Designation: E2259 – 03a (Reapproved 2011)

# Standard Guide for Archiving and Retrieving Intelligent Transportation Systems-Generated Data[1]

This standard is issued under the fixed designation E2259; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

## INTRODUCTION

This guide has been developed within the framework of developing standards of ASTM International and thus the format, structure, as well as the review, balloting, and approval processes conform to those specified by ASTM International. These processes may differ in their particulars from those used by other standards development organizations associated with developing ITS standards. However, their general intent is the same–that of having a representative group of interested and knowledgeable stakeholders develop standards that can be used by a wide variety of public and private organizations in developing Archived Data Management Systems (ADMS), which can be used to facilitate data sharing and interoperability among systems.

This guide has been prepared with various ITS data stakeholder groups in mind, that is, data users, ADMS policy makers, ADMS developers, and ADMS administrators. To data users and ADMS policy makers, it can provide a general understanding about technical approaches to archiving and retrieving ITS-generated data. For the ADMS developer group, which includes data application software developers, this guide can be a bench-marking reference against which existing ADMS could be further refined or improved with broader perspective. Further, it is also intended to bring greater awareness and consistency in the use and understanding of concepts and use of terminology by the ADMS developers. Finally, the ADMS administrators, whose main function includes collecting, archiving, managing, and distributing ITS data, can find practical guidance and approaches from this guide for effectively operating their ADMS.

It should be noted that there is a subtle distinction in the use of the acronym ADMS throughout this guide. When it is being used in the context of the National ITS Architecture the *S* stands for a *subsystem* of the National ITS Architecture. When the acronym ADMS is being used in the context of an Archived Data Management System in general, or a particular deployment of one, then the *S* stands for *System*. It is also noted that Appendix X1 to this guide presents relationships of the guide to several other activities and standards.

This guide follows the intent of all guides prepared within the ASTM International framework. In particular, it suggests approaches, offers an organized collection of information, or proposes series of options or instructions that give direction without recommending a specific course of action. Guides are also intended to increase awareness of potential users of the standards of the available techniques in the subject area, while at the same time providing information from which subsequent practices can be derived. This document, however, cannot replace a broad-based education or pertinent experience and should be used in conjunction with professional judgment. Not all aspects of this guide may be applicable in all circumstances. This guide is not intended to represent or replace the standard of care by which one judges the adequacy of a given professional service, nor should this document be applied without consideration of the many unique aspects of a project. The word *Standard* in the title of this document means only that the document has been approved through the ASTM International consensus process.

## 1. Scope

1.1 This guide covers desired approaches to be considered and followed in planning, developing, and operating specific ADMS for the archiving and retrieval of Intelligent Transportation Systems-generated data. The scope of this guide anticipates incremental or modular implementation of an ADMS, which over time and with a series of investment of resources will approach or exceed desired practice. However, it is recognized that programmatic constraints of time and budget resources do not always allow practitioners to follow a more desirable course of action and that during interim periods the ability to implement a particular fully functioning system may be less than desired.

1.2 The desired approaches described in this guide are foundational and are not intended to be all-inclusive. Users of this guide are allowed, and indeed encouraged, to exceed the desired practices in one or more of several ways. An example of one way is that to address and satisfy the particular needs and requirements of some of the intended users and stakeholders for a particular implementation may necessitate exceeding the desired practice. Another example is that some implementations may want to foster innovations and research into new methods and procedures related to the overall implementation of a particular ITS activity. Part of that may be the recognition that specialized archiving or retrieval processes, or both, would facilitate such innovations or research, or both. A third example is that some organizations may simply have more resources to invest in activities such as archiving and retrieval systems and may choose to have more quantities or higher quality of data and information available to their planning or operations units to use in their day-to-day activities.

## 2. Referenced Documents

2.1 *ASTM Standards:*[2]

E867 Terminology Relating to Vehicle-Pavement Systems

2.2 *IEEE Standard:*[3]

IEEE 1489 Standard for Data Dictionaries for Intelligent Transportation Systems

## 3. Terminology

3.1 *Definitions:*

3.1.1 The ASTM publication of the Committee on Definitions; Terminology E867 maintains a complete list of terminology some of which are applicable to this guide and by convention are not repeated here.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *aggregate, v*—to process or combine like items into a category; for example, adding together 30-s traffic volume counts, or averaging speeds from lane-by-lane detectors that are parts of traffic data counting station to be the total traffic volume or average speed at that traffic counting station for a 5-min time period.

3.2.2 *aggregation, n*—the resultant set of aggregated data associated with an aggregating process.

3.2.3 *archive, n*—the organized collection of data and information derived from selected ITS-generated data flows and other data sources.

3.2.4 *archive structure metadata, n*—descriptive data about the structure of the data archive itself and of the data and information in the archive that facilitate use of the archive.

3.2.5 *archived data administrator, n*—the person who is responsible for the day-to-day operations and long-term management of an ADMS.

3.2.6 *archived data management subsystem (ADMS), n*—a subsystem of the National ITS Architecture that provides a means for several organizations to collect, store, and subsequently, analyze and retrieve data from ITS data sources, usually by way of one or more ITS centers.

3.2.7 *archived data management system, (ADMS), n*—this is a system that is a specific implementation of an ADMS within the context of a local, regional, or statewide ITS architecture.

3.2.8 *archived data user service, (ADUS), n*—one of the ITS user services that defines the scope of the National ITS Architecture with regard to archiving and retrieving ITS-generated data.

3.2.9 *archiving, v*—collecting and actively managing original source and other data with the intent of saving the data as well as associated metadata.

3.2.10 *data, n*—a quantitative or qualitative representation that is observed, measured, collected, or gathered that characterizes some static or dynamic attribute of the physical world or the use of it by individuals or groups of people and that is suitable for communication, interpretation, or processing by humans or machines.

3.2.11 *data collection system metadata, n*—data about the conditions and procedures under which original source data were observed, surveyed, measured, gathered, or collected, as well as about the equipment that was used.

3.2.12 *data dictionary, n*—an information construct that describes the particular data stored in a database typically in terms of a common set of attributes that include the meaning, concept, and use; see IEEE 1489.

3.2.13 *data element, n*—a data item that is a basic building block of a data dictionary that is a formal representation of some single unit of information of interest with a singular instance value at any point in time about some entity of interest.

3.2.14 *data logging, v*—to capture a stream of near-real-time data as it passes through a transportation management center.

---

[1] This guide is under the jurisdiction of ASTM Committee E17 and is the direct responsibility of Subcommittee E17.54 on Archived Data User Service - (Disbanded 06/11).

Current edition approved June 1, 2011. Published July 2011. Originally approved in 2003. Last previous edition approved in 2003 as E2259 - 03a. DOI: 10.1520/E2259-03A(2011).

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] Available from Institute of Electrical and Electronic Engineers, Inc., 345 E. 7th St., New York, NY 10017.

3.2.15 *data mart, n*—a National ITS Architecture market package in which an archiving system collects and archives operational data from one organizational source; it is analogous to a library whose collection is acquired from a single publisher.

3.2.16 *data quality, n*—the fitness of data for all purposes that require it. Examples of data quality measures include accuracy, completeness, coverage, and timeliness.

3.2.17 *data sources, n*—the systems that provide data.

3.2.17.1 *Discussion*—Traffic count data can be retrieved from traffic management centers or air temperature data can be retrieved from surface transportation weather service providers.

3.2.18 *data warehouse, n*—a National ITS Architecture market package in which physically distributed data sources (*a*) contribute to a central clearinghouse, where each data source is locally managed and (*b*) exchange multiple transportation-related data; it is analogous to a library whose collection is acquired from multiple publishers.

3.2.19 *database, n*—a collection of related data typically organized in a computerized record keeping system that is part of a system whose purpose is to maintain the data and information derived from it so it can be made available for use.

3.2.19.1 *Discussion*—In addition, as noted by Smith (**1**)[4], it should (*a*) represent some aspect of the real world, organization, or enterprise, (*b*) that is logically related and has inherent meaning and (*c*) be designed, built, and populated for a specific purpose.

3.2.20 *database management system (DMBS), n*—a set of programs that manipulate and maintain a database while providing independence from an application.

3.2.21 *equipment package, n*—the building blocks of ITS subsystems that perform a specific function.

3.2.22 *erroneous data, n*—one or more pieces of data in a data set that has incorrect values due to a variety of reasons such as improper functioning of the data collection device, miscommunications, or shortcomings in subsequent processing programs.

3.2.23 *expired data, n*—the status assigned to a set of data, summary statistics, or information that is no longer current or valid and needs to be replaced.

3.2.24 *imputation, n*—the act of putting onto a data set estimated values for that data to fill in for missing values or to replace erroneous values.

3.2.25 *information, n*—the result of processing one or more pieces of data to produce a meaningful and useful statistic or indicator for users.

3.2.26 *integration, n*—the result of blending compatible data sources into a composite data set has a unity or wholeness for purposes of analysis, summarization, and retrieval of specific data or information.

3.2.27 *intelligent transportation systems (ITS), n*—systems that apply modern sensing, communication, computing, analysis, or display technologies, or a combination thereof, to one or more aspects of the operations, management, and use of transportation systems.

3.2.28 *log file, n*—a usually free form set of text or data, or both, that sequentially lists or enumerates events, many of which may be independent, while some may be chained or interrelated

3.2.29 *market package, n*—a service-oriented perspective to the National ITS Architecture, including subsystems (with applicable equipment packages) and architecture flows.

3.2.29.1 *Discussion*—The market packages relevant to ADUS are ITS Data Mart, ITS Data Warehouse, and ITS Virtual Data Warehouse.

3.2.30 *metadata, n*—loosely has been defined as data about data but is more tightly defined as the detailed description of instance data, including the format and characteristics of instance data where instance data is defined by Tannenbaum (**2**) as "that which is input into a receiving tool, application, database, or simple processing engine."

3.2.30.1 *Discussion*—Three categories of metadata are defined in this guide as (*a*) archive structure metadata (*b*) processing documentation metadata and (*c*) data collection system metadata.

3.2.31 *missing data, n*—one or more fields in a data record that has no valid values due to a variety of reasons such as improper functioning of the data collection device, failed communications, or shortcomings in subsequent processing programs.

3.2.32 *National ITS Architecture, n*—a document prepared through the sponsorship of the U.S. DOT that provides a common structure for the design of intelligent transportation systems giving a framework around which multiple design approaches can be developed by defining: (*a*) the functions that must be performed to implement a given user service (*b*) the physical entities or subsystems where these functions reside (*c*) the interfaces/information flows between the physical subsystems and (*d*) the communication requirements for the information flows.

3.2.33 *near-real-time data, n*—data that is collected, processed, aggregated very close in time, usually within seconds or just a few minutes, to the actual time period or instance when the phenomenon being measured or observed actually occurred.

3.2.34 *original source data, n*—data as received by a center that is a source for an archive.

3.2.35 *persistence history, n*—documentation of one or more modifications made to a set of data that can be used to trace back the specific procedures that were applied and when that was done and that becomes part of the processing documentation metadata.

3.2.36 *processing documentation metadata, n*—information that describes the processes applied to data from original source data through to storage in an ADMS.

[4] The boldface numbers in parenthesis refer to the list of references at the end of the standard.

3.2.37 *quality control, n*—a system or process for obtaining, maintaining, and verifying a desired level of quality in a data set, process, or service by careful planning and design, the use of proper equipment, continued inspections, and taking corrective actions where and when required.

3.2.38 *retrieving, v*—extracting from the ADMS one or more data values or information sets of interest to a user where the format of that may be one or more sets of text, tables, charts, graphs, schematic diagrams, or maps, using one or more varieties of media such as print, electronic files, electronic mail, or storage media such as CD ROMs or DVDs.

3.2.39 *resolution, n*—the level of spatial or temporal detail or administrative classification at which a particular data element is collected, archived retrieved, or both, from an ADMS; that is, lane-by-lane versus detector station or intersection approach; a 20-s versus a 15-min time period; or a particular named city versus the generic term of city.

3.2.40 *stakeholder, n*—a person, party, or private or public agency or organization that is interested in one or more aspects of how ITS-generated data is collected, archived, analyzed, or retrieved from an existing or proposed ADMS.

3.2.41 *storage media, n*—the systems upon which the archive or data set, summaries, or displays extracted from it are kept for use by the ADMS or for distribution to various users or stakeholders.

3.2.42 *terminators, n*—functions that basically take place outside of the National ITS Architecture but which interfere with one or more subsystems of the National ITS Architecture by sharing data or information, or both, with them.

3.2.43 *transformed data, n*—values created by combining data with other data or subjected to a methodology or mathematical process such as logarithmic transformation, averaging; for example, annual average daily traffic, average speed by segment, congestion indices, or equivalent single-axle loads.

3.2.44 *truth-in-data, n*—provision of adequate and appropriate metadata that should be given to users of an ADMS that a particular information set they are using is not original source data or information but rather is transformed data; that is, some transportation management centers (TMC) estimate/calculate average traffic speed by detector for each time period based upon algorithms that use volume and lane occupancy as original source data, that could also apply to some of the data being imputed.

3.2.45 *user service, n*—the National ITS Architecture defines various user services that are groups of ITS related functions that generally meet the need and requirements of a group of stakeholders for a particular type of service–the archived data user service was the thirty-first user service to be defined in the National ITS Architecture following a set process and period of interaction of the architecture team with the stakeholder community and the development and approval of a set of documentation of the need and characteristics of that user service.

3.2.46 *virtual data warehouse, n*—a National ITS Architecture market package in which physically distributed ITS archives exchange multiple transportation-related data even though the warehouse is not a center that directly collects its own ITS-related data, and it is analogous to a library whose collection is acquired from other libraries or from a central library system.

3.3 *Acronyms:*

| | | |
|---|---|---|
| AASHTO | = | American Associaton of State Highway and Transportation Officials |
| ADMS | = | Archived Data Management Subsystem or Archived Data Management System |
| ADUS | = | Archived Data User Service |
| DOT | = | Department of Transportation |
| FHWA | = | Federal Highway Administration |
| IEEE | = | Institute of Electrical and Electronics Engineers |
| ITE | = | Institute of Transportation Engineers |
| ITS | = | Intelligent Transportation Systems |
| JPO | = | Joint Program Office (for Intelligent Transportation Systems) |
| NCHRP | = | National Cooperative Highway Research Program |
| NTCIP | = | National Transportation Communication for ITS Protocol |
| NEMA | = | National Electrical Manufactures Association |
| NMEA | = | National Marine Electronics Association |
| SAE | = | Society of Automotive Engineers |
| TCIP | = | Transit Communication Interface Profile |
| TRB | = | Transportation Research Board |
| U.S. Dot | = | United States Department of Transportation |

## 4. Summary of Guide

4.1 This guide intends to help a range of practitioners better understand and keep their practice current, with this rapid pace of change, by disseminating information and general direction about this evolving state-of-the-art, which may be out ahead of the state-of-the-practice. ADUS standards are to be developed for several sources of ITS-generated data with traffic monitoring data being the first of these sources. This guide will help form the conceptual basis for developing these specific standards. This guide is to foster the timely development and use of many ADMSs that could benefit many stakeholders who are interested in using archived data and information from these expanded systems.

## 5. Significance and Use

5.1 This guide recognizes that activities associated with the archiving and retrieval of ITS-generated data must satisfy two classes of considerations – technical considerations and institutional considerations. Many aspects of such processes are analytic ones that need to address and satisfy various mathematical, statistical, data base management, information technology, or other similar technical concerns. On the other hand, many aspects of such processes are ones that need to address and satisfy various intra-agency, intra and inter-organizational, management approaches, decision-making, public-private participation, public involvement, or other similar such institutional concerns. This technical-institutional duality usually must both be satisfied for a particular archiving or retrieval process to be efficient and effective. That may also require a change in focus of the group responsible for the ADMS work to meeting the needs of their customers. Thus, in

planning, designing, or operating the functions and interfaces of an ADMS, appropriate technical and institutional approaches must often be explored concurrently in order to have an overall efficient and effective archiving or retrieval of ITS-generated data, or both. However, even though this guide gives importance to the institutional considerations when planning, designing, or operating an ADMS, this guide gives more weight and attention to the technical considerations of the functions and interfaces associated with an ADMS. This approach of focusing on function will give more latitude to institutions to be freer to arrange institutional matters on their own.

5.2 This Guide also deals in general with the questions of what data and information in particular should be considered for archiving as well as what data and information in particular should be considered for being retrieved. While more in depth consideration is to be given in the companion specifications, the following is a general characterization of what data and information should be considered for archiving. These include the following:

5.2.1 A structure of data storage based upon individual core data elements,

5.2.2 The data elements are stored in electronically readable media,

5.2.3 Documentation is included regarding the data structure and quality, and

5.2.4 Metadata is part of the archive relating to aspects such as the quality of the data and whether quality control adjustments have been made.

5.3 The particular original data sources and data elements that will be in an archive are expected to be unique to that archive. Which data sources and elements are used will depend upon a combination of the needs and requirements of potential users as well as the availability of particular data sources from other systems. For example, a particular ADMS may have some emergency management data on incident management responses, some traffic station sensor data from a traffic management subsystem, and some pavement temperature data from a roadway weather information system.

5.4 With regard to the *what is to be retrieved*, this guide recommends flexibility to provide for the needs of a diverse set of stakeholders. Some will want to retrieve short tabular summaries while others will want longer and more complex ones; some will want graphs and charts while others will want thematic maps of variations in the data. Accounting for different time periods and combinations of data and information quickly leads to an enormous range of possible data retrieval queries and approaches. While this guide does not recommend one retrieval approach being better than another because what is best will be very user specific, general guidance about the retrieval of data and information can be found in 10.16.8.

## 6. Concepts Related to Data Archiving, Retrieval, and Warehousing

6.1 *Distinctions Between Data Logging and Data Archiving*—Data logging is seen as a process where a transportation or TMC captures a stream of real-time data as it is passing through the TMC or as transmitted by the TMC. Various metaphors have been used to describe such logging processes such as turning on a spigot, or attaching a fire hose to the original source data. The data being captured through such a logging process typically gets saved in some data storage medium using some structured file or database system, and may actually subsequently go through a data archiving process. The term data archiving implies using some more comprehensive processes as well as longer-term storage than data logging. Extending the analogy to water flow, the archiving may do things such as the following: (*a*) put the data into various buckets or containers for later use (*b*) filtering or distilling it in some way to remove impurities and improve its quality or (*c*) redirecting, or aiming, the stream or containers of data at current or near-term applications that can use the data or derived information, or both to quench a need.

6.2 *The Terms Data Archive or Data Warehouse*—A data archive is a lot more than a collection or storage of original source data. Generally speaking, there is (*a*) on-going operational monitoring of some combination of the use, performance, characteristics, or events related to all or parts of a transportation system (*b*) some sort of system for continually, or very frequently, capturing the data into the archiving process (*c*) some quality control checks on the data (*d*) information about the data that can be helpful to people trying to use the data and (*e*) processes to get the data and information out of the archive and into the hands of users in ways that are meaningful to them. The use of the term warehouse to describe the archiving implies–drawing from a commerce analogy–a larger scale, more wholesale type of operation, rather than a more retail-oriented, limited-sized store. However, just as in the more contemporary styles of commerce, there can be megastores that do serve larger markets and often act as a blend between retail and wholesale. More discussion of a data warehouse is given in Appendix X2.4.4.

6.3 *The Growing Importance of Data and Information Retrieval*—This guide places a high degree of importance on the retrieval of data and information from the data archiving process. However, it is recognized that data and information retrieval has probably been the least developed aspect of the ADMSs that are in operation. As such there are proportionately fewer terms and concepts referenced in this guide that directly deal with retrieval activities. Perhaps the experiences, lessons learned, and future practice will result in additions to the terminology that relate more to the analysis and retrieval of data and information than to the storage of the data. Concepts and terms related more to temporal analysis of the data such as analyzing trends over time, comparative analysis, visualization of data, predictive applications, and very short-term forecasts may be among those that will need to be added. Broader approaches to data warehousing and data mining, which are

applied in many other fields and disciplines beyond just that of transportation, may become a source of such future references and terminology.

## 7. Providing Systems to Archive and Retrieve ITS Generated Data

7.1 *Reasons for Providing Systems*—There are many reasons why organizations should provide approaches and systems to archive and retrieve ITS generalized data. Some of the ones given here are from a report prepared by Turner (**3**).

7.1.1 *Maximize Cost-effectiveness of Data Collection Infrastructure*—Data archiving permits transportation agencies to maximize their investments in data collection infrastructure by reusing the same data for numerous transportation related functions such as planning, design, operations, and research. Two key goals related to this are (*a*) collect data once and store it once and (*b*) use the data for multiple purposes many times.

7.1.2 *Established Business Practice in Other Industries*—The retention and analysis of operational data is an established practice in most competitive industries that use data to manage their business activities. For example, the retail sales industry data warehouses full of customer transactions and inventories to better understand the basics of supply and demand in numerous markets around the world.

7.1.3 *Additional Information to Plan, Manage, Operate, and Evaluate the Transportation System*—Archived ITS data can be directly applied to a variety of transportation applications. These include:

7.1.3.1 *Operational Strategies*—The first step in proactive operations-oriented management is to know where problems are likely to occur before they actually do, then preventing or mitigating the impacts of those problems. Archived operations data can be used to (*1*) predict when and where problems may occur again (*2*) evaluate alternative strategies for preventing or mitigating problems (*3*) provide feedback for existing operations strategies, including short-term performance measures and in-depth evaluation studies and (*4*) develop sophisticated operations products such as real-time travel time predictions.

7.1.3.2 *Performance Measurement, Mobility Monitoring, Congestion Management Systems, and Research Studies*—There are a number of other transportation planning related activities that can be better carried out when archived data is available as an information source, which include performance measurement, mobility monitoring, congestion management systems, and independent research studies. The need to incorporate ITS archived data into such efforts may help foster the more timely development of an ADMS. Data archiving and retrieval can be significantly less expensive than having a planning or design workshop re-collect even a small percentage of the data using conventional manual methods or special studies.

7.1.3.3 *Support Governmental Reporting Systems*—Those reasons are primarily locally focused on the value of developing an ADMS for a specific locality or region, functional agency, or business organization. In addition, archiving can help more inclusive organizations have better information–such as states having information from their metropolitan areas, or the U.S. DOT having better information from the states.

## Planning, Development, and Operations of Archived Data Management Systems

## 8. Issues and Challenges in ADMS Development

8.1 Instituting an ADMS in practice raises several issues and challenges for developers of such systems. Specifically, developers should be aware of the following before embarking on ADMS development.

8.1.1 *Storage Requirements*—Real-time ITS generated data, the source of primary data for an ADMS, is voluminous. Developers should make realistic assessments of storage requirements and balance these with the needs of archived data users. Delineating offline versus online storage for different types of data–and the procedures associated with using these–are parts of this assessment, and will aid in the optimum design of storage of data processing.

8.1.2 *Choice of Operating System Environment and Data Base Management System*—It is expected that in managing the archive, the archived data administrator will select an industry standard operating system and database management system that are generally available. That will facilitate the sharing of data and the gathering and retrieval of data and information. While the use of uniquely designed systems is discouraged, it is anticipated that each ADMS would tailor or configure a database management systems to best meet their specific needs and requirements.

8.1.3 *Coordination with Existing ITS Data Systems and Sources*—ADMS development must account for the structure of existing systems, particularly those that are responsible for collecting and processing ITS-generated data (for example, TMC). Design cannot be done in isolation, but rather in conjunction, with those of these systems.

8.1.4 *Outside Contracting versus In-house Development*—Public agencies wishing to implement an ADMS need to determine whether system development can proceed with agency personnel, whether outside firms need to be contracted for the work, or perhaps whether an explicit blend of responsibilities for development can be established. There are many issues that will need to be addressed including those of (*a*) system security (*2*) ownership of source code and (*c*) knowledge retention.

8.1.5 *System Development Process*—In general, the system development process should adhere to sound information design principles. Foremost among these is a detailed specification of system requirements prior to actual implementation (see 10.2), followed by a detailed design. If agencies have decided to use outside contractors for system development, then one approach of many for handling the requirements definition and design is a design/build type of contract. Under such a contract, requirements definition can be done either in-house or as part of the design process.

8.1.6 *Uniqueness of Many ITS Data Sources*—System developers need to recognize that several of the ITS data sources that form the basis of an ADMS are uniquely different from traditional transportation-related data sources. While the underlying nature of the data elements may be fundamentally the same as traditional transportation data (for example, traffic counts, passenger loadings, incident characteristics), it is the

combined temporal and spatial relationship among the data elements that tends to be unique. Most traditional transportation-related data sources are either a sampling at a few locations very intensively over time, or are an extensive sampling of many locations during a smattering of different short intervals of time. It is this extensive concurrent spatial sampling at many locations for intensive short intervals of time that takes place over long periods of time that is unique to ITS-generated data. Structuring the ADMS to maintain that uniqueness can be very important to facilitate the effective retrieval of pertinent information from the ADMS that benefit a variety of different users.

## 9. Timing of ADMS Planning and Development

9.1 *Coincident Development*—In the ideal situation, an ADMS is identified and implemented as a system as part of a much broader ITS deployment in a city, region, or state. That is, it is designed and deployed simultaneously, or coincident to other ITS systems. This avoids the problem of retrofitting an ADMS to existing ITS deployments and is seen as more cost-efficient. However, situations are not always ideal and it may be more important to independently proceed with developing an ADMS, roughly in parallel but perhaps off-set in time, than to necessarily wait for the needs and requirements for other systems to be sufficiently defined and ready to start their development cycle.

9.2 *Inclusion in the Transportation Planning and ITS Planning Processes*—To achieve coincident development, the ADMS should be fostered by its proponents into the ongoing planning and programming processes in place at transportation agencies. Several opportunities should be pursued, and the following are some of the typical planning and programming documents and processes that may be used to achieve the objective of developing an ADMS in a timely manner:

9.2.1 *Long-Range Transportation Plans*—ITS strategies and needs are being increasingly identified in regional long range transportation plans. Where this is the case, identifying an ADMS in conjunction with other ITS deployments will bring the issue into future programming considerations. However, many of these processes are still just capital projects oriented and may not be set up to guide operations oriented activities, such as development of an ADMS.

9.2.2 *Short-Range Transportation Programs*—Regional or statewide coordinated transportation programming processes provide for an explicit set of capital projects or operating programs, or both, within a given time period and a set of fiscal and human resources. Identification of the deployment of an ADMS as a specific project or program, or as an element of a more inclusive project or program, is often a prerequisite to being able to actually proceed with the development of an ADMS on a particular time schedule.

9.2.3 *Annual Capital or Operating Budgets*—Most agencies use annual capital or operating budgets, or both and an ADMS may need to be explicitly included in order for development to proceed. In some circumstances, an agency may have sufficiently broad programming authority to enable an ADMS to be developed without the necessity to explicitly identify it in their budgets.

9.2.4 *ITS Strategic Plans and Other Pre-Deployment Documents*—In some cases, planning for ITS deployment occurs outside of the traditional transportation planning process. In that situation many prospective users of an ADMS may not be initially included in such strategic planning or processes. It is suggested that the responsible parties make an effort to involve prospective users from the outset or the earliest appropriate time.

9.2.5 *Statewide and Regional ITS Architectures*—Proponents of an ADMS should be directly involved in the development of statewide and regional ITS architectures as these are meant to be the specific set of systems that agencies and private sector organizations will provide for when deploying their individual ITS projects. Potential ADMS users should get involved early in the process and the developers of these architectures should seek out the participation of such potential users.

## 10. Guiding Principles for ADMS Development

10.1 This section provides a detailed discussion of guiding principles that should be used in developing an ADMS. A total of eleven overarching principles are identified. The first nine of these eleven principles, are organized around the functions found in the National ITS Architecture, although their titles and sequence have been modified slightly in some cases in order to meet the needs of this guide. The eleven guiding principles are the following:

10.1.1 Get archived data from other centers.

10.1.2 Integrate selected other transportation data including roadside data collection.

10.1.3 Manage the archive to account for data quality.

10.1.4 Specify and maintain metadata to support the ADMS.

10.1.5 Manage the Interfaces of the Archive Data Administrator.

10.1.6 Interact with other archives and monitor other standards.

10.1.7 Process user requests for data.

10.1.8 Support analysis of the archived data.

10.1.9 Prepare data for government reporting systems.

10.1.10 Reliance on user needs and requirements processes.

10.1.11 Provide security for the ADMS.

10.1.12 Figure 1 presents a grouping of ten of these principles that shows that they can be grouped primarily according to whether their focus is on (*a*) acquiring data (*b*) actually managing the ADMS or (*c*) retrieving data and information. The activities associated with those ten principles take place within an overall context and backdrop of the first principle of reliance on user needs and requirements processes. This guide does not view these principles as being sequential with respect to one another. Rather all of these are seen as being more organic in their interrelationships, with each interacting simultaneously with the others.

10.2 *Reliance on User Needs and Requirements Processes*—A general principle that should be followed in the establishment of an ADMS is to rely upon user needs and requirements processes that involve the range or sample of stakeholders that are anticipated as the providers and as the users of the archived data. This may involve determining

**TABLE 1 Illustration of Groupings Among the Guiding Principles for ADMS Development**

| Reliance on User Needs and Requirements Processes | | |
|---|---|---|
| Acquiring Data | Managing the ADMS | Retrieving Data and Information |
| | Manage the archive to account for data quality | |
| Get archived data from other centers | Provide security for the ADMS | Process user requests for data |
| Integrate selected other transportation data including roadside data collection | Specify and maintain metadata to support the ADMS | Support analysis of the archived data |
| | Manage the Interfaces of the Archive Data Administrator | Prepare data for government reporting systems |
| | Interact with other archives and monitor other standards | |

which agencies or internal workgroup, or both, should have primary responsibility for operating and maintaining the data archive, as well as facilitating retrieval of data and information from the archive. Having an explicit user needs and requirements process is seen as an essential step in achieving interdependency between archiving and retrieval of ITS-generated data. This guide strongly recommends that in planning, developing, and operating an ADMS that the needs and requirements of as diverse a group as possible of stakeholders be considered and accounted for to the extent practical. The ADUS is a service and a customer service perspective and is seen as being important and critical to the ultimate success of the archiving and retrieval processes.

10.3 *Providing for the Diverse Needs and Requirements of Different Stakeholders*—Work carried out by Margiotta (**4**) and others in establishing the ADUS showed that there is a very broad and diverse group of stakeholders with interest in the archiving and retrieval of ITS-generated data. These stakeholders groups tend to have somewhat different needs and requirements with regard to what data sources get archived, the detail at which they are archived, and ways in which they retrieve the data and information. A challenge is to not let the needs and requirements of one particular group necessarily set, control, or constrain the archiving or retrieval processes that will be available to the other groups. Rather, to the extent practical and affordable, the resulting archiving or retrieval processes, or both, should be inclusive and provide for as much of the full range of user needs and requirements of all groups. In providing the stakeholders' needs, the following areas should be addressed:

10.3.1 Potential users.

10.3.2 Types of data and information desired.

10.3.3 Format for receiving the resulting data and information.

10.3.4 Format for getting the data.

10.3.5 Frequency for wanting the data or information.

10.4 *Types of Stakeholders Based upon Functions*—Work performed by Winick (**5**) in examining the archiving and retrieval needs and requirements, associated with an updating of the CHART traffic management center of Maryland, built upon the list of stakeholder groups associated with ADUS and combined them into categories based upon similar organization functions and responsibilities. Those four categories are modified somewhat further here as follows (*1*) planning function, (*2*) operations function, (*3*) safety and regulatory functions, and (*4*) research functions. In addition, some additional stakeholder groups were identified or renamed, or both. In many cases a particular group may consist of more than one level of government or also include private or educational institutions in addition to the public sector. A listing of 17 stakeholder groups combined into these four functional categories is given in the following list:

10.4.1 *Planning Functions:*

10.4.1.1 State, regional, or local transportation planning.

10.4.1.2 Transportation system monitoring.

10.4.1.3 Air quality analysis.

10.4.1.4 Freight and intermodal planning.

10.4.1.5 Land use regulation and growth management.

10.4.1.6 Transportation administration and policy analysis.

10.4.1.7 Transit planning.

10.4.2 *Operations Functions:*

10.4.2.1 Traffic management.

10.4.2.2 Incident management.

10.4.2.3 Transit management.

10.4.2.4 Construction.

10.4.2.5 Maintenance

10.4.2.6 Traveler information.

10.4.2.7 Transportation demand management organizations.

10.4.3 *Safety Functions:*

10.4.3.1 Safety planning and analysis.

10.4.3.2 Commercial vehicle operations.

10.4.3.3 Emergency management.

10.4.4 *Research Functions:*

10.4.4.1 Transportation research.

10.5 *The Potential for Additional Stakeholder Groups and Accounting for Within Group Diversity*—This guide recommends using the categories and groupings of stakeholders in 10.4 as a starting-point in seeking out user needs and requirements for an ADMS. However, it is recognized that this listing of stakeholder groups or categories may still not be all-inclusive, and that for a particular application of this guide other groups may be added. In addition, for any particular group there can be numerous individual stakeholders, and there will be a diversity of needs and requirements within the group. For example, there can be many local traffic operators in an area with each interested in archived data from the freeway management system of the state DOT, but still having a somewhat different set of needs and requirements in either archiving or retrieving data, or both, from that source.

10.6 *Outreach and Participatory Involvement of Stakeholders*—This guide recommends the use of outreach and participatory involvement processes to sample the typical needs and requirements of a broad range of the stakeholder groups within each of the categories. Transportation agencies and researchers at the national, state, regional, and local levels have been doing work and research on how to carry out

stakeholder involvement processes so that they are effective for the stakeholders as well as for the organization conducting the involvement process. While such work has been aimed at transportation in general, the methods, techniques, and involvement processes can also be applied to the ongoing and future development of ITS and the establishment of an ADMS. In these instances it becomes more of an interagency and inter-organizational involvement process rather than a public involvement process. In addition, such stakeholder involvement activities can have the side benefit of promoting documentation that might otherwise be left undone in such data archiving efforts. Organizations seeking to refine or establish an ADMS are encouraged to use such involvement methods, techniques, and processes.

10.7 *Balancing Archiving and Retrieval Needs of Different Users*—This guide recognizes that there is interdependency between the archiving of data and information and the retrieval of data and information. A main premise of this guide is that attention needs to be given to both functions–archiving and retrieval–for the overall ADMS to have true value. Each function is necessary but neither is sufficient by itself. For example, simply having an archive system that just stores data without anyone to access and use the data would seem to be a waste of time, energy, and money. Conversely, having some user who would like to retrieve operational data that is not being archived requires them to somehow have the resources, knowledge, and permission to gain access so as to tap into an on-going stream of operational data, which may also divert resources of the operations staff. While some systems have functioned in such a latter fashion for a limited number of users, they often have been cumbersome and costly to use and the retrieved data may have gaps and not have a sufficient range of associated information, such as concurrent incident or weather conditions.

10.8 *Flexibility to Accommodate Future Unforeseen Needs*—The needs and requirements process should anticipate that no matter how thorough and well intentioned, there will always likely be some future unforeseen needs that will materialize. The ADMS should be designed and operated with such an expectation in mind. However, it is also recognized that how much flexibility is provided is a judgment call that may be handled differently by different archived data administrators.

10.9 *Get Archived Data from Other Centers*—ITS deployments have systems, subsystems, and terminators with functions that are performed in part through the flow of data among them. Those data flows can be used to support an ADMS. Thus, support of an ADMS should be based on, but not limited to, existing data flows between these ITS subsystems. Table X1.1 identifies data sources by subsystems and terminators associated with the National ITS Architecture. In getting archived data from other centers, a number of particular items need to be considered, including the following ones.

10.10 *Anticipate a Variety of Data Sources*—Archiving systems may accommodate a wide variety of existing data sources. That means including not only various quantitative data sources but also ones that are more qualitative, such as administrative records or log files from incident management, service patrol, or construction work zone activities. Some data sources may not be available in near-real-time, such as data gather using automated vehicle location (AVL) technology installed on part or all of a transit fleet that has the data downloaded one bus at a time when the buses return to the garage at the end of the day. It also means anticipating the inclusion of data from emerging and new technologies such as (*a*) transformed data based upon anonymous geo-locations from the energy patterns of cell phones in use or (*b*) traffic flow characteristics collected by unmanned drone mini-airplanes using video camera detectors, optical pattern recognition technology, and communications to an operations center in near-real-time.

10.11 *Rely upon Incremental Approaches – Think Long-term But Act in Phases*—In getting archived data it may be appropriate in many instances to start in terms of phases but think long-term – an incremental approach to getting data into the archiving system. That may mean beginning with a modest prototype archive focused on a single source of few types of data, such as freeway or arterial street detector system, or both, and then over a period of time expanding the archive to be more comprehensive in the types of data included in the archive. Over time the acquisition of archived data may emphasize different sources of data in a sequence, for example the archiving of (*a*) highway traffic related data (*b*) transit related data (*c*) multi-modal incident related data and (*d*) commercial vehicle related data. The cumulative effect of such an approach could be the on-going archiving of as much data as can be obtained for archiving to meet user needs, but staged over time to satisfy the requirements of working within a limited set of resources at any given time.

10.12 *Facilitate Data Access Across Agencies and Locations*—By its very nature and design, an ADMS should be intended for use by multiple users. An intention of this guide is to get beyond *stove-piped* architectures, which serve single agencies or single purposes, or both, but rather to facilitate data sharing and integration over multiple agencies, broad areas of geographic coverage, as well as across different functional or purposes categories. Such an approach would also, by extension, facilitate the establishment of regional architectures that could link together using distributed database technologies between two or more ADMS to function as a virtual data warehouse. The NTCIP center-to-center communications technologies as well as an integrated geographic reference system can greatly facilitate data exchange among an ADMS and other centers as defined in the National ITS Architecture.

10.13 *Not all Data Sources Need to be for Near-real-time Data*—Some data sources to be obtained by an ADMS may not be needed by a transportation operations center for their near-real-time operations. For example, the needs and requirements might call for data on traffic volumes gathered by a permanent count station to be included in the archive, which may be polled and retrieved not on a near-real-time basis. Another requirement may call for having processed information from such traffic volume counts; for example, estimates of annual average daily traffic could be available as one or more transportation management centers may want to access that

processed information from time-to-time for their own operational planning purposes.

10.14 *Provide for Some Experimentation*—In deciding which data sources should be sought for inclusion in an archive, there should be allowance for some experimentation and provision for some yet to be defined use and users. While most of the data sources should be based upon known stakeholder needs and requirements, there may not always be full anticipation, identification, and consideration of all potential stakeholder needs and requirements. The gathering of limited amounts of additional data based on judgment or research, or both, should be provided for without concern as to who are the known likely users or whether the data source is a non-traditional one. This may also be akin to literally installing a *data faucet*, but not necessarily turning it on initially but waiting for when circumstances are appropriate. This guide views the getting of data for archiving in the ADMS as being permissive, to allow for many diverse possibilities of use of what some may view as disparate data sources. For example, to get at issues of cause and effect in the overall analysis of crash records, and being able to link them through an ADMS may provide some crucial pieces of data or information in order for some users of the ADMS to test out hypotheses of causality

10.15 *Integrate Selected Other Transportation Data Including Roadside Data Collection*—The systems to support the ADUS should take advantage of the data flows inherent in ITS to provide information in usable form for stakeholder applications. In managing the integration of selected transportation data from data flows among subsystems, including data from roadside data collection, a number of particular items need to be considered, including the following ones.

10.15.1 *Selection of Particular Data Elements for Archiving*—Procedures should be established to enable the administrator or management oversight committee to decide what particular data elements should be archived. That would include specifying for what period of time the data should be archived as well as whether it will be short-term or permanent. To the extent possible, consideration of which data elements to archive should be based upon user needs and requirements process.

10.15.2 *Integration of Data from ITS and Non-ITS Sources*—A major use of ITS-generated data is to support existing stakeholder applications that currently rely on their own data sources and information systems. ITS-generated data can supplement—and sometimes replace—the data in existing information systems that support these applications. Therefore, integration of ITS-generated data into existing ("legacy") transportation information systems is a key function of the ADMS. The integration does not need to take the form of a single integrated system. Rather, the data in the ADMS should be stored in such a way as to be compatible with existing information systems. Compatibility of data elements between ITS and non-ITS sources should account for such issues as:

10.15.2.1 *Data Definitions*—Existing data elements should be capable of being derived from ITS-generated data either by adapting similar definitions or developing standard translations between data elements.

10.15.2.2 *Location Referencing*—It is possible that data from ITS sources are based on different location referencing systems than those used by existing systems. It may be beyond the scope of the ADMS to reconcile different location referencing systems, but at a minimum, ADMS developers should:

(1) communicate the discrepancies to personnel responsible for ITS and non-ITS data sources and

(2) develop translation schemes to allow existing systems to access ITS-generated data.

10.15.3 *Retention of Original Source Data*—Original source data from incoming flows should be maintained at its most detailed spatial and temporal resolution for a period of time specified by individual system designs. The retention time will vary for individual systems, but the retention capability is considered to be a major feature of an ADMS. Further, systems should strive for flexibility by providing options for retention. The question of how original source data is retained and the level of spatial, temporal, and categorical resolution is much more of a policy decision in establishing an ADMS than it is a technical issue. The ADMS capability should not be encumbered by technical designs for the system, selected just by system developers, resulting in aggregation of some data that could have readily been available at finer resolutions.

10.15.4 *Storage of Original Source Data*—Depending on the needs of a particular ADMS, original source data may be permanently stored either online or offline. In some cases, original source data may be purged completely, but this practice is strongly discouraged. If this practice is in place, provisions should first be made to enable interested users to save the original source data to their own system before that data is removed from the ADMS. However, exact replication of the data structure is not required, particularly for extraneous data related to communications that are not required by users. Instead, the intent of this recommendation is to preserve the technical content as well as the temporal and spatial level of detail in the original source data. This should also apply to the metadata that may be available. If practical, it is not only data that should be kept for a period of time, but also some hardware that may be necessary to *read* the data. However, as storage media technology changes, the storage of original source data that is being retained ought to be migrated to the newer technology.

10.15.5 *Retention of the Resolution of the Original Source Data*—Some users may be interested in the data at the most detailed spatial or temporal resolution possible, thus ADMS may only retain data for a limited time period before it is distilled or aggregated to a less detailed resolution. However, to the extent possible, an ADMS should keep the data at the spatial and temporal resolution at which it was gathered and if it is to be aggregated, provisions should first be made to enable interested users to save the detailed data to their own system before that data is aggregated. In case a center prefers to aggregate data before making it available for archiving, an effort should be made by the archived data administrator to see if the more detailed resolutions could be retained.

10.15.6 *Sampling of Original Source Data*—In some cases, local system design or operating policies may dictate that not all original source data be retained. Depending on the needs of

individual users, samples of original source data may be adequate for these cases. If sampling is used to retain original source data, metadata must be specified indicating the details of the sampling procedure.

10.15.7 *Anticipating Retrieval in Blending and Data Integration Activities*—In managing the integration and blending of the selected transportation data system, developers should anticipate retrieval needs and provide for processes that would facilitate retrieval of information. That would include items such as: *(a)* structured data storage for individual data elements that is more oriented to ease of retrieval by a variety of users than it is towards data management efficiency *(b)* electronically readable formats that to the extent possible follow widely available commercial formats and *(c)* documentation of the data structure that facilitates finding aspects such as data types, transportation system functional classes, corridor and route locations, and jurisdictional summaries.

10.15.8 *Integrate Data Flows from Roadside Subsystems when Appropriate*—Some ADMS may find it more efficient, at times, to directly gather data from roadside data collection devices and integrate that data with other data being gathered by the ADMS usually through one or more traffic management centers. That may include other types of data, such as environmental data, parking data, or commercial vehicle inspection data collected by roadside devices. Having the ability to either poll or routinely gather roadside data collection needs to be coordinated with the organizations that own, operate, or maintain those roadside data collection devices. Such roadside devices may also include those of traffic monitoring or survey groups, such as the office that runs the data collection from permanent automated traffic recorders, and who are often associated with the planning function within an agency rather than an operations function.

10.16 *Manage the Archive to Account for Data Quality*—This guiding principle deals with establishing management systems that will enable the effective functioning of an ADMS.

10.16.1 *Data Quality and Providing for User-Specified Quality Control Procedures*—Indicators of data quality, collection conditions, and type of data source should be developed and documented. The ADMS should accept user-specified data quality control and editing procedures where quality control methods are used to flag or remove suspect, erroneous, or duplicative data from the data archive. However, a *truth-in-data* perspective should be fostered and maintained in managing the ADMS that alerts users to when they are not using original source data. In addition, whenever these types of procedures are applied to specific data, a permanent record or persistence history should be made of the results in the metadata portion of the archive or data dictionary, or both.

10.16.2 *Maintaining Version Control in the ADMS*—In most ADMSs it is anticipated that processing procedures will be applied to the original source data or that summary statistics and information will be generated from the data, or both. These processed data, information, or statistics can become obsolete for several reasons, such as after-the-fact quality control decisions or acquiring more data after statistics are generated. Since the earlier statistics may have been used for some purpose, an effective archive system needs to maintain status information for all statistics. If the system allows the incoming original source data to be replaced, these data should also have status information. For example, a useful form of status information maintains both *effective date/time* and *expiration date/time* for all of the on affected records. The statistics or data being replaced should not be deleted, but *expired*. This means setting the expiration date from null (or some date several thousand years into the future) to the date and time when the processing routines replaced the information with new information. The new information is added as a new record with an effective date equal to the expiration date of the replaced data or statistics. The expiration date should be either null or some very future date.

10.16.3 *Responsibility for Data Quality of the Original Source Data*—Data owners (that is, organizations responsible for the equipment that collect data generated by ITS subsystems) should, in general, have the responsibility to review, edit, and flag data prior to the data being permanently archived. This is not intended to be an *unfunded mandate* and flexibility is needed. The organization making an effort to collect data is probably in the best position to review and assess the data quality prior to the archiving of the data. However, if the need for a particular degree of quality by users of the archive data is much higher than that of the organization that collected the data, then perhaps sharing of responsibility by the users for meeting the costs of having the extra quality may be warranted.

10.16.4 *Provide Security for the ADMS*—One of the responsibilities of the archived data administrator is to provide security for the ADMS. Security should be provided, to the extent feasible, so as to minimize compromising the ability of users to access and use the system. ADMS security is multifaceted and includes issues such as data integrity, data confidentiality and usage rights, computer and communication systems protection, and open access. This guide further divides these ADMS security concerns into areas of physical security, legal access, and data security and maintenance.

10.16.4.1 *ADMS Physical Security*—The archived data administrator should consider and provide for the physical security of the ADMS, including factors such as the following:

*(a)* Safeguarding hardware and media against damage, theft, and misappropriation,

*(b)* Providing multi-level hardware and media backup, including, in general, provision for computer system hot-swap, backup power, and off-site duplication of data stores,

*(c)* Routine hardware maintenance and media rotation/replacement,

*(d)* Authorized accessibility twenty-four hours per day, seven days per week, and

*(e)* Designation and control of secure components that are distinguished from insecure or public components.

10.16.4.2 *ADMS Legal Security*—The archived data administrator should be aware of, prepare for, and comply with specific legal and policy restrictions, including items such the following:

*(a)* Data confidentiality – data that must only be assessable to specifically authorized users,

*(b)* ITS data privacy – following ITS privacy principles developed by ITS America and precluding the possibility of identifying or tracking individual citizens,

*(c)* Legislated data privacy requirements,

*(d)* Data copyright – data whose distribution and use is restricted by the owner or contributor, or both,

*(e)* Data accessibility – free and open access to public data,

*(f)* Data license requirements – to support public-private partnerships for data collection and management (such as granting access to public rights-of-way to a private data collector in exchange for access to the data),

*(g)* Data retention (and destruction) requirements, and

*(h)* Data requested by legal process, such as subpoenaed data.

10.16.4.3 In general, it is suggested that an ADMS for ITS-generated data avoid and exclude legally restricted data to the maximum possible extent and that an archived data administrator carefully consider the implications of admitting any legally restricted data into the archive. Some implications include the following:

*(a)* Legal repercussions and liability from inadvertent breach of covenants,

*(b)* The practical necessity and inevitability of the archived data administrator effectively having complete access to all data, including legally restricted data, access control schemes, and passwords,

*(c)* Potential liability for the improper actions of restricted-data providers, including their failure to provide access to data as required by law or policy, and

*(d)* Potential conflicting assertions of copyright and legal covenants now and in the future.

10.16.4.4 *ADMS Data Security and Maintenance*—The archived data administrator should ensure the long-term integrity of the ADMS, particularly with respect to the following considerations:

*(a)* Archived data should not change unless adequate metadata description of the changes are maintained along with the data,

*(b)* Archived media should be routinely monitored for readability and re-written to currently readable media and formats as required by changing information technology or physical deterioration,

*(c)* Data ownership and authority should be maintained as metadata along with all data,

*(d)* All data should be accurately time-stamped at the time of entry into the ADMS,

*(e)* All data should be location stamped where appropriate,

*(f)* As much as possible, data elements should not be duplicated within the ADMS, and special attention should be paid to automated data sources, data filters, and *processing daemons* that may inadvertently duplicate or alter data elements within the archive, and

*(g)* A system of user identification level of access, passwords, and other similar security features should be provided.

10.16.5 *Establish and Maintain Metadata*—To manage the data archive it is necessary to establish metadata, which is a set of adequate documentation about the archive and the data residing in it. For the purpose of this guide, three subcategories of metadata are defined that respectively relate to how the data can be used, how it has been processed, and where the data come from. These categories generally follow, elaborate, and rename information noted by Turner **(6),** in discussions about adequate documentation for data archives and data collection systems. These three categories of metadata for purposes of this guide are termed as (*a*) *archive structure metadata*, descriptive data about the structure of the data archive itself and of the data and information in the archive that facilitate use of the archive (*b*) *processing documentation metadata*, information that describes the processes applied to data from original source data through to storage in an ADMS and (*c*) *data collection system metadata*, data about the conditions and procedures under which original source data were observed, surveyed, measured, gathered, or collected as well as about the equipment that was used.

10.16.5.1 *Archive Structure Metadata*—This form of metadata is the most common one found in the current literature. Colloquially referred to as *data about* data, archive structure metadata is typically thought of as dataset descriptions, where a dataset may be an entire database, a grouping of data elements (for example, a table in a relational database), or individual data elements. Archive structure metadata deals with the structure of the archive itself including the data elements, definitions, and data dictionary attributes that are available, the various databases and tables that are maintained, and other items that might typically fall under a data schema. Archive structure metadata are analogous to a library card catalog that contains information about books and periodicals such as, subject, keywords, accession number, place of printing, author, titles of articles, and so forth. In this analogy, the books themselves are the data or datasets and are not part of the archive structure metadata. The descriptions typically found in a data dictionary (for example, definition, size, source, valid values) are also archive structure metadata. It is anticipated that the organization providing the original source data will find it in their interest to contribute appropriate archive structure metadata. It will be the responsibility of the archived data administrator to incorporate that, with necessary modifications, into a catalog for the ADMS. The ITS data registry may be a source that can be used for some metadata attributes. Commercial-off-the-shelf database products generally have the ability to report such archive structure metadata information once the metadata is entered into the system. Carrying out these responsibilities on a continuing basis can be a major long-term commitment of manpower and resources that will need to be sustained once it has begun. Archive structure metadata have several purposes, which like those of a library card catalog, are oriented towards serving the needs of the users of the system. These purposes include the following:

*(a) Summary,* to summarize the meaning of the data,

*(b) Finding,* to allow users to search for the data,

*(c) Advisement,* to allow users to determine if the data is what they want,

*(d) Selection,* to help users decide which instance of the data should be retrieved (if multiple formats are provided),

*(e) Retrieval,* to retrieve and use a copy of the data (that is, where does one go to get the data),

*(f) Restriction,* to prevent some users from accessing data,

*(g) Interpretation,* to instruct users on how to interpret the data (for example, format, encoding, encryption),

*(h) Specifications,* to give information that affects the use of data (for example, legal conditions on use, its size, or age); terms and conditions for use of an object (an access list of who can view the object, a Conditions of Use statement; a schedule of fees for use of the object; or a definition of permitted uses of an object),

*(i) Data Linkages or Relationships,* to give specifications about the relationship between objects, for example, linkages between data element definitions that may have changed over time, and

*(j) Data Structure,* to list the logical components of complex or compound objects and how to access those components (that is, table of contents or the list of components of a software suite).

10.16.5.2 *Processing Documentation Metadata*—Correctly specifying this category of metadata is among the most critical things that an ADMS needs to do well. Thus this guide has been developed to separate out, define, and highlight this category of metadata, which is central to establishing an ADMS. Data will typically undergo a series of processing steps along their path from collection to archival. The purpose of processing documentation metadata is to provide an audit trail of these processes, including documentation of any modifications the data may have undergone since they were collected at the source or received by the archive. Such processing level metadata is focused on the internal management of the data and the development of related information from the data within the archiving process. Processing documentation metadata may be stored as unique data elements as part of the ADMS database or they may be stored as archive structure metadata if the process being documented applies to all data in a set (or database). For example, a quality control procedure of a specific type may be applied to all data in a set. In this case, it would be unnecessary to add a new data element to capture this condition for it could be documented in the archive structure metadata. Examples of processing documentation metadata include the following:

(*a*) *Data Quality Control Procedures,* the type of editing/quality control used to process the data,

(*b*) *Data Quality Control Results,* the results of the data editing/quality control (for example, error flags),

(*c*) *Data Imputation/Replacement,* any imputation used to fill in missing data (for example, for aggregated data items) or the replacement of erroneous data, as determined by quality control procedures,

(*d*) *Data Administration,* to give specifications for the management of an object within a server or repository, such as date of last modification, date of creation, and the administrator's identity,

(*e*) *History,* a complete summary of the changes, or provenance of data, that the data has gone through since being original source data and any subsequent transformations, such as filtering, decimation, summarization, aggregation, and so forth,

(*f*) *Data Processing Status,* indicates the extent of processing that has occurred to create the data element, with possible categories including the following:

*(1) Original Source Data:* data measured directly by collection equipment as received by a non-archival center,

*(2) Aggregated Data:* information created as simple summaries of a category, that is, peak hour or daily traffic based upon observed traffic volume data, and

*(3) Transformed/Computed Data:* information created by combining data with other data or subjected to a methodology such as averaging, that is, annual average daily traffic, average speed by segment, congestion indices, or equivalent single-axle loads.

(*g*) *Aggregated/Summarized Data Statistics:*

*(1) Method,* a textual description of the aggregation method,

*(2) Number of Observations Used,* the actual number of observations used in the summary exclusive of imputed or altered data,

*(3) Observation Units,* the level from which the aggregation is made (for example, hours for daily summaries, days for annual summaries),

*(4) Distribution Statistics,* standard deviation and selected percentiles (e.g., minimum, 5th, 25$^{th}$, 50th, 75th, 95th, and maximum) from the data from which the aggregation is made,

*(5) Precision,* where possible, a calculated precision for the aggregated data assuming only sampling bias, and

*(6) Precision method,* the methodology used to calculate the precision.

(*h*) *Transformed Data Statistics*—An ADMS should be capable of archiving transformed data as defined by local option. Where transformed data are archived, the original 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.

*(1) Method,* a textual description of the computational method,

*(2) External data name,*

*(3) External data value, and*

*(4) External data source.*

10.16.5.3 *Data Collection System Metadata*—This form of metadata is used to describe the conditions under which the data were observed, surveyed, measured, gathered, or collected. Such data collection system metadata is focused on the external measurement of the original source data at the source. Note that some forms of data collection metadata may be collected for ITS operations or management uses and need not

be duplicated as metadata for archiving purposes. Examples of data collection metadata include the following:

*(a) Initial Collection Source,* the source of the data as collected by the initial ITS equipment, including the type, make and model of equipment, or other identification of the detectors,

*(b) Equipment Location and Identification,* details of the location of fixed equipment such as roadway sensors, equipment identification numbers, and whether and when the location of the detector may have changed,

*(c) Collection Conditions,* for field equipment, a description of the physical environment under which the data were collected. This may include such things as weather and special events as well as a time-stamp of the actual data collection/observations, which may differ somewhat from the time stamp when the data is received by a center.

*(d) Data Collection Equipment Status and Log,* the working status of the equipment used to collect the data as described by personnel responsible for maintaining the data collection system, which may be summarized in a log file of events affecting system availability, as well as whether the system was on-line or off-line for particular time periods,

*(e) Equipment Self-diagnostics,* the results of tests made internally by data collection equipment,

*(f) Equipment-assigned Edit/quality Check,* error flags assigned by data collection equipment, and

*(g) Equipment Calibration,* items such as calibration date, method, equipment, algorithm, field level processing, and results.

10.16.5.4 *Maintain and Update the Metadata*—To facilitate on-going use and access by stakeholders, the metadata developed for the archive, including the data dictionary and schema of the systems, should be maintained and updated. That enables the overall metadata on each data element to remain current and will improve consistency.

10.16.5.5 *Development of a Data Catalog*—The metadata should be used to establish a formal data catalog that describes the data in the ADMS. The data catalog should be developed in conjunction with the data catalogs of other ITS data sources in accordance with procedures in the National ITS Architecture. This integration is essential since many of the data elements in the ADMS will be derived from data generated from ITS sources.

10.16.6 *Managing the Interfaces of the Archive Data Administrator*—This is a guiding principle that focuses on the of the archive data administrator to carry out responsibilities, including maintaining the metadata. In interfacing with the archive data Administrator and maintaining metadata, a number of specific items need to be considered, including the following ones.

10.16.6.1 *Follow Agreed Upon Architectures*—To accommodate existing ITS activities as well as the incremental deployment of new ITS, the systems to support the ADMS should use agreed upon architectures as developed at the state, regional, or local levels. For example, there may be cost efficiencies and common design features that would allow easier access and use of the archived data and information.

10.16.6.2 *Responsibilities of the Archived Data Administrator*—The overall responsibility of the archived data administrator is to manage the archive on a day-to-day basis and administer the on-going and long-term efficiency and effectiveness of the ADMS. This includes establishing user authentication controls for the archive, sending data into the archive, and helping users retrieve appropriate information.

10.16.6.3 *Provide for User and Customer Feedback*—The usability of the data interfaces and the understandability of the metadata by a broad range of users is an important concern during the maintenance of the ADMS by the Archive Data Administrator. Various customer-oriented feedback processes should be provided for to find out how users perceive the ease and effectiveness in using the ADMS in general, as well as particular types of data sources or, or both information contained in the ADMS.

10.16.7 *Interact with Other Archives and Monitor Standards*—This guiding principle should help the archive data administrator with the exchange and sharing of data and information as well as the maintaining of a data archive that would facilitate use by a wide range of users. Oftentimes the sharing may not be reciprocal in that some archives may receive information from one archive but share more information with a different one. The coordinated sharing of data and information with other archives can support analysis, data fusion, and data mining of archived data and information across functionally dispersed archives. In coordinating with other archives and standards, a number of specific items need to be considered, including the following ones.

10.16.7.1 *Maintain Awareness of the Development and Applicability of Other Standards*—Monitoring other ITS standards efforts is crucial because most archived data originate at sources not under the direct control of ADUS stakeholders. Changes in such parallel activities may be able to provide opportunities for the more effective or efficient management of an ADMS.

10.16.7.2 *Temporal and Spatial Standards are Critical*—Two key aspects associated with most data generated by ITS activities is where the data was observed and what period of time the data covers. Thus monitoring other standards for spatial and temporal referencing is particularly important for an ADMS. That is especially the case when data and information is being gathered from many sources and systems that can cover a wide area on an extensive set of transportation networks over long time periods.

10.16.7.3 *Protection of Privacy*—While not a standard per se, the protection of privacy is an important consideration in the coordination with other archives and standards. The privacy of individuals and firms must be maintained, except in cases where an organization explicitly collects data for other purposes, such as an accident investigation associated with incident management activities, enforcement of traffic laws, regulations related to commercial vehicle operations, and in toll or transit fare card administration. In those and similar situations, the organization collecting data and information about individuals or firms, or both, are obligated to protect their privacy. However, that organization should be able to coordinate the sharing of appropriate data and aggregated information

with an ADMS, if that data and information meets the needs and requirements for being archived and used by various users of that ADMS.

10.16.7.4 *Follow ITS America Privacy Principles*—Permanent or temporary storage of data within the systems to support the ADUS should preclude the possibility of identifying or tracking individual citizens and should follow the ITS privacy principles developed by ITS America.

10.16.7.5 *Rely Upon Surrogate Identifiers*—Unique system-developed identifiers may be assigned to stored data that do not distinguish individuals. Public domain identifiers, such as Social Security Nos. and license plate numbers, should not be tagged or cross-linked with the stored data.

10.16.8 *Process User Requests for Data and Information*—This guiding principle focuses on supporting a full cross-section of stakeholders and users, including potential users of the ADMS. In processing user requests for data and information, a number of specific items need to be considered, including the following ones.

10.16.8.1 *Provide for Broad Access to the Data and Information*—Develop the data archiving system in a way that permits ordinary users with typical devices available to users to access and analyze the data. To the extent practical, major parts, if not all of the data archives, should be oriented more toward typical computer users if the archived data and information is be most effectively retrieved and utilized by the agencies that need the data. Automated or on request services ought to be available from the archiving system to provide for basic query mechanisms as well as some typical and frequently requested summaries. However, the archived data administrator may want to establish *communities of users* so that appropriate system security procedures can be established and followed.

10.16.8.2 *Use Widely Available Media*—One way to foster retrieval of data and information from an ADMS is to provide access to and distribution of archived data and information through the Internet, CD-ROMs, or other similar widely available media such as DVDs.

10.16.8.3 *Include Appropriate Metadata with the Response to the Request for Data*—As a standard practice, users requesting data or information should automatically get the appropriate metadata about their requested data or information, perhaps similar to the message sets that are used to facilitate and document center-to-center communications.

10.16.8.4 *Process On-Demand Archive Requests*—The ability to process on-demand requests will be important to serving the needs of infrequent users as well as frequent users who have a particular deadline requiring the needed data or information. In processing on-demand requests, a number of specific items need to be considered. Potential users of the archived data should be provided reasonable access to the archived data either through enabling them to effectively search and summarize appropriate data and information, or by the data administrator providing a service to them that responds in a timely manner to their request.

10.16.9 *Support Analysis of the Archived Data*—This is a guiding principle that helps to provide users with the ability to perform activities such as data mining, data fusion, summarizations, aggregations, and various types of transformations. In supporting analysis of the archived data, a number of specific items need to be considered, including the following ones.

10.16.9.1 *Facilitate Retrieval of Data and Information*—Given the potentially very large and extensive nature of the data, the archive data administrator should support systems that enable and facilitate analysis and summaries of the archived data. This guide views facilitating the retrieval of archived data and information from the archive as a core function of the ADMS, which includes supporting analysis of the archived data. That may include *(a)* developing pre-coded or standardized queries *(b)* providing for ad hoc queries and data searches and *(c)* structuring data so it is readily available in ways that typical users may want to review it, such as by time sequence or periods, geographic and political boundaries, and functional class categories. In addition, a user should be able to retrieve a catalog of the data and information that is available, and then request the desired format and data.

10.16.9.2 *Feedback on the Operations of the Detectors and Sensors*—Use of the ADMS by operating agencies to perform diagnostic reviews of the reliability and operational availability of specific detectors and field equipment should be encouraged with provision for analysis.

10.16.10 *Prepare Data for Periodic Government Reporting Systems*—A major use of ITS-generated data is to support periodic updates to government reporting systems, that is, traditional transportation information systems. Government reporting systems include those maintained by the U.S. DOT as well as many state and local systems. A few examples at the federal level include: HPMS, FARS, National Transit Database, MCMIS, and HMIRS. ADMS developers are referred to a recent report by FHWA for more guidance. In managing an ADMS, a number of specific items related to government reporting systems should be considered including the following ones.

10.16.10.1 *Recognize Potential Links to Governmental Reporting System Data Needs*—The data that reside in the ADMS should be scrutinized for their potential to be linked to government reporting systems. This involves a detailed comparison of the data dictionaries to reconcile basic definitions and valid values. Some data elements will be defined in exactly the same way in both systems.

10.16.10.2 *Prepare Translation Techniques When Appropriate*—For data elements that are similar but differ in their details (for example, different sets of valid values), translation schemes or *cross-walks* may be developed to bridge the gap between the ADMS and government reporting systems. Exercise care to maintain the integrity of the data in government reporting systems by limiting the amount of assumptions necessary to perform the translation. In fact, ADMS developers should make a frank assessment of whether the quality of the data in government reporting systems will suffer due to a

particular translation scheme. In some cases, this may mean forgoing the translation altogether.

10.16.10.3 *When Feasible, Develop Automated Approaches for Report Updating*—Once common data have been defined, an automated process of updating government reporting systems should be developed.

10.16.11 *Quality Matters*—Approaches to the archiving and retrieval of ITS generated data should be concerned with the quality of the data and information going into the ADMS as well as the summaries, analyses, and other information being produced by the ADMS. Data quality management is a basic responsibility of the archived data administrator that may also be shared with data providers. For example, the archived data administrator may be primarily responsible for monitoring, evaluating, and reporting incoming data quality, whereas data providers may be primarily responsible for the completeness and accuracy of data from field equipment. General principles of effective data management encourage feedback loops on concerns for quality to original sources at the earliest time possible. The ADMS and the operation of it need to be part of the overall solution to the total quality management of organizations. An effectively working ADMS can help an overall organization better achieve overall goals for quality. In the past, data quality has not been a large concern for ITS operators: their applications only required the identification of major traffic problems. In contrast, many archived applications require much greater resolution in the quality of the data. However, as operational applications become more sophisticated, their quality requirements will increase. Therefore, a focus on providing quality data will enhance both archived and real-time applications of the data.

10.16.12 *Funding and Financing Considerations for Administering an ADMS*—All this work to provide for an ADMS requires appropriate levels of financing and funding sources. This guide does not prescribe any particular institutional arrangement for financing or funding. However, it does recognize that issues that may be addressed in managing the archive are *(a)* to what degree should key stakeholders participate financially in the funding of the operation of the ADMS or *(b)* should all users participate somehow in the cost arrangements or *(c)* should just the organization sponsoring the ADMS bear all of the costs. Part of those considerations could also be the marginal costs to the data providers to possibly collect additional data for metadata purposes at the ADMS, but that they may not need for the actual operations. Each local application of an ADMS will need to evaluate and prioritize among options such as those, as well as concerns of local importance. Agreeing to necessary and sufficient financing and particular funding arrangements is an essential component of establishing an ADMS.

10.16.13 *Anticipating Revising the Standard Guide*—There is a range of considerations covered by this guide that reflect the evolving state-of-the-art and state-of-the-practice. Consideration will be given to making revisions to this guide in the foreseeable future. That is because first, it is expected that rapidly changing technology may provide better tools, lower costs, and lower other barriers that may lead to the opportunity for different approaches to the archiving and retrieval of ITS-generated data. However, a review of the need for revision would not be likely until after the first few specifications have been developed. ASTM International procedures require that an adopted standard be reviewed every five years and subsequently reapproved, revised, or withdrawn. ASTM International reserves the right to automatically withdraw a standard after eight years of inactivity. If in the interim any interested individual or organization has reason to believe that the time has come to seek revision to this guide, they should contact either ASTM or the Chair of the Subcommittee on the Archived Data User Service.

10.16.14 *General Activities Outside the Scope of this Guide*—There are a number of considerations that are not covered by this guide, that while necessary for a functioning archiving and retrieval system, are more the province of fairly available off-the-shelf applications of technology, or are covered elsewhere by other ITS related architecture or standards activities. Included are items such as *(1)* database management software, practices, design and applications and *(2)* communication protocols for entering data to the archive or retrieving data or information from it. While some items are adequately and directly covered by the National ITS Architecture, such as type of archiving approaches to use, that is, data mart, data warehouse, virtual data warehouses, some discussion of them are given in this guide in Appendix X2 for ease of reference for the readers. It is also anticipated that there will be a separate guide that will be oriented to items such as *(a)* applying this standard to contracting specifications and designs, *(b)* implementations and operations using the standard and *(c)* case study summaries of using the ADUS requirements from the National ITS Architecture and related standards.

## 11. Keywords

11.1 archived data; archived data management systems (ADMS); archived data user service (ADUS); data archiving; ITS data archiving; metadata; retrieving ITS data; warehousing ITS data

# APPENDIXES

**(Nonmandatory Information)**

## X1. RELATIONSHIPS TO OTHER ACTIVITIES AND STANDARDS

### X1.1 Relationship to the Strategic Plan for the Development of ADUS Standards

X1.1.1 The document prepared by Cambridge Systematics (**7**) for the Federal Highway Administration (FHWA) anticipated that this Guide would become its initial key element. Such strategic plans have been developed for each standard effort associated with the National ITS Architecture. That strategic plan creates the framework for the development of standards related to the implementation of the ADUS, including identifying the general types of standards that are required and relationships with other standards.

X1.1.2 The strategic plan serves as the basis for the activities of ASTM Subcommittee E 17.54. It also recommended that ASTM International and Subcommittee E 17.54 should seek the support of the U.S. DOT and the other standards development organizations in accelerating the development of standards related to ADUS. A series or *waves* of ADUS standards was envisioned in the strategic plan where, for each category of ADUS standard, two types of standards would be prepared, *(1)* guidelines for processing, storage, and retrieval, and *(2)* specifications for data dictionaries. The Strategic Plan also anticipated the development of a series of project plans that would seek specific support for specific time frames.

X1.1.3 The first of the project plans proposed the concurrent development of two ADUS related standards, this guide and a companion *Specification for Archiving Travel Monitoring Data*, the latter to be the first of a series of standard specifications. The project plan proposal also changed the thrust of the guidelines component for the different types of archiving to be more general and to cover all types of archiving at the same time. Thus, this guide has taken on a broader perspective beyond the archiving of travel monitoring data and is intended to also be a guide for archiving other sources of ITS-generated data. The initial work in developing the companion *Standard Specification for Archiving Travel Monitoring Data*, has taken on a more particular bent so as to focus just on traffic monitoring data and it will be referred to using the term *traffic* rather than *travel* in the remainder of this guide.

### X1.2 Relationship to the Standard Specification for Archiving Traffic Monitoring Data and Other Similar Standard Specifications

X1.2.1 As noted above a parallel standards development process is being done to prepare a Specification for Archiving Traffic Monitoring Data, which will be detailed specification for monitoring data of vehicular traffic flows. That type of data is seen as a subset of a broader class of data related to the monitoring of travel by people. The latter is more complex and difficult than the monitoring of vehicular traffic flows. The specifications for traffic monitoring data are being prepared first primarily because:

X1.2.1.1 Traffic monitoring data are the most common source of ITS-generated data currently available.

X1.2.1.2 Traffic monitoring appears to have the widest range of use in applications.

X1.2.1.3 In various user needs and requirements studies, the potential stakeholder users seem to have a higher degree of interest in traffic monitoring data relative to other sources of ITS-generated data.

X1.2.2 The strategic plan anticipates that a series of specifications will be prepared after the initial one for the archiving of traffic monitoring data. It is anticipated at this time that subsequent specifications will be developed for several other sources of ITS-generated data. While these specifications will probably be developed in the order listed here, subsequent program plans may recommend and result in a somewhat different sequence being followed. Those other sources include: *(1)* transit *(2)* incident and safety *(3)* commercial vehicle related, as well as other sources such as the phasing of signal cycles for traffic control systems.

### X1.3 Relationship to the National ITS Architecture and Data Sources

X1.3.1 This guide is intended to build upon and augment functions developed in the National ITS Architecture (**8**) that defines the functionality of an ADMS and that interfaces with subsystems and *terminators*. Examples of other subsystems include traffic management, transit management, or parking management while examples of terminators include intermodal freight depot, surface transportation weather service, or other data sources. Terminators represent functions that basically take place outside of the National ITS Architecture but which interfaces with one or more subsystems of the National ITS Architecture by sharing data or information, or both, with them. The standards requirements package developed for the ADUS identifies a list of some sixteen sources of archived data, which are grouped in Table X1.1 by whether they come from other subsystems or from terminators.

**TABLE X1.1 Data Sources for the Archived Data User Service Listed by Subsystems and Terminators Associated with the National ITS Architecture**

| Subsystems as Data Sources | Terminators as Data Sources |
|---|---|
| Commercial vehicle administration | Asset management |
| Emergency management | Intermodal freight depot |
| Emissions management | Multimodal transportation service provider |
| Information service provider | Other data sources |
| Maintenance and construction management | Surface transportation weather service |
| Parking management | Weather service |
| Roadway | |
| Toll administration | |
| Traffic management | |
| Transit management | |

X1.3.2 In addition, these data sources provide data that can be grouped into other categories that were used during the development of the ADUS:

X1.3.2.1 Freeway and toll collection data,
X1.3.2.2 Arterial and parking management data,
X1.3.2.3 Transit and ridesharing data,
X1.3.2.4 Incident management and safety data,
X1.3.2.5 Commercial vehicle operations data,
X1.3.2.6 Environment and weather data, and
X1.3.2.7 Vehicle and passenger data.

X1.3.3 Alternatively, such data sources can also be broken down into many data elements, which should be the data elements being used in specific data flows from either other subsystems or terminators, or both. In developing an ADMS, there may be other data sources and data elements that are associated with that system that are above and beyond those identified in the National ITS Architecture. In work done for examining the needs and requirements for establishing an ADMS for Maryland's CHART system, Winick (**9**), adapted tables of data sources prepared by Margiotta (**4**). That work identified some main 30 data sources and over 110 data elements. Table X1.2 presents a portion of one of those tables for just the freeway and toll collection data categories, where the data elements shown have common names rather than those used in the National ITS Architecture or other standards, such as the Standard for Functional Level Traffic Management Data Dictionary.

**TABLE X1.2 Example of Potential Data Sources for Archiving Freeway and Toll Collection Data Types**

| ITS Data Source | Primary Data Elements |
|---|---|
| Freeway and Toll Collection | |
| Freeway traffic flow surveillance data and | Volume |
| | Speed |
| | Lane occupancy |
| | Origin/destination |
| Automated traffic recorders | Vehicle classification |
| | Vehicle weight |
| Ramp meter and traffic signal preemption | Time of preemption |
| | Preemption duration |
| Ramp meter and traffic signal cycle lengths | Begin time |
| | End time |
| | Cycle length |
| Visual and video surveillance data | Specific time |
| | Specific location |
| | Queue length |
| | Vehicle trajectories |
| | Vehicle classification |
| | Vehicle occupancy |
| Vehicle counts from electronic toll collection | Time interval/lane type |
| | Vehicle counts/classes |
| | Pricing in effect |
| TMC generated traffic flow metrics (forecasted or transformed data) | Link congestion indices |
| | Stops/delay estimates |

## X1.4 Relationship to Other ITS Related Standards

X1.4.1 This guide also draws upon relationships with other ITS Standards; a number of the key ones have already undergone initial development. Some of these other standards have begun processes of refinement and updating to reflect circumstances such as *(1)* the dynamic nature of the ITS program *(2)* changing information technology *(3)* changes in the technologies that underlie ITS activities or *(4)* changing institutional relationships or perceived needs, such as an increased emphasis on security as an aftermath of the tragedy of September 11, 2001.

X1.4.2 This guide and the subsequent specifications are intended to build upon, coordinate with, and interact with the on-going development of other ITS related standards. In particular, various definitions of data elements involved in data flows among subsystems and terminators, frameworks identifying interrelationships among data elements, and communication protocols for interoperable uses of the data have been among the important activities associated with the other standards.

X1.4.3 However, other standards have not necessarily dealt with concerns such as metadata—data about the data—or with what this guide and the associated material for the *Standard Specification for Archiving Traffic Monitoring Data*, are now terming as *data processing documentation*. It is expected that revisions to the other ITS related standards will begin to incorporate or refer to these aspects of the ADUS related standards and these types of concepts in the on-going updating and refinement of their standards.

# X2. HISTORY OF SELECTIVE ARCHIVING

## X2.1 Archiving and Retrieval Process Prior to Establishment of the ADUS

X2.1.1 While not very widespread, there have been many examples of the selective archiving of ITS-generated data prior to the establishment of the ADUS in 1999. Turner et.al. (**10**) present a fairly detailed summary of the state-of-the-practice prior to the establishment of the ADUS that gives case studies of ITS-data archiving practices for some sixteen different agencies or locations, or both. Turner (**11**) also reports that implementation and analysis of operations data archives in the U.S. have been somewhat limited as a typical practice, despite the fact that several early pioneers have been archiving and analyzing operations data from traffic control sensors and detectors for at least twenty years. He cites several examples of such early pioneering efforts at archiving ITS-generated data as well as some of the applications to which the retrieved data and information has been applied. He also identifies several reasons why the archiving and analysis of operations data are not more widely implemented, which include the following:

X2.1.1.1 Some operating workgroups/agencies are focused on crisis management and do not see the utility of anything other than real-time data.

X2.1.1.2 Operating workgroups/agencies see data archiving as the responsibility of planning workgroups/agencies, who they feel are the primary beneficiary of archived data.

X2.1.1.3 Planning workgroups/agencies are typically not involved in the operational data collection, thus they are not aware of or are not comfortable with the quality of the data to be archived or the quantities of data that need to be dealt with.

X2.1.1.4 Data archiving was not considered an essential component of traffic control and management software during system development.

X2.1.1.5 There may be data ownership, maintenance, or control issues that cannot be resolved between workgroups/agencies that collect data and archive data, including concerns for security and data privacy.

X2.1.2 It is noted that the establishment of an ADMS can address many of these issues. For example, with an ADMS in operation it may be able to take on the responsibility for data maintenance and control of historic data relieving the TMC operating group of that burden. Turner continues his description of experiences with the retrieval component of archiving systems and notes that in some cases even when ITS-generated data has been archived, the data or information from the system have not been widely distributed or analyzed for several reasons, which include the following:

X2.1.2.1 Proprietary data formats and data storage devices (for example, magnetic tape cartridges) hinder archived data distribution.

X2.1.2.2 Distributing archived data to users sometimes places an unreasonable burden on operations personnel (if the distribution is not automated).

X2.1.2.3 Before the Internet and CD technology arrived in the early 1990s, it was difficult to distribute the large quantities of data that were typically stored in proprietary data formats and data storage devices (for example, magnetic tape cartridges).

## X2.2 Archiving and Retrieval Process After Establishment of the ADUS

X2.2.1 During the late 1990s, a series of activities raised the awareness of the opportunities and importance of establishing or improving the archiving and retrieval of ITS-generated data. Those activities lead to the creation of the ADUS as a new user service for the National ITS Architecture, and then to the subsequent amendment to the National ITS Architecture that included the definition of an ADMS and its interfaces. The effort to develop standards for the ADUS was initiated at about the same time. However, this has resulted in an interim period of heightened awareness of the need for archiving and retrieval of ITS-generated data but a lack of standards to help guide organizations interested in establishing or refining archiving and retrieval systems.

## X2.3 Archiving and Retrieval Process Prior to Establishing these Standards

X2.3.1 During this interim time period, a number of organizations throughout the nation have been proceeding to plan, design, implement, and operate a new or refined ADMS and interfaces to meet their own purposes and objectives. Generally speaking, such organizations have relied on a combination of the material prepared for *(1)* requirements established in the ADUS and *(2)* an amendment to the National ITS Architecture that can be used as a framework, general guidance, or specific direction in establishing or refining an ADMS.

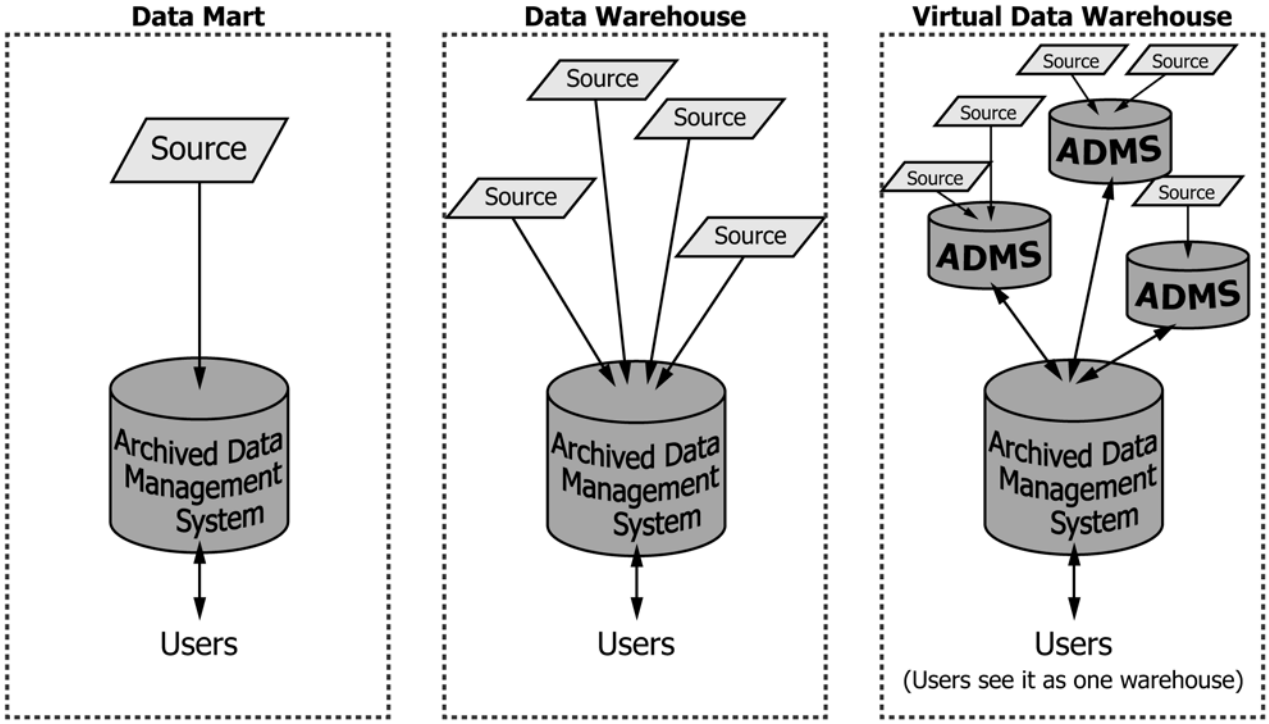
X2.3.2 For instance, work performed by Maki et.al. (**12**) for the Nevada Department of Transportation for the Freeway and Arterial System of Transportation (FAST) for the Las Vegas area is an example that was funded by a grant from the U.S. DOT. Maki et. al. followed a needs and requirements process and identified components of the system design including *(a)* data collection and storage *(b)* archival period *(c)* data retrieval *(d)* archive database description and *(e)* the ADUS system components. In the data retrieval activity they used data warehousing, data mining, and meta-data concepts and approaches.

## X2.4 Archiving Concepts Included in the National ITS Architecture

X2.4.1 The theory of operations of the National ITS Architecture presents a broad level narrative and technical description of the overall architecture to serve transportation professionals involved in ITS planning or implementation, or both. The theory of operations uses the concept of market packages to provide an accessible, deployment-oriented perspective to the National ITS Architecture. Market packages collect together one or more equipment packages that need to work in conjunction with each other to deliver a given transportation service. In particular, three archived data management market packages are identified: ITS Data Mart, ITS Data Warehouse, and ITS Virtual Data Warehouse.

X2.4.2 The similarities and differences among these three ITS market packages are schematically illustrated in a simplified fashion in Fig. X2.1. One emphasis of the National ITS Architecture is to partition function into subsystems and clearly identify the information flowing into and out of those subsystems. The three market packages were offered to represent different approaches to archiving and retrieval of data, and can be further modified as needed by deployments. This guide can be used to complement and build upon those market packages by being more concerned with the quality of the content of the data being stored in the archives and specification of particular approaches for metadata and retrieval. The discussion in the theory of operations also uses an analogy to traditional library systems to help distinguish among the three market packages. The purpose of the ADMS is to provide information to the archive users much the same as a library serves the library users. Using this analogy, differences between the market packages can be seen as the source of information for the archive and where and in what media/format that information is physically kept.

X2.4.3 *ITS Data Mart,* a National ITS Architecture market package that houses data collected and owned by a single agency, private sector provider, research institution, or other organization, typically including data covering a single transportation mode and jurisdiction collected from an operational data source; and providing basic data quality, privacy, metadata management, and general query and report access to users.




NOTE 1—*Based on a diagram by Paul Marrero*

**FIG. X2.1 Similarities and Differences among the Three ITS Market Packages for Data Archiving**

X2.4.3.1 An ITS Data Mart is an archiving system that collects and archives operational data from one source, such as traffic volume, speed, and lane occupancy at detector stations along a toll road or at a major bridge or tunnel that may be operated by a single purpose authority. It is analogous to a library whose collection is acquired from a single publisher. It is anticipated that an ITS Data Mart market package would be used in situations where a particular subsystem or terminator wishes to archive its own operational data, and then make that resulting archived data available to external users or government reporting systems. This is more like a private library or special purpose collection that may or may not be open to the general public at the discretion of the library administrator.

X2.4.4 *ITS Data Warehouse,* A National ITS Architecture market package that includes all the data collection and management capabilities provided by a data mart plus enhanced functionality and interface definitions following collection and maintenance of data sources from multiple agencies and across jurisdictional boundaries such as additional transformations, metadata management, on-line analysis, and data mining.

X2.4.4.1 An ITS Data Warehouse can be composed of physically distributed ITS archives, each locally managed, exchanged transportation related data. It is analogous to a library whose collection is acquired from multiple publishers. It is more like a public library that has many collections and serves many customers, generally from a specific region or area. Such an ITS Data Warehouse might be administered by a regional agency, such as a metropolitan planning organization or by a regional or central office of the state transportation department. There may be some efficiencies of physically having the ITS data warehouse at one location, for example, a regional office of the state transportation department may operate 24 h per day, seven days per week. However, the decision as to the physical location is often more of an institutional issue rather than a technical issue and there appears to be no universally best answer as to where to house an ITS data warehouse that serves a particular area. In this market package, much like the public library, more attention should be paid by the librarian to how the information is stored to ease access to users to find it, usually on their own, while still protecting the collection by setting up various access permissions.

X2.4.5 *ITS Virtual Data Warehouse,* A National ITS Architecture market package that includes all the features and functionality of a data warehouse, but with enhanced functionality and interfacing that extends interoperability among physically distributed ITS data sources that are each locally managed, that is, using remote data access methods over a public network.

X2.4.5.1 An ITS Virtual Data Warehouse is constructed as physically distributed ITS archives that exchange transportation related data even though the warehouse is not directly associated with an ITS center. An example might be an archiving system set-up at a research university that may exchange data with ITS data archives of several traffic management centers. An ITS virtual data warehouse is analogous to a library whose collection is acquired from other libraries or to a central library system that often physically moves items available in one location to another to serve the needs of a particular user. Benefits of this market package are that it can offer more of a *one-stop-shopping* experience for users as well

as make it more convenient for users to get information in which they are interested. In many cases the users may be part of a much larger community of interest and not be directly associated with the region of the source archive. The availability of useful and effective metadata, like a good card catalog that adequately describes what is available, becomes even more important to users of such an approach to ITS data archiving.

## REFERENCES


**(1)** Smith, Brian L., *Database Management Systems: What Transportation Professionals Need to Know*, University of Virginia, 2001.

**(2)** Tannenbaum, A., "Metadata Solutions: Using Metamodels, Repositories, XML, and Enterprise Portals to Generate Information on Demand," Addison-Wesley, 2002.

**(3)** Turner, S. M., "Guidelines for Developing ITS Data Archiving Systems," Texas Transportation Institute, Texas A&M University Systems, September 2001, pp. 7-8.

**(4)** Margiotta, R., "ITS as a Data Resource: Preliminary Requirements for a User Service," Office of Highway Information Management, FHWA, April 1998, pp 3-5.

**(5)** Winick, R. M., "An Assessment of Architecture Approaches for Data Integration and Archiving," Maryland Department of Transportation, State Highway Administration, November 1999, Addendum of Exhibits, p 14.

**(6)** Turner, S. M., "Guidelines for Developing ITS Data Archiving Systems," Texas Transportation Institute, Texas A&M University System, September 2001, pp 17-18.

**(7)** Cambridge Systematics, Inc., "Strategic Plan for the Development of ADUS Standards," FHWA, May 2000.

**(8)** National ITS Architecture Team, "The National ITS Architecture: A Framework for Integrated Transportation into the 21st Century, Version 4.0." FHWA, April 2002, CD-ROM and available online at www.iteris.com/itsarch.

**(9)** Winick, R. M., "An Assessment of Architecture Approaches for Data Integration and Archiving," Maryland Department of Transportation, State Highway Administration, November 1999, p 5.

**(10)** Turner, S. M., Eisele, W. L., Gajewski, B. J., Alpert, L. P., and Benz, R. J., "ITS Data Archiving: Case Study Analyses of San Antonio TransGuide Data," Texas Transportation Institute, Texas A&M University System, August 1999, pp 5-54.

**(11)** Turner, S. M., "Guidelines for Developing ITS Data Archiving Systems," Texas Transportation Institute, Texas A&M University System, September 2001, pp 8-10.

**(12)** Maki, K., Fares, G., and Osborne, D., "Archived Data User Service," Kimley-Horn and Assocs., Las Vegas, NV, June 2000.