Research Report
KTC-07-25/SPR279-03-1F

KENTUCKY TRANSPORTATION CENTER

ARCHIVED DATA MANAGEMENT SYSTEM IN KENTUCKY

UNIVERSITY OF KENTUCKY
College of Engineering

OUR MISSION

We provide services to the transportation community
through research, technology transfer and education.
We create and participate in partnerships
to promote safe and effective
transportation systems.

OUR VALUES

Teamwork
Listening and communicating along with
courtesy and respect for others.

Honesty and Ethical Behavior
Delivering the highest quality
products and services.

Continuous Improvement
In all that we do.

Research Report
KTC-07-25/SPR 279-03-1F

ARCHIVED DATA MANAGEMENT SYSTEM IN KENTUCKY

By

Mei Chen, Ph.D.
Associate Professor
Department of Civil Engineering

and

Jingxin Xia
Graduate Research Assistant
Department of Civil Engineering

Kentucky Transportation Center
College of Engineering
University of Kentucky
Lexington, Kentucky

in cooperation with
Transportation Cabinet
Commonwealth of Kentucky

and

Federal Highway Administration
U.S. Department of Transportation

September 2007

1. Report No. KTC-07-25/SPR 279-03-1F	2. Government Accession No.	3. Recipient's Catalog No
4. Title and Subtitle Archived Data Management System in Kentucky	5. Report Date May 2007 6. Performing Organization Code	
7. Author(s) Mei Chen and Jingxin Xia	8. Performing Organization Report No. KTC-07-25/SPR 279-03-1F	
9. Performing Organization Name and Address Kentucky Transportation Center College of Engineering University of Kentucky	10. Work Unit No. (TRAIS) 11. Contract or Grant No. P279	
12. Sponsoring Agency Name and Address Kentucky Transportation Cabinet 200 Mero Street, Frankfort, KY	13. Type of Report and Period Covered Final 14. Sponsoring Agency Code	
15. Supplementary Notes		

16. Abstract

Archived Data User Service (ADUS) was added to the national ITS architecture in 1999 to enable multiple uses for ITS-generated data. In Kentucky, ARTIMIS and TRIMARC are collecting volume, speed, occupancy, length-based classification, and incident data on a continuous basis. This study developed the Kentucky Archived Data Management System (ADMS) to archive and disseminate data collected by sensors in ARTIMIS and TRIMARC systems. This system is designed to be the prototype of ITS data clearinghouse in Kentucky.

The Kentucky ADMS operates on a web server that allows users to retrieve traffic data (i.e., volume, speed, lane occupancy) in 15 minute increments from a GIS-based map. Some freeway performance measures such as AADT are also provided. Data quality control and assurance were also performed to identify the erroneous data items and/or records and to provide statistically more accurate estimates.

17. Key Words Archived Data Management System; Intelligent Transportation Systems		18. Distribution Statement Unlimited	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 34	22. Price

EXECUTIVE SUMMARY

The Kentucky Archived Data Management System (ADMS) is developed under the framework of the Archived Data User Service (ADUS) which became an official component of the National ITS Architecture in 1999. The goal of the system is to archive and disseminate traffic and incident data generated by sensors of the two ITS deployments in Kentucky, ARTIMIS and TRIMARC. The data has been collected on a continuous basis and offers rich information on traffic characteristics and operation efficiency. Since many ITS sensors (especially those in early deployments) are designed and calibrated for traffic monitoring applications rather than data collection, the quality of the data they generate needs to be examined.

This traffic monitoring data had not been archived and used in the past. The Kentucky ADMS aims to serve as the clearinghouse of the ITS-generated data in Kentucky by providing a web-based interactive access to the original data as well as processed/aggregated performance measures of the roads. The Kentucky ADMS applies a set of quality control criteria to the archived traffic data; records that fail the criteria are flagged in the database. Imputation is performed to provide statistically sound estimates of flagged values. The users can decide whether to use the imputed data to generate results for their queries.

The web-based interactive GIS interface of the Kentucky ADMS provides a user friendly access to the archived traffic and incident data. The data has been used by various planning agencies and consultants.

The success of the Kentucky ADMS requires continuing support from the KYTC and other stakeholders. At a minimum, an annual update of the data archive shall make the latest traffic data available. Furthermore, the system shall be able to expand its coverage to include other ITS projects in the state.

TABLE OF CONTENTS

Executive Summary 1
Table of Contents i
List of Figures ii
List of Tables ii
Chapter 1 Background 1
Chapter 2 ITS Data in Kentucky 3
2.1 Data Source 3
2.2 Data Items 3
2.3 Data Needs Survey 3
Chapter 3 Data Archiving 4
3.1 Database Design 4
3.2 Data Quality Control 4
3.3 Data Quality Assurance 4
3.3.1 Historical Average 4
3.3.2 Temporal Interpolation 4
3.3.3 Spatial Interpolation 5
3.3.4 Hybrid Algorithm 5
3.3.5 Artificial Neural Network 5
3.3.6 Implementation Strategy 5
Chapter 4 Web-Based Data Dissemination 7
Chapter 5 Implementation 12
References 13
Appendix A Database Design 14
Appendix B Data Dictionary 19
Appendix C Data Quality Screening Criteria 29

LIST OF FIGURES

Figure 3-1 Imputation algorithm implementation strategy 6
Figure 4-1 Kentucky ADMS homepage 7
Figure 4-2 Sample data query pages and output 9
Figure 4-3 Kentucky ADMS system function 10

LIST OF TABLES

Table 2-1 Number of detector stations by year 3

CHAPTER 1 BACKGROUND

Sensors in various intelligent transportation system (ITS) deployments are collecting a large amount of traffic data every day. Such data contain vehicle volume, speed, travel time, vehicle classification, incidents, traffic signal control, etc. For example, incident, vehicle-length-based classification, speed, and hourly volume in 30-sec increments are among the data collected by the ARTIMIS center in northern Kentucky and the Greater Cincinnati area. These data offer great potential in long-term transportation and operations planning, in addition to their primary use in short-term operations management (by transportation authorities) and trip management (by private individuals and industry).

The archiving of such data has been receiving increasing attention around the country. A variety of data archiving activities have been underway, such as the TransGuide® system in San Antonio, TX (1), the SmartTrek system in metropolitan Seattle, WA (2), the DataMart system of San Francisco Bay Area, CA (3), the TRANSCOM project in the tri-state area of NY, NJ, and CT (4), the Smart Travel data collection and analysis project in Virginia (5), and the Freeway Performance Measurements Project (PeMS) in California (6).

The archived ITS data have several features that distinguish them from traditional data sources in that (i) ITS data are temporally intensive – they can be collected in very short intervals; (ii) ITS data meet some major data gaps that can not be met using traditional methods due to resource limitation; and (iii) all ITS data are in electronic format, thus expediting data analysis and dissemination. Some of the potential areas of application of ITS data are summarized as follows (7).

Planning Applications

- Urban travel demand modeling & planning
- Data reporting
- Air quality analysis

Operations Applications

- Assessment of highway security vulnerability
- Monitoring performance measures
- Traffic delay management
- Adaptive signal time strategies

Highway Safety Applications

- Intersection safety
- Work zone safety
- Evaluation of speed management strategies

Transit Applications

- Operations management and planning
- System vulnerability assessment

Potential users of the archived ITS data may include MPO and state transportation planners, traffic management operators, transit operators, air quality analysts, freight and intermodal planners, safety planners and administrators, maintenance personnel, commercial vehicle enforcement personnel, emergency management services, transportation researchers, and private sector users (8).

In Kentucky, two regional ITS deployments, ARTIMIS and TRIMARC, have been collecting various traffic information since their deployment. How to effectively and efficiently archive these data and integrate them with data from traditional sources such as traffic monitoring systems remains an unsolved issue.

An integrated ADMS is an urgent need in Kentucky in order to provide information to a variety of data users. For example, a study in delay analysis for highway maintenance work needs queuing data to evaluate the level of service at and around the site. Also, important information such as travel time can be obtained from the proposed system.

The objectives of this research are to

- Identify the data needs of stakeholders in Kentucky;
- Develop an ADMS that supports online query;
- Estimate the required maintenance effort (and the associated cost) of the proposed ADMS;
- Examine the quality of data and develop error detection algorithms.

At a minimum, the ADMS should collect and process volume, length-based classification and speed data.

CHAPTER 2 ITS DATA IN KENTUCKY

2.1 Data Source

The data used in this study come from two regional ITS deployments in Kentucky, ARTIMIS and TRIMARC, between the years of 2002 and 2005. The ARTIMIS and TRIMARC operations centers archive the sensor data at 15 minute increment into text files in zipped format. The data was then provided to the ADMS research team via FTP or CD. The numbers of detector stations within Kentucky that reported data in each year are listed in Table 2-1.

Table 2-1 Number of detector stations by year

Year	ARTIMIS	TRIMARC
2002	27	53
2003	17	71
2004	17	95
2005	11	92

2.2 Data Items

The most commonly available data items from the Kentucky ITS sensors are volume, speed, and lane occupancy at 15 minute increments at each detector station. Other data files such as those in FHWA Record 3 format, as well as those in the Record S and Record C formats are also available from certain stations.

2.3 Data Needs Survey

A survey was conducted at the beginning of the study to assess the data needs of various transportation data stakeholders in and around Kentucky. Personnel from KYTC Divisions of Planning and Operations, MPOs, and FHWA offices responded to the survey.

Data needs of these stakeholders are assessed through a series of choices on their desired data items, their priorities, as well as aggregation levels. The results indicated that people in traffic operations are more interested in data aggregated in shorter intervals than those in the planning area, who generally place high priority on the accuracy of volume data.

CHAPTER 3 DATA ARCHIVING

3.1 Database Design

A relational data schema was designed for the database archiving traffic data from ITS sensors. The design specifications and data dictionary are detailed in Appendix A and B, respectively. The platform used in this study is an open source database management tool named mySQL.

3.2 Data Quality Control

Preliminary analyses showed that the counts generated by the ITS sensors at some stations could be quite different from those produced by the automatic traffic recorders (ATR) at the same locations. The reliability of ITS sensors in producing accurate counts and other estimates have long been questioned because the primary purpose of those sensors were to monitor traffic rather than to record traffic data for future use, especially in early ITS deployments.

A set of data quality screen rules developed as part of the Mobility Monitoring Program of FHWA were used in this study for data quality screening (9). These rules, shown in Appendix C, check for inconsistencies using basic relationships between traffic parameters. Records that fail the quality screening are flagged in the database.

3.3 Data Quality Assurance

Imputation was performed to those data item/records flagged in the quality screening effort. The purpose is to provide estimates of the missing or erroneous values that are statistically more accurate.

Missing data imputation requires investigation of temporal and spatial characteristics of the traffic pattern at each detector station. Through extensive literature review, several imputation methods are adopted in this study. They are discussed in the sections below.

3.3.1 Historical Average

This method uses the average values from the unflagged records for the same time of day and day of the week to impute the missing values.

3.3.2 Temporal Interpolation

This method imputes a missing value using the average of its preceding and succeeding values at the same detector station. This method is rather accurate provided that traffic does not change over a short period of time. However, it depends upon the availability of the data in the previous and successive 15 minute periods.

3.3.3 Spatial Interpolation

This method uses the average value between the data obtained from the immediate upstream and downstream detector stations to impute a missing value. However, this method may not be reliable when there is a ramp between these stations.

3.3.4 Hybrid Algorithm

A hybrid algorithm was also developed for missing data imputation. It integrates the historical average and time series analysis technique. A triple exponential smoothing procedure was constructed for volume data.

3.3.5 Artificial Neural Network

Using the enormous amount of data generated by the ITS sensors as the training set, neural network models were also developed. The basic idea is to simulate the information processing paradigm inspired by human nerve system without an explicitly formulated functional relationship between the inputs (e.g., location, time, traffic parameter at adjacent locations) and output (such as traffic parameters at the time and/or location in question).

3.3.6 Implementation Strategy

The performance indices for these imputation methods are defined below.

Mean Absolute Percentage Error (MAPE): $MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_i^*}{Y_i}\right| \times 100$

Root Mean Squared Error (RMSE): $RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(Y_i - Y_i^*)^2}$

where, n is the number of imputed values, Y_i is the observation value, and Y_i^* is the imputed value. Based on their applicability and performance, an implementation sequence of the methods is recommended, as shown in Figure 3-1 (10).

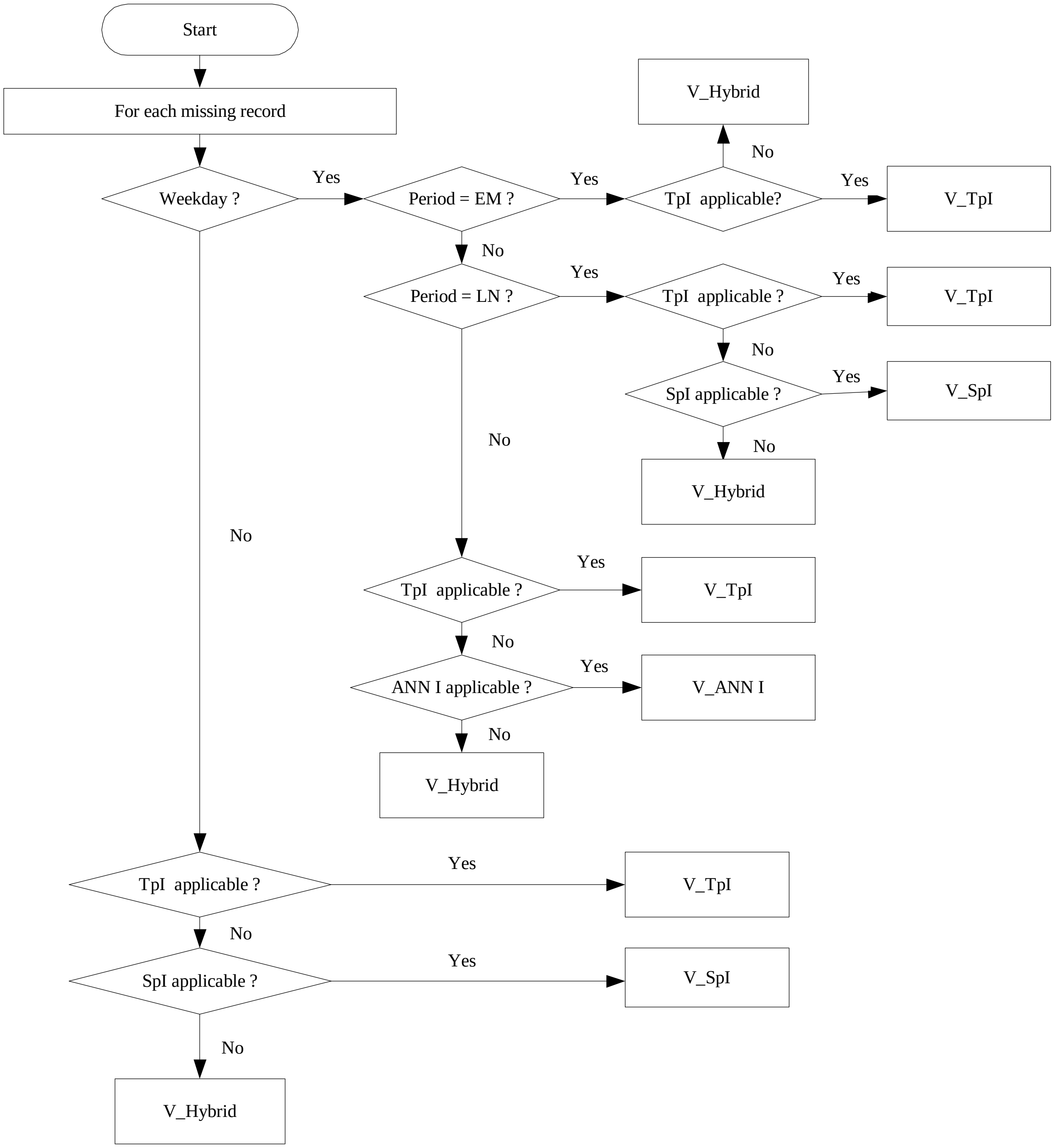


Figure 3-1 Imputation algorithm implementation strategy

All missing and erroneous values in the detector station data files were imputed following the above strategy. It should be noted that the original recorded values were kept intact in the database. It is up to the users of the Kentucky ADMS to determine if they want their query results returned based on imputed or original values.

CHAPTER 4 WEB-BASED DATA DISSEMINATION

The Kentucky ADMS site was launched in 2006 at the URL of http://adms.uky.edu. The site contains an ArcIMS-based GIS service that allows for dynamic map and data delivery via the web. Due to limited resources the map server, database server, and web server all reside on a single PC running Windows XP. A screen shot for the homepage is shown in Figure 4-1. The Kentucky ADMS web site is best viewed using Microsoft Internet Explorer.

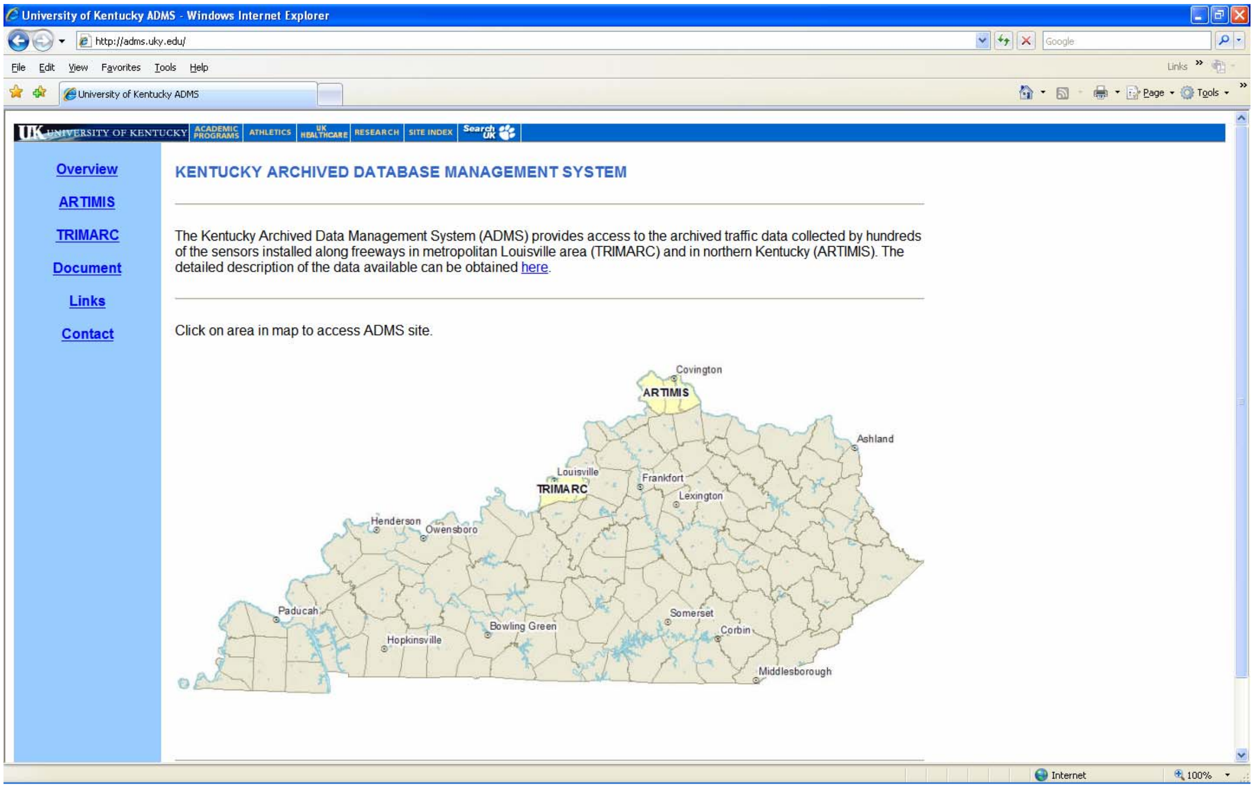
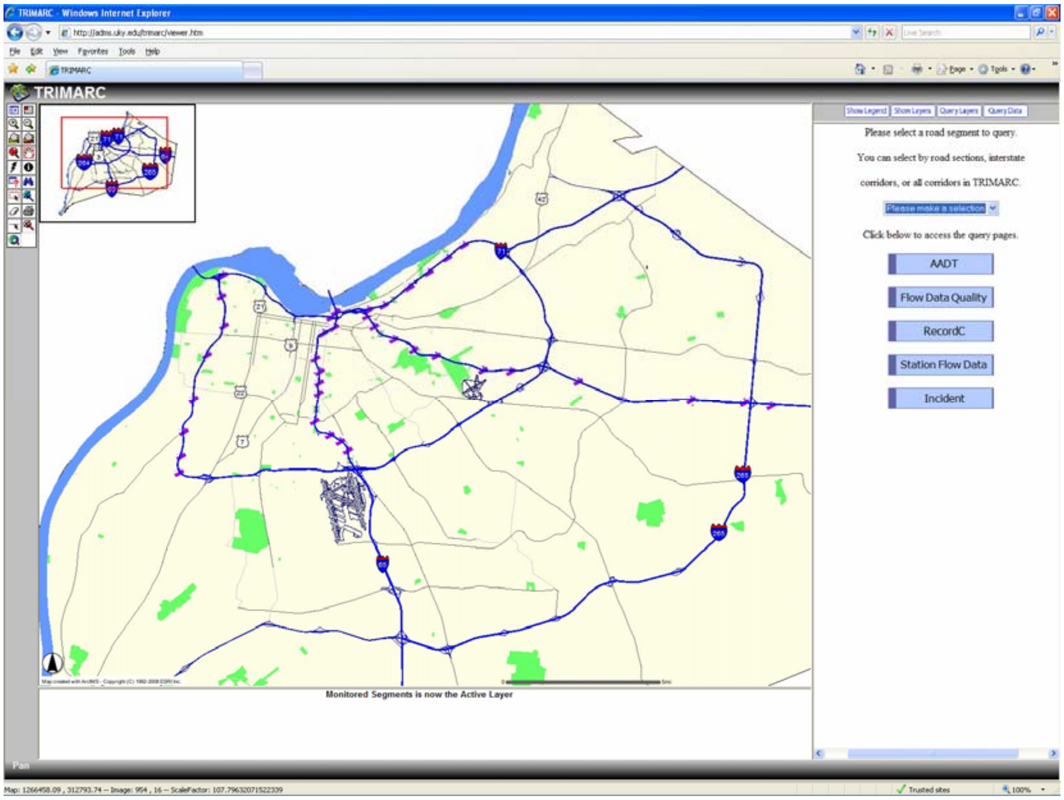
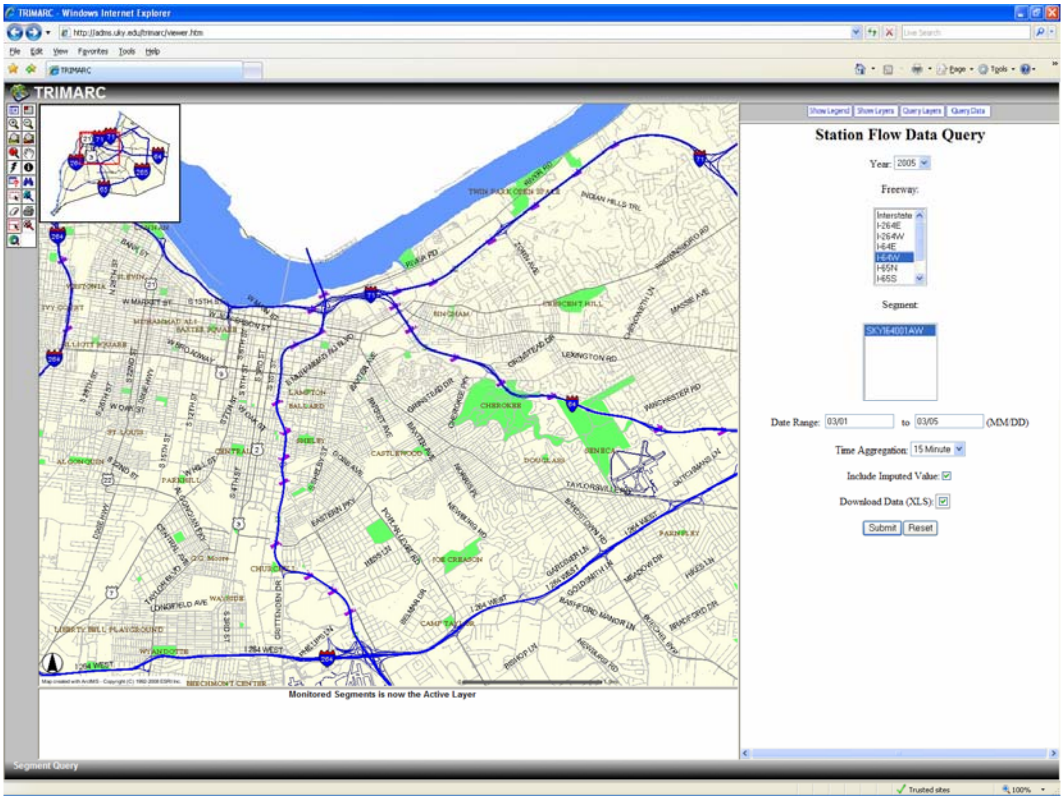


Figure 4-1 Kentucky ADMS homepage

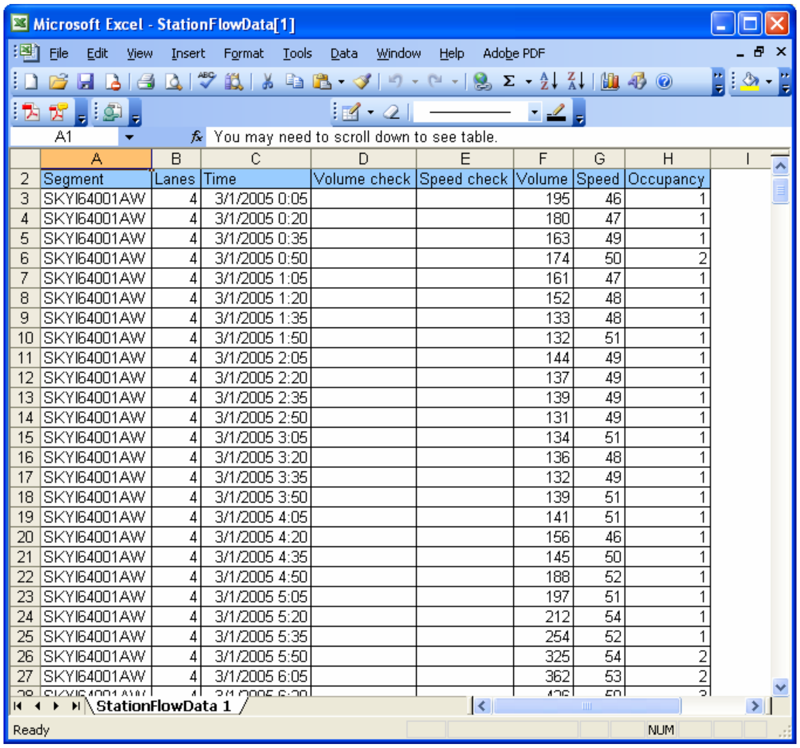
The Kentucky ADMS provides a variety of information based on data from each detector station. In addition to the display of detector stations (point layer), monitored segments, and freeway corridors (line layers) on a map of Jefferson County, the system also provides query interface for various items, including AADT and station flow data (i.e., 15-minute volume, speed, and lane occupancy). The data quality report can also be retrieved for a given segment. The report contains information on the data completeness index (i.e., the actual number of records over total number of records that should have been recorded in a given period of time) and the percentage records passed the screening criteria. All query results can either be displayed in the web browser or saved to a Microsoft Excel file. Some sample data query pages are shown in Figure 4-2.



(a) TRIMARC overall data query page



(b) TRIMARC detector station data query page



Microsoft Excel - StationFlowData[1]

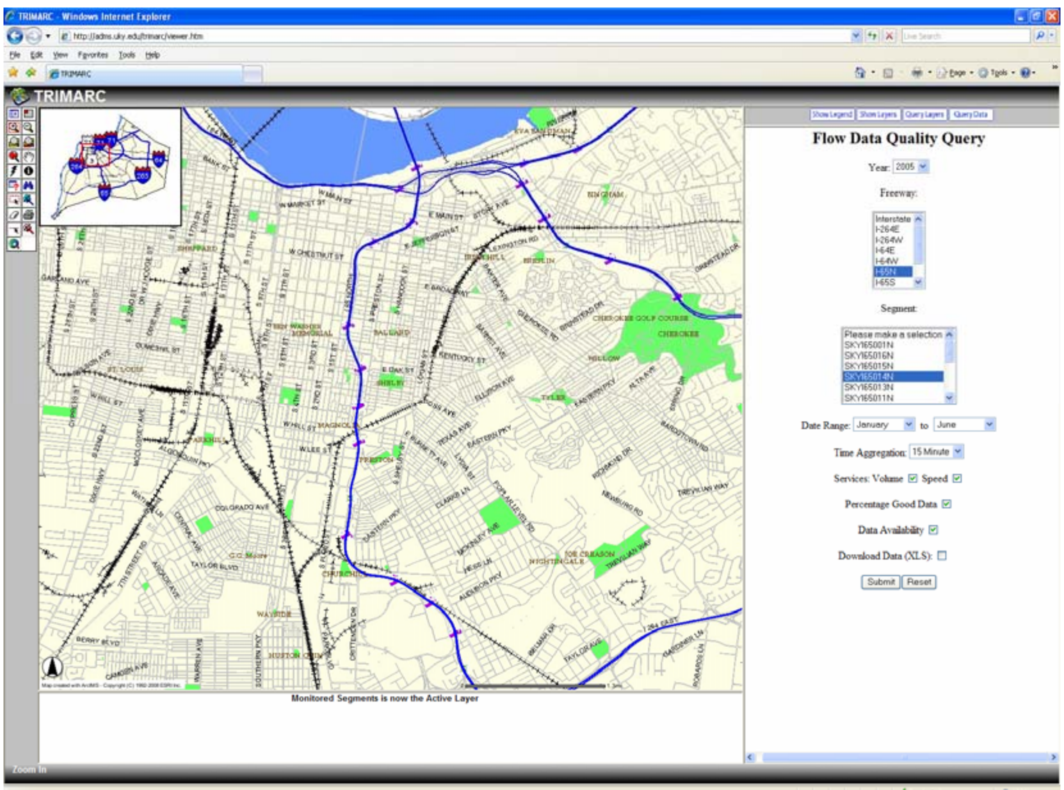
A1 You may need to scroll down to see table.

2	Segment	Lanes	Time	Volume check	Speed check	Volume	Speed	Occupancy
3	SKYI64001AW	4	3/1/2005 0:05			195	46	1
4	SKYI64001AW	4	3/1/2005 0:20			180	47	1
5	SKYI64001AW	4	3/1/2005 0:35			163	49	1
6	SKYI64001AW	4	3/1/2005 0:50			174	50	2
7	SKYI64001AW	4	3/1/2005 1:05			161	47	1
8	SKYI64001AW	4	3/1/2005 1:20			152	48	1
9	SKYI64001AW	4	3/1/2005 1:35			133	48	1
10	SKYI64001AW	4	3/1/2005 1:50			132	51	1
11	SKYI64001AW	4	3/1/2005 2:05			144	49	1
12	SKYI64001AW	4	3/1/2005 2:20			137	49	1
13	SKYI64001AW	4	3/1/2005 2:35			139	49	1
14	SKYI64001AW	4	3/1/2005 2:50			131	49	1
15	SKYI64001AW	4	3/1/2005 3:05			134	51	1
16	SKYI64001AW	4	3/1/2005 3:20			136	48	1
17	SKYI64001AW	4	3/1/2005 3:35			132	49	1
18	SKYI64001AW	4	3/1/2005 3:50			139	51	1
19	SKYI64001AW	4	3/1/2005 4:05			141	51	1
20	SKYI64001AW	4	3/1/2005 4:20			156	46	1
21	SKYI64001AW	4	3/1/2005 4:35			145	50	1
22	SKYI64001AW	4	3/1/2005 4:50			188	52	1
23	SKYI64001AW	4	3/1/2005 5:05			197	51	1
24	SKYI64001AW	4	3/1/2005 5:20			212	54	1
25	SKYI64001AW	4	3/1/2005 5:35			254	52	1
26	SKYI64001AW	4	3/1/2005 5:50			325	54	2
27	SKYI64001AW	4	3/1/2005 6:05			362	53	2

StationFlowData 1

Ready NUM

(c) Volume, speed and occupancy for station SKYI64001AW in 15-minute increment



(d) Data quality query page

Figure 4-2 Sample data query pages and output

The traffic data system function of the Kentucky ADMS is shown in Figure 4-3.

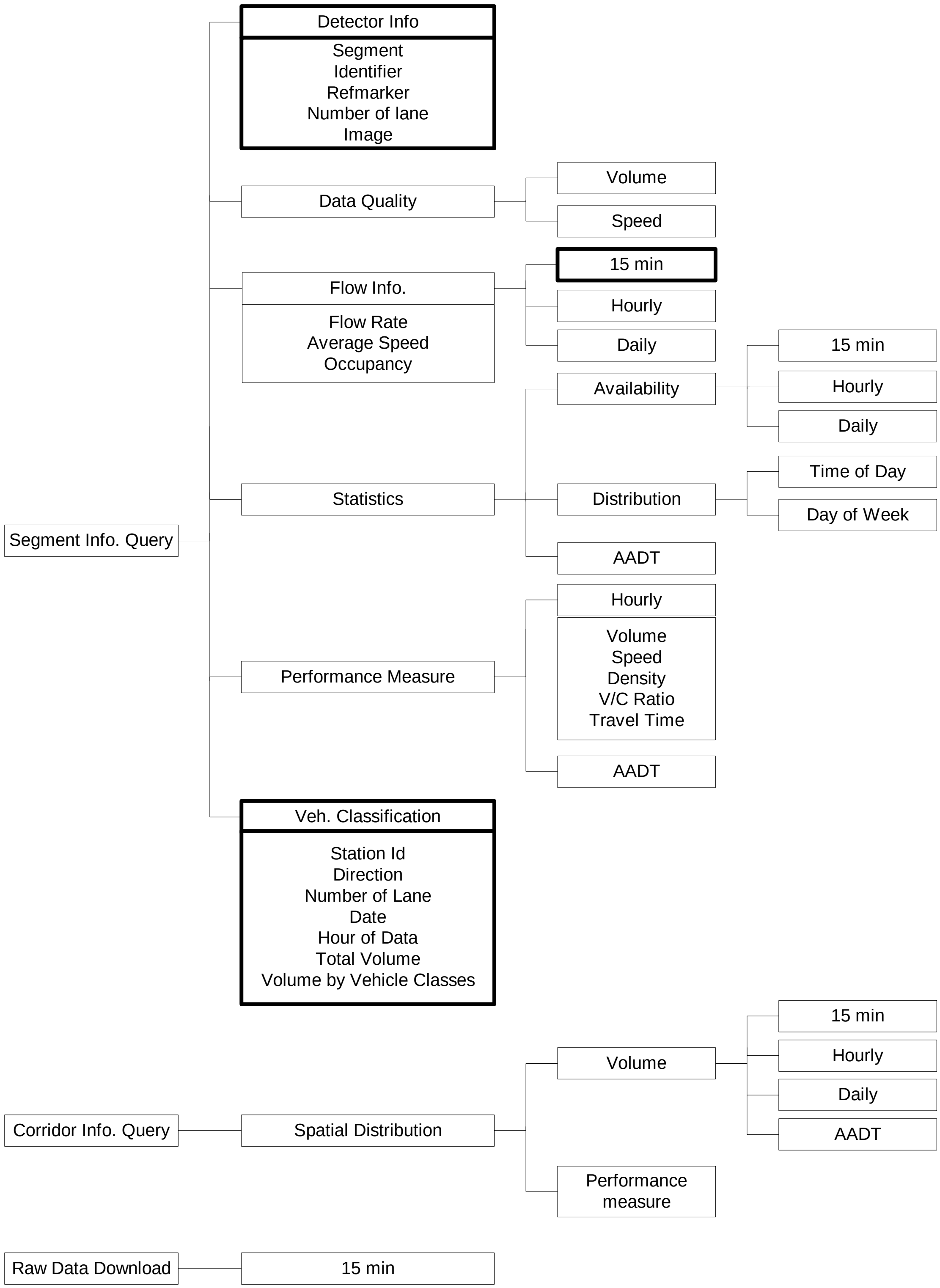


Figure 4-3 Kentucky ADMS system function

Starting in 2004, the incident information is incorporated to the Kentucky ADMS. The source of this information is the incident logs at the two ITS deployments. Users can query the incidents by date, time, and milepoint range.

CHAPTER 5 IMPLEMENTATION

The Archived Data User Service (ADUS) was added to the national ITS architecture in 1999 to enable multiple uses for ITS-generated data. The Kentucky ADMS is a prototype application of ADUS aiming to become the ITS clearinghouse for Kentucky. The goal is to archive and provide access to traffic operations data that were either previously not available or very expensive to collect.

It is a national consensus that ITS data have great potential in a number of applications including operations planning, performance monitoring and evaluation, traveler information services, safety evaluation, air quality analyses, as well as long-range planning. It is essential to make the Kentucky ADMS a continuing effort with at least annual updates when new data becomes available. If future data amounts correspond to the amount of data in 2005, the cost of annual updates will be approximately $8000 a year. This estimate includes personnel time for data processing and maintenance of database, map, and web servers. Additional expenses will be incurred if system and equipment upgrade becomes necessary in the future.

The FHWA has been promoting efforts of linking planning and operations; it encourages the involvement of universities in developing data sharing applications (11). Potential applications of Kentucky ADMS shall be identified and their data requirements should be defined. The concerns (mostly regarding data quality) of using ITS data in these applications need to be addressed.

REFERENCES

1. Texas Transportation Institute (1999) ITS data archiving: case study analyses of San Antonio TransGuide® data. FHWA –PL-99-024. http://tti.tamu.edu/product/catalog/reports/FHWA-PL-99-024.pdf

2. http://www.smarttrek.org/

3. http://www.mtc.dst.ca.us/datamart/index.htm

4. http://www.xcm.org/

5. http://smarttravellab.virginia.edu/

6. http://pems.eecs.berkeley.edu

7. Hu, P.S., Boundy, R.G., Truett, L.F., Chang, E.C., Gordon, S.R. (2002) Cross-Cutting Studies and State-of-the-Practice Reviews: Archive and Use of ITS-Generated Data. ORNL/TM-2002/120, Oak Ridge National Laboratory, Oak Ridge, TN.

8. Turner, S. (2001) Guidelines for developing ITS data archiving systems. Report 2127-3, Project Number 0-2127, Texas Transportation Institute, College Station, TX.

9. Battelle, Cambridge Systematics, and Texas Transportation Institute (2004) Traffic data quality measurements, Final report, Federal Highway Administration.

10. Chen, M., J. Xia, and R. Liu (2006) Develop a strategy for imputing missing volume data, *Journal of Transportation Research Forum*. Vol.45, No.3, 57-75.

11. Federal Highway Administration (2004) Getting more by working together: opportunities for linking planning and operations.

APPENDIX A
DATABASE DESIGN

The database is one of the most important parts of a data archiving system. The design of the database including the selection of database software, analysis of demand and supply, conceptual design, logic design, and physical design subprocesses is the key process that leads to the success of a data archiving system.

(1) Analysis of Demand and Supply

Demand and supply analysis is one of most important steps in database design; it is the foundation for the following steps. It mainly includes collection of materials, analysis and processing, development of data flow chart and data dictionary.

Collection of materials

Materials collection mainly consists of two steps: [1] understand the procedure of current demand and supply and collect all data materials including original data, involved data for the prediction of AADT, reports, etc. [2] analyze user demand including information demands, process demands, and demands of security and integrality of data. Analysis of information demands is the procedure to figure out information the users would like to obtain from the system. The properties and relationship between data should be clearly demonstrated. Process demands are the analysis of function and operation the system will provide. The demands of security and integrality are the requirements the system corresponds to the operation.

Based on the original data sets value for hourly volumes, daily volumes, and AADT will be estimated. Hourly volume includes data elements of segment name, date, hour, and hourly volume. Daily volume includes data elements of segment name, date, and daily volume. And AADT includes data elements of segment name and AADT. Because there is a corresponding relationship between detector and segment name, detector information will also be included in the system. Detector information includes items like segment name, detector identifier, state, highway, reference marker, and year.

Analysis of data and data flow chart

Due to malfunctions of field detectors and communication systems, some of the data were neither corrected nor collected. Therefore, data should be checked with quality for volume. Criteria used for this study were initially developed in the Mobility Monitoring Program performed by the Texas Transportation Institute (TTI). When checking the volume data records, only those relevant criteria were applied in this study. Through the volume quality checks, those records identified as suspicious or erroneous data will be flagged. Then the hourly volumes and daily volume will be estimated. Finally, AADT can be estimated. In this system, a procedure in FHWA Traffic Monitoring Guide will be used to estimate AADT. Figure 1 shows the flow chart for AADT prediction.

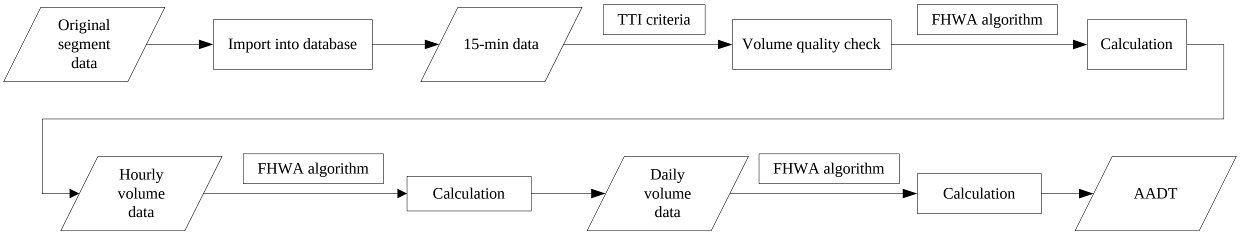


Figure 1 Procedure of AADT prediction

Data dictionary

The data dictionary mainly includes detailed elements of data sets and involved data elements when using the algorithm to predict AADT. Since AADT is a frequently requested piece of information and since extensive amounts of time are involved in its estimation, the intermediate results including hourly volume, daily volume, and AADT are stored in separate tables. In addition, archived data collected by TRIMARC include Record 3 data, and Record C data. These data will all be stored in the database for further application. Base on this, the data dictionary is listed in the Appendix B, and it will be revised and updated in the full procedure of the database design.

(2) Conceptual Design

Conceptual design is the first step when designing a database system, and it is the basis of the logic design and physical design of a database. The purpose of conceptual design is to establish the conceptual modules for the data archiving system on the basis of the analysis of demand-supply. Methods that can be used include description, User-interface model, E-R model, etc. E-R model is the most popular methods, where “E” refers to the entity, and “R” refers to the relationships between entities. Therefore, conceptual design of database includes two aspects. One is the analysis of entities, and the second one is the analysis of relationships, which includes local relationship analysis (local view) and integration of relationship (full view).

Entities in this system currently mainly include detector and segment. Figure 1 shows their properties using E-R model. E-R of segment is established based on the properties over 15 minutes.

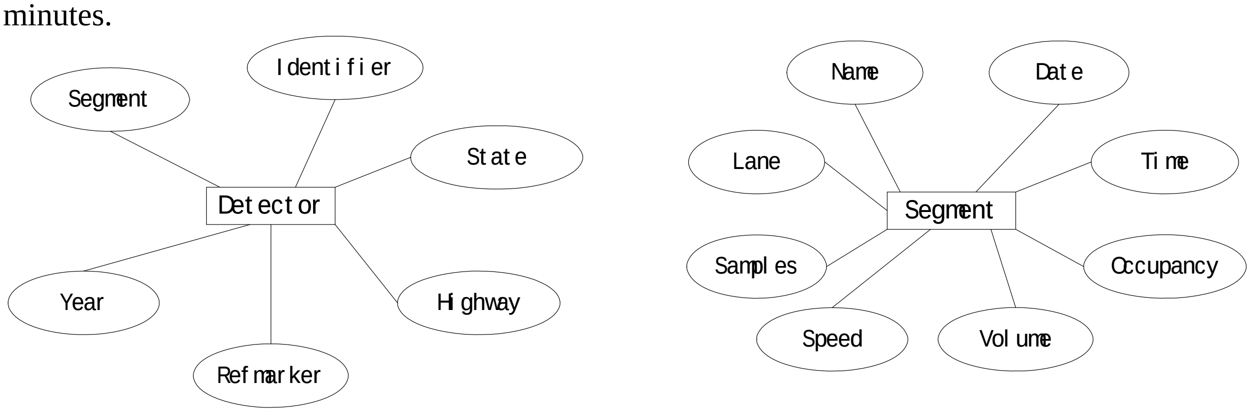


Figure 1 (a) E-R of detector (b) E-R of segment

This figure shows the conceptual properties of a segment over 15 minutes. Due to the demands of the system to calculate the hourly volume, daily volume and AADT, the segment also has some other properties based on different time periods, and they are shown in figure 2.

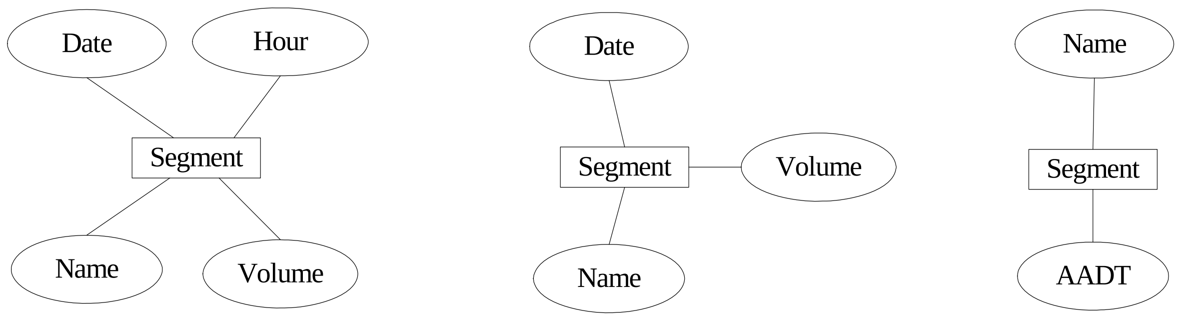


Figure 2 (a) Hourly-volume E-R of segment (b) daily-volume E-R (c) AADT E-R

On the basis of these local views showing the relationship of entities, full view can be obtained as shown in Figure 3. In addition to the relationships between each individual entity, relationships between entities are also established. There is a 1-1 or 1-2 relationship between detector and segment, which means that each detector may correspond to one or two segment(s).

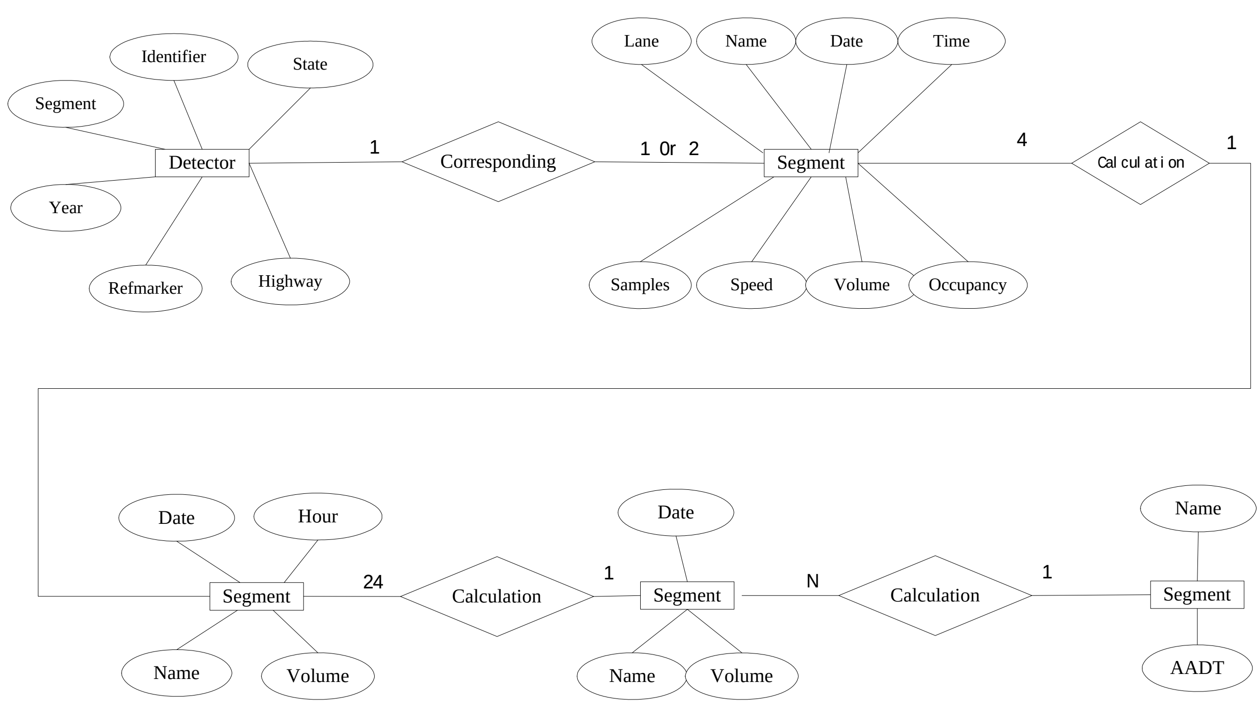


Figure 3 System E-R (full view)

(3) Logic Design

Conceptual design only establishes a conceptual model independent of the database management system (DBMS). The task of logic design is to transfer conceptual modules of the system to a data model that is supported by a specific DBMS. In the development of the data archiving system in Kentucky, mySQL was selected as the DBMS.

Criteria used to transfer the conceptual design to logic design include: (1) one entity is converted into a relationship module, properties of the entity are the properties of the relationship, and primary key of the entity is the primary key of the relationship; (2) a 1:1 relation is converted into a relationship, and each primary key is the candidate primary key of the relationship; (3) an N:1 relation is converted into a relationship, and the primary key of all entities is the primary key of the relationship; (4) an N:M relation is converted into a relationship, and the combination of all primary keys of all entity is the primary key of the relationship; (5) relationships that have the same primary key can be united.

Using the original segment as an example, the entity is the segment, and the corresponding relation is the segment (name, datetime, number of lanes, samples, speed, volume, occupancy). Using the detector as an example, the entity is the detector, and the corresponding relation is the detector (segment, identifier, state, highway, reference marker, year). The primary key of the segment and detector is Segment + Datetime, and segment, respectively.

After converting the relation of entities, standardization is applied in the design of the table structures. Detailed information is listed in the Appendix B.

(4) Physical Design

Physical design of the database is a procedure to select an application environment for the physical structure. It includes the design of store structure and store path.

APPENDIX B
DATA DICTIONARY

1. Original Table

Table name: original_data_2002
Function: store one-year data over 15 minutes intervals collected by TRIMARC, and provide for data quality check

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary Foreign	Yes (Series)	No	No	No	No	Name of segment
Date_time	Varchar	30	Primary	Yes (Series)	No	No	No	No	Date and time that collect record
No_lane	Number	2	No	No	No	No	No	No	Number of lanes
Samp	Number	5	No	No	No	No	No	No	Number of samples
Speed	Number	5	No	No	No	No	No	No	Average Speed (mph)
Vol	Number	10	No	No	No	No	No	No	Volume (Veh)
Occ	Number	5	No	No	No	No	No	No	Average Occupancy (%)
Status	Varchar	1	No	No	Yes	No	No	No	Quality status
Volume_new	Number	10	No	No	Yes	No	No	No	Factor up to 30 samples
Check_1	Varchar	1	No	No	Yes	No	No	No	Illegal data time
Check_2	Varchar	1	No	No	Yes	No	No	No	Volume, speed, and occupancy all equal to zero
Check_3	Varchar	1	No	No	Yes	No	No	No	Maximum volume > 750
Check_4	Varchar	1	No	No	Yes	No	No	No	Occupancy > 80
Check_5	Varchar	1	No	No	Yes	No	No	No	Volume = 0, speed <> 0
Check_6	Varchar	1	No	No	Yes	No	No	No	Volume, speed = 0, and Occupancy <> 0
Check_7	Varchar	1	No	No	Yes	No	No	No	Occupancy =0, Volume > (2.932 * 15 * 60) / 600
Check_8	Varchar	1	No	No	Yes	No	No	No	
Check_9	Varchar	1	No	No	Yes	No	No	No	
Check_10	Varchar	1	No	No	Yes	No	No	No	8 consecutively identical Volume

2. Detector Information

Table name: Detector_info_2002
Function: Store one-year detector information of TRIMARC

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary Foreign	No	No	No	Yes	No	Name of segment
Identifier	Varchar	15	No	No	No	No	Yes	No	Identifier of detector
State	Varchar	5	No	No	No	No	No	No	State
Highway	Varchar	10	No	No	No	No	No	No	Highway ID
Refmarker	Number	10	No	No	No	No	No	No	Milepost
T_year	Number	4	No	No	No	No	No	No	Year within which that data were collected

3. Hour volume information

Table name: Hourly_vol_2002
Function: Store one-year hourly volume information, calculated from 15-min data records

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary	Yes (Series)	No	No	No	No	Name of segment
D_Date	Date	10	Primary	Yes (Series)	No	No	No	No	Date
H_hour	Number	2	Primary	Yes (Series)	No	No	No	No	Hour in which hourly volume is calculated
Vol	Number	10	No	No	No	No	No	No	Calculated hourly volume
C_1	Number	50	No	No	No	No	No	No	Comments

4. Daily Hour volume distribution information

Table name: Hour_distribution_2002
Function: Store daily hour volume distribution information for all segments, calculated with completed data set. It can be used to predict daily volume

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary	Yes (Series)	No	No	No	No	Name of segment
H_hour	Number	2	Primary	Yes (Series)	No	No	No	No	Hour information
H_percent	Number	10	No	No	No	No	No	No	Percentage of volume

5. Daily volume information

Table name: Daily_vol_2002
Function: Store one-year Daily volume information, calculated from Hourly volume

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary	Yes (Series)	No	No	No	No	Name of segment
D_Date	Date	10	Primary	Yes (Series)	No	No	No	No	Date
Vol	Number	10	No	No	No	No	No	No	Calculated daily volume
C_1	Number	50	No	No	No	No	No	No	Comments

6. Weekly volume distribution information

Table name: w_perc_2002
Function: Store weekly volume distribution information, calculated as the average of daily volume for all weekdays.

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary	Yes (Series)	No	No	No	No	Name of segment
W_day	Number	1	Primary	Yes (Series)	No	No	No	No	Week day
Perc	Number	10	No	No	No	No	No	No	Percentage of volume

7. Monthly day-volume information

Table name: m_vol_2002
Function: Store monthly day volume information, calculated from daily volumes and weekly volume distribution information.

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary	No	No	No	No	No	Name of segment
M_number	Number	2	Primary	No	No	No	No	No	The number of month
Vol	Number	10	No	No	Yes	No	No	No	Monthly day volume
C_1	Varchar	30	No	No	Yes	No	No	No	Comment of month 12

8. AADT information 1

Table name: AADT_1_2002
Function: Store AADT information, calculated from Monthly day- volume information.

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary	No	No	No	No	No	Name of segment
AADT	Number	10	Primary	No	No	No	No	No	AADT
C_1	Varchar	30	No	No	Yes	No	No	No	Comment of month 1
C_2	Varchar	30	No	No	Yes	No	No	No	Comment of month 2
C_3	Varchar	30	No	No	Yes	No	No	No	Comment of month 3
C_4	Varchar	30	No	No	Yes	No	No	No	Comment of month 4
C_5	Varchar	30	No	No	Yes	No	No	No	Comment of month 5
C_6	Varchar	30	No	No	Yes	No	No	No	Comment of month 6
C_7	Varchar	30	No	No	Yes	No	No	No	Comment of month 7
C_8	Varchar	30	No	No	Yes	No	No	No	Comment of month 8
C_9	Varchar	30	No	No	Yes	No	No	No	Comment of month 9
C_10	Varchar	30	No	No	Yes	No	No	No	Comment of month 10
C_11	Varchar	30	No	No	Yes	No	No	No	Comment of month 11
C_12	Varchar	30	No	No	Yes	No	No	No	Comment of month 12

9. Monthly factor information

Table name: m_factor_2002
Function: Store monthly factor information, obtained from transportation planning department

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
F_1	Number	10	No	No	Yes	No	No	No	Factor: month 1
F_2	Number	10	No	No	Yes	No	No	No	Factor: month 2
F_3	Number	10	No	No	Yes	No	No	No	Factor: month 3
F_4	Number	10	No	No	Yes	No	No	No	Factor: month 4
F_5	Number	10	No	No	Yes	No	No	No	Factor: month 5
F_6	Number	10	No	No	Yes	No	No	No	Factor: month 6
F_7	Number	10	No	No	Yes	No	No	No	Factor: month 7
F_8	Number	10	No	No	Yes	No	No	No	Factor: month 8
F_9	Number	10	No	No	Yes	No	No	No	Factor: month 9
F_10	Number	10	No	No	Yes	No	No	No	Factor: month 10
F_11	Number	10	No	No	Yes	No	No	No	Factor: month 11
F_12	Number	10	No	No	Yes	No	No	No	Factor: month 12

10. AADT information 2

Table name: AADT_2_2002
Function: Store AADT information, calculated from Monthly day- volume information and monthly factor.

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Segment	Varchar	30	Primary	No	No	No	No	No	Name of segment
AADT	Number	10	Primary	No	No	No	No	No	AADT
C_1	Varchar	30	No	No	Yes	No	No	No	Comment of month 1
C_2	Varchar	30	No	No	Yes	No	No	No	Comment of month 2
C_3	Varchar	30	No	No	Yes	No	No	No	Comment of month 3
C_4	Varchar	30	No	No	Yes	No	No	No	Comment of month 4
C_5	Varchar	30	No	No	Yes	No	No	No	Comment of month 5
C_6	Varchar	30	No	No	Yes	No	No	No	Comment of month 6
C_7	Varchar	30	No	No	Yes	No	No	No	Comment of month 7
C_8	Varchar	30	No	No	Yes	No	No	No	Comment of month 8
C_9	Varchar	30	No	No	Yes	No	No	No	Comment of month 9
C_10	Varchar	30	No	No	Yes	No	No	No	Comment of month 10
C_11	Varchar	30	No	No	Yes	No	No	No	Comment of month 11
C_12	Varchar	30	No	No	Yes	No	No	No	Comment of month 12

11. Original information of Record 3

Table name: original_Rec3_2002
Function: store one-year record S data collected by TRIMARC in one-hour intervals for further use

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Record_type	Varchar	1	No	No	No	No	No	No	Record type
State_code	Number	2	No	No	No	No	No	No	State code
F_class	Number	2	No	No	No	No	No	No	Functional classification code
Station_id	Varchar	6	No	No	No	No	No	No	Station identification
Direct_travel	Number	2	No	No	No	No	No	No	Direction of travel code
Y_year	Number	4	No	No	No	No	No	No	Year
M_month	Number	2	No	No	No	No	No	No	Month
D_day	Number	2	No	No	No	No	No	No	Day
DayofWeek	Number	1	No	No	No	No	No	No	Day of week
H_hour	Number	2	No	No	No	No	No	No	Hour
Vol	Number	10	No	No	No	No	No	No	Traffic count
Footnotes	Varchar	20	No	No	No	No	No	No	Footnotes

12. Original information of Record C

Table name: original_RecC_2002
Function: store one-year record C data collected by TRIMARC in one-hour intervals for further use

Field Name	Data Type	Size (bytes)	Key	Index	Null	Default Value	Check Criteria		Comments
							Unique	Other	
Record_type	Varchar	1	No	No	No	No	No	No	Record type
State_code	Number	2	No	No	No	No	No	No	State code
Station_id	Varchar	6	No	No	No	No	No	No	Station identification
Direct_travel	Number	2	No	No	No	No	No	No	Direction of travel code
No_lane	Number	2	No	No	No	No	No	No	Lane of travel
Y_year	Number	4	No	No	No	No	No	No	Year
M_month	Number	2	No	No	No	No	No	No	Month
D_day	Number	2	No	No	No	No	No	No	Day
H_hour	Number	2	No	No	No	No	No	No	Hour
Total_vol	Number	10	No	No	No	No	No	No	Total traffic count
Vol_1	Number	10	No	No	No	No	No	No	Traffic count of class 1
Vol_2	Number	10	No	No	No	No	No	No	Traffic count of class 2
Vol_3	Number	10	No	No	No	No	No	No	Traffic count of class 3
Vol_4	Number	10	No	No	No	No	No	No	Traffic count of class 4
Vol_5	Number	10	No	No	No	No	No	No	Traffic count of class 5
Vol_6	Number	10	No	No	No	No	No	No	Traffic count of class 6
Vol_7	Number	10	No	No	No	No	No	No	Traffic count of class 7
Vol_8	Number	10	No	No	No	No	No	No	Traffic count of class 8
Vol_9	Number	10	No	No	No	No	No	No	Traffic count of class 9
Vol_10	Number	10	No	No	No	No	No	No	Traffic count of class 10
Vol_11	Number	10	No	No	No	No	No	No	Traffic count of class 11
Vol_12	Number	10	No	No	No	No	No	No	Traffic count of class 12
Vol_13	Number	10	No	No	No	No	No	No	Traffic count of class 13
Vol_14	Number	10	No	No	No	No	No	No	Traffic count of class 14
Footnotes	Varchar	20	No	No	No	No	No	No	Footnotes

APPENDIX C
DATA QUALITY SCREENING CRITERIA

Rules	Sample Code with Threshold Values	Action
Logical consistency tests • Typically used for date, time and location. • Caused by various types of failures.	If DATE={valid date value} If TIME={valid time value} If DET_ID={valid detector location value}	• Write to off-line database and/or remove records with invalid date, time or location values.
Controller error codes • Special numeric codes that indicate that controller or system software has detected an error or a function has been disabled.	If VOLUME={code} or OCC={code} or SPEED={code} where {code} typically equals "-1" or "255"	• Set values with error codes to missing/null, assign missing value flag/code.
No vehicles present • Speed values of zero when no vehicles present • Indicates that no vehicles passed the detection zone during the detection time period.	If SPEED=0 and VOLUME=0 (and OCC=0)	• Set SPEED to missing/null, assign missing value code • No vehicles passed the detection zone during the time period.
Consistency of elapsed time between records • Polling period length may drift or controllers may accumulate data if polling cycle is missed. • Data collection server may not have stable or fixed communication time with field controllers.	Elapsed time between consecutive records exceeds a predefined limit or is not consistent	• Use flow rate rather than actual counts for VOLUME.
Duplicate records • Caused by errors in data archiving logic or software process.	Detector and date/time stamp are identical	• Remove/delete duplicate records.
Maximum volume • Traffic flow theory suggests a maximum traffic capacity.	If VOLUME > 750 per lane	• Assign QC flag to VOLUME, write failed record to off-line database, set VOLUME to missing/null.
Consecutive identical volume values • Research and statistical probability indicates that consecutive runs of identical data values are suspect. • Typically caused by hardware failures.	No more than 8 consecutive identical volume values	• Assign QC flag to VOLUME, OCCUPANCY and SPEED; write failed record to off-line database; set VOLUME, OCCUPANCY and SPEED to missing/null
Maximum occupancy • Empirical evidence suggests that all data values at high occupancy levels are suspect. • Caused by detectors that may be "stuck on."	If OCC > 80%	• Assign QC flag to VOLUME, OCCUPANCY and SPEED; write failed record to off-line database; set VOLUME, OCCUPANCY and SPEED to missing/null
Minimum speed • Empirical evidence suggests that actual speed values at low speed levels are inaccurate.	If SPEED < 5 mph	• Assign QC flag to SPEED, write failed record to off-line database, set SPEED value to missing/null
Maximum speed	If SPEED > 80 mph	• Assign QC flag to SPEED, write failed record

Rules	Sample Code with Threshold Values	Action
• Empirical evidence suggests that actual speed values at high speed levels are suspect.		to off-line database, set SPEED value to missing/null
Maximum reduction in speed • Empirical evidence suggests that speed reductions greater than some maximum value are suspect.	If $SPEED_{n+1} < (0.45 \times SPEED_n)$	• Assign QC flag to SPEED, write failed record to off-line database, set SPEED value to missing/null
Multi-variate consistency • Zero speed values when volume (and occupancy) are non-zero • Speed trap not functioning properly	If SPEED = 0 and VOLUME > 0 (and OCC > 0)	• Assign QC flag to SPEED, write failed record to off-line database, set SPEED value to missing/null
Multi-variate consistency • Zero volume values when speed is non-zero. • Unknown cause.	If VOLUME = 0 and SPEED > 0	• Assign QC flag to VOLUME, write failed record to off-line database, set VOLUME to missing/null
Multi-variate consistency • Zero speed and volume values when occupancy is non-zero. • Unknown cause.	If SPEED = 0 and VOLUME = 0 and OCC > 0	• Assign QC flag to VOLUME, OCCUPANCY and SPEED; write failed record to off-line database; set VOLUME, OCCUPANCY and SPEED to missing/null
Truncated occupancy values of zero • Caused when software truncates or rounds to integer value • Calculate maximum possible volume (MAXVOL) for an occupancy value of "1":	If OCC = 0 and VOLUME > MAXVOL where MAXVOL=(2.932*ELAPTIME*SPEED)/600	• Assign QC flag to VOLUME, OCCUPANCY and SPEED; write failed record to off-line database; set VOLUME, OCCUPANCY and SPEED to missing/null

For more information or a complete publication list, contact us at:

KENTUCKY TRANSPORTATION CENTER

176 Raymond Building
University of Kentucky
Lexington, Kentucky 40506-0281

(859) 257-4513
(859) 257-1815 (FAX)
1-800-432-0719
www.ktc.uky.edu
ktc@engr.uky.edu