

Final report of ITS Center project: Transportation Data Clearinghouse

A Research Project Report

For the Center for ITS Implementation Research

A U.S. DOT University Transportation Center

TRANSPORTATION DATA CLEARINGHOUSE

Principal Investigator:

Dr. Aaron Schroeder

Kathy Laskowski

Virginia Tech Transportation Institute

3500 Transportation Research Plaza (0536)

Blacksburg, VA 24061

Tel.: (540) 231-1505

Fax: (540) 231-1555

E-mail: hrakha@vt.edu

July 2003

Disclaimer:

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

TRANSPORTATION DATA CLEARINGHOUSE

FINAL REPORT

By
Dr. Aaron Schroeder
Kathy Laskowski

Submitted to: ITS Implementation Center—RSPA

July 2003



Introduction

In late 1999, the Virginia Tech Transportation Institute (VTTI), at the request of the Virginia Department of Transportation, undertook the process of examining the feasibility of expanding Virginia's Operational Database Management System (ODMS) to include data from neighboring states along the I-81 Corridor. This expanded system was intended to become a multi-state repository for "real time" traffic data that could be used for decision-quality level information for travelers from Bristol, Virginia to Harrisburg, Pennsylvania. This data would also be archived for later use as a planning tool for transportation professionals.

To achieve these goals, VTTI went through the process of designing an expanded ODMS and of determining if the other states' data could be used to fulfill the designs requirements. The first task in the design process was to perform interviews with traffic control professionals along the corridor. The information regarding their data source quality, format, connectivity, and other details was collected and used to create the larger technical design plans.

The information gained from these interviews provided the backbone to create the following plans for the expansion of the database. The design plans included:

- Conceptual System Model
- Clearinghouse Functional Requirements Plan
- Clearinghouse Conceptual, Logical and Physical Data Model Plan
- Functional Design Plan

These documents were then used to assess the different data sources according to technical integrity, institutional barriers, timeliness of data, and overall data quality to determine if the information was of a high enough quality to meet the needs of an expanding ODMS.

Task One: Interviews

A series of interview sessions were conducted with representatives from Virginia, West Virginia, Pennsylvania and Maryland Departments of Transportation, the State and Highway Administration and Traffic Administration officials. Questions posed to interviewees were tailored to their individual expertise in order to inventory specifics about operations in each participating state.

Information culled from these interviews ranged from data sharing procedures and processing to the specifics of data collection equipment and the reports that the equipment can provide. The final product of these interviews resulted in a thorough inventory of the infrastructure and data processes in place in each state.

The Conceptual Architecture for the system (Figure 1) is based on a high level data flow and control model across each of the participating states. The I-81 Corridor Conceptual System Model illustrates the decomposition of the input (or source) terminators for the information flow into and out of the system. The conceptual architecture is illustrated as a single function to demonstrate the external inputs later decomposed to address each state's data sources. The lowest level of decomposition will be the actual process that specifies the data flow constructed from an input of data, processes or subsystems. Each of the four states is depicted in detail with respect to their existing devices and traffic monitoring systems, road weather conditions, and incident management. The model provided a foundation upon which to devise strategies for disseminating information among prospective end-users through the data clearinghouse.

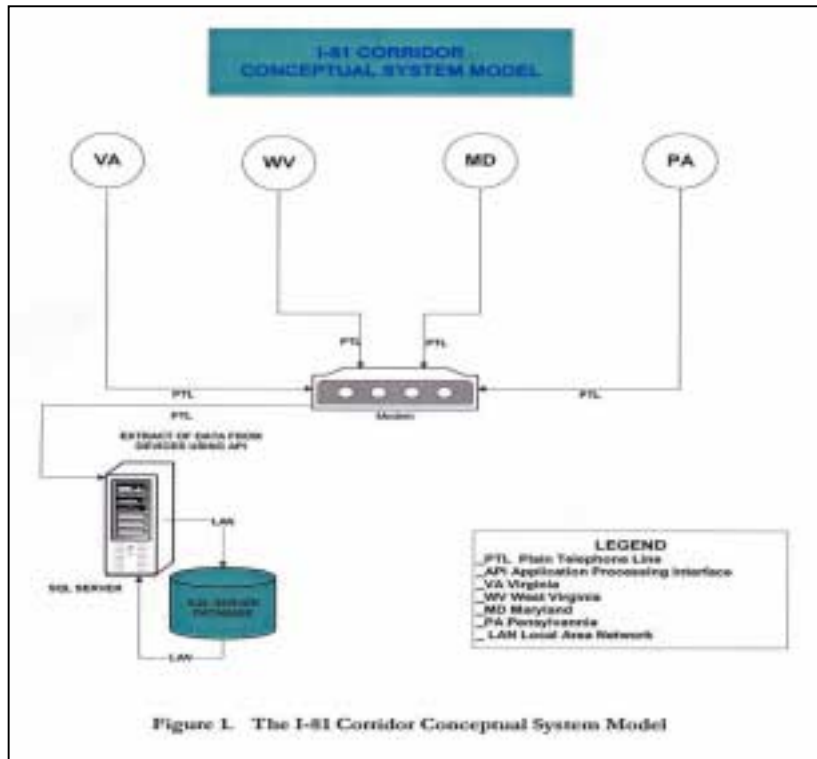


Figure 1. Conceptual Architecture.

All of the states had a variety of automated traffic recorder devices that collected traffic flow, counts and vehicle classification by type, information. This data is aggregated and used at a later time for planning purposes.

Real-time information varied in quality and quantity by state. This information is used by DOTs to better manage the roadway, as in the case of road weather information systems (RWIS), and may be relayed to the traveling public through various means, including changeable message signs(CMS), dynamic traffic alert signs(DTAS), and highway advisory radio (HAR).

The following table represents the existing data sources and whether their format is electronic or paper. Regardless of format, all data arriving at the clearinghouse would need to be processed by ITS operators and entered into an interface before being made available to the public.

Table 2. Purposed Data Sources for the Data Clearinghouse

State	VA	WVA	MD	PA
Incidents	State Police - CADS Computer Aided Dispatch DOT – VOIS VA Operations Information System	County 911 Dispatch Faxes	State Police Website	State Police Faxes
Traffic Flow	No available real- time data	No available real-time data	No available real-time data	No available real-time data
Weather	VDOT RWIS	-	-	-
Other	-	-	DTAS	VMS

This real time information was collected and validated via manual processes, meaning that these systems are not automated and needed to be initiated by ITS operators.

Task Two: Clearinghouse Functional Requirements Plan

Based on the information collected during the interviews, a clearinghouse requirements plan was created. This plan provided the specification for a functional infrastructure that is necessary to support the data clearinghouse database. Processes involved in the infrastructure pertain to the obstruction-free acquisition, transformation, and integration of the data to be collected from the designated device types along I-81 across Virginia, West Virginia, Maryland, and Pennsylvania. The specified infrastructure also includes the network, hardware, and software functions needed to support these processes. By conforming to these requirements, the database management system will have the foundation through which travelers can be informed of traffic, weather, and road conditions from place to place along the I-81 Corridor.

The data clearinghouse database represents the primary point of storage for event data associated with traffic incidents and weather and road conditions. Information in the database will be accessed at the discretion of VTTI end-users through an Online Analytical Processing (OLAP) application. In cases in which sources permit real-time extraction, information stored in the database will be current and detailed in terms of its status. The database design will support the consolidation of the real-time Intelligent Transportation System (ITS) data from designated information systems as well as data manually captured by an ITS operator. Information stored in the Data Clearinghouse will include material from sources such as the Weather Management Systems (WMS) and Variable Message Signs (VMS), as well as emergency operator logs, e-mails, and faxes. The design for the data clearinghouse will provide VTTI with both a logical and a physical construct of the model.

Clearinghouse Architecture:

The following is the functional decomposition of the data clearinghouse system. The interfaces and their task are described in detail to provide a hierarchical perspective of the chosen architecture.

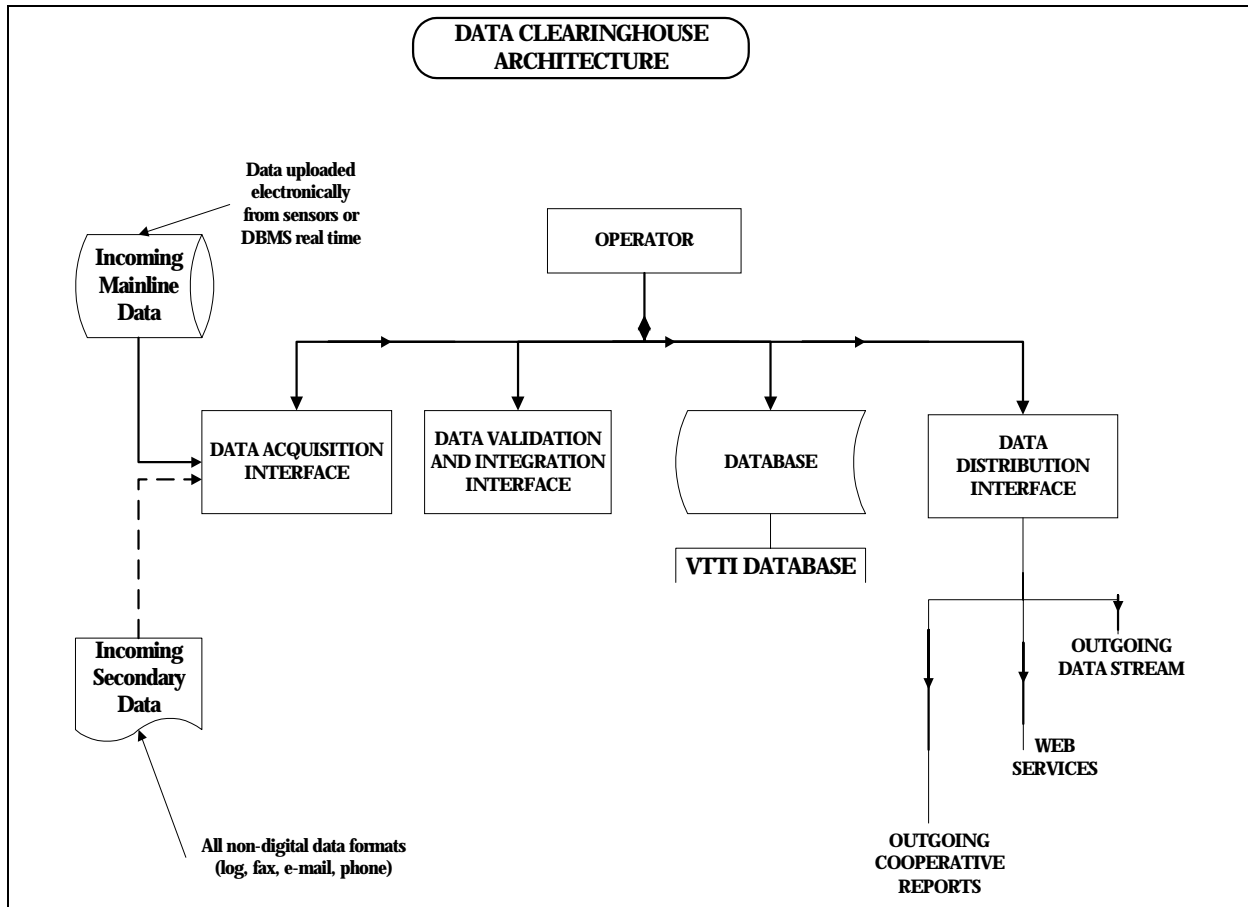


Figure 2. System Architecture.

Data Acquisition Interface

The Data Acquisition Interface is used to collect data from data sources. The interface connects to and extracts data from information systems, providing direct electronic transfer of data. The interface also facilitates operator-led input for indirect data streams, such as faxes and e-mail. Furthermore, the Data Acquisition Interface communicates with the Data Validation and Integration Interface to report communication fault and failure. Within the architecture, the Data Acquisition Interface is composed of one or more VTTI SQL Server engines, such as Transact-SQL, distributed queries, and command-line applications.

Data Validation and Integration Interface

The Data Validation and Integration Interface assesses the integrity of the data received from the Data Acquisition Interface. The interface applies established transformation rules to the data types and values of the data and then integrates the successfully validated data. The Data Validation and Integration Interface uses SQL Server Data Transformation Services (DTS) Designer to perform the transformation tasks by selecting the source data and mapping the data columns to a set of transformations. The transformed data is then sent to its target database in the Data Clearinghouse.

Database

The primary responsibility of the I-81 Data Clearinghouse database is to allow for the collection, organization and redistribution of traffic data. Collection of the data from the various state agencies is contingent upon the agencies' authorization for VTTI to acquire the data. The information collected in the database is designed to inform the public through traveler advisories available via the Internet (and other types of channels) on an as-need basis. The I-81 database design will be a combination of both relational and multidimensional design, where feasible, to enable support for incoming and outgoing data transactions. This database design will support data access and multiple-transaction processing.

Data Distribution Interface

The Data Distribution Interface will be located on a SQL server, thereby enabling reports and outgoing data streams to be generated through queries to the database. Upon receiving a request, the Data Distribution Interface uses an access interface to communicate with the database. It then retrieves the pertinent data and presents it to the end-user.

The plan also provides a high level review of system requirements for networking, hardware, and software to work in conjunction with existing systems and plans for the further expansion of the system.

Task Three: I-81 Data Clearinghouse Conceptual, Logical and Physical Data Models

Data modeling is a practical method employed for identifying information to collect and manage in a database. It involves analyzing informational resources, extracting the elements significant to the organization, and organizing those elements into the design of a database structure that is efficient and effective for information storage and retrieval. The conceptual, logical and physical data models present in this document represent the progressive steps performed in designing the data clearinghouse database. In turn, the database will act as the primary point of reference for informing travelers of real-time traffic, weather, and road conditions along the corridor.

After the initial investigation of the existing data sources employed in the participating states, the currency of the data from those resources and the feasibility of gaining access to and extracting data from those resources, the following guidelines were put into place regarding the content and data sources of interest for the Data Clearinghouse.

Table 3. Actual Data Sources for the Data Clearinghouse.

State /Content	VA	WVA	MD	PA
Incidents	State Police - CADS DOT – VOIS	none	CHART Web Site	State Police Fax Layout
Traffic Flow	-	-	-	-
Weather	VDOT RWIS	-	CHART Web Site	none
Other - DTAS	-	-	DTAS	
Other – VMS	-	-	-	None
Construction	-	-	CHART Web Site	PA Turnpike Web site, Penn DOT District 8 work zone web site

Conceptual Data Model

The conceptual data model for the Data Clearinghouse (Figure 3) is the first step in the design of the target database. The Data Clearinghouse conceptual data model is an entity-relationship diagram that presents the candidate objects of interest concerning travel-related data from the designated data sources.

An entity is the significant object or idea about which information is stored in the available sources. An example of an entity is *Incident*, which can be described by such characteristics as the location, time, and details of a traffic incident.

Relationships indicate any association that may exist among these entities and that may represent information of relevance to the Data Clearinghouse. An example of a relationship in the conceptual model is the association between the vehicle accident and the set of weather conditions existing at the time of the incident.

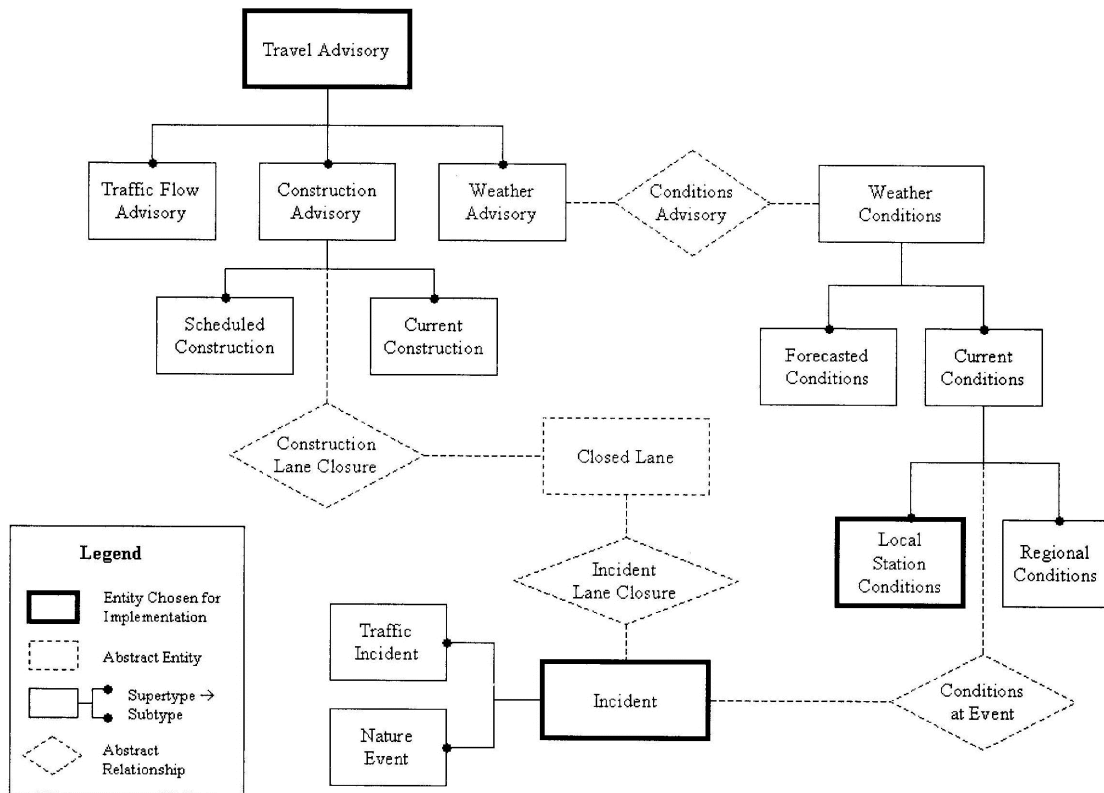


Figure 3. Conceptual Data Model for the Data Clearinghouse.

The description of the relationship contained in the conceptual data model are as follows:

Table 4. Relationships in the Conceptual Data Model.

Relationship	Description
Construction Lane Closure	The relationship wherein a construction activity (as detailed in a construction advisory) causes closed lanes. Aspects of this relationship include, for example, the effective period (start and expected end date) of the lane closure.
Incident Lane Closure	The relationship wherein an incident causes closed lanes. Aspects of this relationship include, for example, the effective period of the lane closure (when the lanes were closed and the expected time of reopening)
Conditions Advisory	The relationship between a weather advisory and the weather conditions manifest at that particular time.
Conditions at Event	The relationship between an incident and the weather conditions manifest at the time of the incident.

Logical Data Model

The logical data model (Figure 4) is the next step towards finalizing the database design. It translates the entities and relations from the conceptual model into a model that emphasizes the structuring of the data into a design for the database, independent of the particular target platform (such as SQL Server for Windows NT). Such structuring includes the tables (or entities), applicable attributes with meaningful names, logical data types (e.g., character, number and date), a particular schema type (such as entity-relationship model or a dimensional model), and other such logical considerations.

The logical data model for the Data Clearinghouse is designed to emphasize the three primary entities (Incidents, Travel Advisories, and Weather Conditions) of the integrated clearinghouse, suppressing unnecessary decomposition accommodating typical reporting on these entities. An exception is weather-locations information, which is broken out of the weather-conditions entity and formed into weather-locations ref in order to take advantage of the relatively constant nature of weather-sensor locations in contrast with the attributes pertaining to weather conditions. This division removes the redundancy of repeating the same location information when weather conditions change and history is tracked in the database.

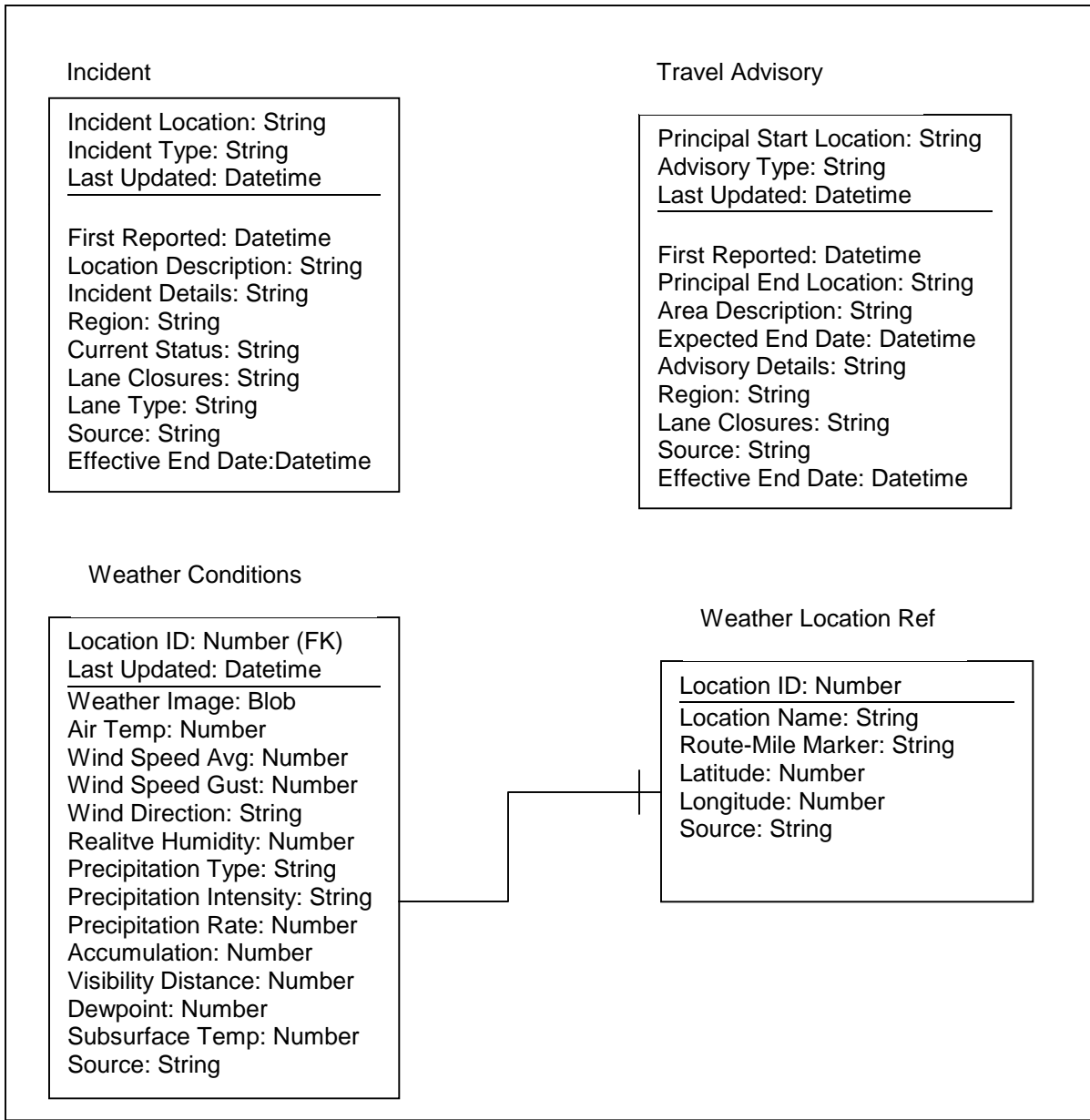


Figure 4. Logical Model for the Data Clearinghouse.

Another objective accomplished in the logical design is meaningful attribution of the entities. The entities in the model contain as many meaningful attributes as the data sources can accommodate and as many of those attributes that pertain to the needs of the database for informing the public on travel conditions. For example, the Incident entity contains attributes that answer such questions as what, when, where, and – to some extent – how and why with respect to the incidents. The question of who was involved in an incident is confidential and, therefore, is not deemed relevant to the needs of the database.

Another important feature of the logical design is the standardization of location referencing. Standardization is necessary for the identification of exact location points and boundaries relative to a geographical referencing system and relative to one another. Although location information from the sources tends to be broad, it appears that enough detail is provided for at least one location point at or near the boundary of an area to be identified. Therefore, the logical design imposes at least one location point on the entities, requiring the ITS operator to isolate the location point from source information. The attribute “Principle Start Location” in the entity “Travel Advisory” is an example of a location point imposed on the database.

Physical Data Model

The physical data model (Figure 5) is the final step in the database design, providing the specifications for implementing the data structure on a particular database platform. In the case of the I-81 Data Clearinghouse, the physical data model is designed for implementation on a SQL Server database running on a Windows NT platform.

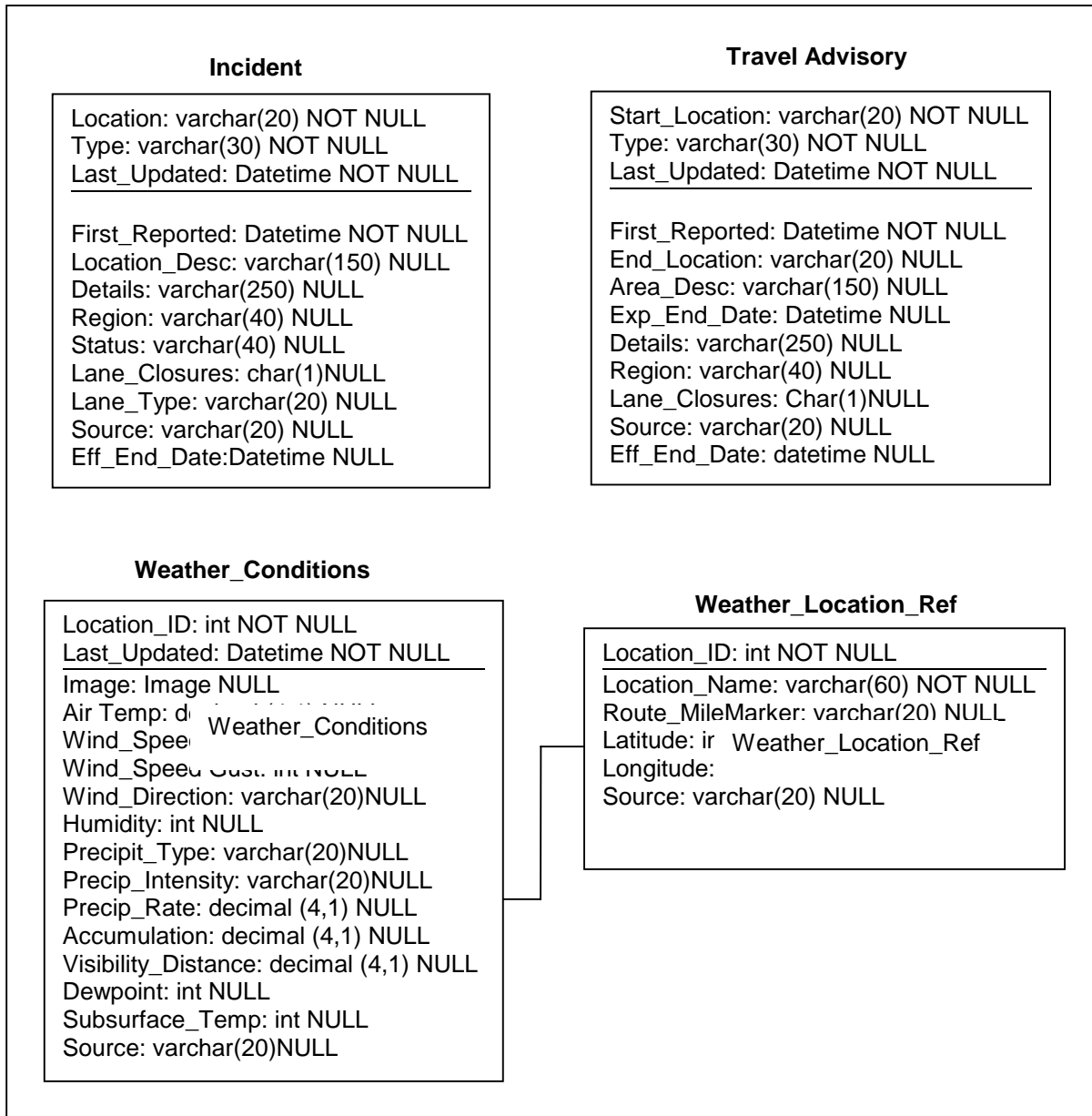


Figure 5. Physical Data Model.

The physical model uses abbreviated attribute names and physical data types to optimize storage on the SQL Server database.

Task Four: Data Clearinghouse Functional Design

The Functional design of the Data Clearinghouse database is based on the supporting architecture established in the Functional Requirements document.

Some key assumptions involved in the functional design are:

- The target Data Clearinghouse database is a downstream database, populated *via a processing cycle* that transfers, transforms/validates, and loads data into the database. Hence, the database is not the direct data store for source data, and Data Clearinghouse operators do not perform direct updates to data in the database.
- A data store is needed, either as an established source system or as a staging area, for collecting source data in an electronic format as input to the processing cycle.
- Data acquisition interfaces are needed to input source data into the data store.

Most of the Data Clearinghouse's designated data sources contain unstructured or free-form information that requires manual examination, selection, and entry of essential data into the data acquisition process. Because there is no input data store nor data acquisition interfaces in place or specified to represent the structured result of the manual distillation process, the functional design includes some assumptions regarding data that will be provided from the unstructured source content or that will be provided by the data acquisition operators. The document provides alerts to the reader where data ATIS operators must evaluate unstructured information and select the essential elements as inputs to the Data Clearinghouse.

Additional features highlighted in the functional design include a plan for the tracking of data history and a review of how the data processing cycle should function.

Tracking Data History

While the primary interest in this project was the collection of current, real-time data, there was also an interest in retaining history. History is useful for evaluating timeliness of information reporting and for satisfying simple ad-hoc queries pertaining to past events. History is not needed for analytical purposes requiring data warehouse or other specialized data structures for organizing historical data.

Below are the desired characteristics for tracking history:

- Be able to retain history for up to six months at a time.
- Be able to query specific attributes to minimize trial-and-error searching.
- Exclude camera images from history.
- Be able to retrieve desired historical data within a day of initiating the search.
- Separate historical data from current data into its own data store.

To accommodate these criteria for tracking history, a separate database is envisioned for the Data Clearinghouse architecture. The tables forming the database schema can simply be the same as those for the Data Clearinghouse database, except that the camera image attribute is removed from the counterpart to the *Weather Conditions* table to comply with the fourth criterion mentioned above.

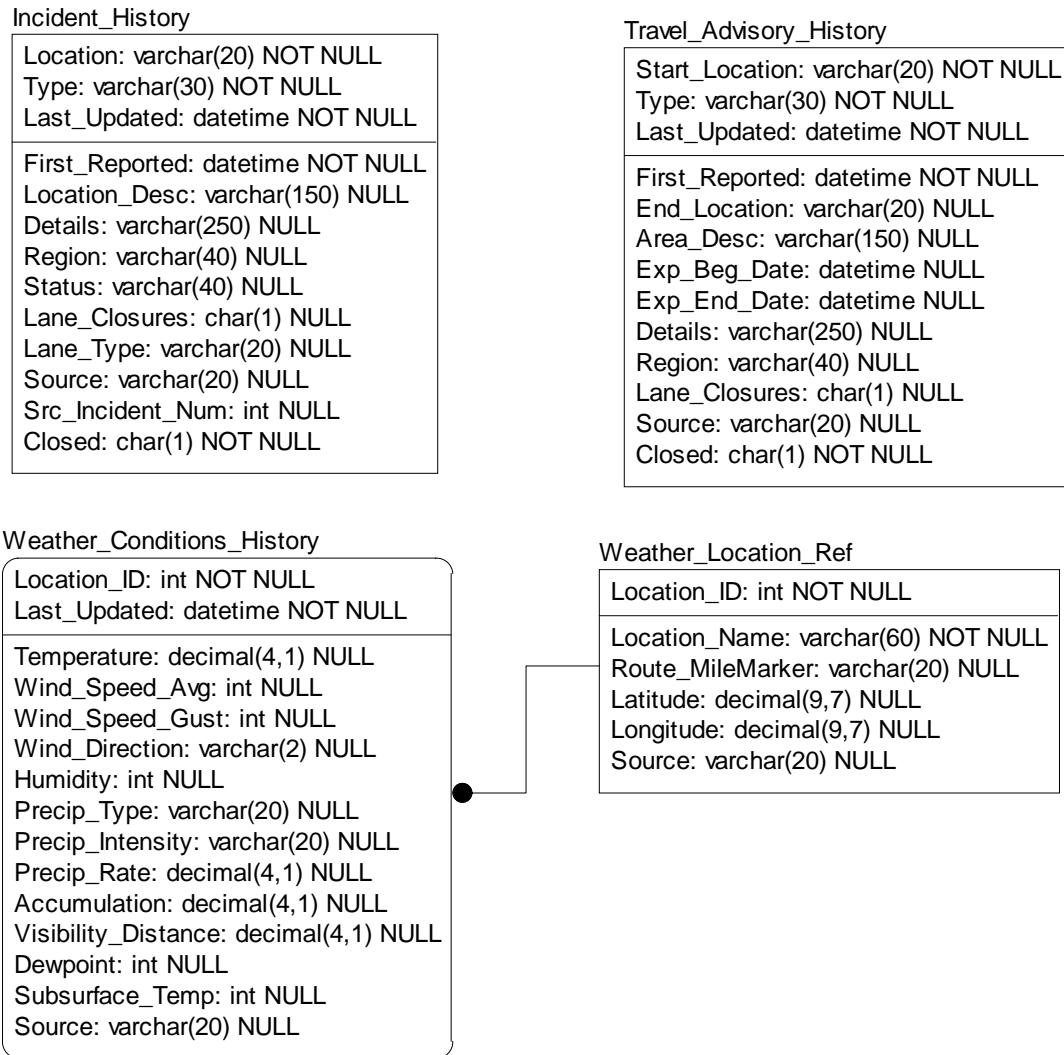


Figure 6. Historical Database Model.

The history database will contain only changed data in order to avoid large volumes of redundant unchanged information. The strategy for populating the database is described in the next section as part of the Data Clearinghouse processing cycle.

Data Processing Cycle

The processing cycle for populating the Data Clearinghouse follows the course presented in Figure 6. It consists of the interfaces described in *The I-81 Data Clearinghouse Functional Requirements* – the Data Acquisition Interface and the Data Transformation and Integration Interface – as well as an interface for loading the Data Clearinghouse and historical databases – the Changed Data Capture Interface. The sequence of processing follows.

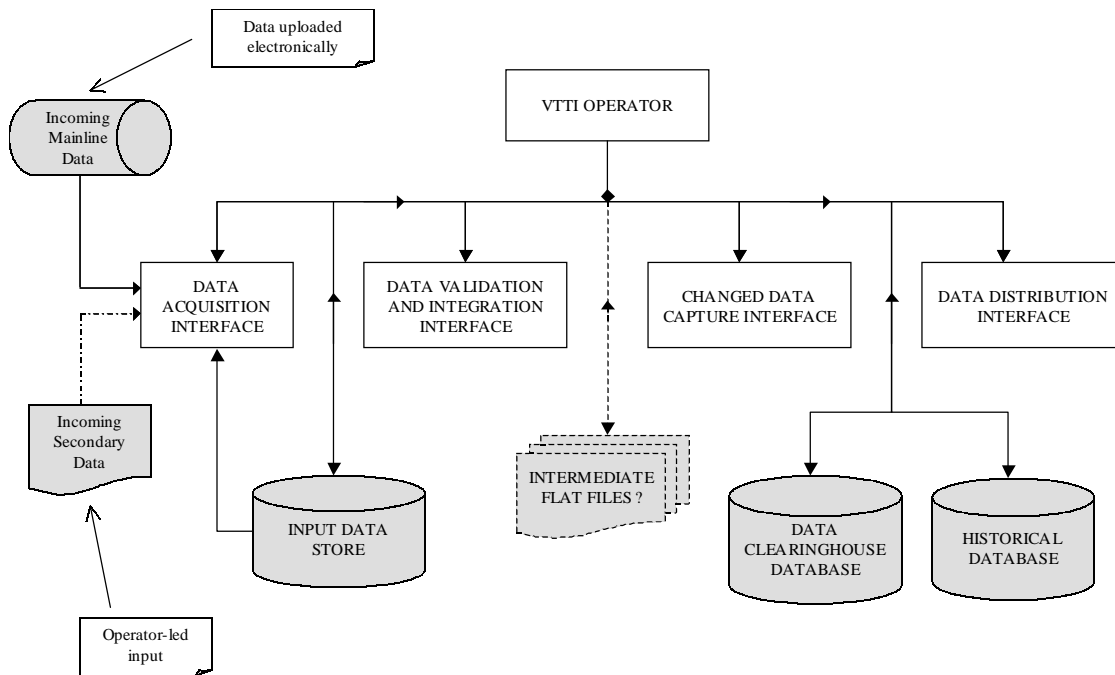


Figure 7. Data Clearinghouse Processing Cycle.

The cycle begins with the acquisition of source data via the Data Acquisition Interface. As the interface gathers source data, it stores the data in an input data store. Once the interface accumulates a set of inter-related current data, it passes the data set to the Data Transformation and Integration Interface, where it transforms, integrates, and validates the data as input to the Changed Data Capture Interface.

At this point, the Changed Data Capture Interface compares the incoming data against the existing data in the target table of the Data Clearinghouse database. In doing so, the interface checks for unchanged, new, and changed data. The interface ignores unchanged data, while it adds the records containing new and changed data to the target table. In the process, the interface writes the expired records of the database table to the associated table in the historical database.

The processing cycle, as described above, is event-driven. Data is pushed through the system from interface to interface once the data enters the system. In this way, data propagates through the system in a timely manner so that the Data Clearinghouse can store data, satisfying the projects real-time requirement.

Conclusions

The final output of this project was a technically complex plan for an Operational Database Management System for the I-81 Corridor . This plan was intended to integrate seamlessly and to structurally enhance VTTI's existing database that operates a rural advanced traveler information system in Virginia.

After a complete evaluation of the plans the following conclusions were drawn:

- **Technical Integrity:** The information from West Virginia, Maryland and Pennsylvania was available and accessible with no insurmountable technical barriers.

- Institutional Issues: The participating agencies were all willing to share the information.
- Timeliness of Data: The timeliness of the data was a large concern because VTTI would be relying on other manual processes performed by the other states to relay information the database. For example, Pennsylvania Police handwrite reports regarding traffic incidents and would then need to fax that report to the ITS operators at VTTI who would then input the data and make it available to the public.
- Overall Quality of Data: After several months of receiving trial data from the states, VTTI quickly learned that the information was neither timely nor relevant in many cases. Information on non-related traffic events were often included in the data transmissions, and traffic information was often in areas not relevant to the I-81 Corridor.

In the end, it was determined that the expansion of this database and rural ATIS would be ill advised at this time. When the other states are more able to guarantee the quality and timeliness of their data, VTTI will re-evaluate the situation.

The plans will still be able to provide a framework upon which further expansion will be possible, and they have allowed VTTI the chance to evaluate the level of data availability and quality within the I-81 region.