3. TECHNICAL AND INSTITUTIONAL ISSUES FOR IMPLEMENTATION

3.1 Introduction

Beyond defining the Archived Data User Service for inclusion in the National ITS Architecture, it is critical to outline the serious technical and institutional issues affecting its implementation. In fact, since most of the relevant data flows have already been defined in the National ITS Architecture, overcoming the technical and institutional issues is the major challenge for implementation. The institutional issues are particularly onerous since the beneficiaries of the data are primarily personnel not directly involved in ITS operations, leading to such questions as who bears the cost, who maintains the system, and who owns the data. This Chapter identifies these issues and the recommendations in Chapter 4 have been made with them in mind. However, complete guidance for resolving these issues will require case studies of current efforts as well as future efforts specifically targeted to the use of data generated by ITS for multiple purposes.

3.2 Technical and Institutional Issues

3.2.1 Development, Operation, and Maintenance Costs

Operating agencies such as state DOTs and local traffic engineering agencies are the purveyors of ITS. This means that funding for ITS comes from their budgets and unless they are convinced of the value of archiving the data, they may balk at including the archiving function in their systems. As an information management system, an implemented archived data system requires data base administration, backup procedures, routine operation of quality control and summarization programs, responding to special users, maintaining existing code, and developing new code for new applications. Even if the data "owners" are convinced of the data's value, staff resources for operation and maintenance may be slim. Moreover, because there is little precedent in the field, the costs of building, operating, and maintaining an archival system are largely unknown. Finally, the sharing of costs among stakeholder groups, while in theory could help defray costs to any one group, is a problem that needs to be worked out locally. It is therefore crucial that these costs be explicitly addressed in future funding of ITS deployments. Local agencies should also be free to pursue innovative approaches to paying for archived data systems, including the use of non-ITS funds (e.g., state and local planning allocations).

3.2.2 System Access

For the data to be of use to stakeholders, they must be easily accessible. This is largely a technical issue of how to design the communications and interconnects in the final system. However, access to the data by remote users must be accounted for in the system design. Mechanisms for data distribution depend on the arrangements made for access; they can range from free and open access to the data via direct communications or the Internet to periodic release of the data on CD-ROMs or other media. A related issue is the degree to which unverified data or data that have failed quality control checks will be released to certain users.

3.2.3 Ownership

If ITS operating agencies are responsible for the systems that archive data, and if those data are not really used or are not seen as having value by them, there is a concern that data quality and completeness will suffer. Ownership may also affect access to the data if the "owners" are sensitive to releasing data to

others (even if the data are seen as high quality). In some cases, data in a central repository may have multiple owners, further complicating the institutional arrangements for data sharing. For example, freeway traffic surveillance data may be "owned" by the state DOT and transit ridership data may be "owned" by the local transit authority. Therefore, issues of ownership need to be addressed prior to field implementation.

3.2.4 Data Quality

Even if the issue of ownership does not affect data quality, development of quality control (QC) procedures for data generated by ITS has been limited. Data generated by ITS collected by field surveillance equipment are prone to errors due to equipment malfunctions, calibration "drift", and communication disruptions. Therefore, it is clear that raw data from the field must be subjected to quality control and editing procedures at some point. Two options exist: (1) perform QC and editing before they are sampled, summarized, and stored for later use, and (2) simply store all data as they are received from the field and leave QC up to the individual end users. The first option is preferred because "owners" of the data understand their data and are in the best position to perform QC. At a minimum, the conditions under which the data are collected and possible sources of error must be communicated to end users so they can make their own judgments; this dictates that full disclosure of equipment status and environmental conditions be documented.

An even more complex problem with data quality is the mechanism for correcting problems with data collection equipment, given that QC procedures have identified the problems. In some cases, data may be ancillary for control purposes. In others, the level of accuracy required for control may not be as high as for other applications. For example, for detecting freeway traffic breakdowns for VMS control, operators need to know if speeds are either free flow (around 50 mph) or forced flow (less than 30 mph). However, many models used by transportation planners require greater resolution of speeds, and researchers want speeds at even higher accuracy. The care and maintenance of field data collection equipment is therefore seen as an integral part of archived data systems process and must be explicitly addressed.

Other issues surrounding data quality include: detection procedures for questionable data; categorizing and flagging errors in the data; the use of missing or questionable data in the data aggregation process; geographic coverage of the data; full understanding of what the data truly represent; and the imputation and replacement of questionable data using *post hoc* quality control procedures. The stakeholder groups identified here are not used to reviewing, cleaning, and editing data and they will require guidance on this matter. Because little experience exists for multiple uses of data generated by ITS, quality control is a new facet for data owners as well.

3.2.5 Data Management

Data management issues relate to how much data should be kept and at what level of summarization or sampling should be present. The needs of various stakeholder groups are quite different in this respect. A specific example involving freeway traffic surveillance data was provided as Figure 2.1. It shows how these data, which are reported from field controllers at 20-second intervals, can be used for different purposes depending on the level of aggregation. For example, transportation planners may require roadway surveillance data (volumes and speeds) at 1-hour intervals summarized over all lanes of the facility. On the other hand, researchers investigating microscopic traffic flow may need surveillance data at the smallest possible intervals for individual traffic lanes. The temptation is to save the data at its

lowest level of aggregation so that all stakeholders can be accommodated, but that may not be cost-effective and may discourage use for some applications. Alternative solutions to this problem include specifying different levels of aggregation in the archive system and storing sampled data at the lowest level of aggregation rather than saving it all.

A related data management issue is the degree to which transformed data should reside within the archived data system structure. (See Section 1.2.1 for a definition.) The example shown in Figure 2.1 is not transformed data -- it merely shows how data may be summarized to different levels of aggregation. Because transformed data are usually used for highly specific purposes, specification in the Archived Data User Service may not be cost-effective -- it is more properly a responsibility of the end user. However, some local circumstances may require that certain transformed data be permanently archived.

Another data management issue is how to link data from disparate sources. To some extent, the National ITS Architecture and the various standards development efforts (e.g., NTCIP) will help to provide conventional data definitions. However, a major problem still remains with the development of a common location referencing system. The fundamental problem is how to link the geo-coordinates (latitude/longitude or other x/y system) to points on the transportation network. For systems where the location of the data collection is fixed -- as with loop detectors or AVI (probe) readers -- location referencing is less of a problem than with systems that rely on GPS or other geo-coordinates. Fixed data collection equipment can be easily mapped one time with existing referencing systems (e.g., mileposts, distance to nearest intersection). However, as the use of advanced technologies such as in-vehicle navigation systems become more widespread, the location referencing problem takes on a greater significance. Even though many nonreal-time applications of the data do not require a great level of precision (e.g., locations of traffic counts for planning uses could be accurate to the nearest major intersection), the translation of geo-coordinates to points on the highway network is still highly problematic.

Finally, the mechanism for how the system for maintaining archived data generated by ITS will interface with legacy systems must be resolved. Standards for data definitions and communications can foster the interface but there are still large technical problems to be addressed, especially since many legacy systems were developed using dated data base management systems.

3.2.6 Data and Communications Standards

Standards and protocols are highly important because they foster interoperability of field equipment, aid retrieval of data, and allow data comparisons across areas. Wherever possible, standards for archived data should conform to existing standards for transportation data (e.g., the *Traffic Monitoring Guide*). Standards should include the specification of metadata which not only describe how the data are represented but identify the conditions under which they were collected. Also, many standard development efforts are currently underway that could affect archived data systems. Coordination with these efforts is essential to successful implementation.

3.2.7 Privacy Concerns

Some forms of data generated by ITS have the potential for violating individuals' privacy. This becomes a more sensitive issue if the data, which now may disappear after immediate use by Traffic Management operators, are permanently archived. Examples include video surveillance and probe vehicle data. In the case of video surveillance data, image processing can be used to extract key information (e.g., volumes,

speeds, vehicle classifications) without concern for the identify of individual travelers. Probe data can either be summarized over time without regard to individual vehicles or a "pseudo" identification tag can be assigned prior to permanent storage. In no case should a cross-link between the "pseudo" tag and the actual identity be maintained.

3.2.8 Data Analysis

Although not strictly related to implementing the Archived Data User Service, data analysis is a major issue affecting its success. Operations, planning, maintenance, and administrative staffs may not have the resources to manipulate the data. Even if they did, there are no guidelines for how the data could be used (e.g., spot speeds from loop detectors). Lack of a common geographical referencing system also hinders analysis of the data. Data analysis can be facilitated to some degree by effective data management and access. For example, if the data are pre-screened for accuracy and summarized to levels commonly used for traditional analyses, these steps can be avoided by end users. Specifically, less sophisticated software can be used (e.g., conventional spreadsheets rather than expensive statistical packages) and working storage requirements for end users can be greatly reduced.

3.2.9 Coordination With Other Data Collection Efforts

Data generated by ITS can be used to supplement or replace existing data collection programs, such as statewide traffic monitoring, ISTEA-related management systems, and HPMS. The most effective structure for an archival system would include methods of integrating the various data collection efforts in use throughout stakeholder agencies. In this way, duplication of effort can be avoided and data programs will become more cost-effective. Careful review of what ITS can provide to existing programs is in order -- stakeholders must be aware that ITS may not have the level of geographic coverage required to meet all of their needs. For example, only instrumented highway segments can report traffic surveillance data; there is still a need to continue traffic monitoring on other segments.

3.2.10 Liability

Litigation has been a large issue to the transportation profession for many years. Archiving some forms of data generated by ITS may actually create problems for transportation agencies involved in lawsuits (e.g., video surveillance data). Public records statutes may force release of archived data.

3.2.11 Confidentiality of Privately Collected Data

Particularly within CVO applications, some data generated by ITS are collected by private companies for internal use. There may be reluctance to share these data on the grounds it might give a company's competitors confidential information on operations or that the government will take advantage of increased opportunities for regulation and taxation (e.g., weight-distance taxes). Therefore, any specification of data for the Archived Data User Service from private sources should conform to the "Fair Information Principles for ITS/CVO" now being developed by ITS America.

3.2.12 Incremental and Uncoordinated ITS Deployments

ITS deployment in metropolitan areas evolves as part of many separate projects over time, rather than being developed as part of a single, coordinated project. Multiple agencies are usually involved in ITS deployment and coordination with other existing or ongoing ITS projects is not guaranteed. Therefore, the nature of the ITS deployment process will make it difficult to coordinate the archival of data for later use. For this reason, inclusion of the Archived Data User Service in the National ITS Architecture is crucial because of its importance in guiding local ITS development.

3.2.13 Retrofitting vs. New Development of Systems

Once ITS systems are deployed, it is more arduous to retrofit an archiving function rather than building it directly as part of initial system development. The fact that many ITS projects have and are being deployed without having the advantage of the National ITS Architecture means that retrofitting will have to be accommodated if the Archived Data User Service is to be realized. The process may be aided as ITS integration is promoted -- the infrastructure for achieving ITS integration (e.g., central servers, communication "backbones") can serve as a basis for implementing the Archived Data User Service.

3.2.14 Data Not Defined by the National ITS Architecture

Although the National ITS Architecture is the basis for the Archived Data User Service, ITS have evolved without the benefit of its guidance. Further, future ITS may include features not foreseen by the Architecture. Examples of data not currently defined by the National ITS Architecture that may be used for multiple applications include: door-to-door origin/destination patterns and tracking intermodal freight containers. Resolution of these matters is vital: either the National ITS Architecture must be amended to include these new data flows or it must be flexible enough to allow inclusion of them in field-implemented systems.

3.2.15 Conformance With Metric Conversion Standards

The National ITS Architecture does not now address metrication; units are specified in the English system. The degree to which metrication will influence the construction of data elements in the National ITS Architecture is still an open issue. Similarly, metrication must be resolved for the Archived Data User Service as well.

3.2.16 Training and Outreach

There will be a continuing need to train personnel not familiar with ITS in the fundamental concepts and methods used in ITS technologies. This will foster not only the development of the Archived Data User Service but the "main streaming" of ITS in general. In a similar fashion, training and outreach activities for ITS operators should be geared to understanding the needs and duties of other transportation agencies.