4. BASIC REQUIREMENTS

4.1 Introduction

The basic requirements for the Archived Data User Service are presented in this Chapter. These requirements are based on integrating the needs of the stakeholders with data that are either now available from ITS or could be easily achieved through ITS. A review of the general data types given in the National ITS Architecture -- as previously presented in Table 1.3 -- is the basis for identifying available data. This information, along with the background and issues presented in the earlier Chapters of this document, can be used to open detailed discussions with the stakeholder groups for the purpose of specifying the Archived Data User Service within the National ITS Architecture. The information is also of sufficient detail that it can also be used as a foundation for implementing the Archived Data User Service in real-world ITS applications. It must be noted, however, that *the recommendations regarding specific data elements identified in Section 4.2 are not meant to be exhaustive, merely a starting point for opening discussions with stakeholders.* In addition, several recommendations regarding the implementation and use of the Archived Data User Service by stakeholders are made in Section 4.3.

As previously stated, this User Service is fundamentally different from the 30 User Services currently defined by the National ITS Architecture in two key aspects. First, the number of stakeholder groups is very large, necessitating a high degree of coordination. Second, many stakeholder functions lie outside of the realm of ITS. For example, MPO transportation planners are concerned with a wide array of issues, with ITS only being a subset of those issues. It is not the intent of the National ITS Architecture to re-engineer these functions. Further, it is clear that many stakeholder needs can never be met with the existing structure of the National ITS Architecture or even with minor modifications to it. Rather, given that the functions already exist and that data from ITS can be used to assist those functions, the central question then becomes: *how can data from ITS be used to facilitate stakeholder functions.* The data needs of stakeholders have been well documented in past studies (as summarized in Appendix A). Stakeholder functional requirements, as they relate to available data from ITS, were given in Tables 1.3 and 2.1, although they have not been stated in formal system requirements terms. The reason for this is that the stakeholder requirements lead to specific recommendations about what data elements should be considered and what the structure of those elements should be (Table 4.1). In essence, many of the formal system requirements are imbedded in the recommendations concerning data elements. The next steps in the process from the point of view of the National ITS Architecture are to: (1) ensure that recommended data elements meet stakeholder needs by opening direct discussions with stakeholders and (2) work out the details of how the data specified in Table 4.1 can be accommodated with the National ITS Architecture.

4.2 Guiding Principles for the Requirements Definition

The following principles are general in nature and should serve as the basis for developing a more detailed set of requirements for the National ITS Architecture.

1. The systems to support the Archived Data User Service should take advantage of the data flows inherent in ITS to provide information in usable form for stakeholder applications. In doing so, the systems should consider the needs of all the stakeholder groups identified in Table 1.1 and should provide for data generated by ITS to meet those needs wherever possible.

2. The systems to support the Archived Data User Service should be based on, but not limited to, existing data flows within the National ITS Architecture. As new data flows are added to the National ITS Architecture -- and as additional uses of existing data flows are identified -- they should be examined for their inclusion within the systems to support the Archived Data User Service. The systems should also be flexible enough to accommodate data flows unique to individual ITS deployments that may not warrant a change to the National ITS Architecture. They should also be capable of handling data from existing data collection programs that may not be deemed as being in the ITS realm.

3. To accommodate both existing ITS and the incremental deployment of new ITS, the systems to support the Archived Data User Service should encompass two types of development:

   a. Type 1: Decentralized Structure - Each subsystem possesses its own archiving function with a minimum of interconnects with other subsystems but with compatible data definitions.

   b. Type 2: Centralized Structure - Relevant data flows from each subsystem should be captured in a central repository either directly or "virtually" through the use of appropriate distributed technologies and standards.

   Type 2 should be the preferred level of development and should be consistent with all definitions and standards within the National ITS Architecture. Type 1 should be defined in such a manner that it can be applied to ITS deployments that are not based on the National ITS Architecture (e.g., pre-existing systems). There should be a clear migration path from Type 1 to Type 2.

4. Three distinct subsystems are defined to support Type 2 systems to support the Archived Data User Service development. These subsystems should be implemented as part of any new deployment that incorporates the principles of the Archived Data User Service as part of its system design:

   a. Data Processing - the receipt and processing of incoming data from other ITS subsystems. The processing includes data quality control and editing as well as aggregation.

   b. Data Storage - both online and offline storage of raw and processed data.

   c. Data Retrieval - the interface between the data repository and stakeholders.

   These functions should either be developed as separate National ITS Architecture subsystems or should be incorporated into a single subsystem. The functions are enumerated as part of other requirements, listed below.

5. The functions or subsystem(s) specified in {4} should be linked through data flows with the appropriate subsystems in the National ITS Architecture. As new subsystems are added to the National ITS Architecture, new data flows generated by them should be considered for inclusion in the systems to support the Archived Data User Service.

6. The functions of the existing Planning Subsystem should be integrated into the systems to support the Archived Data User Service. It is recommended that the Planning Subsystem be renamed so that it is more clear that its focus is the support of multiple uses of data.

7. The systems to support the Archived Data User Service should accept user-specified data quality control and editing procedures. Whenever these procedures are applied to specific data, a permanent record should be made of the results in the metadata portion of the data dictionary (see {8} below).

8. The systems to support the Archived Data User Service should have the ability to perform the following data processing functions:
    a. Store data in the same structure as received from the field.

    b. Accommodate levels of aggregation of the data flows, depending on the type of data represented.

    c. Sample raw data flows for permanent storage in accordance with user specifications. Permanent storage of the sampled data should be either online, offline, or both.

9. To facilitate use and access by stakeholders, the data dictionary and schema of the systems to support the Archived Data User Service should contain **metadata** on each data element including: basic data dictionary attributes, the source of the data (including make and model of equipment), the conditions under which the data were collected (e.g., weather), status of the data collection equipment and/or tests of sensor output (where appropriate), error flags assigned by field equipment, the type of editing/quality control used to process the data, any imputation used to fill in missing or erroneous data (for aggregated data items), and the results of the data editing/quality control (including error flags).

10. The raw data from incoming flows should be capable of being stored online for a period of time specified by individual system designs.

11. Data owners (i.e., personnel responsible for the equipment that collect data generated by ITS) should have the ability to review, edit, and flag data prior to them being permanently archived.

12. The systems to support the Archived Data User Service should be consistent and should work with all applicable standards and protocols for both ITS and non-ITS programs as specified by ITS America, American Association of State Highway and Transportation Officials, Institute of Transportation Engineers (ITE), Society of Automotive Engineers (SAE), and USDOT, including:

    a. Location Referencing Specification
    b. NTCIP and TCIP
    c. applicable electronic data interchange standards (e.g., ANSI's ASC X.12)
    d. *Traffic Monitoring Guide*
    e. Highway Performance Monitoring System
    f. Fair Information and Privacy Principles for ITS/CVO

13. Permanent or temporary storage of data within the systems to support the Archived Data User Service should preclude the possibility of identifying or tracking individual citizens and should follow the ITS Privacy Principles developed by ITS America. Unique system-developed identifiers may be assigned to stored data that do not distinguish individuals. Public domain identifiers -- such as Social Security Numbers and license plate numbers -- should not be tagged or cross-linked with the stored data.

14. The systems to support the Archived Data User Service should accommodate, at a minimum, the data and associated structures shown in Table 4.1. The primary data elements shown in Table 4.1 should be matched to specific data flows in the National ITS Architecture. If the data flows do not currently exist in the National ITS Architecture, the National ITS Architecture should be modified (i.e., new data flows should be defined) to account for them.

15. The systems to support the Archived Data User Service should be capable of archiving *transformed* data as defined by local option. (See Section 1.2.1 for a definition.) Where transformed data are archived, the source data should also be archived for a period of time and keyed to the transformed data, along with metadata describing the calculation methods, assumptions, and external data used to perform the transformations.

16. The systems to support the Archived Data User Service should be compatible with and should take advantage of data specified the various ITS data dictionaries now being developed. These include: the Advanced **T**raffic **M**anagement Systems **D**ata **D**ictionary (TMDD) being produced by ITE, the Advanced Travelers Information Data Dictionary being prepared by SAE, and efforts by the Transportation Research Board Committee on Traffic Flow Theory and Characteristics to support data needs for advanced ITS modeling.

4.3 Specific Recommendations Regarding the Data Structure for the Archived Data User Service

To promote the adoption of the Archived Data User Service, a preliminary definition of the data structure is provided in Table 4.1. Either the primary data element or record type is given, often with a description of supplemental data fields that should be considered in the data structure. Time and location of the data collection are crucial companions to all the primary data elements shown and must be accounted for when systems are designed. (The issue of location referencing is beyond the scope of this document. However, there are several ongoing studies dealing with this issue that should be incorporated into final system designs.)

Table 4.1 Requirements for Archived Data from ITS for Multiple (Nonreal-Time) Uses

<table>
<tr><th>Primary Data Element or Record Type</th><th>Definition</th><th>Units</th><th>Internal Data Structure and Data Reduction Cycle</th><th>Level of Accuracy</th></tr>
<tr><td>Mainline traffic volume</td><td>Count of vehicles, during a given time period, past a point on the highway that is not influenced by a traffic control device or intersection. Includes volume counts from Electronic Toll Collection equipment.</td><td rowspan="3">Vehicles per unit time</td><td rowspan="9">Raw data from field sensors should be stored online for at least one day in the form in which they are received from the field. Data reduction should allow for permanent offline (or online) storage of the raw data. Data reduction should store the data online in multiple levels for each highway location where the data are collected:

1) 5-minute summaries by lane and direction -- online storage for at least one month
2) 1-hour summaries by direction -- online storage for at least one year
3) 24-hour summaries by direction -- permanent online storage

In performing the aggregation, locally-specified rules for handling missing or erroneous data should be applied. (The rules should be part of the metadata for these data elements.) Volumes in the aggregation are the simple totals for the time period; loop occupancy and density are simple averages; and speed should be the volume-weighted average.

For vehicle classification and weights, categories should conform to those in the most recent Traffic Monitoring Guide. Both classification and weight data will be stored for each highway location where the data are collected. Classification data should be stored as 1-hour summaries for each direction and lane and should be kept online for at least one year. Vehicle weight data should be accumulated online 24-hour intervals, at which time they should be summarized to the same structure as Level 3 of the Long-Term Pavement Performance Traffic Data Base (Appendix B). Permanent offline storage of both classification and weight data received directly from the field should be provided.

For vehicle headways, the type of lead and following vehicle should also be indicated.</td><td>+/-5%</td></tr>
<tr><td>Traffic control device approach volumes</td><td>Count of vehicles during a given time period on the approach to signalized intersections or at ramp meter controls.</td><td>+/-5%</td></tr>
<tr><td>Signalized intersection turning movements</td><td>Count of vehicles during a given time period for each turning movement at a signalized intersection.</td><td>+/-5%</td></tr>
<tr><td>Vehicle speed</td><td>Average speed of vehicles past a point on the highway during a given time period</td><td>Kilometers per hour</td><td>+/-5%; low speeds more important than high speeds</td></tr>
<tr><td>Loop occupancy</td><td>Average percent time that inductance loops sense vehicles during a given time period</td><td>Percent</td><td>+/-5%</td></tr>
<tr><td>Density</td><td>Average density of vehicles occupying a segment of highway during a given time period</td><td>Vehicles per lane-kilometer</td><td>+/-5%</td></tr>
<tr><td>Vehicle headway</td><td>The distance between two vehicles in the traffic stream, measured from the front bumper of the lead vehicle to the front bumper of the following vehicle.</td><td>Meters</td><td>+/-5%</td></tr>
<tr><td>Vehicle classification</td><td>Count of vehicles in pre-defined categories during a given time period</td><td>Vehicles per unit time by category</td><td>+/-10%</td></tr>
<tr><td>Vehicle weight</td><td>Weight of individual vehicles, axle groups for individual vehicles, or axles for individual vehicles</td><td>Kilograms</td><td>+/-10% for total vehicle weight</td></tr>
<tr><td>Traffic control device queue detection</td><td>Presence or absence of a queue located at significant setbacks from traffic control devices.</td><td>Yes/No</td><td>Time and setback from traffic control device should be stored on this record. Data should be stored at the level received from field controllers.</td><td>Accurate reading 90% of time</td></tr>
<tr><td>Traffic control device preemptions</td><td>Number of times traffic control devices (ramp meters and signals) have their timing preempted by transit, HOV, or emergency vehicles</td><td>Number</td><td>No special summarization required; save all data in raw form.</td><td>95-100% accuracy</td></tr>
</table>

Table 4.1 Requirements for Archived Data from ITS for Multiple (Nonreal-Time) Uses

Primary Data Element or Record Type	Definition	Units	Internal Data Structure and Data Reduction Cycle	Level of Accuracy
Traffic control device cycle lengths, phasing, and offsets	Time allocated for each phase (traffic signals only) and cycle (ramp meters and traffic signals). Offsets for traffic signals immediately upstream and downstream of the signal being inventoried.	Seconds	No special summarization required; save all data in raw form.	95-100% accuracy
Visual-based queue length	**Freeways:** Length of a platoon of vehicles where front-to-rear headways between vehicles are less than 25 feet, measured over a given time period. **Nonfreeways:** Length of a platoon of vehicles where front-to-rear headways between vehicles are less than 15 feet, measured over a given time period.	Kilometers	These data are determined from video or still-photography by either image processing or manual coding. Data should be permanently stored for each highway location at time intervals no smaller than 1-minute for signalized intersections and freeway ramps, and no smaller then 5-minutes for freeway mainline segments (requires computing average queue length for each time interval). The data should indicate the downstream point (beginning) and upstream point (end) of the queue using the local standard for location referencing.	+/-500 feet
Locally-derived traffic flow metrics generated by TMCs	Indices and other measurements used by local agencies to define congestion at either points on the highway (volume-to-capacity ratio) or highway segments (travel rate, accessibility index). These types of metrics are calculated from measured data (e.g., spot speeds used to calculate travel times).	Locally determined	Storage of speed and travel time data should follow the recommendations listed under the appropriate entries in this table. Locally-derived metrics should use these as a guide for their storage structure. Metadata must contain definitions and a thorough description of methods used in the calculations.	Unknown
Parking lot utilization	Proportion of parking spaces in use at a given time for a given locations	Percent	Stored data should contain not only the percent utilization but the total number of spaces available at the parking location being inventoried. Stored data should be summarized by 15-minute intervals for each parking location.	+/-10%
Transit vehicle boardings	Number of individuals paying transit fares upon entering a transit vehicle at specific times and locations (applies to both fixed-route and paratransit vehicles)	Number	Data should be permanently stored at the level that they are collected by electronic fare payment systems. Fields for identifying vehicle and route should be included in the structure.	+/-5%
Transit vehicle locations and times	The time that fixed-route transit vehicles arrive at scheduled stops and transfer points	Time	Data should be permanently stored at the level that they are collected by automatic vehicle location technologies. Fields for identifying vehicle and route should be included in the structure. Supplemental data should also include if an advisory for a route deviation was issued.	Unknown
Rideshare requests	The origin and destination of rideshare patrons by time of the request	Prevailing location referencing system	Data should be permanently stored by individual request.	95-100% accuracy

Table 4.1 Requirements for Archived Data from ITS for Multiple (Nonreal-Time) Uses

Primary Data Element or Record Type	Definition	Units	Internal Data Structure and Data Reduction Cycle	Level of Accuracy
Key times for incident specification	Incident start - time the incident occurred Incident notification - time the incident was reported from the field to a central operator Incident verification - time the incident was verified and recorded by a central operator Incident dispatch - time an EV was dispatched to the scene (multiples allowed) Incident scene arrival - time an EV arrived at the incident scene (multiples allowed) Incident lane clearance - time when each lane blocked was re-opened to traffic (for lane blockage incidents; multiples allowed) Incident shoulder clear - time when all shoulders were cleared of the blockage **and** EV Incident return - time each EV left the incident scene (multiples allowed)	Time	Data should be permanently stored for each incident reported from the field, whether they are verified or not. Times should be uniquely keyed to individual incidents. Supplemental data must include whether the times were actually measured/reported or estimated by system operators.	+/-5 minutes
Incident type	Category of incident	Formatted codes	Incident types should include at a minimum: (1) traffic crash, (2) debris (not water), (3) disabled/stalled vehicle (not crash-related), (4) fire on or adjacent to roadway (not related to a crash), (5) flooding or excess water on roadway (6) other weather-related (dust storm, tornado).	95-100% accuracy
Incident extent	Extent of traffic lane and shoulder blockage	Number	Number of lanes and/or shoulders blocked by the incident, including the times the blockage started and ended (to allow for multiple phase incidents).	
Incident hazardous material category	Hazard class and U.N. numbers (where appropriate) from the placard (multiple entries allowed)		Supplemental data on the incident; data should be permanently stored at the level that they are collected from the field.	
Incident hazardous material release	Amount of material released (multiple entries allowed)	Must be specified depending on container type		
Police accident report (PAR) reference	If the incident is a crash, the PAR reference number	Number		
Construction and work zone extent	Number of lanes and shoulders blocked by the construction or work zone activity	Number	Beginning and ending times **and** locations of each activity also need to be specified. Supplemental data could include a description of the activity.	95-100% accuracy
Train arrivals at HRIs	The beginning and ending times that HRIs are blocked by trains	Time	Data should be permanently stored at the level that they are collected from the field.	95-100% accuracy

Table 4.1 Requirements for Archived Data from ITS for Multiple (Nonreal-Time) Uses

Primary Data Element or Record Type	Definition	Units	Internal Data Structure and Data Reduction Cycle	Level of Accuracy
Emergency vehicle dispatch times	EV dispatch - time an EV was dispatched to the scene EV scene arrival - time an EV arrived at the scene EV clear - time emergency personnel reported the case "cleared" EV leave - time an EV left the scene EV destination arrival - time an EV arrived at its return location (e.g., hospital for medical service)	Time	These data are relevant for **all** emergencies, not just incidents. Data are usually collected by individual agencies through computer-aided dispatch systems.	+/-5 minutes
Emergency vehicle locations during response	EV origin - location of the EV when it was dispatched EV destination - location of the case or incident EV intermediate location - location of the EV at selected time intervals between origin and destination	Prevailing location referencing system	These data are used to track the routes taken by EVs in responding to cases or incidents. Intermediate locations should be recorded at 1-minute intervals between the times the vehicle was dispatched and it arrives at the scene. Because of the volume of the data generated, permanent online storage is optional if the data are stored offline.	Unknown
Commercial vehicle cargo type	The SIC code for the type of cargo being transported	SIC code	These data are collected by CVO systems, usually field sensors that detect the passage of individual trucks. The data should include time, location, and a vehicle identification code. Archiving data from every truck would probably not be cost-effective; however, provision to permanently store a sample of data should be made.	90-95% accuracy
Commercial vehicle origin and destination	For the shipment being made by this vehicle, the first point of origin and last destination	Prevailing location referencing system		Unknown
Intermodal container cargo type	The SIC code for the type of cargo being transported and the type of container.	SIC code	Same as for commercial vehicle cargo and O/D.	90-95% accuracy
Commercial vehicle origin and destination	The first point of origin and last destination for the container.	Prevailing location referencing system		Unknown
Hazardous material cargo type	Hazard class and U.N. numbers (where appropriate) from the placard (multiple entries allowed)			95-100% accuracy
Hazardous material pre-planned shipment route	The specified route to be taken for hazardous material shipments that require such treatment	Highway routes (as determined by the issuing agency)		Unknown
Commercial vehicle driver log	Selected locations and dates/times to determine hours of service for drivers	Prevailing location referencing system	These data are collected from on-board safety systems that are downloaded to field sensors. Archiving data from every truck would probably not be cost-effective; however, provision to permanently store a sample of data should be made. Supplemental data include vehicle identification and cumulative vehicle mileage. Privacy concerns may preclude the collection of these data.	95-100% accuracy

Table 4.1 Requirements for Archived Data from ITS for Multiple (Nonreal-Time) Uses

Primary Data Element or Record Type	Definition	Units	Internal Data Structure and Data Reduction Cycle	Level of Accuracy
Commercial vehicle subsystem status	Type of subsystem and status of its operation	(N/A)		Unknown
Roadside emission concentration	Volumetric concentration of pollutants measured by roadside sensors	Grams per unit volume (HC, CO, NO_x, SO_x)	Data should be saved for a minimum of one day online in the form that they are received from the field. These data should be aggregated and permanently stored for 15-minute time intervals (average concentrations for 15-minutes).	Unknown
Roadside temperature	Air temperature	Degrees Celsius	These data are collected by roadside weather sensing equipment. Data should be aggregated to no longer than 15-minute summaries (total precipitation, average temperature, predominant wind direction, average speed of wind in predominant direction) for permanent storage.	Unknown
Roadside precipitation	Type and amount of precipitation	Cubic centimeters (liquid volume)		Unknown
Roadside light conditions	Light level at roadside	At a minimum, should follow the codes specified in the Fatal Accident Reporting System		Unknown
Roadway surface condition	The surface condition of the roadway in terms of amount and type of moisture			Unknown
Roadside wind conditions	Direction and speed of wind	Kilometers per hour (speed)		Unknown
Segment travel times from probe vehicles	The time for a probe vehicle to traverse a given roadway segment	Seconds	For permanent storage, probe information (times at given points on the highway system) should be converted to total seconds. The data should be permanently stored online as 5-minute summaries (total probes counted, average travel time). A supplemental data item for permanent storage is the segment length. The raw probe data may be stored offline if actual vehicle identification is not included.	+/-10%
Transit vehicle times and locations	Data from AVI- or GPS-equipped transit vehicles	Seconds; prevailing location referencing system	For permanent storage, arrival times at pre-determined stops should be recorded along with vehicle and route identification. If transit vehicles are used as general travel time probes, they should be included as a special category under "Segment travel times from probe vehicles".	+/-10%
Traveler message content	Actual text of message displayed on a VMS or to travelers via personal devices.	(N/A)	All VMS messages (along with time and location) should be permanently stored online. Messages to personal devices (e.g., in-vehicle signing) may be sampled prior to permanent storage.	Close to 100% accuracy

Primary Data Element or Record Type	Definition	Units	Internal Data Structure and Data Reduction Cycle	Level of Accuracy*
Vehicle trajectories	Time and location of individual vehicles; measured for very short time intervals (1-10 seconds)	Prevailing location referencing system	Vehicle trajectory data can be collected through either GPS or advanced video image processing. The type of vehicle should also be indicated (see scheme under "vehicle classification"). Permanent online storage of both vehicle trajectory and origin/destination data are not recommended because of the sheer volume of data. Therefore, data should be aggregated to locally-defined geographic zones (e.g., traffic analysis zones, Census block groups) by activities. However, data may be accumulated online for short periods of time and stored offline for future use. Origin and destination activities (GPS-collected data) may be either collected directly or inferred from GIS base information (in advanced deployments.)	Unknown
Traveler origins and destinations	Origin - the point at which the trip began Destination - the point at which the trip ended Origin/Destination Activity - type of activity engaged in by the traveler at the origin and destination	Prevailing location referencing system		Unknown
Route guidance	Starting Point - the location of the trip at the time the guidance was given Ending Point - the desired destination Recommended Route - the route segments recommended	Prevailing location referencing system		Unknown
Variable facility pricing	Amount charged for a parking facility or toll for a highway segment where congestion pricing is in effect	Dollars	All recorded changes to pricing should be permanently stored.	Unknown

Notes: *This is a general indication of the desired accuracy.

(1) In addition, metadata in accordance with the principles put forth in Section 4.2 must be specified for each data element.

(2) time and location referencing must be considered in constructing systems around these data elements.

ABBREVIATIONS

HRI: Highway Rail Intersection
EV: Emergency Vehicle
TMC: Traffic Management Center
VMS: Variable Message Sign

Unless otherwise specified, the data specified in Table 4.1 should be saved online in their recommended forms for a minimum of one year, at which time they should be copied to offline storage media. The data structure is based on the notion that some data will be more useful to a wide variety of users if it is aggregated (summarized) and/or sampled prior to permanent storage. The aggregations described are simple totals or averages of lower data levels. Transformations of the data -- such as converting speeds to travel times -- are not specified for the basic structure of the Archived Data User Service. However, this does not preclude state and local agencies from incorporating data transformations in their implementations of the Archived Data User Service.

The information in Table 4.1 can be easily converted to User Service Requirements as developed for the other User Services in the National ITS Architecture; an example is provided in Table 4.2. Before this is done, however, it is important to reach consensus on the full list of data elements and their attributes through stakeholder discussions. The structure shown in Table 4.1 is thought to be more effective for stakeholder interaction than the traditional requirements format.

Table 4.2. Example Derivation of User Service Requirements

1.0	ROADWAY SURVEILLANCE DATA
1.1	MAINLINE TRAFFIC VOLUME
1.1.0	ADUS shall provide data on the count of vehicles, during a given time period, past a point on the highway that is not influenced by a traffic control device or intersection. This shall include volume counts from Electronic Toll Collection equipment.
1.1.1	Raw data from field sensors shall be stored online for at least one day in the form in which they are received from the field.
1.1.2	Data reduction shall allow for permanent offline (or online) storage of the raw data.
1.1.3	Data reduction should store the data online in multiple levels for each highway location where the data are collected.
1.1.3.1	5-minute summaries of total volume shall be provided by ADUS
1.1.3.1.1	5-minute summaries shall be for each lane and direction of traffic
1.1.3.1.2	5-minute summaries shall be stored online for at least one month

The structure shown in Table 4.1 can also be used as a basis for constructing a detailed data dictionary and schema for the National ITS Architecture. It is recommended that the entire structure be incorporated into the systems developed around the Archived Data User Service. The structure can also be used as guide for system developers who wish to retrofit data archiving on top of legacy systems, but development cost may prohibit full implementation. If this is the case, developers may wish to prioritize the retrofitting in accordance with the guidance provided below. (These general priorities are offered for guidance only. Local conditions may dictate otherwise.)

<u>High Priority</u>

- Traffic surveillance data (volumes, speeds, densities, and loop occupancies)
- Truck monitoring data (classifications and weights)
- Probe vehicle data
- Incident-related data
- Transit vehicle boardings

<u>Medium Priority</u>

- Transit vehicle locations and times
- Rideshare requests
- Construction and work zone extent
- Commercial vehicle cargo type and origin/destination data

Low Priority

- Traffic control device cycles, phasing, offsets, and preemptions
- Queue length
- Locally-derived traffic flow metrics
- Parking lot utilization
- Train arrivals at HRIs
- Emergency vehicle dispatch times and locations
- Pre-planned routes for hazardous material shipments
- Commercial vehicle driver log
- Commercial vehicle subsystem status
- Roadside emission concentration
- Roadside weather conditions
- VMS content
- Vehicle trajectories
- Route guidance
- Variable pricing

It is important to recognize that stakeholders must be free to develop their own priorities depending on local conditions. Specification of the system architecture is for guidance only.

4.4 Applicable Technologies

The Archived Data User Service is essentially an information management system, therefore, several information technologies (IT) apply to its development. System designers should make full use of these technologies when implementing the Archived Data User Service. These include (1) relational and distributed data base design; (2) data warehousing; (3) data mining; (4) expert systems, and (5) geographic information systems (GIS). Relational and distributed design and data warehousing are IT concepts that aid in the management and retrieval of data. Data mining can be applied to the analysis of the data to identify trends that otherwise might be buried in the huge volume of information that exist. Expert systems can be used as an aid in data quality control. GIS is an obvious technology for managing and displaying data generated by ITS.

4.5 Other Recommendations for Implementation

As stated many times in this document, simply defining the data that should be maintained in the Archived Data User Service is insufficient to achieve successful implementation. Therefore, several other recommendations are offered that will promote development and use of the Archived Data User Service in the field.

- "Best Practice" procedures for performing quality control and editing on Archived Data User Service data should be developed. Because ITS operations are a new phenomenon, little is known about how to identify and adjust questionable data received from field equipment. Data quality is a particular concern for those data elements that are aggregations of raw data: what should be the "rules" for handling not only questionable but missing data in the aggregation process. An ongoing research effort jointly funded by several states is examining QC and editing procedures for vehicle classification and weight data. Consideration should be given to performing similar studies for not only other forms of traffic surveillance data (e.g., volumes, speeds, densities) but for other types of Archived Data User Service data as well. Once the procedures are established, there is an additional need to develop *automated tools* to facilitate quality control and editing.

! Similarly, "Best Practice" procedures for analysis of archived data generated by ITS would provide guidance to stakeholders and would promote the use of Archived Data User Service. Demonstration of analytic methods (including graphical displays) would be extremely valuable to stakeholder groups. This is especially important because the sheer volume of data may be daunting to some groups not accustomed to working with large data sets. Assistance could take many forms: providing software and sample data, case studies of how other stakeholders use data, or documenting analysis techniques for specific applications (e.g., congestion monitoring, bus route planning, TDF model input preparation).

! Coordination with ongoing data dictionary efforts is crucial to the future development of the Archived Data User Service. There are several ongoing efforts to develop data dictionaries for various subcomponents of ITS (e.g., TMDD). Because these efforts are specifying data structures, they are highly relevant for the Archived Data User Service. Therefore, at a minimum, the stakeholders identified here should have input to these efforts. Further, consideration should be given to developing an Archived ITS Data Dictionary in accordance with the guidance provided by this document and the input of stakeholders. The endeavor of creating a data dictionary will force stakeholders and system architects to resolve many of the technical issues raised here.

! The needs of stakeholder groups should be represented in all forms of standards development that may have an impact on the relevant data. Other efforts have the potential for influencing the nature of data used in ITS technologies and are relevant to the Archived Data User Service. Therefore, input from Archived Data User Service stakeholders to these efforts will ensure that data are compatible.

! The Archived Data User Service should be integrated into other Federal, state, and local data collection programs. Although the Archived Data User Service is only one source of data for stakeholders, it can be used to supplement or replace many existing data programs. Examples include submittals of certain data to HPMS and statewide traffic monitoring. Full consideration should be given to how the Archived Data User Service fits into a comprehensive data collection program, including data sharing and standards.

! Training the various stakeholders in each others' needs is seen as an ongoing requirement. Personnel not directly involved in ITS operations currently have a limited working knowledge of the National ITS Architecture and of ITS in general. Likewise, many personnel who come from a systems engineering background rather than a transportation background do not yet have an appreciation for the breadth of traditional transportation functions. Education and outreach activities need to be increased for all transportation professionals. Several immediate options are available including: additional training under the Professional Capacity Building effort, strong recruitment of all stakeholders for the upcoming Regional Architecture Workshops, and development of a special short course for transportation managers.

- A concentrated field effort to demonstrate the implementation and use of the Archived Data User Service should be undertaken. Similar to the concept of ITS Field Operational Tests, the Archived Data User Service demonstration would provide a model for how to perform system development as well as how the data may be put to use. A key part of the demonstration would be to document the value of the increased information provided by the Archived Data User Service to local decision making and operations. Full documentation on the institutional issues as well as the technical hurdles to developing such a system must also be addressed.

- Funding for the Archived Data User Service should be implicit in future ITS deployment. Once the Archived Data User Service is specified in the National ITS Architecture, special emphasis should be placed on it when Federal ITS funding is offered to state and local governments. The cost of implementing the Archived Data User Service should be included when budget levels are determined.