely through the Internet). Potential users may include transportation operations / planning agencies (who also provide data to the system), academic researchers, private organizations (who conduct value-adding services), and the general public. Generally, the archive system intends to provide the general public with access to data, though probably with different levels of authorized data-retrieving capabilities. However, it will be mandatory that all users register on the system through established procedures, so that the usage of the system for each user can be tracked and made available for revenue generating or system improvement purposes and for system security.

4.2.2. System Layer

The following part explains in more detail the system layer and the proposed workflow.

4.2.2.1 Data Cleansing (Component 2)

The analyses and results that the archive system can provide are dependent on the raw input data. If the data entering the system contains errors, the output information will also contain errors and the analysis results will be unreliable. Unfortunately, data from loop detectors, RTMS units, and other devices, which are the main source of real time traffic data, are not guaranteed to be error free. For example, at any given time there are typically some detectors that have been noted by the agencies as having a potential problem and are being investigated. For California PeMS, the percent of problematic detectors can be as high as 40%

Hence, a quality control component is mandatory for the data archive system. Various processes and checks (e.g., based on established state-of-art data mining algorithms, range checks, and artificial intelligence approaches) should be conducted at this layer to ensure that the data entering the system are consistent and reliable. For example, this component should be able to detect missing records in the incoming data and abnormal values of numerical entries, as compared with historical records. When the data are found to satisfy certain quality control requirements, they are permitted to enter into the system and are stored in the predefined data marts in the system. On the contrary if the data are found to be erroneous, the data is not accepted into the system and a feedback warning is sent to the data suppliers notifying them of the detail of the errors. The suppliers should accordingly perform data cleansing procedures, which may include imputation or correction algorithms, to remove the sources of the errors and to improve the quality of the data. Throughout this process, the suppliers may make use of existing data in the archive system (i.e., historical records), to check for “regular” patterns of the data and to facilitate data cleansing. The improved data should then be sent back to the archive system, and the quality control process is to be repeated.

The data archive system should also have back-up data capabilities that will check data quality even in case the data provider fails to fulfill the data cleansing function. In that case, a further warning message and permanent record may be generated for future reference, and the data provider may be penalized based on established business rules; see Chapter 4 “Business Model” for further information.

In summary, the quality control measures in this component have two main functions: to keep the data in the system accurate, and to help the data providers improve the quality of the data.

4.2.2.2 Data Marts (Components 3-10)

Components 3-10 in the schematic represent the data marts, which are subsets of the complete archive that each holds specific information that can be related to information in other data marts for distillation and analysis. Each data mart in the system can store data that is usually relevant to one particular type. Often, however, more than one data provider may access and feed data into the same data mart. Data from different sources may be redundant or even inconsistent at times; the data is to be sorted out and fused before storage in a data mart.

GIS Based map (Component 3)

All the information in this data mart is related to some physical dimension, e.g., real-world road networks and detector locations. A GIS-based map showing different stretches of roads, detector locations, or field equipment locations is central to the functioning of the data archive. All the other data stored in the system should reference this data mart spatial relationship. Various maps (e.g., for different geographical regions, from different sources, with different resolutions, and using different location referencing systems) should be compiled and integrated into a comprehensive map for northeastern Illinois. This integrated map system would be central for location referencing, spatial matching and effective presentation of all archived information. The archive may be able to adopt spatial integration techniques or compilation engines from the Gary Chicago Milwaukee (GCM) corridor information program. The data stored in this data mart would be generally static and non-volatile. Apart from roadway networks, locations of detectors, transit routes, weather stations, and Traffic Analysis Zone (TAZ) boundaries would be included in this data mart.

Traffic data (Component 4)

The traffic data mart holds all the real-time traffic information that is generated by sensors such as loop detectors. The data usually includes traffic speed, flow, and occupancy in addition to temporal information. Such data may be fed to the data archive in real time, or may be uploaded to it regularly, e.g., once a day. In either case upon arrival of raw data, information is aggregated and sorted (usually 5-minute or 15-minute averages) before storage. The data in this mart may have multiple levels of granularity.

Weather data (Component 5)

The weather data mart holds data from meteorological stations for different geographical regions in northeastern Illinois. The weather data usually includes temperature, precipitation, humidity, surface conditions, weather alerts (if any), and other information. The data in this mart are likely to be received quite frequently, probably many times, in a single day.

A potential source of the weather data for the archive system is the CLARUS system initiated by FHWA (http://www.clarusinitiative.org). The CLARUS system aims to provide real-time weather data and weather forecasts across all the road infrastructure of the country. A comprehensive set of weather data including the atmospheric condition and the related performance of the road surface (at the roadway segment level) is available through this initiative, hence this system could be an important source of weather data in addition to other independent sources.

Construction data (Component 6)

The construction data mart holds data on any construction activities in the region. Information regarding spatial locations, date/time, the number of lanes closed (if any), type of work, and other relevant information is to be stored. Details about moving closures and the real-time location of vehicles could be incorporated to give details to the users about the location of closures at specific times. The data in this mart is event-driven, i.e., a record is generated and stored only when an activity occurs.

Incident data (Component 7)

The incident data mart holds information about any of the incidents (including special events) on the roadways. Incidents include crashes, vehicle breakdowns and other similar data. Information regarding spatial location, date/time, severity, and other relevant items is to be stored. Data from traffic management centers (TMCs) may include all incidents but might not be exhaustive about the details/nature of the incidents. Meanwhile, the statewide crash database only includes crashes but not other incidents (e.g., sports events) that may impact traffic operations, which may be restrictive for the intended applications of the archive system (e.g., traffic analysis). Hence, information from TMCs and the statewide crash database should be merged carefully so that the input to the archive system is complete but without duplicates. The unique Case-IDs of the incidents/crashes may be good record identifiers for the merge process. Date/time and location filters may also be used to screen for duplicates. Such data are also event-driven.

Parking data (Component 8)

The parking data mart holds the data regarding parking practices, locations of meters or lots, number of spaces available, and number of spaces used. The data here are mixed with some static in nature (e.g., location of lots and number of total spaces) while others are dynamic (e.g., spaces used and, to a lesser extent, number of available spaces).

Socio-economic data (Component 9)

The socio-economic data (e.g., definition of Traffic Analysis Zones, demographical information such as population, incomes, gender, etc.) are stored in this data mart. Such information is relatively static in nature although potentially voluminous, although it will be updated on an annual or longer basis.

Transit data (Component 10)

The transit data mart holds all data on transit routes, frequency of dispatches, revenue, ridership and other related data. The data here are mixed with some static in nature (e.g., transit routes and location of stations) while others are dynamic (e.g., passenger loadings, fare collections). Additionally, the data from CTA and PACE moving sensors (i.e., CTA and PACE busses with on-board sensors) may provide data on real-time locations and speeds of the buses.

4.2.2.3 DATA WAREHOUSE (Component 11)

The data warehouse is the collection of all the data marts including the means to access, manage, and analyze each data type. If the warehouse is compared to a cabinet, then the data marts are the different drawers in the cabinet. Data are stored in different data marts so that there is clarity and separation in the types of data that are dealt with in the system.

Data stored in multiple data marts must have a clear hierarchical structure for easy access. The proposed regional data archive is to be a relational database system where data are stored in the form of tables. Many tables reside in each data mart, and all marts are part of the data warehouse.

4.2.2.4 Query Processing Layer (Component 12)

The query-processing layer includes programs for summarizing relevant data from the data warehouse to answer user-generated inquiries. Other analyses such as delay computation and travel time estimation may also be built in this layer. The query processing programs should be activated in two ways. The user could request data either via a GIS map-based interface (Component 3) by clicking on the target of interest, or via SQL queries to request data directly from the data marts (Components 4-10). Either way, an additional query verification component is needed, as described below. Over time, the data archive would be expected to build up a library of standard queries for commonly requested information derived from custom queries that users complete.

4.2.2.5 Query Verification (Component 13)

The queries made by the users may be invalid (e.g., when the users request data for road links or detectors that do not exist in the time period of interest), or may have inconsistencies (when new detectors were added during the time period requested). The query verification component checks if the query is consistent with the existing data and if meaningful results can be generated. If yes, the query-processing layer is contacted to run the necessary analyses. If the query is found to be invalid, a warning message is sent back to the user. The warning message may be either fatal (such that the query would not be executed) or nonfatal (such that the query is still run, but the user is notified of the potential issues).

4.2.2.6 Backup Data Warehouse (Component 14)

The backup data warehouse retrieves and stores an extra copy of all original data in the data warehouse, backing-up another copy preferably at a different physical location to prevent damage and loss of data from possible threats. The backup data may be updated every day or other suitable time intervals. If disruption to the data warehouse (Component 11) occurs, the data in the backup warehouse must be quickly retrieved and utilized for data recovery.

4.2.2.7 Gary Chicago Milwaukee Corridor Information Program

The existing Gary Chicago Milwaukee (GCM) corridor information program could be an important source of data for the northeastern Illinois regional data archiving system. The GCM, similar to the proposed system, is assimilating traffic and some related data from different sources and supplying this information to officials and the public. The key differences are that only a limited portion of GCM data is archived, and the GCM does not cover the range of data envisioned for the archive. Since the setup for collecting and disseminating the data is already available, this setup could supply a large portion of the data to the data archiving system. The GCM corridor information program could either continue to operate independently even after the implementation of the data archiving system, or it could be integrated into the data archiving system depending on the circumstances. In either case, it is important that the needs of both the data archive and the GCM corridor information program be jointly considered and developed in a coordinated manner.

4.2.3. Potential Applications of the Archive

The applications of the data archive system and its usefulness act as driving forces for its functional design, creation and maintenance. Since the data archive system is to contain a variety of data from different sources, there are numerous possible applications. In general, the GIS- based map would form the basis for spatially connecting information from different data marts, and the data marts would provide a wealth of temporal data. Some examples of the potential applications of the data archive system are briefly described below. For the following examples, it is assumed that the data archive has been populated with information from the last 10 years. It is also assumed that the archive has built-in graphing, correlation, and statistical analysis tools that can be easily used to compare and contrast data from different data marts.

4.2.3.1 Safety Analysis

Automobile crashes are one of the major causes of deaths in Illinois. Over the years major initiatives have been carried out for safer transportation. The data archive system would be an effective tool for planners and highway authorities to study how safety has evolved over time on different stretches of the roads, and how best to invest resources in safety improvement projects. The data archive could also understand the impact of new safety measures on the occurrence of crashes.

For example, suppose a user wants to study safety improvements on I-80 as a result of major reconstruction by both IDOT and the Tollway. The user may start with the GIS map (via the archive interface, Component 15) to select the road segment of I-80 between the Indiana state line and the I-294 split. The analysis will need traffic data for this part of the highway, which will be available in the traffic data mart (Component 4). Note that such data may be originally from both the Tollway and IDOT (through Components 1a and 2), but the user would not need to know this when starting the query. The analysis will also need the crash records that are stored in the incident data mart (Component 7), which may originally come from different Illinois State Police districts. The user may also want construction dates from the construction data mart (Component 6).

Appendix C illustrates a possible user interface, with which the user does not need to know the exact jurisdictions involved. First, the user may highlight the segment of I-80 in question and selects check boxes for traffic and accident data. The system asks the user for what dates, what type of traffic data (ADT vs. hourly volume/speed) and what type of crash information. The user selects ADT for each year and number, type, and injuries from the accidents. The system then gives the user a message that the ADT information is from two different sources as well as the accident information sources. The user notes having read and understood this warning, or checks a box for more information. Finally the user selects the output method and waits for the information. The format of the output data likely will not be one simple graph or report, but possibly a series of tables that needs to be further manipulated. In this case, the output might come in two tables: one a list of segments with their start and end mileposts and their ADTs over the time period; the other a list of crashes by mile post with date/time, type of accident, and number and type of injuries. The user might then request more detailed information on construction, weather, or cross-classified accident information by type and severity. These items might then be considered together to determine whether or not the improvement has had the desired safety effect.

4.2.3.2 Ramp Metering Effectiveness

Ramp meters help in reducing congestion in freeways by regulating the entry of traffic from the ramps into the freeways. The data archive system could help in studying the impact of ramp metering on reducing congestion and by improving traffic throughput on freeways.

For example, suppose a user wants to study the impact of the 'recent' installation of a ramp meter on I-94 near downtown Chicago. The user could start with the GIS map as described in Section 4.2.3.1. However, the user may have knowledge of the highway network and jurisdictions, such that direct queries can be entered via the archive interface (Component 15). An illustrative query interface is shown in Appendix C, second part. The user may select the road segments of I-94 immediately upstream and downstream of the ramp meter in question (entering milepost ranges), and select time periods before and after the metering installation. The analysis will need traffic data (ADT vs. hourly volume/speed) for this part of the highway, which will be available in the traffic data mart (Component 4). The analysis may also need crash records from the incident data mart (Component 7), and construction information from the construction datamart (Component 6) if they impact comparative results. The user selects the output method and waits for the output data, which will likely include a brief report and a series of tables (data regarding upstream and downstream of the ramp meter, before and after the installation) for further manipulation. Detector data may need interpretation and analysis to measure both congestion and throughput, since congestion measures are derived from raw volume, speed and occupancy data. As with the first example, the user might run a series of queries and reports to address all aspects of the basic issue.

4.2.3.3 Operations Use

Operations have different needs relative to a data archive. Often the needs are immediate, such as a major incident. For example, a hazardous material spill has occurred on the outbound Stevenson Expressway (I-55) at 2 pm on a Wednesday. IDOT knows it will take at least 6 hours to get the road cleared, or the entire rush hour. An operator quickly types in the limits of interest for I-55 and the two adjacent expressways – the Eisenhower (I-290) and the Dan Ryan (I-90/94) – and asks for a report titled “next 12 hours.” For each expressway, the results show the hourly volumes that are expected on a Wednesday in the current month based on the last three years for each expressway. The report also generates the hourly volumes from the last Wednesday for the same 12 hours of each expressway. From this information, IDOT can determine about how much traffic has to be diverted to the adjacent expressways. Ramp metering rates are changed. Potentially signal timing along parallel surface streets could be modified to help move the additional traffic. The media are alerted to help get some people in downtown to either leave work early, or stay longer.

In this situation, access and use of the archive must be quick and easy so the information can be rapidly acquired without distracting from the incident management activity at hand. The TMC operator would not normally have time to conduct follow-up queries, so the first round must provide all the critical information. Standard queries, developed and refined from actual operating experience, would play a big part in making the archive a very valuable tool under such circumstances.

4.2.3.4 Transit Performance Analysis

Transit is one of the major modes of transportation, especially in urban areas. But the popularity and user acceptance of transit services depend on the quality and convenience of the level of service. The data archive system could help facilitate studies on the effectiveness of transit services, which may help transit agencies improve service quality.

The data archive system could provide sufficient data not only on transit routes and operational records (e.g., ridership, headways, etc.), but also on the highway network (e.g., traffic, construction, incident, parking availability) in the geographical proximity. Such integrated data could be retrieved from the archive either via the interactive GIS map interface or directly through queries based on topological or jurisdiction information. The user might select the geographic section and request all bus routes in the area (Component 10), any arterial traffic information (ADTs or daily/hourly volumes; Component 4), and park-and-ride information (Component 8). The system might let the user know that several bus routes are available, but only ADT information is available on the selected arterials. Several tables would be presented to the user. Questions addressed might be ridership and major boarding/alighting points by time of day, travel time, and relationships to the adjacent road system. In this regard, information from local agencies (cities and counties) might play a key role to help illuminate conditions on roads that buses use. In the future, surface street systems may provide performance measures such as travel times and degree of congestion. Such information, if archived, could then be a tool for transit agencies to assess their services and consider route or operating changes.

4.2.3.5 Long Term Land Use Planning

A sound plan on metropolitan land use and future transportation infrastructure investment is essential for the future development of the region. Suppose the user is interested in studying how the provision of transportation infrastructure affects the socio-economic development in a corridor. The user would select a region graphically. From the socio-economic data mart (Component 9), the user would request housing and population data. Additionally, the aggregated traffic data (e.g., yearly, Component 4) on all selected roads would be requested. This would also include information on the highway network (Component 3). Also in Component 3, transit routes over time could be requested along with passenger loadings on those routes from Component 10. The data archive would provide the data in several tables, noting gaps in data, where none is available (e.g., minor roads) and any other issues related to the query. The user would then likely construct a series of queries to correlate demographic growth trends to infrastructure changes over time.

CHAPTER 5 BUSINESS MODEL

5.1 INTRODUCTION

A business model describes various aspects of a business system, such as the purpose of the system, operating policies, infrastructure needed, and the costs and revenue sharing mechanism. Once the purpose of the system, the infrastructure needed and the cost estimate have been prepared, the next step is to decide upon a suitable cost and revenue sharing mechanism that is acceptable to all stakeholders. This chapter deals with the cost and revenue aspects of the business plan for the Northeast Illinois Regional Data Archiving & Management System (i.e., “the Illinois System”).

As part of the efforts to develop the Illinois System, Section 5.2 first reviews the cost and revenue sharing mechanisms that are in place for some of the existing data archiving systems across the nation and details one specific case. Then, Section 5.3 estimates the initial setup and annual costs for the Illinois System under a few possible implementation scenarios. Section 5.4 discusses probable revenue sources and preliminary ideas on data provision and pricing mechanisms, followed by recommendations in Section 5.5.

5.2 REVIEW OF SIMILAR DATA-ARCHIVING SYSTEMS

5.2.1 Existing Archive Costs and Revenue Sharing

The earlier literature review (Chapter 2) covers several existing data archive systems in the nation. This section goes beyond the technical details and focuses on these systems’ cost and revenue sharing mechanisms. Almost all of the established data archive systems are funded by their corresponding state DOTs or the FHWA as societal benefit projects. Some of the systems are funded jointly by partnering agencies. Some of the important benefits to the society and the partnering agencies include the following: (i) improved traffic operations, (ii) ability to evaluate existing or planned transportation operations, and (iii) increased safety practices.

From the review of the existing systems, it was observed that the data for these systems come mainly from different government agencies, predominantly from the state DOTs and other public agencies concerned with traffic flow and road safety. None of the reviewed systems have private organizations as significant data suppliers. The one potential exception is Traffic.com. While they are a private organization that deploys their own sensors, they are largely funded through an FHWA grant, and they only archive their own sensors. Traffic.com typically operates as a service provider to a public agency (the Illinois Tollway in the case of northeastern Illinois) providing data that can be used and shared by the agency to some degree, subject to some proprietary contract limitations.

Another important observation is that most existing systems do not consider themselves to be an important revenue generator. Revenue from these systems is generally not stated as a primary means through which the maintenance, operations, and upgrading costs are covered. Furthermore in some cases, the system is being operated and maintained by a partnering university who contributes to the operational and maintenance costs. Table 5-1 summarizes the reviewed systems in more detail.

Table 5-1. Cost and Cost Sharing Mechanism of Several Existing Data Archiving Systems (U.S. DOT, Intelligent Transportation Systems, 2005)

System	Cost Sharing Mechanism
Washington TRAC	\$ 70,000 for initial deployment of ADMS[1] \$ 15,000 for data preparation/distribution via CDs \$ 250,000 to \$ 300,000 – biannual budget All costs covered by WSDOT
King County Metro, Seattle	Funded by FHWA under the Metropolitan Model Deployment Initiative Program
Caltrans PeMS	Initial setup cost - \$ 8 Million \$150,000 to \$200,000 for annual upgrades All costs paid by Caltrans System was originally hosted at UC Berkeley, but now subcontracted to Berkeley Transportation Systems
Michigan MITS	Costs are shared by MDOT & FHWA
Minnesota TMC	Cost for system development paid by MnDOT Cost for hosting the system paid by UM Duluth
Maricopa County RADS	Funded through Federal Congestion Mitigation and Air Quality (CMAQ) Funded through a cooperative effort of Maricopa County, Arizona DOT, and Maricopa Association of Governments.
Virginia ADMS	Initially funded by FHWA and VDOT as an operational test Maintenance and upgrades by VDOT \$ 300,000 set aside every year by VDOT

5.2.2 EXAMPLE: ADMS Virginia

The recently developed Virginia Archived Data Management System (ADMS) is reviewed in this section for information regarding the funding, implementation, and operations of the system. The following is a summary of the noteworthy details.

5.2.2.1 Partners

The Virginia ADMS is operated by a partnership among government agencies, universities in the region and private firms. The partnering agencies are (University of Virginia 2008):

- Virginia Department of Transportation (VDOT)
- Virginia Transportation Research Council
- University of Virginia (UVA) Center for Transportation Studies
- Open Roads Consulting, Inc.
- George Mason University
- Federal Highway Administration (FHWA)

5.2.2.2 Traffic Management Centers and Initial Data Archive Development

The traffic control and monitoring operations of the system are controlled from various traffic management centers (TMC) across the state, as follows:

- Northern Virginia Smart Traffic Center
- Hampton Roads Smart Traffic Center
- Smart Travel Lab at UVA.

Each of the traffic management centers is responsible for particular regions and operations of the system. For example, the Northern Virginia Smart Traffic Center handles the operations of congestion mitigation, incident management, managing HOV lane operations and traffic planning efforts (VDOT, 2008a). The Hampton Roads Smart Traffic Center monitors and evaluates interstate traffic flow, communicates information to the public and sends assistance when necessary (VDOT, 2008b). The Smart Travel Lab at UVA provides researchers at the university an opportunity to perform studies on the data and related phenomena (University of Virginia, 2008b).

According to the official *Concept of Operations* document used to prototype and develop the system, “The Smart Travel Laboratory (STL) is the designated official ITS archived data management system (ADMS) for the state of Virginia (University of Virginia, 2004). Another official document on TMC applications states that “VDOT formally established the Virginia Smart Travel Laboratory (STL) as the facility responsible for archiving ITS data. The STL, established in 1998, maintains a real-time data and video connection to all VDOT STCs (Smart Traffic Center)” (University of Virginia, 2002). In the first several years of ADMS development, the STL at the University of Virginia housed the data archiving system with connections to various TMCs. The archive database was centralized at the STL while the operations of traffic control and monitoring were distributed across different TMCs.

5.2.2.3 Implementation of the System

Nine legacy systems of the transportation department had operated on four different database systems before the ADMS system was developed. Furthermore, there had been three different hardware platforms and three operating systems in use; the ADMS bridges these different systems into a unified one, forming a single standard for operations throughout the state. “The system under development at VDOT will allow for the screening, quality control, and summarization of the data as it is input into the data warehouse” (TTI, 1991). It appears that the new ADMS was planned to replace the archiving function of the legacy systems with all data transferred to the new system. While each TMC uses a common protocol to exchange data, the systems themselves remain on their various legacy equipment. The ADMS was built in a series of incremental steps implemented at different locations progressively, with each incremental step increasing the functionality and the coverage of the system. The production implementation, handed over from the STL to VDOT in 2006, is essentially a centralized archive, housed in the information technology center in Richmond.

5.2.2.4 Costs Sharing

"The Federal Highway Administration (FHWA) and the Virginia Department of Transportation (VDOT) *originally funded* ADMS Virginia as an operational test. VDOT *continues to sponsor* this project and is leading the effort with the team members of the Virginia Transportation Research Council (VTRC) and the University of Virginia (UVA) Center for Transportation Studies (CTS). UVA has subcontracted the software development part of the project to Open Roads Consulting, Inc. (ORCI) (U.S. DOT, 2005). The maintenance and upgrading of the system are funded by VDOT, which has set aside an amount of $300,000 for each year for a support contract (U.S. DOT, 2005).

Annual costs are incurred for the operation and maintenance of the system which include expenses towards staff, office space, utilities etc. Recent conversation with ADMS developers confirmed that the annual operating expenses are on the order of $300,000 per year; the bulk of this is from fully burdened salaries (Smith, 2008). Revenue from the system is not a primary source of funding. In summary, the funding for the initial setup of the system and for the subsequent maintenance and operations has been taken care of by public agencies.

5.2.2.5 Additional Information: Lessons Learned

The follow-up conversation with the Virginia ADMS developers also yielded other vital information relative to the deployment of an ADMS. When discussing the staffing required for such a system, the need for a champion who understands the data and its usage is crucial. It is critical for the system manager to be able to converse with the data suppliers and users and to understand their needs, restrictions, and desires. As such, while IT staff are critical to the development and maintenance of the system, they typically do not have the right skill set to manage it. Instead, a person with an understanding of the installed field equipment and an appreciation for traffic operations is the type of individual needed.

Also emphasized was the required coordination between the data archive and the TMCs that supply most of the data. Virginia now recognizes that a more integrated implementation that combines the archive with on-going operations in a more user-friendly manner would make it a more frequently used and valuable tool. Data quality checking and validation have been mentioned before, but the relationship is more symbiotic than this. Both operations and planning help each other improve the way they function. It is important that these two main users be recognized. As stated, the manager must be able to communicate effectively with both operators and planners, and the system should be capable of addressing the needs of both.

Virginia has found that the ADMS has led to a much higher level of performance monitoring and evaluation than existed previously. Further as a result, there is a demand to maintain a large network of detectors at a much higher level than ever before.

5.3 ESTIMATED ANNUAL COSTS FOR THE ILLINOIS SYSTEM

To lay out the business model for the Illinois System, the annual costs likely to be incurred need be estimated. As a part of this cost estimation process, it is helpful to compare possible options and decide how the system should be implemented and operated. Hence, four system implementation alternatives are considered and the corresponding costs are estimated in this section.

Section 4.2.2.7 has discussed from a technical point of view the potential to adopt or utilize the GCM Gateway hardware or software (e.g., future enhanced versions) for the proposed data archive. In the following cost analysis for all business options, however, we do not specifically consider such possibilities so that the cost estimate would be on the conservative side.

5.3.1 Option 1: System Hosted By a Partnering Agency

The first option is that one of the partnering agencies is responsible for hosting, operating and maintaining the system. The assumption is that the data archiving is a new unit within the host agency, not combined with existing responsibilities. It may or may not be co-located within the host agency. As of August 2008, only one agency within the region expressed any desire and had the capabilities to assume new staff – the Chicago Metropolitan Agency for Planning (CMAP). CMAP is currently developing a related data warehouse that will have some of the desired features of the data archive discussed here, though with not as much real-time information flowing in. The following analysis attempts to address the issues from a more generic standpoint, but issues specific to CMAP will be addressed.

The details about how this agency would recover the costs are explained in detail in the subsequent Section 5.4. The different cost components for setting up and running the system, such as expected hardware infrastructure (including workstations, servers and peripherals) as well as system operating costs (including staffing, space, utility, and maintenance costs), are estimated. The costs have been split up into four components: initial hardware set-up costs, initial software costs, recurring operational costs, and recurring maintenance costs.

5.3.1.1 Hardware Set-up Costs

The Illinois System is likely to require the following hardware infrastructure:

(i) Workstations
The number of professionals that would most likely be required to operate the data archiving system is about the equivalent of 1.5 full-time employees who would require two workstations. The employees for initial system loading will be considered a part of the software set-up costs. It is expected that each employee would have a dedicated workstation. Additionally, two monitors would also be required to monitor the activities of the servers. As the system will likely have an interface with the agency LAN, additional workstations may be available from other staff within the agency at no additional cost.

(ii) Database server
The database server would run a suitable Database Management System (DBMS) and would respond to the queries of the users. It would house the data archive system and would have the expected storage requirements.

(iii) Traffic data server
The traffic data server would receive the real-time data from the detectors in the field. It would maintain records of the state of detectors and run preliminary operations (data checking and aggregation) on the incoming data before forwarding it to the database server.

(iv) Backup servers
The backup servers would contain duplicate copies of the important information in the system so that critical data is not lost if some malfunction of hardware or software occurs. It is recommended that the backup servers be placed in another location.

(v) GIS map server
The GIS map server would house all the map-related data of the system, and would cater to the operations that require map representation of data.

(vi) Web server

The web server would be the interface for participating agencies and the general public to access the system through the Internet. This server contains the GUI components. Hence it would also address the security needs and firewalls to prevent the system from any external malicious attacks.

(vii) Wall display
Wall-mounted displays would be provided to display archived data in an easily readable format.

A rough estimate of the system requirements has been made based on preliminary knowledge of the nature of the data to be stored in the system. According to the stakeholder survey, it is generally agreed that at any time, the system should store data for the last ten to twenty years. The total storage space required for ten years of data is estimated to be around 8.7 Terabytes (TB) (1 Terabyte = approximately 1 Million Megabytes) [*]. This would be the minimum storage capacity required for the database server. The requirements of other servers have been estimated based on the nature of the data to be stored. The storage devices can be purchased at the initial setup or progressively over the years of operations. The estimates have been made assuming that all storage devices would be purchased at the initial setup (and each subsequent hardware replacement/upgrade to be completed approximately every five years or so). Therefore, the initial storage requirements could easily be reduced to the expected 5 or 10 year capacity with the assumption that the equipment will be upgraded in later years to better address the long term storage requirements. While it is easiest to budget equipment if portions of the system are replaced every year, it is likely that several components will be changed at the same time. This is especially true if the operating system or key software components have to be upgraded.

A summary of the requirements and the estimated costs of the hardware components have been listed in Table 5-2. The costs are computed by considering the current prices of the components at Dell's website, as of April 29, 2008.

* The detailed responses from the stakeholders on a data storage requirement survey are shown in Appendix D.

Table 5-2. Estimated Costs for Initial Hardware Set-up (rounded up)

Hardware item	Quantity	Storage capacity	Price range ** ($)	Total cost ($)
Workstations	2	250 GB	1400 – 1600	2800 – 3200
Workstation monitors	4	-	400 – 600	1600 – 2400
Database server	1	10 TB	8600 – 8800	8600 - 8800
Traffic data server	1	4 TB	4400 – 4600	4400 – 4600
Backup server	2	10 TB	8600 – 8800	17200 – 17600
GIS map server	1	2 TB	2800 – 3000	2800 – 3000
Web server	1	2 TB	2800 – 3000	2800 – 3000
Wall-mounted displays	2	-	1400 – 1600	2800 - 3200
TOTAL	14			42,400 - 46,400

Expected total hardware set-up costs = *$ 42,400 - $ 46,400.* For estimation purposes, we take the high end of the above range which is **$46,000** per year.

5.3.1.2 Software Set-up Costs

In addition to the purchase and installation of hardware, the software also needs to be developed at the start of the system. The software would define the way in which the data are stored in the tables in the database, actual loading of data into these tables during the initial phase, creation of initial standard queries and the creation of web interfaces through which the users would interact with and query the system. Although an accurate estimate of the costs involved in this process is not available at this time, an approximate estimate has been made based on the study made on existing similar systems (e.g., PeMS). The software set-up costs are expected to be in the range of $700,000-$1,000,000. This would mainly be the compensation provided to the developers. Additionally, the development contract may possibly include the first year or two of operations. This facilitates addressing ongoing fixes and initial upgrades.

5.3.1.3 Operational Costs

The main components of operational costs include labor, power and office space.

(i) Labor cost

Conservatively, we assume that the equivalent of 1.5 full-time employees is initially required to operate the system. The full time senior employee will most appropriately be an intelligent transportation system expert who is very familiar with the road network and can oversee data collection, processing and manipulation, and promote/market the archive system. The equivalent of one half-time employee may ideally be an information technology expert (or two quarter-time employees) who is (are) capable of supporting the operations of the archive.

** Cost estimates are from Dell's website. Standard configurations of Dell Precision Workstations and Dell PowerEdge Tower Servers (with added storage capacity) are considered.

Assume the baseline average annual salary for a full-time employee is $85,000, then *Annual direct labor cost = $85,000* x *1.5 = $127,500.*

Hence, the estimated total direct labor cost for the system is $127,500 per year. Depending on the agency fringe benefits, the real cost could be approximately 1.5 – 2.5 times this amount. Considering this, the total labor cost could be up to $127,500 x 2.5 ≈ $320,000 per year.

(ii) Power cost

We assume that the power consumption rate of each hardware component is 175 Watts and that all the components will work throughout the whole year. For simplicity of analysis, we assume the HVAC for the computer systems consume a similar amount of power to that of the computer equipment[1]. If the average electricity cost in Chicago, as of May 2008, is $ 0.11/kW-hr (Source: http://www.think-energy.net/electricitycosts.htm), then we have

Annual power cost
= *2* x *0.175* x *14(components)* x *24(hours)* x *365 (days)* x *0.11 ≈ $ 4700.*

Thus, the total power cost for the system would be approximately $4,700 per year. Regardless of where the ADMS is located, this is the expected additional cost that the agency would have to bear.

(iii) Communications cost

Communication infrastructure is required for data transfer and exchange, regardless if the data is continuously fed into the archive (e.g., for traffic data) or if the data is periodically transferred in bulk quantities (e.g., for incident data). The investment should cover at least the following:

- Any leased lines between centers
- Access to an internet portal

We currently assume that the hosting agency and data suppliers/users will use their established communication infrastructure, and hence we assume that the owning agency picks up the cost of any ongoing work/upgrades.

Annual Communication cost = 0.

Thus, there are no communications cost assumed for the system. Depending on where the Data Archive is housed and how it is connected to the various sources, this number could change.

(iv) Office space costs

We estimate that three office rooms, each with a size of 300 square feet, would be needed to house the system equipment and personnel. Given that the current average rental cost is about $22 per square foot per year in suburban Chicago, or $32 in the downtown CBD (Source: http://www.chicago-office-space.com/), then for the total space (900 square feet).

[1] Normally, HVAC would consume more energy than computer units; however, the impact of our assumption on total annual costs is very slight.

Annual rental cost = 22 x *900 = $ 19,800*   *(Suburban Chicago)*

or

Annual rental cost = 32 x *900 = $ 28,800*   *(Downtown Chicago)*

The total rental cost will be $19,800 to $ 28,800 per year. For estimation purposes, we take the average of the above two costs which is $24,300 per year.

If the Data Archive is to be housed at an existing agency, the space may already be available. Even though this would still represent an opportunity cost (less space for staff, etc.) the space costs may be effectively $0. It should also be noted that if any new facilities have to be constructed in the future, the one-time set-up cost for office space would be much higher than the entire Data Archive, but will be assumed to come from a separate source of funding.

Hence, the total operational cost including labor, power, communications, and office space is approximately $349,000 per year.

5.3.1.4. Maintenance Costs

Assuming that the maintenance routines for the hardware and software are outsourced to an on-call maintenance solution provider, the cost would be approximately **$10,000 to $15,000** per year (Caltrans 2004). This service would generally not include hardware upgrades but would include general program debugging and troubleshooting. It should be noted that this suggested cost is for preventative maintenance; in case of unexpected outages or other severe problems, the cost for maintenance and recovery may be much higher.

In addition, it is reasonable to budget for annual software upgrades to improve coverage and performance. The upgrades are estimated at 5-7% of initial development cost, or about $50,000.

Hence, the total maintenance and software upgrade cost is about **$65,000** per year.

5.3.1.5 Life Cycle Equipment Costs

Computer systems require occasional upgrades and replacement. A conservative estimate would be that all of the equipment is replaced every five years. The computer industry has been relatively unusual in that equipment costs typically do not rise over time. Conversely, capabilities increase and costs are relatively stable. Therefore, if we divide the total replacement costs over each of the five years, this will result in an equipment replacement budget of:

Annual equipment replacement = $46,000 / 5 ≈ **$9,000 per year.**

5.3.1.6. Summary

The costs for this option are summarized below. It should be noted that all costs are in constant (non-inflating) 2008 dollars.

Item	Initial Purchase Cost	Annual Costs
Hardware (including set-up) with a life cycle of 5 years	$ 46,000	$ 9,000
Software	$ 700,000 – $ 1,000,000	(included in Maintenance)
Operational Costs	New Building?	$ 349,000
Maintenance Costs		$ 65,000
Total	**$746,000 - $1,046,000**	**$423,000**

The major advantage of hosting the archive system at one or more partnering agencies (e.g., CMAP) is that the existing databases and communication architecture can probably be partially integrated into the proposed archive system. This may potentially save hardware procurement and software development costs. Also, the staffing costs may be absorbed at least partially by existing manpower programs. Data transfers will be required among the agencies, and agencies other than the host may support computer and communication costs.

The major challenge to the hosting agency is the capability of assuming the new responsibility associated with the proposed archive system. For example, the agency has to reconcile any potential conflicts that may arise between existing IT systems and the proposed archive, regarding aspects such as staffing, hardware procurement, and management. The archive will require a new commitment of staff and other resources.

The partnering agency host (e.g., CMAP) may have ongoing planning projects that can provide cost sharing opportunities that support a portion of the operational costs (e.g., labor and maintenance). This will be a very desirable advantage for hosting the data archive at one of our partnering agencies.

5.3.2. OPTION 2: SYSTEM CONTRACTED TO A UNIVERSITY

5.3.2.1 Features

Almost all of the existing data archiving systems in the nation have some sort of collaboration with a university in the region; in some cases the system is hosted, operated, and maintained by a university. Accordingly, the Illinois System could also be hosted by a university in the northeastern Illinois region. In such a case, the operational and maintenance costs would be borne by the university as overhead, and the university would have to provide office space to house the system. The revenue from the system or funding from the partnering agencies would in turn compensate the costs borne by the university. Professional staff members and students in the university could supervise the operations of the system. The university would also stand to benefit in terms of the opportunity created for the researchers and the students to access the system.

There are several universities in the greater Chicago region that have previously expressed interest in this project. It would be expected that under this option, an RFP would be created and distributed to the various Illinois universities to allow them to competitively bid on this project.

The estimation of the costs for the system would be very much similar to those of *'Option 1,'* as discussed in Section 5.3.1. Minor differences would be that the office rental cost and labor cost might be cheaper (see Section 5.3.1.3.). For example, the office rental cost may be on the low end (most likely in the suburban area), which is about $19,800 per year. There are ways to reduce staffing costs too. For example, it is possible that university personnel can donate some of their time to the project (e.g., paid only part-time from the project, reduced overhead charges), such that only one full-time staff is needed to oversee the system; this reduces the labor costs (including benefits) by about 1/3 to approximately $320,000 x 2/3 ≈ $213,000 per year. In this case, the total operational costs reduce to $238,000 per year. In addition, the staffing costs may be further reduced if part of the salaries can be billable to other related projects.

Other cost estimates are expected to be similar to those in the previous section.

5.3.2.2 Summary

The costs for this option (in non-inflating 2008 dollars) are summarized below.

Item	*Initial Purchase Cost*	*Annual Costs*
Hardware (including set-up) with a life cycle of 5 years	$ 46,000	$ 9,000
Software	$ 700,000 – $ 1,000,000	(included in Maintenance)
Operational Costs	New Building?	$ 238,000
Maintenance Costs		$ 65,000
Total	**$746,000 - $1,046,000**	**$312,000**

The major advantage of contracting out the archive system to a university is the abundance of manpower (e.g., graduate students) and IT technical expertise. The support staffing costs may be significantly reduced. The need for a champion as manager remains.

The major challenge is that the databases and communication architecture for the proposed archive system probably needs to be established from the ground up.

5.3.3. Option 3: System Contracted To a National Lab

5.3.3.1 Features

Alternatively, a national lab in the region could host the system. Similar to the previous cases, the national lab would incur operational costs such as labor costs and rental costs. The possible revenue from the system and funding from the partnering agencies should in turn compensate the expenses borne by national lab. The main difference between national laboratories and other hosts (e.g., university) is their connection with the federal agencies for procuring funding, equipment, and services.

There are two national labs in the greater Chicago region. Argonne National Lab is quasi-public and managed through the University of Chicago. The other national lab is Fermi Lab that is private. If this option is chosen, there exists the possibility of requesting proposals from both labs. The larger issue is identifying sources for funding this option.

In terms of the costs involved with this option, there is an assumed savings in the initial hardware setup costs, since the hardware should be already in place at the national lab. But the labor costs would be similar to *'Option 1'*, since the personnel working at the national lab would be full time employees. Hence for 1.5 full time professionals, an annual labor cost of about $320,000 would be reasonable. Rental costs would be mostly similar to that of *'Option 2,'* with an approximate estimated cost of around $19,800 per year (considering that the national lab is most likely to be located at a suburban location). Hence the annual operational costs are of the order of $ 345,000. Other cost components would be consistent with the other options discussed above.

5.3.3.2 Summary

The costs for this option (in non-inflating 2008 dollars) are summarized below.

Item	Initial Purchase Cost	Annual Costs
Hardware (including set-up) with a life cycle of 5 years		$ 9,000
Software	$ 700,000 – $ 1,000,000	(included in Maintenance)
Operational Costs		$ 345,000
Maintenance Costs		$ 65,000
Total	**$ 700,000 - $ 1,000,000**	**$ 419,000**

The major advantage of contracting out the archive system to a national laboratory is the abundance of hardware, computing resources, and IT technical expertise. The hardware and software maintenance/upgrade tasks may be contracted to the laboratory with somewhat lower costs.

The major challenge is that some of the communication architecture for the proposed archive system probably needs to be established from scratch.

5.3.4 Option 4: System Outsourced To a Service Provider

5.3.4.1 Features

Instead of developing the data archiving system from scratch, the entire development and operation could be outsourced to an outside organization that specializes in data archiving systems (e.g., some operational tasks of the California PeMS system is outsourced to Berkeley Transportation System, Inc.). The advantage of pursuing this option is that experts who have successful experience in this regard would be able to contribute their knowledge and expertise to the Illinois System. Per conversation with BTS, recent development of PeMS allows integration of arterial data and transit data to freeway sensor data, which will probably satisfy the needs of the Illinois system (Skabardonis, 2008; Petty, 2008).

The website of BTS, Inc. (http://www.bt-systems.com/products.html) states the following: "BTS sells PeMS in two modes. First, you can purchase a standalone system which can be installed at your local site and integrated into your other transportation management systems. Alternatively, you can purchase PeMS as a service. In this scenario, BTS would collect your freeway sensor data in real-time and transport it to our facilities. Once there, we would filter it, apply our diagnostic routines, perform the proper level of imputation, and aggregate across various spatial and temporal dimensions. The results are then presented to the users via a series of dynamic web pages. Full user account control is given to the customer, with the ability to give different users permission to see different parts of the site." The costs involved for either PeMS outsourcing option (standalone system or service) are explored. Per conversations with BTS, if Illinois purchases a standalone system, the cost components include those for software licensing, system/hardware setting-up, customization, and optional hardware purchase. The purchasing price for a system with thousands of detectors will be on the order of several hundred thousand dollars. BTS can also provide training and maintenance service (hardware, database, data cleansing / processing software), which will incur costs for approximately one full-time employee. Purchasing standalone system can be regarded as the "software procurement" part that is part of the previous three business options.

Alternatively, if Illinois purchases PeMS as a service, there will be no hardware/software accommodation needs. The data will be stored at a remote location (probably outside of Illinois) and PeMS will provide real-time access to the data while meeting certain level-of-service requirements in the contract. Normally BTS does not claim ownership of the data, and off-line back-ups can be provided. With this outsourcing option, the initial purchasing price for system implementation and set-up will be lower. However, there will be a higher annual cost for service (i.e., operations and maintenance). BTS could not provide more detailed cost estimates without more detailed functional requirements.

Another potential supplier is RoadStats, LLC, the developer and host for www.gcmtravelstats.com (see Chapter 2). This locally based company today archives 5-minute sensor and travel time data from northeastern Illinois freeways (Saville, 2008) and provides a way for travelers to obtain both current and historical travel time data on any instrumented freeway segment. RoadStats has worked with the data feeds from the GCM Gateway Traveler Information System (see Chapter 2) to develop the travel statistics web site. This working relationship highlights the need to consider the needs of the Gateway and the northeastern Illinois data archive in combination at the same time, consistent with the "lessons learned" from the Virginia ADMS (see Section 5.2.2.5**)**.

There likely are other similar data archiving service providers in the nation, each of which may have their own business model for pricing the service. Traffic.com is another private sector option that could be contracted to expand their current Tollway-only data archive. The agencies within the region could solicit proposals from such private interests to determine if they have more cost-effective solutions.

5.4.3.2. Summary

The advantage of outsourcing the archive system to service providers is that existing expertise would directly contribute to the Illinois System. The benefits include shorter time and lower costs for system (hardware and software) development and implementation.

A potential challenge is higher annual costs, and a contracting agency must be found and funding sources identified. Also, this option may not be as conducive to having a regional champion for the data archive as the other options. On the institutional side, would the data archive be able to provide real time service to the users? Since there are multiple stakeholders involved in the archive project, preparing the system outsource contract may be challenging. Finally because the outsourcing contract might terminate in the future, the ownership and transfer of archived data must be considered in the initial contract.

5.4 CONCEPTS OF REVENUE & FUNDING SOURCES

After analyzing hardware requirements and operating costs for the Illinois System, the next step is to examine possible funding sources and revenue-generating mechanisms that could offset the costs. The potential funding sources to set up the system and to meet the operational and maintenance costs are listed in Table 5-3.

5.4.1 Funding From Federal Funds or Partnering Agencies

Almost all of the existing data archiving systems in the nation are funded by partnering agencies, or the Federal Highway Administration, or a combination of both, for the initial setup, subsequent maintenance and upgrades. Existing cost sharing mechanisms have been presented in Table 5-1. Ideally, some sort of agreement would be made among the partnering agencies to jointly bear the costs of the system. It will be desirable if some agencies could help fund the Illinois System, at least for the system set-up stage.

In the long-run, if the agencies can not commit financial support for annual operations and maintenance, other sustainable approaches must be sought. In order for the system to be sustainable, the benefits of the system to society and to the partnering agencies (in terms of the available information and the improved management of traffic conditions) should exceed the costs of the system.

Table 5-3. Potential Sources of Funding and Revenue

Source of revenue	Costs covered	Justification of payment
Partnering agencies (Some projects had funding from FHWA)	Initial set-up costs Operation costs	Benefit to society Improved operations and management
Other non-partnering government agencies	Operation and maintenance costs	Data acquisition
Value-adding resellers or private business (e.g., tourism agencies, trucking industry, concierge services or location based service providers)	Operation and maintenance costs	Data acquisition
University or agency hosting the system	Operation and maintenance costs (partial)	Opportunity for researchers and students to gain knowledge
Advertisements	Operation and maintenance costs	Advertisers pay to display their message in the system's website

Historically, the most common approach has been this: several agencies with federal aid manage to complete the initial study and implement the system. After implementation, one or more host agencies agree to assume the operating costs. This approach may not be as viable in the near term. Federal funding continues to dwindle as FHWA continues to promote alternative financing. It may be necessary for not only the operations funding, but also the initial build, to come from sources other than FHWA.

Some of the agencies involved in this project do fund significant ITS and operational projects without any federal aid. But many of these projects are larger legacy efforts, and current local budgets are also constrained. Agencies may find it difficult to secure non-federal money for a significant contribution to this project. Several participating agencies have expressed this concern from the very beginning of this project.

Option 2 or 3 may become the most viable from a funding standpoint if research money is found outside of the current transportation funding. Option 1 may also be viable if the initial build costs are reduced sufficiently so that the partnering agencies can collectively secure the required funding. Under all options, a regular annual funding stream must be obtained if the data archive is to succeed. Possible mechanisms for the federal funding (if any) to support the Illinois data archiving effort shall be explored in the implementation stage of this project.

5.4.2 Data Metering / Data Trading

While there is no short term plan to generate significant revenue, this option is still much more palatable when considered at the beginning of the project than after implementation. To this end, the data metering or the data trading method could be a model for partnering agencies to develop a cost/revenue sharing agreement on generating revenue from the system. If the system is to be hosted/operated by any one of the partnering agencies, the revenue from the system could contribute at least in part to offset the costs incurred. The following mechanism could be used to finance the system.

5.4.2.1 Sharing Mechanism

Each of the partnering agencies holds certain types of data that are essential for the Illinois System, and the system will receive some data from these partnering agencies. The system will gather data from different sources to a single place and process the data into useful information. The value of the data increases due to not only the "pooling" effect, but also the many quality control processes that improve the accuracy, value and relevance of the data pool.

This value adding process provides the justification for the system to charge certain fees from the users based on the amount of data they download. The business rule for these processes could be as follows. For the sake of illustration, let us assume that:

* There are n partnering agencies and users; $i = 1, 2, \ldots, n$.
* Agency i supplies the system with an average of B_i Megabytes (MB) of data per year (after cleansing and quality control) and downloads C_i Megabytes (MB) of data per year. These amounts are committed by yearly contracts. Without losing generality, we allow a certain user (or partnering agency) to contribute B_i=0 MB to the system.
* Let D be the data utilization indicator; i.e., $D = \max\left\{0, \sum_{i=1}^{n} C_i - \sum_{i=1}^{n} B_i\right\}$. This is the net amount of information distributed by the system (the total amount of data downloaded by all users minus the total amount of data uploaded by all agencies). Collectively, the users including agencies will eventually download more (copies of) data than they upload to the system, and hence D would be greater than 0. For the short term, while the system is being created and established, D will likely be close to 0 (i.e., the agencies are supplying data rather than utilizing them).
* The total prorated set-up/operating/maintenance cost of the archive system is \$ X/year, which is to be transferred to the system host.

The following scheme applies to the long term when D>0. For partnering agency (or user) i, it should be allowed to:

(i) freely download data up to a total amount of B_i.
Beyond that,

(ii) if $C_i > B_i$, pay for the remaining download amount of $C_i – B_i$, at the following price

$$P_i = X * (C_i - B_i)/D.$$

(iii) if $C_i < B_i$, receive revenue of

$$R_i = X * (B_i - C_i)/D.$$

Obviously, the more data an agency provides (i.e., the larger the value of B_i), the more likely that agency will receive a positive revenue (R_i>0). It can be easily verified that the above pricing/revenue return scheme will ensure that the system is able to sustain itself (covering the total cost of \$$X$ per year). It shall be noted that the above data trading mechanism assumes that the participating agency will directly receive revenue, in a way that agency participation will be encouraged (although the amount of revenue in the short term is likely to be minimum). A possible alternative is that the hosting agency collects all (or a portion of) revenue in support of system operation.

In some situations like Transcom, the partnering agencies agree to provide a certain amount of subsidies up front according to some agreed upon formula, but then are able to realize some profits from other Transcom activities. The profits are applied back to according to the same formula. The profits are never enough to significantly offset the costs of the subsidies. In the case of northeastern Illinois a similar application might result in a reduction in agency subsidies to support the data archive.

Additionally, it needs to be recognized that the majority of the agencies participating will still use their own data that is maintained at the agency for their own purposes for free. That is, an agency will not send all of its data to the archive and then use the archive to retrieve its own data for traffic analysis. So the primary users of the archive are those who are interested in data from multiple agencies. The number of such users is expected to be small.

In the above scheme, no distinction is made between non-profit organizations (e.g., research institutions) and private data users (e.g., IT firms, consultants). For non-profit organizations, where $B=0$, we can simply exclude them from the above computation and provide them with free data access. For private data users, where $B=0$ too, we may charge an extra price margin for each unit amount of data downloaded. This option would generate additional revenue that can be shared by the data suppliers, proportional to their data contribution (or be stored temporarily in a cash reserve; see Section 5.4.2.3 below). For consultants doing work for an agency, the cost of retrieving data may be covered by the other project funds.

In the above analysis, we treat all types of data provided from all agencies equal, only to simplify the pricing scheme. In general, we may assign different "values" to different types of data (e.g., static roadway data versus dynamic traffic data). The above formulas will still hold by simply adding multiplicative coefficients (determined by experts or retrospectively based on historical records of archive system usage) to the terms C and B.

It is possible that the estimates of C, B, and X are not available at the beginning of the year (e.g., via anticipative commitments and contracts). Since C, B, X are going to be variable throughout the years, price cannot be determined until the end of the year. In this case, a cash reserve (Section 5.4.2.3) will be critical for building a funding mechanism into the archive system. In practice, the pricing mechanism will probably be modified continuously throughout the first several years until the data provision and utilization mechanisms become well established. Possible initial funding support from the federal sources or local partnering agencies, as described in Section 5.4.1, thus may be critical for successful implementation of the archive system.

5.4.2.2 Quality Control

The quality of the data that is supplied by the agencies is as important as the quantity of data. Hence certain agreements would have to be made such that the accuracy of the data is ensured. For example, in the above computation, B_i represents the quantity of "good" data. In addition, certain penalty measures for supplying inaccurate data should be established so as to minimize the waste of system resources on data imputation and cleaning. The actual rates and the penalty clauses could be decided after consulting the partnering agencies. The amount from these penalties would be an additional source of income to the system, which would help meet any unexpected expenditures.

5.4.2.3 Cash Reserve and Dynamic Pricing

According to the data-metering proposal, the exact cost of the system is being shared by the agencies based on their usage of archived data. There will always be fluctuations in the actual data supply or download amount, which in general will generate cash flows only at the end of the year. We have also mentioned extra charges to private data users. Such extra cash flows can be maintained as a common cash reserve (i.e., bank), for the system to meet contingent expenses that may arise in the future. This could also be particularly helpful in case the agencies do not download the expected amount of data, thus generating a smaller amount of revenue for the system. If this cash reserve grows beyond a particular threshold, the overflow fund could always be redistributed among the partnering agencies based on their actual contribution. Similarly, if the case reserve drains below a threshold, the archive system may dynamically increase the price of data accordingly to make up the shortfall.

5.5 BUSINESS PLAN RECOMMENDATIONS

It must be recognized from the beginning that the primary function of this archive is to assist the participating agencies in their operations, planning, and performance evaluation. Therefore:

- It is assumed that revenues will not amount to a significant amount, at least for the short term.
- Some attempt at revenue generation should be made from the beginning.
- Because the data archive must be suitable for operations (in addition to planning), it must have a quick and easy interface and be able to generate results quickly
- The decision on which option to pursue shall be dictated by the partnering agencies
- The business option chosen must also be matched to the technical requirements.
- Regardless of which business plan option is pursued, the development of the northeast Illinois data archive must be considered in step with short- and long-range needs of the GCM Gateway Traveler Information System.
- Regardless of which business plan option is pursued, the development of the northeast Illinois data archive must be consistent and compatible with the existing ITS Architecture and other legacy systems.
- Similarly, pending full development of CMAP’s data warehouse, there needs to be a link from the data archive to the data warehouse to allow the free flow of information between both.

REFERENCES

Al Deek, H. and Elrahman, A., *An Evaluation Plan for the Conceptual Design of the Florida Transportation Warehouse-Phase 1*, Center of Advanced Transportation Systems Simulation, CATSS, University of Florida, 2002.

Caltrans, "Fontana-Ontario Advanced Traffic Management & Information System (ATMIS) Evaluation Report," 2004, http://www.dot.ca.gov/hq/research/showcase/project_reports/fontana_ontario_evaluation_report_final.pdf, accessed May 4, 2008.

Federal Highway Administration, *Archived Data Management Systems, A Cross-Cutting Study*, 2005, Transportation Research Board: http://www.trb.org/news/blurb_detail.asp?id=5826, accessed August 25, 2008.

Federal Highway Administration, Federal Transit Administration, U.S. Department of Transportation, *Cross-Cutting Studies and State-of-the Practice Reviews: Archive and Use of ITS-Generated Data, 2002,* Center for Transportation, Oak Ridge National Laboratory: http://www.osti.gov/energycitations/servlets/purl/814082-cPPOGL/native/814082.PDF, accessed August 25, 2008.

Margiotta, R., *ITS as a Data Resource: Preliminary Requirements for a User Service,* 1998, Federal Highway Administration: http://www.fhwa.dot.gov/ohim/its/itspage.htm, accessed August 25, 2008.

Petty, K., Berkeley Transportation System, Inc., Telephone conversation, August 14, 2008.

Saville, B., RoadStats, LLC, Telephone conversation, August 15, 2008.

Skabardonis, A., University of California at Berkeley. Telephone conversation, July 25, 2008.

Smith, B., University of Virginia Smart Travel Lab, Telephone conversation, August 4, 2008.

Sunguide, Florida's Intelligent Transportation System (ITS), 2008, http://www.sunguide.org/, accessed August 25, 2008.

Texas Transportation Institute (TTI) (Texas A&M University System), "ITS Data Archiving: Case Study Analysis of San Antonio TransGuide Data," 1999, http://tti.tamu.edu/documents/FHWA-PL-99-024.pdf, accessed June 18, 2008.

Turner, S. M., *Guidelines for Developing ITS Data Archiving Systems,* 2001, The Texas A&M University System, http://tti.tamu.edu/documents/2127-3.pdf, accessed August 25, 2008.

Turner, S. M., *Texas Transportation Institute.* Retrieved August 25, 2008, from The Texas A&M University System, 2001, http://tti.tamu.edu/documents/2127-3.pdf, accessed August 25, 2008.

U.S. Department of Transportation, "Archived Data Management Systems, A Cross-Cutting Study," Intelligent Transportation Systems, 2005, http://www.trb.org/news/blurb_detail.asp?id=5826, accessed on: June 3, 2008.

U.S. Department of Transportation, "National Evaluation of the TMC Applications of Archived

Data Operational Test – ADMS Virginia," 2005, http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_TE/14250_files/14250.pdf, accessed June 3, 2008.

U.S. Department of Transportation, Federal Highway Administration, *Freeway Management and Operations Handbook,* Research and Innovative Technology Administration (RITA), http://ops.fhwa.dot.gov/freewaymgmt/publications/frwy_mgmt_handbook/index.htm, accessed August 25, 2008.

U.S. Department of Transportation, *ITS Joint Program Office Home*, 2008, Research and Innovative Technology Administration (RITA): http://www.its.dot.gov/index.htm, accessed August 25, 2008.

U.S. Department of Transportation, *ITS Standards Program, Final Rule/Policy: ITS Architecture and Standards*, Research and Innovative Technology Administration (RITA): http://www.standards.its.dot.gov/finalRule.asp, acessed August 25, 2005.

U.S. Department of Transportation, *Lessons Learned: Monitoring Highway Congestion and Reliability Using Archived Traffic Detector Data*, 2006, Federal Highway Administration - Office of Operations: http://ops.fhwa.dot.gov/publications/lessons_learned/index.htm#toc, accessed August 25, 2008.

U.S. Department of Transportation, *National ITS Architecture Version 6.0*, 2007, http://www.iteris.com/itsarch/, accessed August 25, 2008.

University of Florida Transportation Research Center, *Development of a Central Data Warehouse for Statewide ITS and Transportation Data in Florida,* Florida Department of Transportation: http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_TE/FDOT_BD545_37_rpt.pdf, accessed August 25, 2008.

University of Virginia, "ADMS Virginia - Build 4: Concept of Operations," 2004, http://cts.virginia.edu/adms/docs/Conops_ADMS_B4_Final.pdf, accessed June 7, 2008.

University of Virginia, "ADMS Virginia," 2008b, http://cts.virginia.edu/stl_adms.htm, accessed June 3, 2008.

University of Virginia, "CTS News," 2008a, http://cts.virginia.edu/news_its_proposal.htm, accessed June 3, 2008.

University of Virginia, "TMC Applications of Archived Data Operational Test: Volume 1: Capabilities," 2002, http://cts.virginia.edu/adms/docs/ArchivedData_Proposal_Volume_1.pdf, accessed June 7, 2008.

University of Virginia, *Smart Travel Lab*, Center for Transportation Studies: http://cts.virginia.edu/stl_adms.htm, accessed August 25, 2008.

Varaiya, P., *Creating Transportation System Intelligence Using PeMS,* Computer Vision and Robotics Research - Events, Seminars and Visitors, UC San Diego, http://cvrr.ucsd.edu/events/PeMS-Pravin.pdf, accessed August 25, 2008.

Varaiya, P., *Freeway Performance Measurement System (PeMS), PeMS 6: Final Report for CCIT TO 15,* 2006, http://www.calccit.org/resources/2007_PDF/060219finalreportPeMS6.pdf, accessed August 25, 2008.

Virginia Department of Transportation, “Northern Virginia Traffic Operations Center,” 2008a, http://www.vdot.virginia.gov/travel/smart-traffic-center-nova.asp, accessed June 3, 2008.

Virginia Department of Transportation, 2008b, “Hampton Roads Traffic Operations Center,” http://www.virginiadot.org/travel/smart-traffic-center-hro.asp, accessed June 3, 2008.

Washington State Department of Transportation, *Intelligent Transportation Systems (ITS) Operations, 2006,* http://www.wsdot.wa.gov/biz/trafficoperations/its/default.htm, accessed August 25, 2008.

Wisconsin Traffic Operations and Safety Laboratory, *About the WisTransPortal Project*, 2007, http://transportal.cee.wisc.edu/about.html, accessed August 25, 2008.

APPENDIX A: INTERVIEW QUESTIONS AND RESPONSES

This appendix compiles written and oral responses to a list of questions posed to the interviewees, in tabular format. The first table lists the interview schedule and participating individuals. Following that table are the details of responses by question. As noted in the text of this report, the interview questions were not necessarily all addressed in order. The questions were used as a discussion guide to gather as much information as efficiently as possible.

INTERVIEW SCHEDULE

AGENCY & DATE	ATTENDEES	
	RESEARCH TEAM	AGENCY OFFICIALS
Office of Emergency Management and Communications (OEMC) Monday, 18th, June 10-12pm	Yanfeng Ouyang Assistant Professor Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign Jeff Hochmuth Hochmuth P.E., P.T.O.E. Principal ITS Engineer Wilbur Smith Associates Paul Pei Research Assistant Industrial & Enterprise Systems Engineering University of Illinois at Urbana-Champaign	David Zavattero Deputy Director Office of Emergency Management and Communications
Illinois State Toll Highway Authority Monday, 18th, June 2-4pm	Yanfeng Ouyang Assistant Professor Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign Jeff Hochmuth Hochmuth P.E., P.T.O.E. Principal ITS Engineer Wilbur Smith Associates James L. Powell Senior ITS Manager Wilbur Smith Associates Paul Pei Research Assistant Industrial & Enterprise Systems Engineering University of Illinois at Urbana-Champaign	John L. Benda Manager of Maintenance & Traffic Illinois State Toll Highway Authority Abby Malloy Traffic Operations Center Manager Illinois State Toll Highway Authority
Chicago Metropolitan Agency for Planning (CMAP) Tuesday, 19th, June 9-11am	Yanfeng Ouyang Assistant Professor Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign James L. Powell Senior ITS Manager Wilbur Smith Associates Paul Pei Research Assistant Industrial & Enterprise Systems Engineering University of Illinois at Urbana-Champaign	Claire Bozic Senior Analyst for Programming Chicago Metropolitan Agency for Planning Tom Murtha Senior Planner for Strategic Initiatives Chicago Metropolitan Agency for Planning Greg Sanders Principal Information Architect Chicago Metropolitan Agency for Planning

AGENCY & DATE	ATTENDEES	
	RESEARCH TEAM	AGENCY OFFICIALS
		Craig Heither Research and Analysis Chicago Metropolitan Agency for Planning
Lake County Tuesday, 19th, June 1-3pm	Yanfeng Ouyang Assistant Professor Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign Jeff Hochmuth Hochmuth P.E., P.T.O.E. Principal ITS Engineer Wilbur Smith Associates Paul Pei Research Assistant Industrial & Enterprise Systems Engineering University of Illinois at Urbana-Champaign	Anthony N. Khawaja P.E. Traffic Engineer Lake County Jonathan Nelson ITS Traffic Signal Engineer Lake County Scott Lee Senior Project Manager Delcan Corporation
Regional Transportation Authority (RTA) Wednesday, 20th, June 9-11am	Yanfeng Ouyang Assistant Professor Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign James L. Powell Senior ITS Manager Wilbur Smith Associates Paul Pei Research Assistant Industrial & Enterprise Systems Engineering University of Illinois at Urbana-Champaign	Duana Love, P.E. Oversight & Technology Development Regional Transportation Authority Gerwin Tumbali Manager, Engineering and Technology Regional Transportation Authority
Will County Wednesday, 20th, June 1-3pm	Yanfeng Ouyang Assistant Professor Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign Paul Pei Research Assistant Industrial & Enterprise Systems Engineering University of Illinois at Urbana-Champaign	Patti Killinger, PE Civil Engineer Will County Dept of Highways Micheal Szubryt Civil Engineer Will County Dept of Highways
IDOT District 1 Wednesday, 25th, July 10-12am	Yanfeng Ouyang Assistant Professor Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign Jeff Hochmuth	Charles S. Sikaras ITS Program Specialist Office of Planning and Programming Jeff Galas IDOT Traffic Systems Center Manager

AGENCY & DATE	ATTENDEES	
	RESEARCH TEAM	AGENCY OFFICIALS
	P.E., P.T.O.E. Principal ITS Engineer Wilbur Smith Associates Umit Deniz Tursun Research Assistant Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign	Justin Potts IDOT Regional IT Manager Region One
IDOT Central Operations Monday, 30th, July 2-4pm	Yanfeng Ouyang Assistant Professor Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign Jeff Hochmuth P.E., P.T.O.E. Principal ITS Engineer Wilbur Smith Associates Umit Deniz Tursun Research Assistant Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign	Aaron Weatherholt P.E. Engineer of Traffic Operations Bureau of Operations Joe Hill, P.E. Engineer of Operations Bureau of Operations Mark D. Kinkade Bureau Chief Bureau of Information Processing Mark Yow Project Manager Bureau of Information Processing

DETAILED QUESTION RESPONSES

1. Data Availability and Needs. Responses are by role: Data Provider and Data User.

1 a. Data Provider

AGENCY	What data can you currently provide?
OEMC	▪ Traffic signal data (3000 signals inventory) ▪ ADT (1995 & 2006, 1200 location,15 min intervals, GIS-based) ▪ Crash report (monthly from Chicago police department) ▪ 911 incident data ▪ SunTrack construction, city operations information ▪ Video detector traffic counts (15 min aggregate level) ▪ Contracted weather service (highway advisory radio) ▪ Dynamic message signs (40 portable signs and 6+ permanent arterial VMS controlled by Traffic Management Center (TMC)) ▪ Plow truck information ▪ Redlight enforcement information
TOLLWAY	▪ Incident/crash reports (since 1999) ▪ Traffic data (speed, volume, occupancy, OD plaza to plaza) ▪ Lane closures and estimated impact ▪ DMS messages ▪ Traffic and revenue forecasts ▪ 17 weather sensors in pavement
IDOT District 1	▪ Traffic volume, occupancy, speed, travel times (20-sec, 5-min, 1-hour data for up to

AGENCY	What data can you currently provide?
	48 hrs, daily data for up to one month) ▪ Detector status ▪ DMS message (hourly updates) ▪ Weather data (snow operations) ▪ Metering rates (fixed rates) ▪ Lane closures (7500 records per half year) ▪ Incident reports from IDOT Comm Center ▪ ETOS – ETP reports ▪ Current ETP response database
IDOT Central Operations	▪ Daily statewide road construction map ▪ Incident reports ▪ Bridge restrictions ▪ Accident Datamart ▪ ATMS, ISP incidents ▪ Roadway Weather Information System (RWIS) Data ▪ Winter Road Condition Data
RTA	▪ Transit info (headway, travel time, ridership, level of service) ▪ Performance data by route (ridership for 10-12 years, with 6 months of lag) ▪ AVL reports (every 2 minutes) ▪ Signal priority (signal timing, travel time) ▪ Incident data ▪ Asset inventory
CMAP	▪ Agency has limited traffic data that is mainly used as input to traffic demand modeling ▪ Agency maintains a schematic network model of regional highway and transit systems. Agency also maintains specific inventory data, including a regional highway traffic signal inventory. ▪ Land use data, including a land development database
WILL COUNTY	▪ Construction data ▪ Signal Inventory
LAKE COUNTY	▪ Travel time estimates from video and loop detector data ▪ Traffic volume, speed, occupancy ▪ Already getting data from IDOT and Tollway ▪ Police reports for incident data ▪ Weather data available from airports

1b. Data User

AGENCY	What data do you currently need from other agencies? What data would you like to see available in the near future?
OEMC	▪ Expressway information (congestion, travel times, detector condition and data) ▪ Tollway information (congestion, travel times, detector condition and data) ▪ Basic TMC information from agencies operating within Cook County ▪ Congestion level ▪ Traffic Counts, traffic volumes ▪ Major incident and event information (location, duration, impact, lane and roadway closures) ▪ Transit information (delays, status, reroutes, service disruptions)
TOLLWAY	▪ Data quality improvements from all ▪ Data storage is not currently available but desired by the Tollway

AGENCY	What data do you currently need from other agencies? What data would you like to see available in the near future?
	▪ The capability to do dynamic vehicle/traffic routing from TIMS ▪ Weather information
IDOT Dist. 1	▪ Incident data (ETP) ▪ RWIS ▪ Ramp counts from all agencies that connect to the IDOT expressway system ▪ The Skyway data in real-time ▪ The Tollway data in real-time
IDOT Central Operations	▪ Additional information from all signals ▪ Traffic counts ▪ Illinois Tollway Construction Information ▪ Illinois Tollway Winter Road Conditions
RTA	▪ Performance overview ▪ Real time travel information (Gateway function?) ▪ An advanced version of the existing trip planner that uses archive information
CMAP	▪ Congestion and delay data (highway travel speeds, volumes, travel times) ▪ Data to produce regional transportation system performance measures ▪ Data to calculate Sustainability measures ▪ Ramp metering data (rates and times of operation) ▪ Traffic volumes by time interval for model calibration ▪ Travel times by time interval for model calibration ▪ Transit volumes by time interval for model calibration, by boarding station or by bus location ▪ Transit vehicle travel time information (to calculate reliability, etc) ▪ Most importantly, CMAP is responsible for regional travel model development and creating new system performance measures. Would like disaggregated data that can aggregate and use for these purposes and for operations analyses ▪ CMAP would also use the disaggregate data for technical analyses of various alternative transportation operations strategies.
WILL COUNTY	▪ Traffic information on expressways and arterials ▪ Traffic count information
LAKE COUNTY	▪ Need ADT ▪ Sensor data is desirable ▪ Capacity analysis is needed for 400 intersections (separate issue?) ▪ Preprocessing of data is desirable ▪ Sensor status should be reported ▪ Weather data ▪ Traffic data ▪ Data provider needs to inform archives about changes ▪ Detect and provide feedback about abnormal input data ▪ Licensed vehicles information by geographical area ▪ Environmental data ▪ Demographics data

See Figure 1 in main text for a listing of current data exchanges as identified by the interview agencies.

2. Data Responsibilities.
   What are the main responsibilities within your agency that currently or may require data from the prospective data warehouse? Which internal division/bureau?

Note: ‘Yes’ in the following table denotes that the corresponding agency has expressed needs for data from the Data archive, while detail of the data needs has not been specified.

Scope	Agency	Data Details	Internal Division
Planning/Programming	OEMC	Yes*	
	TOLLWAY		
	IDOT District 1	Impact study, Surface Traffic Security	
	IDOT Central Ops		
	RTA		
	CMAP	Yes	
	WILL COUNTY		
	LAKE COUNTY	Yes	
Traffic forecasting and simulation	OEMC	Signal timing analysis, simulation	
	TOLLWAY		
	IDOT District 1	Yes	
	IDOT Central Ops	Traffic counts and density	Region 1
	RTA		
	CMAP	Multi-modal speeds, volumes, and travel times	Research and Analysis
	WILL COUNTY		
	LAKE COUNTY		
Performance assessment (agency and/or roadway)	OEMC	Arterial performance monitoring, traffic congestion (light, medium, heavy, etc.), travel times/delay	
	TOLLWAY		
	IDOT District 1		
	IDOT Central Ops		
	RTA		
	CMAP	▪ Multi-modal speeds, Volumes, and Travel Time ▪ Ramp metering data ▪ Incident data (duration) ▪ Construction Data (lanes or entrances blocked, etc.) ▪ Weather Data (Conditions, Precipitation, Pavement Condition)	Programming and Operations
	WILL COUNTY		
	LAKE COUNTY	Yes	
Reporting (statistics, charts, etc.)	OEMC	Traffic and transit performance statistics, major incidents and events (location date/time, duration)	
	TOLLWAY		
	IDOT District 1	Yes	
	IDOT Central Ops	Evacuation, traffic modeling	

Scope	Agency	Data Details	Internal Division
	RTA		
	CMAP	Yes	
	WILL COUNTY		
	LAKE COUNTY	Yes	
Advanced modeling	OEMC	Transportation simulations, evacuation models, network models	
	TOLLWAY		
	IDOT District 1	Commercial traffic/trucking	
	IDOT Central Ops		
	RTA		
	CMAP	(same as "Performance assessment")	Research and Analysis, Programming and Operations
	WILL COUNTY		
	LAKE COUNTY		
Other	IDOT District 1		Public Relations

3. Currently Used Data.
   What types of data do you currently utilize in your agency? How do you currently receive the data?

Type	Agency	Data Details	Method of Receiving Data
Surveillance data (real time traffic/conditions)	OEMC	Operation Virtual Shield CCTV surveillance for traffic conditions, real-time traffic conditions from signal system detectors, arterial performance monitoring	
	TOLLWAY	▪ Speed, volume, occupancy ▪ Lane closures ▪ OD	▪ Shared electronically to and from IDOT and Gateway ▪ Viewed over web by others
	IDOT District 1 IDOT Central Ops	▪ Speed, volume, occupancy ▪ Lane closures	
	RTA	▪ Travel time information	▪ GCM Gateway
	CMAP		
	WILL COUNTY		
	LAKE COUNTY	Detector data from traffic signals	Download detector data from individual intersections
Construction / Maintenance / Condition data	OEMC	▪ Public way permits, road/street and lane closures ▪ Road/street and lane closures for maintenance ▪ Roadway condition,	

Type	Agency	Data Details	Method of Receiving Data
		structure condition, signal conditions / status	
	TOLLWAY	▪ Weather data ▪ Construction and Maintenance lane closures	
	IDOT District 1	▪ Construction and Maintenance lane closures	Shared through www.gcmtravel.com
	IDOT Central Ops	▪ Daily construction map ▪ Bridge Restrictions	
	RTA	▪ Construction Schedule ▪ Environment data ▪ Weather data	National Oceanic and Atmospheric Administration (NOAA)
	CMAP	▪ Limited usage of non-real-time construction data ▪ Highway Performance Management System (mandated by FHWA)	▪ Tollway ▪ IDOT ▪ CMAP does other miscellaneous data collection
	WILL COUNTY		
	LAKE COUNTY	▪ Current work schedule, lane closure status. ▪ Roadway conditions	▪ Phone call
Event-based data (incidents, causalities, etc.)	OEMC	▪ Duration ▪ Notification ▪ Dispatch	▪ Police Report ▪ Real-time report e.g. 911 ▪ Major incidents and events ▪ Crash reports, statistics
	TOLLWAY	▪ Incident, crash data	
	IDOT District 1		
	IDOT Central Ops	▪ Incident	
	RTA	▪ Incident	
	CMAP	▪ Historical crash records for analysis (not geo-coded)	▪ Collect data from IDOT
	WILL COUNTY		
	LAKE COUNTY	▪ Accidents that are affecting the roadway	▪ Data received from connection to Lake County Sheriff Computer Aided Dispatch
Other (parking, public transit, environmental and weather, passenger information)	OEMC	▪ Weather conditions and impacts ▪ Off-street and on-street parking conditions and occupancy ▪ CBD cordon counts, average arterial speeds by roadway /	

Type	Agency	Data Details	Method of Receiving Data
		location / functional class ▪ Airport traffic and parking conditions	
	TOLLWAY		
	IDOT District 1		
	IDOT Central Ops		
	RTA	▪ Limited real time parking information	
	CMAP	▪ Substantial highway and transit facility inventories. ▪ Socioeconomic data	▪ Historical, as modified by programming process ▪ Extensive socioeconomic and land use data collection and forecasting process
	WILL COUNTY		
	LAKE COUNTY		

4. Current Data Format.
   What is the format of the data that you currently utilize?

AGENCY	Fax/ email/ phone	GIS	SQL	XML	ASCII	MS Access	MS Excel	Other (specify)
OEMC	Yes	Yes	Yes					Oracle dbases
TOLLWAY	Yes	Yes	Yes	Yes				
IDOT District 1	Yes		Yes				Yes	
IDOT Central Ops	Yes		Yes	Yes				
RTA	Yes			Yes				
CMAP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
WILL COUNTY	Yes							
LAKE COUNTY	Yes		Yes	Yes				

5. Preferred Data Format.
   What is your preferred format of the data in the centralized warehouse? A few examples include Microsoft Access, XML, direct SQL queries, and spreadsheets.

Agency	Data	Form (e.g., fax, email, voice)	Format (if applicable, e.g., XML)
OEMC			▪ GIS (Locational reference) ▪ XML (most likely) ▪ SQL, Oracle and Oracle triggers ▪ Microsoft Access and Excel
TOLLWAY	**Toll-transponder usage**	▪ RTMS & I-Pass	

Agency	Data	Form (e.g., fax, email, voice)	Format (if applicable, e.g., XML)
	Travel times, speed, occupancy	▪ Shared electronically to another system ▪ Viewed over the internet	▪ XML ▪ SQL ▪ GIS
IDOT District 1	**Travel times, speed, occupancy**	▪ Shared electronically to another system ▪ Viewed over the internet	▪ XML ▪ GIS ▪ SQL (Standard for databases)
IDOT Central Ops			▪ SQL/XML ▪ Should accommodate user query
RTA			▪ XML (for stored data) ▪ SQL (for small data amount) ▪ GIS compatible
CMAP			▪ Depends on data attribute
WILL COUNTY			▪ GIS

6. Accessible Data Types.
   What specific data types would you like to be able to access from the centralized data warehouse (long term)?

Type	Data Details	
Surveillance data (real time traffic/conditions)	OEMC	See list of specific data items in Q#3 for all categories in this table.
	TOLLWAY	Historic speed, volume, and travel times on all roads by hour
	IDOT District 1	Historic speed, volume, and travel times on all roads by hour
	IDOT Central Ops	Speed, Density
	RTA	
	CMAP	Multi-modal speed, volume, travel times
	WILL COUNTY	
	LAKE COUNTY	Travel times, vehicle detector logs
Construction / Maintenance / Condition data	OEMC	See list of specific data items in Q#3
	TOLLWAY	
	IDOT District 1	
	IDOT	▪ Lane Width Restrictions ▪ RWIS ▪ Traffic Speed ▪ Presence of water, ice and salt on pavements and bridges
	RTA	
	CMAP	▪ Construction Data (lanes or entrances blocked, etc.) ▪ HPMS
	WILL COUNTY	
	LAKE COUNTY	Details of other ongoing state / county / municipal construction projects

Type	Data Details	
Event-based data (incidents, causalities, etc.)	OEMC	See list of specific data items in Q#3
	TOLLWAY	
	IDOT District 1	
	IDOT	Data required for traffic modeling
	RTA	
	CMAP	Incident data (type, duration, location/area affected)
	WILL COUNTY	
	LAKE COUNTY	Incidents outside of area covered by Lake County Sheriff
Other (parking, public transit, environmental and weather, passenger information)	OEMC	Parking data is not available but in demand
	TOLLWAY	
	IDOT District 1	
	IDOT	Transit and pedestrian traffic information
	RTA	
	CMAP	Weather (Conditions, Precipitation, Pavement Condition)
	WILL COUNTY	
	LAKE COUNTY	

7. Data Aggregation Level.
   What aggregation level is appropriate for your agency needs?

AGENCY	5 minute	15 minute	Hourly	Daily	Weekly
OEMC	Yes	Yes	Yes	Yes	Yes
TOLLWAY	Yes		Yes	Yes	
IDOT District 1	Yes		Yes	Yes (ADT)	
IDOT	Yes			Yes	
RTA					
CMAP	Yes (Surveillance data)				
WILL COUNTY		Yes (TMC data)	Yes (Condition data)		Yes (Construction & maintenance data)
LAKE COUNTY		Yes	Yes	Yes	

8. Method of Data Access.

AGENCY	How would you prefer to access the data warehouse? (web based, CD based, stand-alone desktop application)
OEMC	▪ Hubspoke-interface interconnected to other servers and periodically updated ▪ Backup on internet (Comment from David Zavattero – "Critical design question of centralized vs. distributed architecture for all or components of a regional data archive remain unresolved. The archive architecture may vary depending on the specific data items, aggregation levels, frequency of need to access, etc. I would argue that a multi-level architecture that includes distributed housing of some data items and centralized housing of other data items may be the most effective approach. And I suggest that you examine the architecture schemes being used by existing (peer example) regional data archives such as PEMS, VaDOT, WashDOT, traffic.com, TTI, etc.")
TOLLWAY	▪ Web based is preferred
IDOT District 1	▪ Automated data dump to archives at certain frequencies (e.g., every hour, once a day, etc.) ▪ Directly connect to archive ▪ Do not need to connect real-time ▪ Access security should be addressed ▪ Internal access should be different than data for public use. Data available to the public should be modified and filtered. ▪ Users should be categorized, such as: industry, academia, public, and access for each should be tailored accordingly.
IDOT Central Ops	▪ Both CD and web based versions are preferable ▪ Data feeding should be at least monthly and more frequently and dynamic is better ▪ Instead of raw data, processed data is preferred ▪ Not willing to host ▪ Outsourcing to Argonne National Lab may be an option to host and facilitate the archive
RTA	▪ Web based is preferred
CMAP	▪ FTP ▪ Gateway
WILL COUNTY	▪ Web based is preferred ▪ Similar to GCM interface is preferable
LAKE COUNTY	▪ Web based is preferred ▪ Users can be categorized as researchers and others ▪ Data provider needs to inform about changes and abnormal data input ▪ Penalty should be imposed on low quality data provider ▪ Archive should be dynamic enough to accommodate changes

9. Data Storage Time Period.

AGENCY	How far back would you like to be able to retrieve the data? If at different levels, please specify the time and data type for each time level.
OEMC	▪ Historic data can be stored in a library ▪ Information about 2 years range should be immediately available
TOLLWAY	▪ Store as much data as possible ▪ For immediate use 4 years range is convenient for Tollway's practices
IDOT District 1	▪ Hourly aggregation level data - 10 years ▪ 5-min aggregation level data - 5 years
IDOT Central Ops	▪ For planning purposes hourly and monthly aggregation level data for 10 years back ▪ For operation purposes 5 min aggregation level data for 10 years back
RTA	▪ Already have a data storage at UIC and 10-12 years span of data is available
CMAP	▪ For crash data all available historical data is desired ▪ For other types of data 2 years is reasonable range ▪ System performance evaluation relies on comparisons made across time to measure progress. Access to historical data will be important for this reason
WILL COUNTY	
LAKE COUNTY	▪ For ADT multiple years of data are required ▪ For other data few weeks/months data may be reasonable

10. Current Data Archiving Practice.

AGENCY	What is your current data archiving practice – separate from system backups?
OEMC	▪ Varies by data item (static data stored permanently), real-time signal data stored 3 days generally
TOLLWAY	▪ In the process of developing a data archiving practice
IDOT District 1	▪ 5 min aggregated level data is available since 1999
IDOT Central Ops	▪ Real time data and aggregated level data are available
RTA	
CMAP	▪ N.A.
WILL COUNTY	
LAKE COUNTY	▪ Only system-wide back ups

11. Future Role.

Can you please explain how you envision the future data archiving warehouse fitting into your current ITS system?

Answers to this question have largely been included in other parts of this memo. Particularly, Lake County and IDOT District 1 have both expressed strong interest in proactively retrieving information from the archiving warehouse and generating monthly reports. OEMC says that the data archive would be used for long-range and operational planning.

Investment Willingness

AGENCY	Is your agency willing to invest in the implementation and development of the centralized data archive?
OEMC	▪ Monetary amount of grant and dedicated staff hour depend on the details and function of the final data archiving plan. ▪ Under hybrid distributed and centralized architecture, it is likely that OEMC would host certain data items and the regional archive “center” would host other items.
TOLLWAY	▪ Monetary support is not guaranteed ▪ Education for participants is desired
IDOT District 1	▪ According to the monetary amount of liability, the issue will be discussed within agency.
IDOT Central Ops	▪ For initial investment monetary support may be possible but for maintenance funding is not possible.
RTA	
CMAP	▪ Interested
WILL COUNTY	▪ Issue has to be discussed within agency before an accurate response about financial aspect is provided.
LAKE COUNTY	▪ Investment amount depends on benefits and cost of the project

12. Expected Benefits

AGENCY	What benefits you expect to obtain from the centralized warehouse?
OEMC	▪ Reduce traffic delay ▪ Decrease travel time ▪ Increase traffic flow speed ▪ Decrease overall cost of travel by decreasing delay time
TOLLWAY	▪ Comprehensive traffic impact in the region ▪ Real data in grand scale in an integrated system ▪ Travel time estimates for commercial trucking companies ▪ Rail movement data ▪ Congestion management
IDOT District 1	▪ To enable faster data retrieval. Current practice sometimes requires up to 16 hours of lead time.
IDOT Central Ops	▪ Better highway operations decisions
RTA	▪
CMAP	▪ Data driven decision making ▪ Make system performance information available to the public ▪ Improve existing transportation models or develop new ones
WILL COUNTY	
LAKE COUNTY	▪ Currently agency has difficulty utilizing the data and expects archiving effort will be helpful

13. Expected Hurdles

AGENCY	Are there any hurdles you can anticipate that would prevent your agency from participating or using a central data archive?
OEMC	▪ Crash data can not be released due to confidentiality ▪ Camera locations can not be revealed due to security purpose ▪ Generally, the data archive should include security, and permissions structure to control access to specific data items based on requirements and constraints of the sourcing agency.
TOLLWAY	
IDOT District 1	▪ Different levels of access to the data warehouse should be granted to different users (industry, academia, public) other than stakeholders.
IDOT Central Ops	▪ Cost
RTA	▪ None
CMAP	▪ There may be some institutional issues
WILL COUNTY	▪ Time is required to get approval and to prepare data for sharing
LAKE COUNTY	▪ Data from the state of Wisconsin is confidential and can not be shared ▪ Previous experience drives the foreseeable needs ▪ Funding details have to be determined

14. Profit Sharing Potential

AGENCY	Is your agency able to participate in profit sharing for any revenue that may be generated from the central data archive
OEMC	▪ Data and statistics can be priced according to demand and usage ▪ Data can be sold to the agencies and firms (public agencies can be charged low as compared to private entities) ▪ Raises much larger policy issues related to data ownership and data sharing. These are being addressed on several fronts including the ITS America PEA Forum d-Bay project
TOLLWAY	▪ Potentially, new system is expected to replace GCM ▪ Educate participating parties ▪ Serve immediate users ▪ Promote for marketing to potential users
IDOT District 1	▪ System should be able to support itself through revenue generation without further maintenance investment by the agency.
IDOT Central Ops	▪ Although IDOT is not currently charging any fee for providing data, for supporting the archive maintenance expenses it may be an option
RTA	▪ Financial issues seem to be not a big problem
CMAP	▪ Revenue possibilities should be identified
WILL COUNTY	▪ Issue has to be discussed within agency before an accurate response about financial aspect is provided.
LAKE COUNTY	▪ General funding is possible

APPENDIX B: POTENTIAL INITIAL DATA EXCHANGES

The following figure presents all potential initial data exchanges for the data archive, representing both those identified in the agency interviews, agency websites, and the *Northeastern Illinois Regional ITS Architecture* via the on-line source, http://dev.tangoinc.net/illinois/Default.aspx.

IDOT

AVAILABLE

Road Network Conditions
Traffic Counts
Traffic Sensor Control
Traffic Flow
Environmental Sensors Control
Environmental Conditions Data
Os/ow permitting information
Credentials information
Credentials status information
Alarm notification and acknowledgment

- **SYSTEM TO SYSTEM DATA SHARED**

Safety Datamart
Surveillance Data

- **SYSTEM TO HUMAN DATA SHARED**

Construction / Maintenance Data
Incident Data
Bridge Restriction Data
Weather Data gathered by weather sensors

IDOT DISTRICTS (District 1)

AVAILABLE

Work zone information
Work plan feedback and coordination
Maintenance and construction resource request & response
Road network conditions
Road network probe information
Resource coordination and deployment status
Multimodal information
Transit incident information
Hazmat information
Equipment maintenance status
Field Equipment status
Traffic images
Resource deployment status
Freeway control status
Reversible lane status
Freeway control data
Roadway information system status
Signal control status
Reversible lane control
Roadway information system data
Signal control data
Video surveillance control
Video images
Traffic information coordination
Traffic counts
Traffic information coordination
Traffic sensor control
Transit emergency data
Transit incident information
Emergency dispatch information
Broadcast advisories
Hri Status
Hri control data
Intersection blockage notification
Intersection status
Local signal preemption request data

- **SYSTEM TO SYSTEM DATA SHARED**

Remote surveillance control
Surveillance Data
Loop Detector Status

- **SYSTEM TO HUMAN DATA SHARED**

Ramp Metering Data from Tollway
Incident Information
Incident information for media
Traveler information for media
Incident response status
Incident report
Incident response coordination
Construction / Maintenance Archive Data
Lane Closure Data
Weather Data

- **UNSHARED DATA**

Com Center Reports
ETP Reports

GCM

- **SYSTEM TO SYSTEM DATA SHARED**

Congestion Data
Construction / Maintenance Data
Lane Closure Data
Travel time Reports
Video Surveillance Data
Skyway Data in real-time (10 min)

CMAP

- **SYSTEM TO SYSTEM DATA SHARED**

Landuse Data
Transit Data
Surveillance Data
Incident Data
Parking Data
Construction / Maintenance Data

AIRPORT WEATHER STATIONS & PRIVATE WEATHER INFORMATION PROVIDERS

- **SYSTEM TO HUMAN DATA SHARED**

Transportation weather information

CITY OF CHICAGO & OEMC (including 311 Center and 911 CAD)

AVAILABLE

Traffic information coordination
Roadway information system status
Roadway information system data
Signal control status
Signal control data
Traffic flow
Traffic images
Traffic sensor control
Intermodal facility incident
Intermod CVO ccordination
Intermodal facility status
Parking lot data request
Parking information
Road network conditions
Road network probe information
Field equipment status
Maintenance and construction resource response
Maintenance and construction resource request
Maintenance and construction resource coordination
Maintenance and construction dispatch status
Maintenance and construction dispatch information
Maintenance and construction vehicle location data
Roadway maintenance status
Work plan feedback
Broadcast information
Broadcast advisories
Transit emergency data
Resource coordination
Hazmat information requests
Hazmat information
Media information requests
Incident information for media
Signal control status
Signal control data
Roadside archive data
Emergency evacuation coordination
Data collection and monitoring control

- **SYSTEM TO SYSTEM DATA SHARED**

Traffic Signal Data
Video Surveillance Data
Video surveillance control

- **SYSTEM TO HUMAN DATA SHARED**

Incident Data
Incident report
Incident response status
Incident response coordination
Construction/Maintenance Data
Landuse Data
Incident Reports (not geocoded)

WILL COUNTY

AVAILABLE

Hazmat information
Incident report
Incident response coordination
Resource coordination

- **SYSTEM TO SYSTEM DATA SHARED**

Construction / Maintenance Data

- **SYSTEM TO HUMAN DATA SHARED**

Weather Data
Signal Inventory

LAKE COUNTY

AVAILABLE

Roadway information system data
Current asset restrictions
Field Equipment Status
Traffic control priority status
Traffic information coordination
Ttraffic control coordination
Traffic sensor control
Environmental sensors control
Environmental conditions data
Signal control data
Signal control status
Local signal preemption request information

- **SYSTEM TO SYSTEM DATA SHARED**

Video and loop detector data
Surveillance Data
Traffic images
Traffic flow
Video surveillance control
Video images

- **SYSTEM TO HUMAN DATA SHARED**

Weather Data

Construction / Maintenance Data
Work Zone Information
Maintenance and construction work plans
Maintenance and construction resource response
Work plan coordination
Work plan feedback
Roadway Maintenance Status
Equipment maintenance status
Incident information
Incident response status

TOLLWAY

AVAILABLE

Hazmat information
Road network conditions
Road network probe information
Resources deployment status
Traffic images
Traffic flow
Traffic sensor control
Video surveillance control
Remote surveillance control
Toll transactions Data
Toll Data
Toll instructions
Toll service change data
Probe Data
Tag data
Field equipment status
Equipment maintenance status
Traffic information ccordination
Traffic control coordination
Freeway control status
Freeway control data
Roadway information system status
Roadway information system data
Traveler information for media
Broadcast advisories

- **SYSTEM TO SYSTEM DATA SHARED**

Surveillance Data
Remote surveillance control

- **SYSTEM TO HUMAN DATA SHARED**

Traffic and Revenue Forecast
Lane Closure Data
Incident information
Incident response status
Incident report
Incident response coordination
Resource request
Resource deployment status
Resource coordination
Maintenance and construction resource response
Maintenance and construction work plans
Roadway maintenance status
Work plan coordination
Work plan feedback
Work zone information
Transportation weather information
Weather information
Environmental conditions data

RTA

AVAILABLE

Road network probe information
Road network conditions
Emergency transit schedule information
Toll data
Traveler archive data
Fare and price information
Transit information requests
Transit and fare schedules
Transit service coordination
Transit traveler information coordination
Transit system data
ISP coordination
Parking availability data
Asset archive data
Archive requests
Archive status
Archive coordination
Selected route information
Transit passenger certification confirm
Transit passenger certification check

- **SYSTEM TO HUMAN DATA SHARED**

Environment Data
Transit Data
Performance Data
Weather Data
Incident Data
Parking Data
Construction / Maintenance Data
Asset Inventory

TRANSIT AGENCIES (CTA, PACE etc.)

- **SYSTEM TO HUMAN DATA SHARED**

Transit Data
Parking Data

KEY

AGENCY INTERVIEW INPUT:

SYSTEM TO SYSTEM: Automated electronic exchange

SYSTEM TO HUMAN: Human interaction required; may include email, hard copy, fax etc.

REGIONAL ITS ARCHITECTURE INPUT:

Information gathered from Northeastern Illinois Regional ITS Architecture is *italicized* and represents Existing Architecture Flows.

APPENDIX C: CONCEPTUAL DESIGN OF USER INTERFACES

This appendix presents the conceptual design of the proposed regional data archive system.

Interactive GIS map Interface

Step 1: The user selects his/her region of interest by zooming in the map

Geographical Area / Location

○ (Geographical area): Select a geographical area from the zoom-enabled map

Step 2: The user specifies the road feature for which information is required

Geographical Area / Location

○ (Intersection or segment): Select roadway or intersection, or detector.

Choose location type of interest

Type of Location: ◉ Intersection ○ Segment

Click on corresponding point(s) on the map to proceed.

Step 3: The user specifies the time range for which information is needed

Time Range

From [▾] to [▾] (dd / mm / yyyy)

Step 4: The user selects the type of information needed

Information Type

- ☐ Geometric Design
- ☐ Traffic Data (ADT) / Hourly Volume (vph)
- ☐ Construction Activities
- ☐ Weather
- ☐ Transit Information
- ☐ Parking Information
- ☐ Incident Data /Crash Number/ Type / Severity

Step 5: Finally, the user selects how the results will be displayed

Display Type

◉ Trend Line ○ Table ○ Pie Chart ○ GIS Map ○ Export Report

[Submit Query]

Query Interface

Step 1: Instead of selecting the region from a map, the user specifies the corresponding details

Geographical Area / Location

- (•) Geographical Area: County: [▾] , City / Township: [▾]
- ( ) Transit route number [▾]
- ( ) Highway Route [▾] [Begin Milepost ( ) End Milepost ( )]
- ( ) Local street [▾]
  from intersection [▾] to [▾]
- ( ) Intersection at [▾] street and [▾] street
- ( ) Detector ID: [▾]

Step 2: The user specifies the time range for which information is needed

Time Range

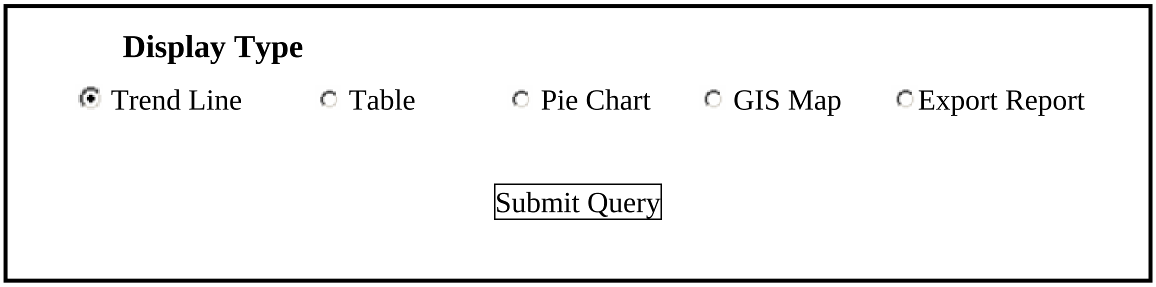
From [▾] to [▾] (dd / mm / yyyy)

Step 3: The user selects the type of information needed

Information Type

- ☐ Geometric Design
- ☐ Traffic Data (ADT) / Hourly Volume (vph)
- ☐ Construction Activities
- ☐ Weather
- ☐ Transit Information
- ☐ Parking Information
- ☐ Incident Data /Crash Number/ Type / Severity

Step 4: Finally, the user selects the mode in which the result has to be displayed



APPENDIX D: AGENCY DATA DETAILS

This appendix presents detailed data types and format from various participating agencies. The information was gathered via a survey of the agencies (April-June 2008).

Summary of IL Regional Data Archive Storage Requirement

Stakeholder	Dataset size per month (GB)	Data types
Tollway	13.44	Incident/Crash, Traffic, Lane Closures, DMS Messages, Weather, Queue
Lake County	14.14	Incident/Crash, Traffic, Lane Closures, DMS Messages
IDOT TSC	8.09	Traffic (mainline detector, ramp detector), Lane Closures, DMS Messages
IDOT District 1	13.75	Incidents, Lane Closures, Traffic (assuming the same number of detectors and record format as Lake County)
RTA	0.01 (13 MB)	PACE AVL(Automatic Vehicle Location), CTA AVL, Bus & Rail Trip and Route (assuming 90-second update interval)
City of Chicago	0.04 (41 MB)	Incident Crash Report, Traffic Count, Traffic Signal
Total	49.47	

More detail of the data available at each stakeholder is listed in the remainder of this appendix.

The total storage space for 10 years of "dynamic" data will approximately be

$$49,47 \times 12 \times 10 = 5936.4 \text{ GB} \approx 5.8 \text{ TB}$$

In addition, the storage for GIS maps (and other static information) would be on the order of a few Gigabytes, which is negligible for storage estimation purposes.
Assuming 50% storage space slack, we estimate that the data archive will require storage of:

$$5.8 \times 1.5 = 8.7 \text{ TB.}$$