

Transportation Metadata: Role of Data and Information Technology Section

1. Introduction

This report is the collaborative product of a workgroup convened under the auspices of Data and Information Technology Section (ABJ00) of the Transportation Research Board (TRB). The Library and Information Science Committee (ABG40) from the Research and Education Section (ABG00) was a full partner. At the Section meeting on January 12, 2005, data committee chairs agreed to examine options for a Section-level focus on metadata. Shortly thereafter, this workgroup was convened. It was comprised of eight members representing the variety of providers, users, and uses of metadata in the larger transportation community (as reflected in the committees of the Data Section and related TRB-committees). The workgroup was charged with recommending a plan of action for the Section. Which type of institutional mechanism, under the Section umbrella, should be given authority for dealing with metadata issues? What should be this entity's mission? This document provides a brief introduction to the definition and value of metadata, identifies challenges in addressing metadata issues, and presents options, as well as recommendations, for the Section focus on metadata.

Why a "Section" Focus on Metadata?

The mission of TRB is to promote innovation and progress in transportation through research, and a key element of its mission is information sharing. Metadata – information about data or datasets -- is one of the keys to successful information sharing. For the last several years, a subcommittee of the Urban Transportation Data and Information Systems Committee (ABJ30) focused on metadata. This subcommittee was ended in 2004 because its membership felt that metadata issues were far bigger than the focus of a single data committee. There is a need to pursue a different mechanism within TRB to provide a focal point for this important topic and involve a broader set of participants.

This workgroup recognizes that, at this moment, there are many different metadata initiatives in North America, and beyond, that are at different stages of fruition and at different levels of structural complexity. The current situation invites the danger of both overlaps and gaps – not to mention the danger of producing incompatible standards and structures building upon these. We believe that the TRB Data Section can and should play a leadership role in fostering consistent and coordinated approaches to transportation metadata efforts. This is very much in keeping with this Section's mission and focus on the data needs of the transportation community.

Key Recommendations

But it is not the role of the Data Section to develop standards. The Data Section has an important role to play in raising awareness of the value of transportation metadata, creating a shared vision of core metadata elements, facilitating the development transportation metadata standards, and shepherding the use of metadata among data producers, data archives, and data users.

The Data Section needs to focus on metadata *now*. For this reason, a staged approach should be used to institutionalize a metadata "group" within TRB. The recommended approach starts with the formation of a Section subcommittee. This subcommittee should be charged with pursuing activities that define a metadata research agenda. This agenda would provide the foundation for the development of a detailed proposal for Task Force formation. If this Task Force proposal is accepted by TRB, the Task Force should focus on strategic goals in areas of organization, standards, outreach, and technical development.

2. Transportation Metadata

This report section provides a cursory briefing on the need for the development of metadata standards and their potential benefits to transportation practice. In general, transportation metadata allow a more

efficient use of data from different sources in different places at different times. This facility is critical as the volume of data producers and data users, as well as the range of types of data, multiply.

Definition and Types

Metadata is “data about data”. These data may be text, numeric, spatial, verbal or visual. Metadata falls generally into four classes -- descriptive, administrative, structural, and rights – as noted below:

- Descriptive: title, creator, keywords, classification, description, publisher, date, type, format, source, language, and coverage (geographic and temporal).
- Administrative Dataset Information: Currency, accuracy, completeness, contact information for persons expert in the details of the data set and its creation and maintenance.
- Data Structure Information: data entities and relationships, data type, file size, software requirements, variable names, coding methods, validation rules.
- Rights: Copyright ownership, privacy, use privileges, restrictions, disclosure, and commercial factors.

Value of Metadata to the Transportation Community

Why use metadata? Using metadata is both cost efficient and time efficient.

- Metadata maintains a transportation organization’s investment in information resources – Metadata provides automated, searchable access to information resources so that employees or clients can find the information they need with minimal time and effort. (*Example: search for GIS coverages in a particular format for a particular area*)
- Metadata provides information necessary for data to be understood and interpreted by a wide range of users -- Thus, metadata is particularly important when the data users are physically or administratively separated from the data producers. Metadata also reduces the workload associated with answering the same questions from different users about the origin, transformation, and character of the data. (*Example: traffic datasets posted on a state DOT website*)
- Metadata enables data to be discovered and used to its full potential – Metadata may also provide information about intended or planned uses (as well as limitations), which can assist the data users in realizing the full potential of the data. (*Example: crash records datasets for different time periods, and using different criteria for what crashes are included*)
- Metadata facilitates the operation of federated database systems (i.e., distributed holdings) -- Metadata enables data to be centrally stored with searchable interfaces; thereby, providing a single access point for varied types of data resources. (*Example: metadata repository that enables access and content details for databases held by data owners who agree to have their data listed*)
- Metadata extends the efficiency and reliability of discovery and utilization processes – By embedding metadata in statistical and modeling elements and clarifying between raw and derived or estimated data elements, the outputs from hypothesis tests, model tests, or other statistical analyses can be more readily compared, evaluated and disseminated (*Example: compilation of information which enables efficient and accurate use of travel survey datasets, e.g., source agency of data, data collected, data quality, caveats, and other descriptive information*)

In summary, metadata can allow data producers to maintain control of how data are used. At the same time, it is a method for users of data to track where data can be found, how it might be accessed, what elements it contains, spatial and temporal timeframes of the data sets and what forms it is in, and thus, whether it is compatible with data user objectives or with other data sets. Metadata is quite valuable, saving data users both time and money.

The full value of metadata is realized when multiple data producers and consumers agree on standard metadata items and formats for presenting this metadata. This enables improved data sharing and opens the door for standard query capabilities to find relevant datasets gathered from a variety of sources. Just as libraries offer a standard set of items on their card catalogues, standard items can be defined for GIS datasets, household travel survey datasets, crash records, traffic counts, and so on. It will not be easy to engender acceptance and agreement on standardized metadata elements.

3. Current Metadata Standard Development Activities

At the moment there are many different contributions to standards development for various industrial sectors, including transportation, at different stages of evolution. While some standards already have a framework that only needs to be completed, others – even though comparable in nature – are quite a distance from any completion. Unfortunately, as multiple groups address the issue of metadata standardization, multiple standards emerge. It will be important to harmonize any future Data Section work in this area with other activities so as not to reinvent the wheel or add to the proliferation of “standards.”

International Standards Organization (ISO)

The ISO has set up a Metadata Working Group to take responsibility for standards applicable to the specification and management of metadata.¹ The workgroup’s scope includes: (1) a framework for specifying and managing metadata; (2) specification and management of metadata elements, structures, and their associated semantics; (3) specification and management of value domains, such as classification and code schemes; (4) specification and management of data about processes and behavior; (4) facilities to manage metadata, for example: data dictionaries, repositories, information resources, dictionary systems, registries, and glossaries; and (5) facilities to exchange metadata, including its semantics, over the Internet, intranets and other media.

Worldwide Web Consortium (W3C) – Resource Description Framework (RDF) Specification

The Resource Description Framework (RDF) is a language for representing information about resources in the World Wide Web.

Ambit eGMS

This is an example of an electronic government initiative for metadata schema of wide potential application (www.govtalk.gov.uk/schemastandards/metadata_document.asp?docnum=768).

Metadata Standards for Archived ITS Data

Under the banner of ASTM International, the Archived Data Management Systems (ADMS) community has published one standard that provided general guidance on developing and designing ADMS implementations, and dealt with the need for metadata in ADMS. A second ASTM standard (expected to be balloted for approval in 2005), titled “Standard Practice for Metadata to Support Archived Data Management Systems”, will be a content standard that describes what information should be provided (as mandatory, mandatory-if-applicable, and optional) to document a data set. A

¹ The American National Standards Institute (ANSI) also has a data representation standard.

third ASTM standard (expected to be balloted for approval in 2006) is currently under development, which deals primarily with metadata specific to traffic monitoring.

Federal Geographic Data Clearinghouse (FGDC) Metadata Elements

The FGDC has developed a number of tools, training materials, and guidelines relating to geospatial data. In addition, it has initiated awareness activities. The FGDC standards cover ten sections or attributes of data: identification, data quality, spatial data organization, spatial reference, entity and attribute, distribution, metadata reference, citation, time period, and contact.

The Australian and New Zealand Land Information Council (ANZLIC)

ANZLIC developed metadata guidelines in 1996 that are widely accepted in government and the spatial information community. The ANZLIC Metadata Guidelines were influenced by the ISO initiative and are committed to converge to it in future developments.

Data Documentation Initiative (DDI)

The Data Documentation Initiative (DDI) developed generalized standards for social science survey data. Five types of metadata are required: (1) the document description, (2) the study description, (3) data files description, (4) the variable description, and (4) other study-related materials. The DDI is currently being launched on the path to being sanctioned by the International Standards Organization (ISO). There is also an effort underway to make the DDI compatible with the ISO metadata exchange standard 11179, "Specification and Standardization of Data Elements," which can serve as a springboard to harmonization with other metadata standards such as Dublin Core and MARC.

Library Science

In the library science area, there are many standards development initiatives, and most are recognized national and international standards. A number are function-specific, such as ONIX for Books Product Information Message, the Text Encoding Initiative (TEI); the Encoded Archival Description (EAD), which are related to publishers, jobbers, and library archivists. The Committee on Scientific and Technical Information (COSATI) sets up cataloging rules and record formats for the descriptive cataloging of technical reports and similar documents. Other metadata initiatives in the library science sector that have broad application are described below.

MARC 21 and MARCXML

MARC (**MA**chine-**R**eadable **C**ataloguing) is a format standard for the storage and exchange of bibliographic records and related information in machine-readable form. All MARC Standards conform to: [ISO 2709:1996 Information and documentation -- Format for Information Exchange](#). MARC 21 is an xml² version of MARC:

Metadata Object Description Schema (MODS)

MODS standards that are somewhat simpler and more modernized than MARC21: As an XML schema, MODS is intended to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than numeric ones, in some cases regrouping elements from the MARC 21 bibliographic format.

² Extensible Markup Language – A World Wide Web Consortium (W3C) Standard for the formal structured description of documents

Dublin Core Metadata Element Set

Dublin Core (DC) provides a standard for cross-domain resource description. It is designed primarily to support discovery and retrieval, by defining semantics but *not* syntax. The value of DC is its simplicity and interoperability, making it usable as a crosswalk between major metadata standards. The Dublin Core Metadata Initiative (DCMI) controls the approved element structure and DCMI is a recognized international standard. DC can be expressed as html and xml and used in a metadata repository. The NTL expresses the metadata from its metadata repository in the html header for its web display.

Metadata Encoding and Transmission Standard (METS)

The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the [XML schema language](#) of the [World Wide Web Consortium](#). The standard is maintained in the [Network Development and MARC Standards Office](#) of the Library of Congress, and is being developed as an initiative of the Digital Library Federation. It offers a standard mode for object “packaging” for preservation, institutional repositories, and other activities

Metadata for Images in XML (MIX)

The Library of Congress' Network Development and MARC Standards Office, in partnership with the [NISO Technical Metadata for Digital Still Images Standards Committee](#) and other interested experts, is developing an XML schema for a set of technical data elements required to manage digital image collections. The schema provides a format for interchange and/or storage of the data specified in the NISO Draft Standard [Data Dictionary: Technical Metadata for Digital Still Images](#) (Version 1.2). This schema is currently in draft status and is being referred to as "NISO Metadata for Images in XML (NISO MIX)".

4. Challenges in Developing Transportation Metadata Standards

In addition to the need to harmonize or coordinate with other standards development activities, there are several major challenges to the development and use of metadata standards within the transportation community that need to be considered in any future course of action by the Data Section.

Challenge #1: Addressing Diverse Needs and Finding Common Ground

There are a diverse set of data types, sources and data users within the transportation community, which means there is much work to be done and a wide variety of perspectives to consider in defining metadata standards for transportation and furthering their use. The term “metadata” means different things to different people. Is it a way to find documents of interest on the internet? Is it a way to make sure people correctly interpret published datasets? Is it descriptive, textual information about how a dataset was assembled and what its limitations are? Is it highly structured, coded information that can be processed by applications as they accept an input dataset? Is it data dictionary information such as variable type, length and validation rules? Is it an agreed-upon set of definitions for commonly used terms? The answer is, all of the above (and more). The challenge is to map out the transportation metadata territory so that it is clear where the common ground is, and where individual efforts can and should proceed more independently. This is essential for defining a coherent and productive approach to metadata work in transportation and to allow for productive communication across people with different perspectives.

Challenge #2: Getting Agreement on Standards

Getting agreement on standards is hard work that must consider variations in methodologies used by different data producers, differences in how commonly used terms (e.g. “trip”) are defined, and

variations in what different data consumers need to know in order to make appropriate use of data. It also needs to recognize practical issues such as the effort to collect metadata, and availability of certain metadata items (particularly for older data sets). Successful development of metadata standards requires several ingredients: people to champion and organize the effort, incentives for participation across the full range of relevant stakeholders, and the knowledge and expertise to navigate the web of existing metadata standards and other relevant data standards (e.g. standard code lists for document formats). There are multiple existing overlapping metadata standards that are under the stewardship of groups including the US Library of Congress, the National Transportation Library, the Federal Geographic Data Committee (FGDC), the Inter-University Consortium for Political and Social Research (ICPSR) and various committees within standards organizations such as ASTM, ISO and IEEE. Many of these standards are purposely defined at a general level in order to be flexible and adaptable. Efforts to get agreement on metadata for a specific type of transportation information need to determine which (if any) of these high-level standards to adopt, and then they must develop extensions or applications of these standards to fit the needs of the relevant stakeholder group. Once this is done, the issue of ongoing stewardship and updates for the new or modified standard must be addressed.

Challenge #3: Overcoming Implementation Barriers

The final challenge is that even where standards exist, there are a host of barriers to be overcome before data producers will routinely build in metadata that adheres to these standards. Processes for collecting, updating, archiving and quality-checking metadata need to be put into place, which can involve considerable effort. For this to happen, managers need to have sufficient incentives to justify the effort required. Both “carrots” (persuasive value propositions) and “sticks” (metadata requirements imposed by data publishers/distributors) can provide these incentives. On the flip side, metadata standards need to be kept reasonable so that they will not be too arduous to implement. There is also an important education component to implementation – current metadata concepts are not well understood by data users and managers. Good examples are needed to provide easily adapted models. A strategy for piloting high value applications of metadata and then publicizing the results and benefits may be required in order to build support and acceptance.

5. Options for Data Section to Address Metadata Issues

This document has presented a basic description of metadata, provided a solid argument for their value to the transportation community, and identified the myriad of challenges that must be considered in addressing metadata issues. It is now time for the Data Section to move forward in leading transportation metadata activities. This report section discusses institutional mechanisms that the Section might pursue to fulfill its role, and outlines potential responsibilities that this new entity might take on.

Institutional Mechanism

Three different institutional mechanisms within TRB were considered for pursuing the development of metadata resources and standards. These mechanisms would be used to pursue activities such as sharing information about current and best practices, identifying research topics and priorities, drafting a plan or guidelines for standards development, and coordinating metadata-related activities. This workgroup recognizes that formal metadata standards development would likely have to be pursued through a standards development organization, such as ISO, ANSI, ASTM International, or others. At the same time and within the transportation community, we see a role for the American Association of State Highway and Transportation Officials (AASHTO) in promoting research, providing feedback, and building support and consensus for metadata approaches and standards.

Subcommittee of a Committee within the Data Section

A metadata subcommittee could be formed under a sponsoring committee within the Data Section. An immediate benefit of this mechanism is that a subcommittee can be instituted quickly – there are virtually no institutional barriers to the formation of new subcommittees. A subcommittee could begin operation as early as August 2005. Subcommittees of committees typically perform specialized committee functions (e.g., paper reviews, session organizers) or advance priority areas of research (i.e., census data, freight surveys). As such, they typically have limited membership (less than ten) that is restricted to those from the sponsoring committee’s members and friends, and a chair who is appointed by the committee chair. Most subcommittees are relatively short-lived. As noted above, there was until recently a metadata subcommittee under the Urban Data and Information Systems committee. This subcommittee could potentially be reconstituted under a different sponsoring committee within the Data Section. However, it would likely face the same issues that forced this initial subcommittee to disband, which was the need for a broader base of participation. A downside, to consider, is that subcommittees have less visibility within the TRB structure than a committee, and could face challenges to getting on the formal annual TRB meeting program, attracting membership, or advancing a research agenda.

Data Section Subcommittee

A relatively new option would be a subcommittee formed under the sponsorship of the Data Section. A Section subcommittee would have a chair appointed by and reporting to the Data Section chair. Membership would be open to members and friends of all Section committees. As precedent, two Data Section subcommittees currently exist both of which are functional in focus (i.e., research agendas, communications). This would be a new subcommittee focusing on a specific topic area of interest. In the case of metadata, there are numerous data user communities (e.g., library science, travel survey, travel monitoring, statewide planning, regional planning, travel demand forecasting, ITS, etc.) that have an interest and are considered key stakeholders. The potential benefit of this mechanism is that it would provide a broader base for addressing metadata issues - the subcommittee priorities, focus, and interest would cross-cut all of the Section committees. As mentioned above, a subcommittee can be initiated with little institutional “red tape”. While a Section subcommittee would encounter similar downsides as mentioned above, these limitations could be mitigated by ensuring that the subcommittee had representation from key individuals that were active in other TRB committees, who could play a strong liaison role. To the degree that the activities of the metadata subcommittee were of priority interest to the Section, a Section subcommittee could engender greater visibility within the TRB structure than a committee subcommittee could. The Data Section subcommittee could facilitate the development and adoption of a cross-cutting research agenda focused on metadata issues.

Task Force, leading to a Committee on Metadata

A TRB task force could be created to refine the purpose and scope of interested metadata stakeholders. If the task force forms a cohesive structure and agrees to long-term goals, the task force would be adopted as a formal TRB committee. Some TRB task forces are formed to address specific problems and issues recommendations—these task forces are often disbanded when the recommendations have been finalized. Others are formed for the purpose of crafting a cohesive vision that leads to committee status. For example, the current Travel Survey Methods committee (ABJ40) began as a subcommittee of the Urban Data committee, moved on to Task Force status for a period of two years, and then was converted to full committee status in 1999 as there was enough common interest and goals in this topic area among the various user communities. A task force structure is advantageous when interested persons come from numerous different user communities and TRB committees, which is certainly the case with metadata. The task force structure would allow

interested persons to further refine guidelines for metadata standards development and long-term goals for metadata standards implementation and preservation. A downside to going the Task Force route is that the formal process for initiating a Task Force takes time. An even greater hindrance is the institutional reluctance to increasing the number of TRB committees. The supporting arguments for committee formation must be compelling, solid, and well-substantiated.

Mission or Charge of the New Entity

In the process of developing this report, the working group encountered the practical realities of preparing a brief written treatment of a complex topic. A new Data Section entity charged with addressing metadata issues faces a challenging mission. While this working group agreed that formal standards development resides outside of the TRB structure and within a formal standards development organization, we agreed that there was a definite role for this new entity in paving the way for standards development. Given this, we recommend the following plan of action for the Data Section metadata group (subcommittee or Task Force):

- Facilitate information sharing about current metadata standards, practices and applications, and work towards coordination across metadata-related efforts within (and outside of) the TRB community,
- Work towards a common understanding of metadata and its value and applications within the TRB community – work with both data producers and data users to demonstrate the advantages of using standardized metadata (adoption by producers of data will be key for metadata initiatives to succeed),
- Promote best practices in the use of standardized metadata among data producers, data users, and data archives – encourage papers, presentations, and publications to stimulate interest in metadata and enhance credibility of metadata issues,
- Serve as a focal point for furthering development of transportation metadata standards, through identification of priority needs, promoting research projects to lay the technical groundwork for standards, supporting and participating in continuing work within standards development organizations and promoting consistency across efforts, and
- Ensure wide usability of standardized metadata by promoting training and tools development opportunities.

6. Recommendations to the Data Section

Transportation metadata is important. Metadata (or data about data sets) facilitates data discovery, sharing, and use in the transportation community. Reliable and consistent metadata use is both time- and cost-efficient. Metadata cultivates a common vocabulary between data producers and data users. The development and implementation of metadata standards within transportation practice is not without challenges. Not only must multiple data producers and consumers agree on standard items and formats for presenting metadata, but also the many different metadata initiatives across the globe invite the danger of both overlaps and gaps unless overarching guiding principles can be developed to inform transportation metadata efforts.

Standards are needed *now* for transportation metadata. The Data Section has an important role to play in facilitating a shared vision among stakeholders in the transportation community to produce stable metadata standards for describing, finding, and using the datasets that underlie much of transportation research, planning, and policymaking. While it is not the role of the Data Section to develop such

standards, it is the role of Data Section to facilitate and guide the development, adoption, and long-term implementation of such standards.

At a *minimum*, this group believes that the Section should focus on facilitating communication and collaboration across transportation data communities and providing opportunities for information sharing about existing metadata standards and applications. *Ideally*, the Section should provide a coherent vision and structure (i.e., a strategic plan) for advancing the use of metadata in the transportation community. A strategic plan would:

- Establish the need for the development of metadata standards and their potential benefits to transportation practice;
- Determine the general types of standards that are required and ascertain the policy implications of instituting metadata standards, especially barriers to acceptance;
- Identify other ongoing efforts that have an impact on the development of metadata standards for transportation data, and recommend how coordination may be achieved;
- Propose a schedule and recommend organizations (and other participants) that should develop and maintain metadata standards; and
- Identify methods to promote the use of completed metadata standards within the transportation community.

A Data Section focus on metadata is needed *now*. For this reason, this group recommends that a staged approach be used to institutionalize a metadata “group” within TRB. Because subcommittees can be formed with little bureaucratic red tape, we recommend that a Data Section subcommittee be formed immediately, with a chair appointed by the Section chair. Members would be drawn from committees within the Data Section and related “stakeholder” committees. Key goals of this subcommittee would be to define a metadata research agenda and to use this agenda to prepare the necessary proposal, with detailed specification of mission, objectives, and activities, for Task Force status. The proposal should lay out goals and guiding principles for the new Task Force as it moves forward over the next two- to-three years. The Task Force should be charged with developing the Strategic Plan noted above – with focus on strategic goals in areas of organization, standards, outreach, and technical development.

Workgroup Membership

Kou-Ann Chiaio, New York Metropolitan Transportation Council

Jim Hall, University of Illinois, Springfield

Frances Harrison, Cambridge Systematics, Inc.

Jonette Kreideweis, Minnesota Department of Transportation

Joyce Koeneman, Research and Innovative Technology Administration

Ralph Gillmann, Federal Highway Administration

Shawn Turner, Texas A & M University

Marcus Wigan, Oxford Systematics

Johanna Zmud, NuStats