6-2000

A Framework for GIS-T Data Sharing

Kenneth Dueker
Portland State University

J. Allison Butler

A Framework for GIS-T Data Sharing

Kenneth J. Dueker
Portland State University

J. Allison Butler
Hamilton County
Tennessee

December 1999
(Revised June 2000)

Submitted to Transportation Research Part C: Emerging Technologies

Center for Urban Studies
Portland State University
Portland OR 97207-0751

www.upa.pdx.edu/cus/

Abstract

This paper develops a framework and principles for sharing of transportation data. The framework is intended to clarify roles among participants, data producers, data integrators, and data users. The principles are intended to provide guidance for the participants. Both the framework and the principles are based on an enterprise GIS-T data model that defines relations among transportation data elements. The data model guards against ambiguities and provides a basis for the development of the framework and principles for sharing of transportation data.

There are two central principles. First is the uncoupling of graphics, topology, position, and characteristics. Second is the establishment of a schema for transportation features and their identifiers. An underlying principle is the need for a common data model that holds transportation features—not their graphical representations—as the objects of interest. Attributes of transportation features are represented as linear and point events that are located along the feature using linear referencing. Sharing of transportation data involves exchange of relevant transportation features and events, not links and nodes of application-specific databases. Strategies for the sharing of transportation features follow from this approach. The key strategy is to identify features in the database to facilitate a transactional update system, one that does not require rebuilding the entire database anew. This feature-oriented enterprise GIS-T database becomes the basis for building separate application-specific network databases.

1 Introduction

The problem with sharing and maintaining Geographic Information Systems-Transportation data (GIS-T data) among applications is the diversity of formats that lead to inconsistencies, inaccuracies, and duplication. This diversity is due to differences among data models that make it difficult to achieve consistent representations of the transportation system. Yet there are legitimate differences in requirements that lead to application-specific definitions and representations of transportation objects and their geometry. This has resulted in multiple and inconsistent digital representations of various parts of the transportation system.

Currently, we have different networks to support applications, such as vehicle navigation systems, overweight truck routing, facility location and address geocoding, emergency management, or for a reference layer of roads for resource management applications. These applications define roads differently. Some include paths and trails, private streets, alleys, and resource roads and some do not. In addition, the level of detail and spatial accuracy differs. The challenge is to establish means of data exchange among these disparate representations that lead to improvements in accuracy, consistency, and completeness. And more importantly in the long run, is the need to share updates to maintain currency in representing the transportation system. At the same time it is important to recognize the legitimacy of specialized representations needed to support different applications.

The following metaphor is offered to make the elusive concepts in this paper more understandable:

> Transportation Features (TFs) are like strings of pasta. A bowl of spaghetti contains TFs of all types, freeways, arterials, local roads, alleys, paths and trails, airports, pipelines, and railroads. Users identify the types of TFs they want and select from the bowl those pieces needed for vehicle navigation, emergency dispatch, pizza delivery, walking, or bicycling. Then using a "clean and build" procedure users can construct application-specific networks.

This paper develops a framework and principles for sharing of transportation data to achieve more accurate representations of transportation data in an intermediate form, from which application-specific databases can be generated. The framework is intended to clarify roles among participants, data producers, data integrators, and data users. The principles are intended to provide guidance for the participants. Both the framework and the principles are based on an Enterprise GIS-T data model that defines relations among transportation data elements. The data model guards against ambiguities and provides a basis for the development of the framework and principles for sharing of transportation data.

This paper assesses the problems of current approaches to sharing of transportation data, and of standards and data models for transportation data. Then the paper describes an "enterprise GIS-T data model" that is designed for data sharing, from which end users of transportation data can generate application-specific databases. This is followed by a discussion of requirements of transportation data participants. This leads to a formulation of principles for sharing of transportation feature data, and discussion of ways to identify transportation features in databases. This is followed by a discussion of issues to foster data sharing to improve accuracy, completeness, and currency of databases used for GIS-T applications.

2 Current Approaches to Sharing of GIS-T Data

Data sharing is more than simply having the ability to occasionally import information from someone else's database. The business needs driving dynamic data sharing include those for multi-agency local government GIS infrastructures, where E-911 (emergency dispatch), public works, and property assessment organizations need to utilize a common database. The property assessor gets a new subdivision plat with recently constructed streets shown on them. The E-911 center needs to know where the new streets are located, the building inspector needs to know all the new lot addresses, and the public works department will need to establish new trash pick-up routes and pavement management segments. Somehow, the information on the plat has to be communicated accurately, efficiently and quickly to these many users, each with a unique linear and/or non-linear location referencing system.

The reality in most cases is that the enterprise approach to GIS is late on the scene, with the different agency-specific GIS applications having been developed in isolation. Until such time as these systems may come to use a single enterprise-wide data infrastructure, these agencies will need to frequently exchange data sets. This paper attempts to set the stage for establishing comprehensive transportation data exchange mechanisms by describing a way to integrate them in an enterprise GIS-T data model.

Sharing GIS-T data is both an important issue and a difficult one. Important because there are many organizations that produce or use GIS-T data, and difficult because there are many ways to segment and cartographically represent transportation system elements. There is a lack of agreement among transportation organizations on defining transportation objects and in the spatial accuracy with which they are represented cartographically. This lack of agreement leads to difficulty in conflating, or integrating, two views of the same or adjacent linear objects.[1]

There are two problems in defining transportation objects: different definitions of roads and different criteria with which to break roads into logical segments. The logical segments become objects in the database that we will refer to as "transportation features." We have selected this term in order to include more than just roads. Roadways, railroads, transit systems, shipping lanes, and air routes are all linear features that utilize the same

[1] Sperling and Sharp (1999) describe conflation of U.S. Bureau of the Census TIGER files with local street centerline files as the "automatic matching and transfer of features and attributes from one geo-spatial database into another."

basic network data model, which utilizes linear travel paths between points of intersection. Since they all use the same basic data model, we will generally restrict our discussion to roadways for simplicity.

Transportation features become the building blocks for specific applications. Persons building vehicle navigation databases need to include private roads that are open for public use and omit "paper streets" not yet constructed and that cannot be navigated. Yet, public organizations responsible for road maintenance follow different rules; they omit private roads and include planned public roads on their maps. Similarly, two organizations responsible for roads on resource lands, the Forest Service and Bureau of Land Management, have quite different definitions of roads.[2] Most organizations that maintain databases of roads break them into logical segments to create discrete transportation features according to some business interests, such as a change of pavement type, jurisdiction, or functional type, or at all intersections.

These differences in original purpose for transportation databases create a difficult arena for sharing data with others. The data sharing arena includes data producers, data users, and, increasingly, data integrators who collect data from the field, legacy databases, or other data producers or users and reorganize it for new uses and/or to maintain currency. A healthy data-sharing environment suggests that data producers embed registration points and feature identifiers in their original data to facilitate importing and registration of foreign or legacy cartography and attribute data.

The need is for standards for data sharing among organizations, both public and private. However, standards are difficult to develop because system requirements of advanced applications of GIS-T technology differ in spatial and temporal accuracy and details of such features as ramps and lanes. Further, differing levels of real-time use of systems dictate response requirements of databases and interfaces.

A more systematic approach to GIS-T data sharing of GIS data calls for relief from the conflation technique that requires the matching of spatial objects of separate data sets. It is a problematic process due to the need to simultaneously match both topological and geometric properties. This is exemplified by efforts of transferring TIGER attributes to more accurate vector files (Brown, et al, 1995, and Tomaselli, 1994). This approach to matching of spatial data of similar scale works well, if the data were captured using the same data model or criteria by which to define and segment roads. If roads are not segmented into similar spatial object, conflation is not a satisfactory way to share transportation data. Some (Sester, et al, 1998 and Walter and Fritsch, 1999) work toward automating the conflation process while Devogele et al (1998) call for the need to develop an integrated schema from data models to facilitate data sharing and/or interoperability.

In addition, sharing of transportation data is not a one-time issue, in a larger context it is a means of disseminating data about changes in the transportation system. Management of

[2] The Forest Service defines roads as any visible track, while the BLM limits roads to tracks that can be traversed by a normal vehicle.

updates to maintain current representations of transportation systems is a growing concern. Up to now the effort has been on building the initial database. Attention is now turning to maintaining currency of databases of increasing detail and complexity. The GIS-T community is in need of guidance on this issue. We do not see sufficient progress on the conflation approach, nor do we see adequate consensus to support schema integration to support the wide range of applications of GIS-T data.

Sester, et al, (1998) identify an alternative to the bottom-up approach of conflation, a top-down approach known as semantic data integration. Our application of this top-down approach to GIS-T requires the aggregation of spatial object primitives into larger (longer) transportation features that can and need be uniquely identified, while maintaining the richness of attribute detail along features by linear referencing. Similarly, the transportation features in our enterprise GIS-T data model are not topological spatial objects. This enables the building of application-specific networks from selections of a single and consistent set of underlying data. Also our enterprise GIS-T data model allows for multiple spatial representations to accommodate the need for both abstract arterial-level data and detailed representations of roads, including freeway ramps and local streets.

Consequently, our enterprise GIS-T data model falls somewhere between one that specifies an end product like TIGER or GDF, and a data exchange standard like SDTS or DIGEST. "Between" implies it is more than a neutral standard and it is more than a format for a database to support a specific application. We call it an enterprise GIS-T data model to convey the notion of a standard approach to maintaining business data in a transportation organization about the transportation system for which they are responsible. One master set of transportation features enables maintenance of current knowledge about the system. Selection of transportation features by type and from the available multiple geometric representations enables building of a number of application-specific networks for functions such as vehicle navigation and emergency, pavement, bridge, and congestion management.

3 Assessment of Data Models Used in GIS-T

There are several GIS data models used for transportation applications. GDF (Geographic Data File Standard, 1995 and 1999), NCHRP 20-27 (Vonderohe, et al, 1998), and our Enterprise GIS-T data model (Dueker and Butler, 1998) have been developed specifically for transportation, while others, Chan, 1998) and TIGER have broader applications. They differ in robustness in representing transportation systems and in handling updates.

Current data models used to represent streets and roads, e.g., TIGER line and ArcInfo geo-relational data model, integrate the cartography, the network link, and attributes of the link into a single linear spatial object. This analysis questions the "integrated data model" and calls for an unbundled approach to facilitate data sharing and maintenance of GIS-T databases.

The tighter the integration of topology, geometry, and attributes, the more information has to be input before the update is usable. The loosely coupled approach of our enterprise GIS-T data model allows each piece of the bundle to stand on its own and be entered separately (perhaps by different organizations). Each piece of the bundle is thus available for use as soon as it is in the database. In this way, the loosely coupled approach requires less business process revisions, it better fits the way people work.

The GDF Aggregated Way feature may be the closest one that matches our transportation feature. GDF, for all intents, is a fully topological data model that requires full specification of cartography, topology, and attributes for any useful data sharing to occur, although pure cartography can be exchanged. There is no explicit support for multi-link objects inside the transmission protocol. The use of Aggregated Way appears to be external, a way to allow agencies to connect their data sets to the GDF objects using a one-to-many relationship. Use of GDF appears to require prior agreement on roadway segmentation.

GDF is designed to operate at two levels; Level 2 consists of complex features, roads and intersections, while Level 1 consists of simple features, road elements and junctions. Road is a topological entity representing a complex feature composed of road elements joining two intersections. A freeway interchange is represented as an intersection node, a complex feature at Level 2, and road elements and junctions to represent ramps and ramp connections at Level 1. Adding intersections in GDF requires the splitting of roads.

This limitation that Road entities are bounded by Intersections is identical to the topological structure of TIGER. The addition of alleys, private roads, pedestrian paths, statistical or political boundaries will require a splitting of topological edges. This increases the difficulty of sharing data among networks that are separately maintained. Alternatively, newer GIS software, such as ArcInfo 8 is based on data models that break this limitation by allowing vector data to be non-topological line features or topological features. Our enterprise GIS-T data model builds on that principle and separates transportation features from topological links. It is an intermediate form from which databases to support applications can be generated.

4 GIS-T Data Modeling Issues

A National Cooperative Highway Research Program project, 20-27(3), is in the process of trying to bring consensus in the area of GIS-T data models (Adams, Vonderohe, and Butler, 1998). Meanwhile, there are two different approaches. Our enterprise GIS-T data model represents a feature (object) database approach. It is best suited for a federated systems environment with legacy data of varying spatial accuracy supplied and used by a wide variety of agencies acting in concert as a single enterprise. This approach relies on the traditional model of linear location referencing systems (linear LRS), which utilize roadway identifiers and linear offset measurements to locate attributes of a highway. The field rules for a linear LRS form its location referencing method (LRM)

An alternative approach can be characterized as a location (geometry) approach (Sutton, 1999). This approach embraces the use of Earth-based, two- and three-dimension positions; such as might be derived from a series of GPS-derived coordinates while traveling down a highway. Thus, location becomes the way in which various data about roads are integrated.

This geometry approach would work well in a strong centralized environment wherein the location of transportation features would be redigitized with high precision using GPS. This would be needed to enable linking by coordinate snapping of spatially accurate tracking or events to a spatially accurate map base. This approach has not been formally stated or tested, however. Issues such as repeatability of GPS positions, how to abstract networks, how to relate to other location referencing systems, and representation at smaller scales have not been adequately addressed.

Our GIS-T data model supports both approaches by employing the anchor point object proposed by Vonderohe and Hepworth (1998), a feature that can be described in both location referencing systems. We endorse the use of these objects as registration points to interface between the two systems. But the fact remains that both approaches must have a way of defining a transportation feature. A series of GPS coordinates do not make a line to describe a roadway centerline unless someone connects the dots and calls it by a common name. Defining a common name for a roadway is also a requirement of the linear LRS approach. Therefore, our foundation business rules address how to combine the roadway pieces into features, and attach the attributes that relate to them.

The NCHRP consensus approach will have to resolve differences in the definitions of basic transportation entities and primitives that exist among the various data models. For example, a Road is a topological entity in GDF, but the Transportation Feature in our enterprise GIS-T data model is not a topological entity. We directly connect the cartography to the transportation feature, not to the linear datum as in the NCHRP 20-27 data model. Making the linear datum the principal entity requires transferring topology to the datum and thence to the cartography. Using the NCHRP 20-27 data model approach, a dynamic segmentation process to display transportation feature attributes requires conversion from network measures to node offsets then to anchor point offsets.

There are missing elements in both approaches that are being addressed in the NCHRP consensus data modeling effort. One is a problem of treating time and the third spatial dimension more explicitly in the data model. The other problem is with GIS data models in general that become paramount in GIS-T. In traditional GIS, a spatial object is defined by its location. Consequently there is no suitable way to represent a moving object, like a vehicle, package shipment, or storm in such a GIS. There needs to be a new dynamic, tracking or moving object class in GIS, especially in a GIS-T. There are three approaches. One is a static object with frequently changing positions. Another is a new object class with location as an attribute rather than part of the definition. Yet another is a moving object construct with starting location and attributes of direction, speed, destination to define a moving object. Emerging object-oriented GIS platforms offer a way to do this by treating location as an attribute of an object.

The next section presents the enterprise GIS-T data model. The model brings consistency to the representation of transportation data and provides guidance to the data-sharing participants. The model also provides the basis for the subsequent development of data sharing principles.

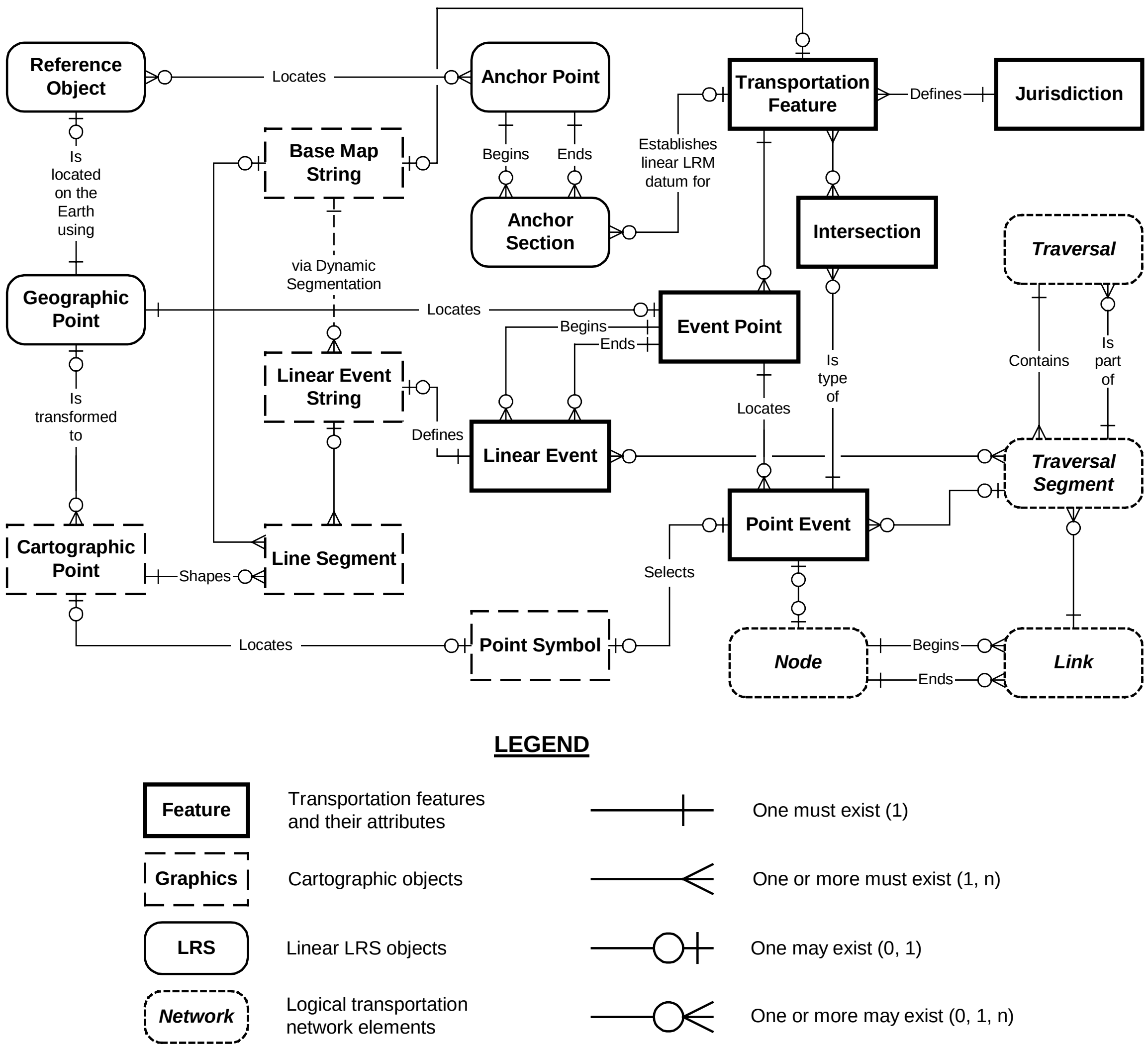
5 The Enterprise GIS-T Data Model

Previous work by the authors (Dueker and Butler, 1998) provides a data model well suited for GIS-T data sharing. It unbundles the geometry, topology, and attributes to facilitate separate maintenance and enables extraction of data for different uses. The original paper provided a general introduction to information system design by offering tutorial appendices on such topics as user requirements analysis, data modeling, and business rule construction. The tutorial information was offered to enhance the reader's understanding of the transportation data model described in the main body of the paper. That model was proposed as a comprehensive description of the entire scope of transportation data that might be housed in a state DOT. While business rules on which the model was founded were presented, the accommodation of implementation details was the main thrust of the work.

One unfortunate outcome of that paper was the general difficulty readers had at taking in the big picture of a comprehensive data model and the myriad of database tables needed to implement it. Reader comments on that paper have motivated us to create a new version, contained herein, that is based on the formal business rules upon which the our Enterprise GIS-T data model is founded. These business rules provide the basis for the ensuing discussion of GIS-T data sharing issues.

The simplified version of the Enterprise GIS-T data model is shown in Figure 1.

Figure 1. *Simplified Enterprise GIS-T data model.*



Entities—the things about which we wish to store information—are shown in boxes, while the relationships between entities are shown using lines. Each entity type has been identified by a special style. Each relationship type has been shown using descriptive text and connection symbols. Each group of entities for a single type can be treated as a stand-alone data set.

An entity is a discrete part of our world, one for which we want to store information. An entity is a basic building block of data models. Entities are related to each other through verb-oriented statements, such as, “A jurisdiction defines one or more transportation features,” and its corollary, “A transportation feature must be defined in the context of a single jurisdiction.” Relationships are the other building blocks of data models. A relationship without explicitly stated verbs generally can be read as an ownership relationship, such as, “A base map string has one or more line segments.” Attributes are part of the entities, not separate entities. For example, a line segment may have width and color attributes, which are aspects of the Line Segment entity, not separate entities owned by Line Segment.

5.1 Transportation Features and Their Attributes

The first group of entities we need to examine is the one that contains our basic transportation elements. The group consists of six entities: Jurisdiction, Transportation Feature, Event Point, Linear Event, Point Event, and Intersection. Let’s start by defining these entities and their relationships.

- **Jurisdiction.** *The political or other context for designating transportation features and their names, which may be merely numerical references unique within the jurisdiction.* A jurisdiction sets the context for defining the extent and name for each transportation feature and is primarily geographic. Jurisdiction carries no other burden; i.e., it does not mean which agency has maintenance responsibility. A state DOT may choose to subdivide the State highway system on a county basis, with each county-specific portion of a roadway having its own identifier. In this instance, “county” would be the value of Jurisdiction. The maintenance jurisdiction in which a transportation feature may be located would be an attribute of the feature, which could be stored as part of the Transportation Feature entity if it applied to the entire feature, or a separate linear event. Jurisdiction need not be the same for all transportation feature types. Airports can be named on a national basis, interstate freeways named on state basis, and local streets named on a zip code basis. *Relationships:* Jurisdiction defines one or more Transportation Feature. Transportation Feature must be defined within the context of Jurisdiction.

- **Transportation Feature.** *An identifiable element of the transportation system.* A transportation feature can be like a point (interchange or bridge), a line (road or railroad), or an area (rail yard or airport). Some transportation features can consist of other features, such as may be the case with bridges and roadways. In a roadway inventory, the bridge is an event that occurs on a specific roadway. The location of the bridge would typically be defined by a linear LRS location description. In a bridge inventory, the bridge is a transportation feature in its own right, and its location may be defined by a set of Earth coordinates. *Relationships:* Transportation Feature may have one or more Event Point. Event Point must be defined on a single Transportation Feature. Transportation Feature may contain one or more Intersection. Intersection must be owned by one or more Transportation Feature.

- **Event.** *An attribute, occurrence, or physical component of a transportation feature.* Attributes include functional class, speed limit, pavement type, and state road number—things that are not tangible but describe a tangible element, such as a road. Occurrences include traffic crashes and projects. Physical components include number of lanes, guardrail, signs, bridges, intersections, and other tangible things that are field-identifiable elements. There are three event subtypes; a given event instance may be expressed as more than one subtype:

- **Point Event.** *A component or attribute that is found at a single location (one event point).* Point events may occur independently or on transportation features of the linear or area form.
- **Linear Event.** *A component or attribute that is found along a segment of a linear transportation feature.* Linear events are defined by two event points (beginning and ending). Linear events may occur only on linear transportation features.
- **Area Event.** *A transportation feature component or a non-transportation entity that affects a transportation feature.* Areas can be explicitly represented as polygons or implicitly represented as to where they intersect transportation features. The implicit option is called an area event and is represented through related linear and point events. For example, an area event could be a city. The city could be expressed by creating a linear event for the portion of a transportation feature located within it, or as point events where the city limits cross a transportation feature. Another example could be a park-and-ride lot, which would be stored as a point event located where the driveway to the lot intersects the adjacent road (transportation feature). Area events may be applicable for any kind of transportation feature. Area Event as a discrete entity is omitted from the simplified model since such events are almost always expressed in a transportation database using linear and point events. The omission of Area Event also lets us drop Polygon, its corresponding cartographic entity.

- **Event Point.** *The location where an event occurs on a transportation feature.* Event points are located on a transportation feature as an offset distance measure from the beginning of the transportation feature. Most transportation databases use event point measures made in units of 0.01 mile. The smaller the measurement unit, the higher the resolution of the database; i.e., the closer two events can be and still be stored as separate events. A resolution of 0.01 mile means that two events within 52.8 feet of each other may be stored as being located at the same position along the transportation feature. Event point locations are stored using field measures in real-world units, not cartographic units. Event points locate events on the transportation system while cartographic points locate representative graphical elements on a map. In addition to locating Event Points by a linear measure along transportation features, they can be located by direct coordinate measurement - by digitizing maps or field measurements from surveying or GPS. Just as linear measures along transportation features can be converted to coordinates by interpolation along shape files representing the transportation feature using dynamic segmentation, coordinates can be snapped to transportation features and converted to linear measures by a reverse of dynamic segmentation. *Relationships:* Transportation Feature may possess one or more Event Points. Event Point must be defined in the context of one Transportation Feature. Event Point may locate one or more Point Event. Point Event must be located [on Transportation Feature] by one Event Point. Event Point may locate the beginning or end of one or more Linear Event. Linear Event must be located by a beginning Event Point and an ending Event Point. Event Point represents Geographic Point. Geographic Point locates Event Point.

- **Intersection.** *A special type of Point Event which may be owned by more than one transportation feature.* From the perspective of a single transportation feature, an intersection has but one owner; i.e., that transportation feature. In actuality, though, an intersection may appear as a point event on more than one transportation feature. Relying solely on the Transportation Feature owns Point Event relationship will result in redundant data storage in that each transportation feature would be required to record the same intersection data. The use of an Intersection entity allows the location of the intersection to be stored as a point event on each intersecting transportation feature, but to store the intersection characteristics once as part of Intersection entity attributes. It may be advantageous to treat any type of transportation system junction or crossing as an intersection. For example, treat all bridges as intersections in order to link bridge height and weight data to the roads going over and under the bridge. *Relationships:* Intersection must correspond to one or more Point Event. Point Event may represent one Intersection. Intersection must involve one or more Transportation Feature. Transportation Feature may possess one or more Intersection.

These entity and relationship definitions are based on the following business rules:

1. Transportation features are contained completely within a single jurisdiction. A transportation feature that exists physically in more than one jurisdiction may be administratively subdivided at the jurisdiction boundary as a data management approach analogous to the tiling of GIS maps.
2. The physical path of a linear transportation feature can be arbitrarily defined, although specific users may adopt rules for defining a transportation feature. The model is independent of these rules.
3. A transportation feature's unique identifier (name) in the database must be independent of its real-world name, which is an attribute of the feature, even if the adopted rule is to follow a named path to define a transportation feature.
4. An external public key identifier may be constructed using a widely accepted naming convention. The public key is independent of the primary (internal) key and the real-world name.
5. Express point and area features relative to linear highway features. For example, consider that a city is an area feature. In a roadway database, the extent of a city is defined by the portion of each roadway within the city. To store the relevant information about the city, define the portion of each roadway inside the city by stating the beginning and ending points (city limits) using linear LRS measures.
6. Store the relationship between roadways at intersections, and efficiently record data about intersections without duplication.

5.2 Network Features and Their Attributes

The second group of entities consists of four entities that contain information about the connectivity (topology) of the transportation network: Node, Link, Traversal, and Traversal Segment. Networks are subdivided into segments called links, which begin and end at nodes. A traversal is a path through the network that is composed of traversal segments, each of which may be formed from one or more links and their attributes.

- **Node.** *A 0-dimension object that represents the topological junction between two or more links, or the endpoint of a link.* In their most common form, nodes will correspond to intersections and other point events. Not every intersection need be represented as a node. It is sufficient for a routing application, for example, to include only those decision points (i.e., where the course may be altered) that have been deemed suitable. *Relationships:* Node may represent one Point Event. Node may begin one or more Link. Node may end one or more Link. In most cases, Node must begin or end at least one Link (may not be true is centroid nodes are included in data set).

- **Link.** *A 1-dimension object that represents the logical connections between nodes.* Links, as used here, are dimensionless in that they only specify the possible connections and not the actual distance between nodes. We could include an attribute for valid directions, but we have chosen to recommend two alternative implementation strategies. One is to order nodes; the other is to only include valid nodes that may be reached from a given node. In the ordered node approach, if one could travel from Node A to Node B or from B to A, as would be the case with a two-way road, then a link table would include the entries A,B and B,A. A corresponding one-way road would omit one of the node pair entries. In the valid destination node approach, a to-node entry for Node A would be B and one for B would be A. If, however, the link A-B were a one-way road with the direction of travel being from B to A, then there would not be an entry for Node B in the node table record for Node A. *Relationships:* Link must begin at one Node. Link must end at one Node. Link may be part of one or more Traversal Segment.

- **Traversal.** *A path or route through a portion of a transportation network consisting of one or more segments.* We have chosen to make a distinction in application between static traversals, such as that formed by a state road crossing many counties, and a dynamic traversal, such as that created to route an overweight truck from origin to destination. However, these are functionally the same thing, only their duration of use is different. Both are constructed by selecting traversal segments with a specified set of attributes. For the state road traversal, this would be those traversal segments with the same road number. For the overweight truck route traversal, this would be those that could support the weight of the truck and possibly match other criteria, such as low traffic volumes. *Relationships:* Traversal contains one or more Traversal Segment. Traversal Segment may be part of one or more Traversal.

- **Traversal Segment.** *A link and its relevant attributes.* A traversal segment is both topological and physical. It combines the connectivity of the link with the attributes of the transportation feature segment it represents. We have elected to utilize this separate entity to store link attributes for two reasons. First, we wanted to keep the purely topological information separate in order to facilitate data sharing and our central concept of unbundling. Second, we wanted to be able to select the attribute set to match the needs of the application while preserving a normalized data structure for transportation features and their attributes. Traversal segments may be viewed as a special linear event that corresponds to the path of the related link. Traversal segment attributes may be derived by finding the desired linear and/or point event records that fall within the same roadway segment. The actual selected values may be minimums (e.g., bridge with the lowest weight bearing capacity, lowest overhead clearance, etc.) or maximums (highest traffic volume, greatest population, etc.). *Relationships:* Traversal Segment may be part of one or more Traversal. Traversal Segment must represent one Link. Traversal Segment may include attributes of one or more Point Event. Traversal Segment may includes attributes of one or more Linear Event. Linear Event may be applicable to one or more Traversal Segment. Point Event may be applicable to one or more Traversal Segment.

These entity and relationship definitions are based on the following business rules:

1. Link-node data structures may be used to express certain attributes that conform to that structure, such as the distances or travel times between intersections (block is a link, intersection is a node).
2. Define static and dynamic paths (traversals) through the highway network. A static path is one that is defined by a slowly changing characteristic, such as state road number or a bus route. A dynamic path is one created to serve an immediate and frequently changing need, such as to route an overweight vehicle to its destination or to direct a rider to the nearest bus stop.

5.3 Linear LRS Objects and Their Attributes

We have provided four entities to tie transportation features to the surface of the Earth in 2- or 3-coordinate space. These entities are derived from those proposed by Vonderohe and Hepworth (1998).

- **Anchor Point.** *A 0-dimension object representing the end or beginning of an anchor section and serving to relate that terminus to an Earth-based location. Position on the Earth and on any related transportation feature are mandatory attributes.* By storing the Earth-based and linear LRS coordinates for an anchor point, the database model provides a registration mechanism to tie the transportation features to the ground. This facilitates cartographic conflation. (We actually recommend storing the linear LRS data in the anchor section records since the position is specific to each terminating anchor section.) *Relationships:* Anchor Point may be located by one or more Reference Object. Anchor Point may begin one or more Anchor Section.

Anchor Point may end one or more Anchor Section. From a practical perspective, Anchor Point must begin or end at least one Anchor Section.

- **Anchor Section.** *A 1-dimension object providing a logical representation of all or part of a transportation feature. Length is a mandatory attribute.* An anchor section begins and ends at an ordered pair of anchor points, the ordering providing an indication of direction of increasing linear LRS measures. Anchor sections and points differ from links and nodes by including more than topological information and an incomplete specification of topology. For example, anchor point locations are defined in possibly several systems, node locations not at all. Ordered anchor point pairs for defining anchor section direction fail to indicate whether traffic can flow in the opposite direction. *Relationships:* Anchor Section may establish the linear LRS datum for one Transportation Feature. Transportation Feature may be defined by one or more Anchor Section. Anchor Section must begin at one Anchor Point. Anchor Section must end at one Anchor Point.

- **Reference Object.** *A physical object or recoverable location to which the position of anchor points may be conveniently related.* Many anchor point locations are likely to be conceptually easy to define but physically hard to find in the field. An example is the middle of an interchange or intersection—easy to recognize, but hard to precisely and consistently locate. Tying the location to a physical object, such as a traffic signal pole, allows the precise location of the anchor section to be readily found using direction and distance from the reference object. Anchor Point and Reference Object may be merged to form a single entity. *Relationships:* Reference Object is located on the surface of the Earth using one Geographic Point (which may have many coordinate descriptions). Reference Object may be used to locate one or more Anchor Point.

- **Geographic Point.** *A 0-dimension object carrying the real-world (Earth-based) location of a reference object.* We have been deliberately restrictive in our definition of Geographic Point as a means of simplifying the model. In reality, every other entity in the real world could be described using a geographic point. What we want to emphasize here is the translation of geographic points, defined minimally for reference objects, to cartographic points as a means of facilitating conflation. *Relationships:* Geographic Point may locate one Reference Object. Geographic Point may be transformed to one or more Cartographic Point (each with its own cartographic datum and coordinate system). Event Point represents Geographic Point. Geographic Point locates Event Point.

The business rules that produced this portion of the data model are:

1. Many transportation features may be efficiently located on the surface of the Earth using non-linear location referencing systems based on GPS and other Earth geode methods. Most typically, these features are of the area and point geometric forms. Point and linear attributes on linear features may be secondarily defined using Earth coordinates. However, this approach will not unambiguously define a point on the

linear feature *from a graphic perspective* due to differences (errors) between the map and the real world, the measured position on the Earth and its equivalent on the map, and between the measured position and the true position.2. Provide the data entities needed to store a linear datum developed to meet positional accuracy and data sharing needs, but do not require that such a datum exist.

5.4 Cartographic Objects.

We use the term 'cartography' to refer to the map or picture produced by a GIS application. There are five cartographic entities that may be used to visually express the location and shape of transportation features, linear events, and point events. The following entity definitions are independent of proprietary software terms.

Cartographic Point. *The internal address reference for placing a single point on the surface of a 2- or 3-dimension map (manifold).* Each mapping environment has its own way of defining single locations; i.e., its cartographic datum. *Relationships:* Cartographic Point may represent one Geographic Point. Cartographic Point may shape one or more Line Segment. Cartographic Point may locate one Point Symbol.

Line Segment. *A straight connection or otherwise defined mathematical path between two cartographic locations.* All line segments must be defined using beginning and ending cartographic locations (points). *Relationships:* Line Segment is shaped by one or more Cartographic Point. Line Segment may be part of Base Map String. Line Segment may be part of Linear Event String.

Base Map String. *A connected non-branching sequence of line segments, usually specified as an ordered sequence of vertices, which cartographically defines the shape of a transportation feature.* Our basic implementation suggestion for mapping transportation features is to start with an equivalency between each entire transportation feature and a single line object. This line object will be composed of one or more line segments. Multiple spatial representations are enabled. A transportation feature can be represented by more than one base map string. *Relationships:* Base Map String may be used to describe the shape and position of one Transportation Feature. Base Map String must consist of one or more Line Segment.

Linear Event String. *A connected non-branching sequence of line segments, usually specified as an ordered sequence of vertices, which cartographically defines the shape of a linear event occurring on a transportation feature.* Our expectation here is that most linear event strings will be created as needed by dynamic segmentation, which uses straight-line interpolation to extract the portion of a base map string that corresponds to the location of a linear event using linear LRS measures. Line style, width, and color may all be used to distinguish the attribute being described. *Relationships:* Linear Event String must represent one linear event. Linear Event String must consist of one or more Line Segment.

Point Symbol. *A cartographic object that is used to show the position and nature of a point-like real-world feature.* Just as linear features can be shown using a line, point features can be shown using a point symbol. *Relationships:* Point Symbol may be located on a map using one Cartographic Point. Point Event may be illustrated by Point Symbol, which may also select the correct point symbol instance to display.

The business rules that produced this portion of the data model are:

1. Separate graphic representations from attribute data to provide scale independence; provides "one database-many maps" functionality.
2. Utilize dynamic segmentation and field-based linear LRS measures to create cartographic objects that correspond to linear events and to position point symbols.
3. Provide a means of relating real-world measures to map locations.

5.5 Applications of Enterprise GIS-T Data Model

More details of the model and examples of implementation are contained in the original presentation (Dueker and Butler, 1998). For example, attributes that are offsets from the centerline, such as signs, guardrails, or number of lanes, are handled by means of point or linear events. Linear events, such as lanes can be reformulated in several ways. A linear event could be represented as a recursive one-to-many relationship for Transportation Feature entity, if we want to look at the lanes as transportation features owned by a larger one. Or each lane could be a stand-alone feature. The data model itself is substantially indifferent to the details of transportation feature construction. We choose not to specify one or the other as the application may drive one over the other.

The data model supports non-planar graphs although we encourage the use of intersections for over- or underpasses to store clearances and restrictions. Although our enterprise GIS-T data model does not support temporal data explicitly, time is an attribute of transportation features or of the point and linear events associated with them.

Whether this data model or another, a common understanding of the transportation system is needed in order to share data effectively. The data model facilitates selection of appropriate transportation features from different databases and spatially registering them and creating new application-specific networks. The GIS open systems concept applied to interoperability of transportation data requires a common feature schema, which is consistent with our transportation features. In addition, the data model separates the update and maintenance of transportation features from the networks used in specific applications. Different networks can be generated from a common set of transportation features and one update process can support a set of applications.

5.6 Comparison of Enterprise GIS-T and NCHRP 20-27 Data Models

The connection between the NCHRP 20-27 data model and the Dueker-Butler data model is considerable and implicitly apparent from the common use of several terms and methods. Recognizing the considerable contribution of 20-27 towards consensus building in the field, we sought to adopt, wherever possible, the conventions, approach, and definitions of the 20-27 data model. A basic component of our model is the concept of transportation features. Transportation features have many of the characteristics of the 20-27 model's traversals. For example, a linear transportation feature can have points located along its length by the use of a linear LRS. However, the concept of transportation feature includes area and point forms, not just linear ones, where the term 'traversal' does not intuitively apply. We have retained the traversal concept for linear transportation features, even to the point of using the same definition as in 20-27, but the primary linear LRS is tied to the transportation feature. (Positions on traversals may be defined using a subordinate linear LRS, including one based on time, e.g., a transit route running on all or part of several transportation features.)

However, there are fundamental differences. First, we do not require transportation features to be composed of one or more links. In other words, topology is not a requirement for transportation features. Second, we relate the cartography to the transportation feature, not to the linear datum. To some degree we also adopt a third difference by embedding the linear LRS in the transportation feature; however, the implementation of this approach and that of the 20-27 data model appear to be identical. A fourth difference is the expansion of datum entities to include reference objects that are more readily recoverable in the field than anchor points. To some degree, our inclusion of reference objects is intended to accommodate the proposed National ITS (Intelligent Transportation System) Datum, which consists of a group of reference objects and the rules to define them.

One of the biggest differences is the fifth one: incorporation of cartography. The 20-27 data model provides support for cartographic conflation in that it can supply position information in the form of Earth coordinates for anchor points. These coordinates can be mapped in the cartographic environment and be used to relate graphical objects to the datum and improve the quality of maps. This approach leaves out many graphical objects and provides little roadway shape information. Cartographic conflation remains a substantially manual process, but one that is a necessary prerequisite for data sharing. The requirement for transferring topology, imposed by the use of links and nodes to connect traversals (a.k.a., linear transportation features) to the datum and thence to the cartography places a still greater burden on data sharing.

6 Transportation Data Participants

The enterprise GIS-T data model provides participants in data sharing a framework for clarifying roles. Departments of transportation or public works are organizations that have ownership and maintenance responsibilities for transportation infrastructure. They are data producers, data integrators, and data users both for internal uses and for other organizations. Internal uses vary considerably. Planning, design, construction, and maintenance divisions re substantially independent of one another. The solution is not a single GIS-T, but an enterprise approach to GIS-T data.

Motorists and the general public are primarily data users. Organizations that use the transportation system in their business, the police, delivery services, etc. often rely on data integrators to provide transportation data in the form of maps and networks for location, path finding, and routing. Increasingly, users are demanding current, logically correct, and spatially accurate transportation data in interoperable digital form for large regions that span many jurisdictions. Currently there are no technical and institutional processes to achieve a single integrated database to handle those diverse needs. Nor is it likely that such processes will be developed and sustained. Rather, principles need to be established to guide data sharing and the development of application-specific transportation databases that can be assembled without costly redundant recollection of source data and updates from the field each time.

There are two participants whose accuracy requirements drive the data sharing process. Others have less demanding needs for temporal accuracy (currency), completeness, and spatial accuracy:

- Emergency management, E-9-1-1, and computer-aided (emergency) dispatch (CAD) have the most demanding need for currency and completeness.
- Vehicle navigation applications, which may include CAD, have the most demanding need for spatial accuracy of street centerline files. This is sometimes referred to as “map matching” of GPS-derived location of vehicles to the correct street or road in the road database. Identifying the correct ramp of a complex freeway interchange on which a disabled vehicle is located is a particularly demanding task. Similarly, electronic (ITS) toll collection applications may require tracking vehicles by lane of multiple-lane facilities.

7 Transportation Data Sharing Principles

Successful data sharing requires a common schema or data model that is flexible enough to handle the needs of diverse participants. The flow of data from providers to users calls for capturing data once and delivery to users in various forms, time frames, and spatial scales.

Principles for successful sharing of transportation data among participants must address a variety of issues: definition and identification of transportation features, cartography and spatial accuracy, generation of application-specific network representations, and interoperability.

Two important principles follow from the GIS-T data model:

- Transportation features are bounded by jurisdictions, not intersections. This is not to say that the underlying cartography could not have other forms, only that the link between attribute data and cartography must occur through transportation features rather than spatial primitives or network topology.
- Attributes of transportation features are represented as linear or point events and are located along the feature using linear referencing.

These two principles enable longer transportation features than is the case in link-based networks, and reduces the number of transportation features that must be maintained to represent the system. Adding network detail or additional attributes does not necessarily increase the number of features. Additional detail can be added by linearly referenced event tables and analyzed and visualized using dynamic segmentation.

Butler and Dueker (1998) also identified important data sharing principles:

- Transportation features must be uniquely identified to facilitate sharing of data among participants. Participants need to identify common features in sharing data.
- Transportation data producers need to include a standardized unique identifier with each transportation feature.

The latter principle leads to subsidiary principles for assignment of identifiers:

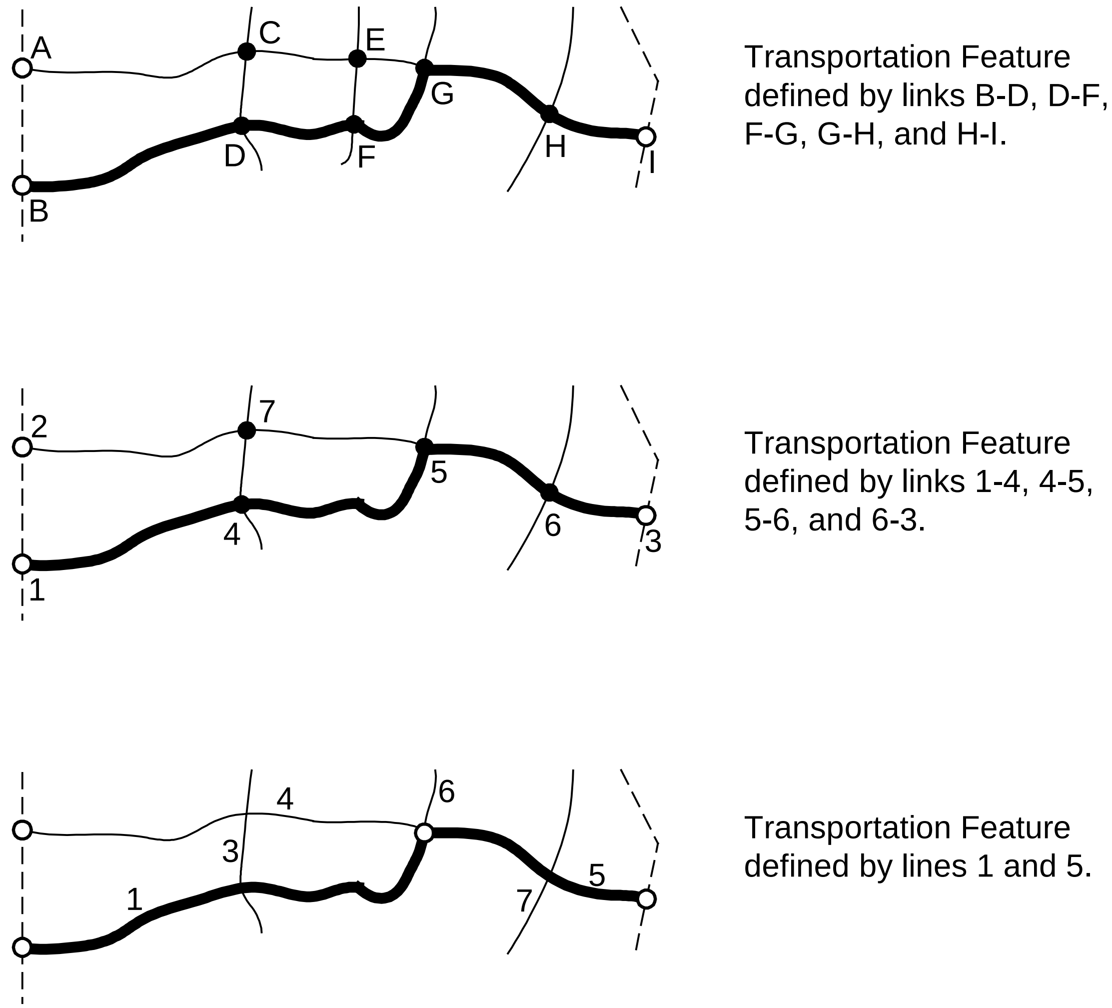
- Segment major arterial facilities at jurisdictional boundaries or major intersections if consistency with traffic assignment networks is desired.
- Collect minor road facilities by street name or route number to minimize the number of unique identifiers. But do not use names or routes as identifiers. They change.

There are several other principles that are offered to reduce the amount of manual coding and conflation, and thereby ease compliance with the data sharing principles. These are offered to avoid the need for simultaneous conflation of cartography and topology with a process to resolve inconsistent segments:

- Exchange attribute data as event tables for logical transportation features, i.e., without shape points.
- Exchange cartography without topology.
- There is no need to code topology, let the GIS generate application-specific networks from a selection of appropriate transportation features.
- Minimize manual coding of transportation feature identifiers by embedding existing identifiers into more global identifiers and using scripts to bulk-assign state and county codes.

Figure 2 shows how using a transportation feature-based identifier to exchange data avoids most of the problems of conflating cartographic and topological objects. In this example, the heavy line represents the common transportation feature in all three schemas. The top and middle versions use topological (link/node) internal data structures but with different sets of links and nodes (the top version has more). The bottom version uses simple line strings. From an implementation perspective, the top two might be Arc/Info-based coverages, while the bottom is representative of line strings used in GeoMedia, ArcSDE, and ArcView. Dependence on conflation to make the cartography and topology the same must proceed the practical sharing of data. However, if all three had previously combined their primitive objects to create the higher-level transportation feature and given it a common identifier, then there would be no need to make the maps and internal structures the same. The shape of the line, the scale of the map, the number of line segments, the topological structure (if any), and other implementation-specific issues would disappear. Data sharing would be direct and immediate.

Figure 2. *Comparison of cartographic and topological model approaches.*



7.1 Issues of Definition and Identification of Transportation Features

A transportation feature can be like a point (interchange or bridge), a line (road or railroad), or an area (rail yard or airport). Nevertheless some applications may restrict their databases to roads, pedestrian paths, or waterways. The important point is to code the type of transportation feature so that the type can be used to select those features of common interest for sharing of data.

Butler and Dueker (1998) proposed an internet-like address identifier for transportation features. Similarly, the Oregon Road Base Information Team Subcommittee (ORBITS) (Wuest, Dueker, and Bosworth, 1998) and the NSDI Framework Transportation Identification Standard (FGDC, 1999) have proposed a roadway identifier schema. The NSDI proposal will likely prevail when it is complete.

The purpose of assigning a stable and unique identifier to each transportation feature is to eliminate or reduce reliance on traditional conflation processes to reconcile different transportation databases. Unique identifiers are used to match transportation features between databases without relying on matching coordinates and links.

A case study was conducted to test methods of assigning both the ORBITS and NSDI identification codes. The ORBITS approach collects or divides roadway features and identifies them with a unique code. In the context of the case study, decision rules for breaking or collecting roadway sections and procedures for bulk assignment of higher level codes to sequenced numbered roadway features were developed. The assignment of NSDI codes to roadway features was similar, except that point identifiers were also assigned to beginning and ending points of roadway features. The ORBITS team chose not to code the topology, leaving that to be generated if needed, as it is too application specific to be of general use.

The ORBITS case study developed different decision rules for assignment of transportation feature identifiers to arterial roads and local roads. Urban arterial roads are segmented at intersections of major arterials. Arterial road identifiers in urban counties are a concatenation of state and county FIPS codes and a concatenation of *i* and *j* traffic assignment network node numbers. In rural counties we would recommend arterial roads be assigned a code that is the concatenation state and county FIPS codes and the state DOT or county DOT road identifier. Portland Metro desired a finer breakdown of uniquely identified major roads than the rural rule would have accomplished.

The decision rule for assigning codes to local roads is to collect connected TIGER lines that have the same name and assign them sequence numbers concatenated with state and county FIPS codes. Some judgment has to be applied to deal with interruptions in connectedness of local streets. When there are minor interruptions, the same code is assigned to the local road with the common name. When interruptions are more than minor, separate identifiers are assigned. Also, there are situations where name changes occur arbitrarily, such as at municipal boundaries where a different identifier may not be needed (See Bender, Bosworth, and Dueker, (1999) for a description of a case study of assigning NSDI identifiers).

Even with the collection of links of legacy databases into larger transportation features, the assignment of identifiers is an onerous task, especially if mandated without assistance. Transportation organizations may not be willing to undertake such an effort in the absence of a payoff or need. A multi-participant project that requires integration of data from various sources may be needed to provide the reason and incentive to assign transportation feature identifiers. The usual approach relies on a single transportation organization to conflate linework from various sources. This may be an adequate solution to build a spatial database, but not for maintenance over the long term.

In the absence of identifiers, projects that involve massive data integration are very difficult to update and maintain. Thus, the identifier approach is particularly crucial to apply to roadway additions, deletions, and modifications. The key to database maintenance is to manage the incorporation of updated transportation features. The fundamental strategy is to identify features in the database to facilitate a transactional update system, one that does not require rebuilding the entire database anew.

7.2 Implications for the NSDI Proposal

The present draft NSDI Framework Transportation Identification Standard (FGDC, 1999) satisfies many of the stated business rules of our enterprise GIS-T data model. But the proposed standard is not based on a formal data model and as a result suffers from ambiguities. Framework Transportation Segments (FTSegs) have a unique identifier and like anchor sections, FTSegs are defined by beginning and ending points (Framework Transportation Reference Points, or FTRPs). Without a formal data model the NSDI proposal lacks guidance on whether length of FTSegs is mandatory or optional, or how to code an interchange.

The implicit topology of FTSegs, created by stating the terminal FTRPs, is equivalent to that in our Anchor Point/Anchor Section structure. The NSDI proposal requires an intermediate FTRP to be defined if there are multiple paths between the two FTRPs defining a given FTSeg. The explicit topology of FTSegs and intermediate FTRPs—which do not create new segments but are located using a distance offset along the segment—is essentially the same as our Point Event. Both intermediate FTRPs and explicit topology can be expressed using our Point Event entity.

A continuing problem with the NSDI proposal is the need to serve both logical and physical descriptions of the transportation network. Some users may need to use a single FTRP to represent an entire interchange that for another user may require a dozen of more FTRPs. We have proposed the use of an Intersection entity to serve as a general, or logical, object for what may be one or more physical position descriptions for all or part of a complex structure.

The NSDI proposal meets the transportation data sharing need for unique and public identifiers. It almost meets the requirements of a national transportation datum, such as has been listed as a prerequisite for many ITS deployments, needing only to make length a mandatory attribute. It almost provides a clear statement of network topology, falling short only in its ambiguity of connection specification.

And, yet, we are not satisfied. Data users need to exchange not only attributes relevant to an entire transportation feature, but also those that apply to a portion of a feature. A proposed standard structure for constructing a universal transportation data exchange format seems necessary for people to be able to exchange entire data sets, not simply identify which road they are talking about. The absence of linework to represent the FTSegs is a serious omission in a “spatial” standard, in the eyes of many potential users. For it is the need for more accurate linework that drive most spatial data sharing needs. But more accurate linework is accompanied by a more detailed representation of transportation features. Yet the NSDI Framework Transportation Identification Standard is a major effort in the process of forging a true data exchange mechanism.

7.3 Cartography and Spatial Accuracy Issues

The problem of sharing transportation data is illustrated by issues states face in constructing a roads layer for a statewide GIS. The problem is stitching together data from various sources and vintages.

Typically, state DOTs have a roadway database for highway inventory. Attributes of roads are recorded by milepoint and associated with a straight-line chart for visualization. Some states have incorporated the linearly referenced data into a GIS and use dynamic segmentation for analysis and visualization. The spatial scale of the cartography ranges from1:24K to 1:100K.

Similarly, the spatial scale of the road layer used by natural resource agencies is from 1:24K or 1:100K USGS sources, but with very little attribution and with uneven currency. However, the USGS digital orthophotography quarter-quadrangle program offers the opportunity for states to update their road vectors and register them using 1:12K imagery.[3] This ought to provide sufficient spatial and temporal accuracy for E-9-1-1 and for vehicle navigation (snapping vehicle GPS tracking data to road vectors). However, these sources may not be sufficiently accurate to distinguish road lanes, which are needed in urban areas for dynamic vehicle navigation and road pricing (e.g., snapping vehicle GPS tracking data to lanes and ramps to relate to lane-specific volumes and speeds from loops, imaging and vehicle probes).

7.4 Spatial Database Completeness and Currency Issues

Vehicle tracking will require proper positioning, both in terms of which road the vehicle is on and where it is on that road. The ITS community has proposed a “cross-streets” profile that provides this information in a message format that includes the name of the current street and the location as the terminal cross streets for the present block on which the vehicle is located. One reason for this approach is to avoid the need to have precise GPS-map concurrence regarding spatial position. Researchers have discovered, however, that current maps do not have sufficiently reliable and complete street name attributes for this schema to be routinely implemented. (Noronha, 1999) Currently, the ITS database community is proposing to use coordinates of intersections as well as street addresses, and may add street type for locational referencing.

If work must be done to populate databases with sufficient information to identify which street and block a vehicle is traversing, then it seems appropriate to develop a more complete approach that avoids the remaining problems, such as differences in spelling that may arise for street names in different databases. The purpose of a transportation

[3] Some regions have more accurate orthophotography.

feature identification schema (perhaps with a redundant cross street index) is to insure completeness and currency of databases. Development of an identification schema requires guidelines or standards for segmenting of arterial roads, insuring completeness of local road segments and collecting them into larger chunks for assignment of an identifier. Similarly, development of standards for coding ramps and lanes is needed. But most importantly, a typology of transportation features needs to be developed to accommodate different definitions of roads and non-road features for different applications. This is particularly important where databases are developed by means of the vertical integration of databases from different organizations for the same geographic area or jurisdiction, say integrating state, county and city data.

Identifiers for transportation features facilitate transactional maintenance, additions, deletions, and changes to transportation features with periodic issuance of new editions for less time-sensitive applications. Real-time users will require dynamic changes issued as linear and point events, or modifications to links and nodes to reflect lane/street closures or construction detours.

7.5 Network Issues

In-vehicle navigation systems will provide the greatest challenge in terms of spatial and temporal accuracy for road map databases. Current technology supports generalized client-based networks for minimum path routing (based on typical speeds or impedances) that produces instructions in terms of street names and turns, based on a road map base that snaps GPS vehicle-tracking data to road vectors.

In the near future we are likely to see detailed server-based dynamic routing based on current network traffic conditions with instructions including ramp and signage details and snapping of GPS vehicle tracking data to lanes and ramps. The coding of topology using formal and widely recognized transportation feature identifiers will allow vehicle routing to be done without reliance on maps.

The chief problem with transportation networks is the perception that there is one base network that will satisfy all applications. We contend this is a false premise, as someone will always want to add or delete links. Networks ought to be application specific, and consequently, the network cannot be the building block of sharable or interoperable transportation data.

7.6 Interoperability Issues

Transportation feature identifiers provide a common key by which to relate data to achieve interoperability among transportation databases. Nevertheless relating databases on the fly may not perform well for real-time applications. Performance may be a problem in relating to a highway inventory database to check underpass clearance for a dynamic pathfinding application for rerouting traffic due to emergency incidents. Instead, clearances may need to be pre-computed and stored as link attributes in the dynamic pathfinding application.

Full interoperability suggests "plug and play," meaning Vendor A's data can be read by Vendor B's system and vice versa. In transportation this will be difficult to achieve because of the varied nature of applications that require data in specific forms, and the size of typical regions for which consistent data would be needed.

Not only is it difficult to achieve consistent and accurate data for a region as large as the nation, consider the temporal data streams that will be created by vehicle tracking systems and by video cameras used to estimate vehicle flow rates. This stream of data will require format standards. (See Dailey *et al.* (1999) for a self-describing method for transfer of data in real time for ITS applications.)

8 In Conclusion

Sharing of GIS-T data poses challenges. This paper identified the issues and developed a framework and principles to address them. The central principle is the establishment of a schema for transportation features and their identifiers. An underlying principle is the need for a common data model that holds transportation features as the object of interest, and that attributes of transportation features are represented as linear and point events that are located along the feature using linear referencing. Until there is agreement on these principles, data sharing and interoperability will not progress well. This lack of agreement stems from the current state of flux with respect to GIS-T data models. This problem should diminish with the completion of the NCHRP 20-27(3) project and final adoption of the NSDI Framework Transportation Identification Standard.

In this context, sharing of transportation data involves exchange of relevant transportation features and events, not links and nodes of application-specific databases. This is a major departure from the existing process of conflation. The exchange of more fundamental features is encouraged in recognition that each application has quite specific requirements for their end-use database, but all users have need for basic transportation features. The major contribution of our enterprise GIS-T data model is this separation of the maintenance database from many applications databases. The term enterprise takes on a broader meaning of a shared stewardship of data about the larger transportation system.

Strategies for the sharing of transportation features follow from this approach. The first is to enlist state and local cooperators to construct transportation features by registering existing transportation vector data from TIGER and local sources to digital orthophotography, such as the USGS orthophotography quarter-quadrangles. A second stage would by to update the vectors using replications of vehicle tracking data. The fundamental strategy is to identify features in the database to facilitate a transactional update system, one that does not require rebuilding the entire database anew.

Although this approach to the sharing of transportation data needs to be refined, it provides a better framework than currently exists. There is no common approach among the communities of ITS, vehicle navigation database vendors, NSDI, state and local transportation organizations, and E-911. We are encouraged, however, by the continuing efforts of these groups to reach a more consistent approach that will facilitate data sharing by reducing the number of inconsistent and duplicative representations.

Transportation features are the fundamental objects in the database and must be uniquely and permanently identified. Similarly, additions, deletions, and modifications to transportation features must be identified to facilitate database updates and maintenance. This approach, or data model, separates the update and maintenance of transportation features from the links of networks used in specific applications. Thus, different networks can be generated from a common set of transportation features and one update process can support a set of applications.

References

Adams, T.M., A.P. Vonderohe, & J.A. Butler, 1998. Multimodal, multidimensional location referencing system modeling issues. Prepared for NCHRP 20-27(3) Workshop on Functional Specifications for Multimodal, Multidimensional Transportation Location Referencing Systems, Washington, DC, December 3-5, 1998.

Bender, P., M. Bosworth, & K. Dueker, 1999. ORBIT: The Oregon Road Base Information Team: the Canby case study report. Center for Urban Studies, Portland State University, Catalog Number PR111, http://www.upa.pdx.edu/CUS/PUBS/contents.html

Bosworth, M., K. Dueker, & P. Wuest, 1998. ORBIT: The Oregon Road Base Information Team; a draft summary report. Center for Urban Studies, Portland State University, Catalog Number PR106, 19 pages. http://www.upa.pdx.edu/CUS/PUBS/contents.html

Brown, J., A. Rao, and J. Baran, 1995. Are you conflated? Integrating TIGER and other data sets through automated network conflation. In GIS-T Symposium Proceedings, American Association of State Highway and Transportation Officials, 220-229.

Butler, J. A., & K. Dueker, 1998. A proposed method of transportation feature identification. Center for Urban Studies, Portland State University, Catalog Number DP97-8, 18 pages. http://www.upa.pdx.edu/CUS/PUBS/contents.html

Chan, K., 1998. DIGEST: A primer for the international GIS standard. CRC Press.

Dailey, D, D Meyers, & N. Friedland, 1999. A self describing data transfer methodology for ITS applications. Preprint CD-ROM. 78th Annual Meeting, January 1999, Transportation Research Board.

Dueker, K., & J.A. Butler, 1998. GIS-T enterprise data model with suggested implementation choices. Journal of the Urban and Regional Information Systems Association. 10,1,12-36.

Devogele, T.,C. Parent, and S. Spaccapietra, 1998. On spatial database integration. International Journal of Geographic Information Sciences. 12,4, 335-352.

FGDC: Federal Geographic Data Committee (Ground Transportation Subcommittee), 1999. NSDI Framework Transportation Identification Standard—Working Draft. May 1999. http://www.bts.gov/gis/fgdc/web_intr.html

GDF: Geographic Data Files Standard, Version 3.0, 1995. Working Group 7.2 of CEN Technical Committee 278, European Committee for Standardization.

GDF: Geographic Data Files Standard, Version 5.0, 1999. ISO/TC 204/SC/WG3. ISO/WD 19990722-2.

Noronha, V., 1999. Spatial data interoperability for ITS. Presentation at Annual Meeting of the Transportation Research Board, Washington DC.

Sester, M., K. Anders, and V. Walker, 1998. Linking objects of different spatial data sets by integration and aggregation. GeoInformatica. 2,4, 335-358.

Sperling, J., & S. Sharp, 1999. A prototype cooperative effort to enhance TIGER. Journal of the Urban and Regional Information Systems Association. 11,2, 35-42.

Sutton, J., 1999. Object-oriented network data structures: the road to interoperability." Presentation at GIS-T Symposium, April 1999, San Diego.

Tomaselli, L., 1994. Topological transfer: evolving linear GIS accuracy. URISA Proceedings. 245-259.

Vonderohe, A., C. Chou, F. Sun, and T. Adams, 1998. A generic data model for linear referencing systems. Research Results Digest Number 218. National Cooperative Highway Research Program, Transportation Research Board.

Vonderohe, A., & T. Hepworth, 1998. A methodology for design of measurement systems for linear referencing. . Journal of the Urban and Regional Information Systems Association. 10,1, 48-56.

Walter, V., and D. Fritsch, 1999. Matching spatial data sets: a statistical approach. International Journal of Geographic Information Sciences. 13,5, 445-473.