

# **A PROPOSED METHOD OF TRANSPORTATION FEATURE IDENTIFICATION**

J. Allison Butler  
Kenneth J. Dueker

January 1998

Prepared for Presentation at the 1998  
Transportation Research Board Annual Meeting

Discussion Paper 97-8

Center for Urban Studies  
College of Urban and Public Affairs  
Portland State University  
Portland, OR 97207-0751  
(503) 725-4020  
(503) 725-8480 FAX  
<http://www.upa.pdx.edu/CUS/>

## **INTRODUCTION**

Geographic information systems (GIS) are being increasingly deployed by transportation agencies to help them display, review, and utilize data. The primary items of interest are transportation facilities and services, which may take the form of highways, airports, bus routes, and seaports, among others. Using GIS software, transportation facilities are represented as geometric shapes; i.e., points, lines, and areas. However, it is increasingly apparent to GIS users in the field of transportation that a geometry-based approach is not sufficient.

The offered solution is to develop a feature-based GIS approach for transportation. The central requirement of such an approach is to have an unambiguously identified set of transportation features and a means of locating related features and attributes on or adjacent to the transportation features. For linear transportation features, the common location referencing system (LRS) is also linear. Usually based on milepoint offsets from a beginning point, linear LRSs are widely used by transportation agencies to locate data stored in numerous legacy databases. Some agencies may have multiple linear LRSs for various applications or topical areas. Data users unfamiliar with linear LRSs have difficulty using highway data provided by transportation agencies, and the correlation of linear and non-linear LRSs is generally quite difficult.

The result is that sharing of digital road map databases within and among organizations is difficult since there are no consistent ways of representing transportation features, and different decision rules exist as to what features to include and how they are identified. Numerous efforts are underway to develop a national standard for linear LRS design and implementation, including a set of field procedures to ensure that needed accuracy is achieved.

This paper will describe a means to unambiguously define transportation features using a simple set of naming rules. These rules will support linear and other forms of LRS for all types of transportation features, including airports, seaports, railroads, highways, transit services, bridges, intersections, signs, and related facilities and services. Such features may be displayed at multiple scales as point, line, and area graphical objects.

The primary naming rules follow the same geographically based logic as do those for internet addresses, with successive portions of the name providing a more refined location identifier. Roads are named on a county basis, with airports, seaports, and other regional types of facilities being named on a state basis. The resulting names can serve as foreign keys in relational databases.

## **THE CONTEXT FOR DEFINING TRANSPORTATION FEATURES**

Managers of highway infrastructure think in terms of routes, road sections, and cartographic strings, while users think of vehicles operating on paths in networks from origins to destinations described. The need is to develop data models to encompass these perspectives of transportation systems. Fletcher, et al. (1996) articulates this community of perspectives for digital road map databases. Gordon (1996) elaborated on intelligent transportation system (ITS) needs for interoperable systems which include a comprehensive framework to handle multiple methods of referencing location in rich heterogeneous databases. Most recently, the authors of this paper have published a proposed universal data model for transportation (Dueker and Butler, 1997). All of these papers address geographic information systems for transportation (GIS-T).

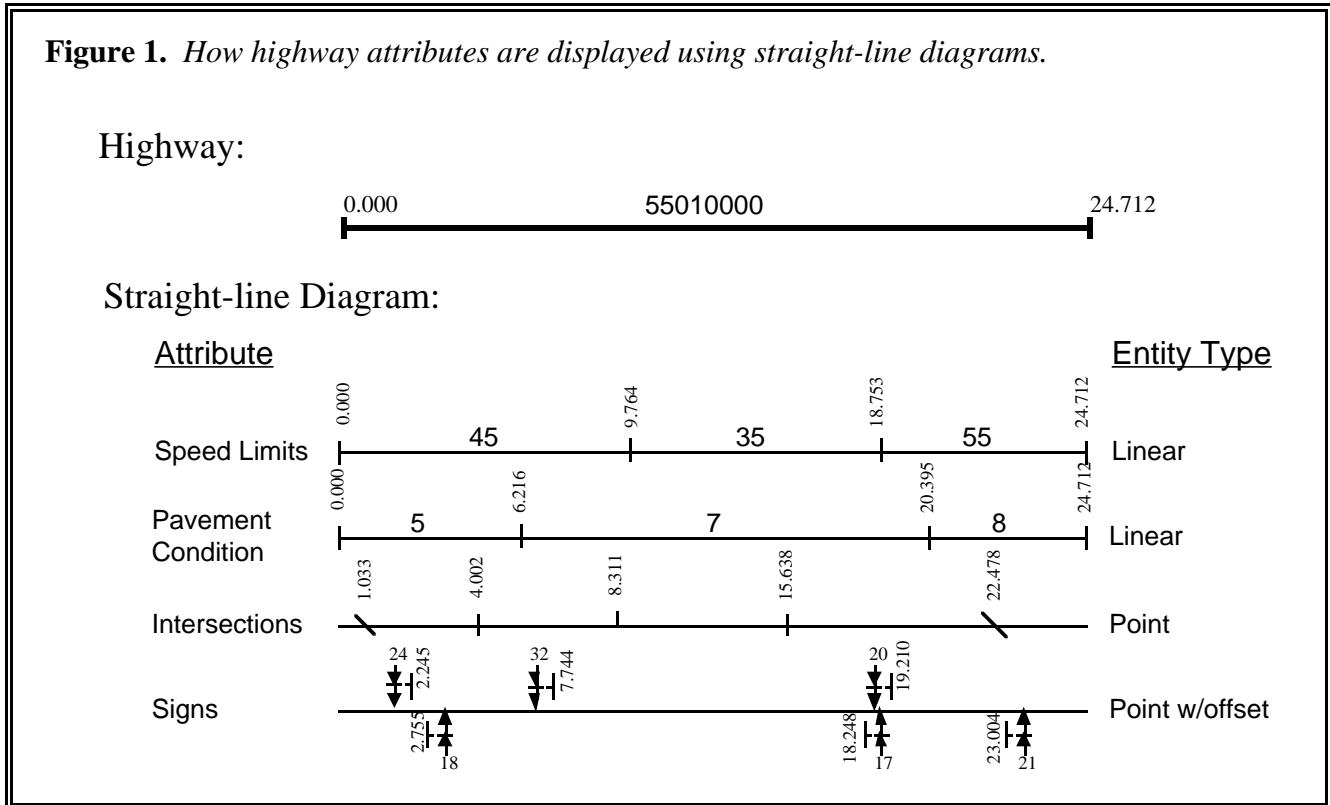
Linear and point events are elements or characteristics of a transportation feature. Elements include tangible objects, such as bridges, signs, guardrails, and intersections. Characteristics include less tangible aspects of a transportation feature, such as a road's speed limit, pavement surface type, and type and width of median; the airlines serving a particular air route; or the trains using a specific railroad track.

Most state departments of transportation use straight-line diagrams and related attribute databases to graphically describe highways. In many ways, state DOTs look at GIS-T as an

evolutionary step that puts true shape into these diagrams and allows connections (topology) between what were previously separate diagrams.

The linear LRS is the glue that binds transportation features to their linear and point events, as well as to the geographic datum of anchor points and anchor sections that place the transportation features on the surface of the Earth. (See Vonderohe and Hepworth, 1996, or Dueker and Butler, 1997, for a discussion of anchor points and sections.) Of course, not all transportation features are on the Earth, with aviation being the primary example. Linear LRSs may be applicable to aviation, but since the air routes between airports are more conceptual than physical, the aviation system may be one mode of transportation to which this glue does not stick.

**Figure 1.** How highway attributes are displayed using straight-line diagrams.



In Figure 1, a highway has been visually described in several ways. The first is simply a straight line representing the transportation feature as a single entity. This entity has been given the identifying number 55010000. To this simple representation, a straight-line diagram will add attributes from a transportation database. The examples shown include point and linear events both on and adjacent to the road. The following tables show the data used to create the straight-line diagram:

*Speed Limit Table*

<u>ROADWAY ID</u>	<u>BEGIN MP</u>	<u>END MP</u>	<u>SPEED LIMIT</u>
55010000	0.000	2.640	55
55010000	2.640	6.486	50
55010000	6.486	9.764	45
55010000	9.764	18.753	35
55010000	18.753	24.712	55

*Pavement Condition Table*

<u>ROADWAY ID</u>	<u>BEGIN MP</u>	<u>END MP</u>	<u>CONDITION</u>
55010000	0.000	6.216	5
55010000	6.216	12.984	7
55010000	12.984	14.539	6
55010000	14.539	18.044	7
55010000	18.044	20.395	6
55010000	20.395	24.712	8

*Intersection Table*

<u>ROADWAY ID</u>	<u>MP</u>	<u>ANGLE</u>	<u>TYPE</u>
55010000	1.033	45	cross
55010000	4.022	90	cross
55010000	8.311	90	left tee
55010000	15.638	90	cross
55010000	22.478	45	cross

*Sign Table*

<u>ROADWAY ID</u>	<u>MP</u>	<u>SIDE</u>	<u>FACING</u>	<u>OFFSET</u>
55010000	2.245	left	westbound	24
55010000	2.755	right	eastbound	18
55010000	7.744	left	westbound	32
55010000	18.248	right	eastbound	17
55010000	19.210	left	westbound	20
55010000	23.004	right	eastbound	21

Some attributes are expressed as annotations (e.g., milepoints), others as geometry (e.g., angles of intersecting streets). Straight-line diagrams may show these attributes and objects on separate lines, as depicted here, or all on one line using various graphical methods. These methods include line width, line pattern, and placing tic marks across the road and labeling both sides of the mark to show what value changes at that point. Additional attributes could be provided for the entities shown; e.g., intersecting street name, sign legend, etc.

GIS-T is a means of providing the data content of straight-line diagrams on a more graphically correct depiction of not just one highway but all the highways in a given geographic region. Dynamic segmentation uses the linear LRS to place attributes on a portion of the cartographic string that describes the entire highway, thereby creating new strings that corresponded to the extents of the component linear attributes. Dynamic segmentation can also be used to place point symbols at the correct relative position on the highway string. Of course, there is a limit on the number of attributes that can be shown on a single map given the need to have each clearly conveyed.

**A Universal Transportation Data Model**

Individual agencies can use whatever internal methods they desire to segment and attribute highways and other transportation features. However, such customized approaches preclude easy data sharing with other potential users. A necessary first step in facilitating data sharing is to adopt a uniform way of identifying transportation facilities. One can translate user needs to a data model as a first step to reach a consensus on transportation database design and data sharing standards.

Data sharing is a concept of decentralized control over data resources. Consequently, there is a need for a robust data model to represent the complex relationships among the components of transportation systems. This data model must be able to support legacy databases, future

enhancements, and database maintenance. The purpose of this section is to examine issues of sharing digital transportation map databases and to propose a data model that can accommodate different applications. What is proposed is best characterized as a GIS-T enterprise data model suitable for organizations responsible for maintaining transportation systems. The basic data model for existing highway databases maintained by transportation agencies using a linear LRS is shown in Figure 2.

**Figure 2.** *The basic non-link transportation data model.*

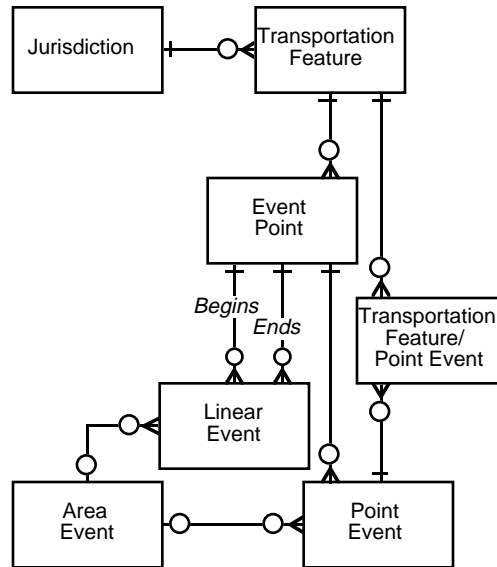


Figure 2 includes seven entities, including three subtypes of a generic event. The definitions of these entities are:

- **Jurisdiction.** *The political or other context for designating transportation features and their names, which may be merely numerical references unique within the jurisdiction.* Jurisdiction need not be the same for all transportation feature types. Airports can be named on a national basis, with streets named on a local basis.
- **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). A more detailed explanation is provided in the next section.
- **Event Point.** *The location where an event occurs.* Event Point is defined initially as an offset distance from the beginning of the 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 guardrailings, 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 location where some transportation feature or attribute occurs as defined by a single event point.* Point events may occur independently, or on transportation features. Examples include bridges, intersections, traffic counting sites, and similar point-like features.

Some of these point-like features may alternatively or additionally be represented as linear or area events; e.g., at larger scales one may choose to show bridges as linear events or airports as area events. Point events may have real-world locations and positions along a linear transportation feature and anywhere else within the involved geographic or cartographic space. Point events occurring as part of transportation features may include those offset from a road edge, such as a sign located at some distance from the road edge.

- **Linear Event.** *An attribute of a transportation feature that has distinct beginning and ending event points (i.e., length), or a means by which to relate attributes to part or all of a transportation feature.* Linear events are defined by two event points (beginning and ending). Linear events may occur only on or alongside linear transportation features. Linear events include such attributes as functional class, speed limit, pavement type, and traffic volume. For such transportation features as air routes, linear events may be air carrier flight numbers, passenger traffic, and traffic control responsibility. Linear events occurring as part of transportation features may include attributes offset from a road edge, such as a guardrail or fence.
- **Area Event.** *A transportation feature component or a non-transportation entity that affects a transportation feature.* Areas can be explicitly represented as polygons or according to where they intersect transportation features. The implicit option is called an area event and is represented through related area, linear, and point events. Area events may be applicable for any kind of transportation feature. For example, an area event could be a city. The city could be expressed by creating a linear event for the portion of a linear transportation feature located within it, or as point events where the city limits cross a linear 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).
- **Transportation Feature/Point Event.** *An entity that resolves the many-to-many relationship between transportation features and point events.* Many-to-many relationships cannot be directly represented in a relational database since each resulting relationship cannot be uniquely identified in a single record. This requires the creation of an entity to store the relationships. The many-to-many relationship occurs because transportation features can intersect and may do so at multiple points. Transportation features may include one or more point events. A given point event may represent the intersection of two or more transportation features. Collectively, then, transportation features and point events can be grouped to form junctions, a term which also needs to be defined:
  - **Junction.** *A location where two or more transportation features cross or connect.* The term includes both generic intersections, such as where two streets cross, and where the unique identifier of contiguous transportation features change, such as at a jurisdictional boundary. The term also includes the connection between different transportation modes, as in the case of an airport being the junction of ground and aerial transportation facilities or services. The concept of junction may also be applied to places where transportation features cross but do not intersect to meet application needs. For example, a bridge over a navigable waterway could be a non-intersection junction between the highway and waterway networks.

The use of point and linear events with dynamic segmentation precludes the need to aggregate attributes in accordance with a rigid transportation feature segmentation schema, e.g., link/node or fixed distance. To relate attributes defined as point and linear events to geometric representations of transportation features requires that the relevant geometric strings have the same end points and length measure (at scale) as the transportation features they represent. They also must have the same transportation feature identifier in order to serve as a foreign key that allows attribute data to be tied to the relevant cartographic objects. This enables interpolation along strings to locate point and linear events using dynamic segmentation functions in software. Some transportation databases support lateral offset measurements for events such as guardrail which may not be adjacent to the road. Others may have data stored by side of road, as in the case of

divided highways, in order to show attributes such as number of lanes, pavement type, and curvature. Dynamic segmentation supports these data practices.

As implied by the junction definition, some transportation system elements "belong" to more than one transportation feature. For example, a bridge could be seen as part of the facility it carries and the one it crosses to form a junction (note that one need not be able to move from one transportation feature to another at a junction with the definition proposed here). Thus, junctions formed by the intersection or crossing of transportation features of different types, such as rail-highway grade crossings, may be viewed as the set of multimodal objects. Paths through the transportation system may move from one mode to the next only at such junctions. Of course, junctions formed by the intersection of transportation features of the same type, such as two intersecting streets, serve to move one along a path through that particular network. The database should be designed to store information about junctions without duplication.

The model shows that transportation features are defined within the context of a jurisdictional coverage, such as of a national, state, or local governmental agency. Event points locate the position of point and linear events—features and attributes—of linear transportation features, typically using a linear LRS with a point of origin defined at the jurisdictional boundary or road beginning point. Linear events are defined using a pair of event points, while point events need only one to be positionally defined. Area events are not directly expressed, but are instead stored as a linear event (segment of road within an area) or as a set of point events (where the area boundary crosses each road). Junctions represent the location of intersections of transportation features. Junctions may be actual intersections or such other key features as bridges. Junctions are represented by a set of event points and point events, one on each involved transportation feature.

### **Sample Database Implementation**

Figure 3 on the next page shows a basic relational database structure that could be used to implement the data model of Figure 2 for a highway inventory application. Each box represents a table in the database, with its name given at the top (shaded area). A table's primary key, or set of data elements that defines a unique address for each row, is underlined. Optional data elements are shown in brackets. Notice that there is not a one-to-one relationship between data model entities and database tables. Multiple Event Tables may be created to serve various purposes.

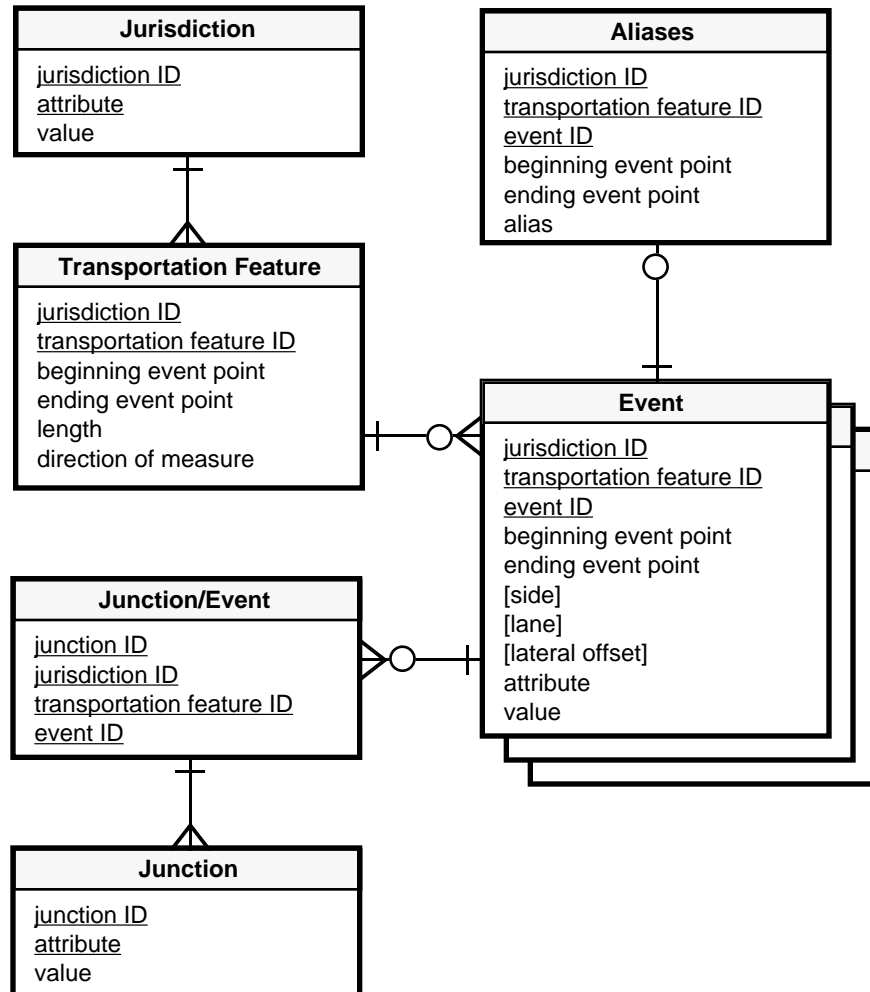
The Jurisdiction Table stores data that apply to the entire jurisdiction in which the specified transportation features are defined. For this example, let's assume the jurisdiction level is county. Thus, the Jurisdiction Table stores information about each county in the state, such as its population, the DOT District in which it is located, and so forth. A table could also be created to store a list of all the transportation features located within each county.

In the same way, the Transportation Feature Table stores information that applies to an entire transportation feature, which in this example is the extent of a road in a given county. Transportation feature ID is a number. Since the transportation feature ID is unique only within a jurisdiction, then both the jurisdiction ID and the transportation feature ID are needed to uniquely define a specific transportation feature. An event point is the linear LRS measure for a location on a transportation feature. The beginning event point is the origin measure, usually zero. The ending event point is the highest possible measure; i.e., the end of the feature. Transportation feature length must be explicitly stated as an attribute since it may not be the simple mathematical difference between the two defining event points. It is also a good data quality check.

A separate Aliases Table has been created to store the various names by which all or a portion of the transportation feature may be known. One could alternatively create a view, or virtual table, using the linear event records storing the names assigned to road segments. The primary key for this table is rather complex since more than one name may begin at a single point, e.g., a county road number and local name may both start at the county line. Note that there was no specific alias entity in the model. The decision to include a table specifically for this kind of linear event is the type of decision one must frequently make to implement a conceptual data model. Since most people know the name of a highway segment, the Aliases Table could be a convenient way to use street name as a foreign key to locate the correct road of interest and its descriptive data.

A foreign key is a data element present in one table that can be used to connect to related records in another table in the database. The Aliases Table can be omitted and Event Table records used, or the Aliases Table can be constructed as an extract (subset) of the Event Table; the latter option is preferable.

**Figure 3.** A sample relational database design implementing the model in Figure 1.



The physical data model in Figure 3 shows that there may be many Event Tables, all with certain common data elements. A common structure facilitates the combination of various event tables to describe a segment of highway, so even point event tables should include an ending point reference column (which will have null values). Some tables may optionally include information on lane, side, and offset to accommodate such attributes as traffic counts (by lane), pavement types (by side of road), and signs (offset from road edge). An event ID is used to help uniquely identify a record in the event tables; event ID is unique only for a given transportation feature. Given this design, a junction would result in a record in the Event Table for each intersecting transportation feature, each with its own event ID.

We have elected to refer to the table associated with the Transportation Feature/Point Event resolution entity as the Junction/Event Table. Normally, all data elements in a resolution entity table should be part of its primary key, the various parts of which will provide foreign keys to relate to the appropriate records in other tables. The Event Table's primary key (jurisdiction ID, transportation feature ID, and event ID) is used in the Junction/Event Table to connect to the point event(s) associated with a junction. The intersecting transportation features may be found (without regard to precise location) by using the Transportation Feature Table's primary key.

A Junction Table stores information about the junction, such as traffic control for a road intersection. A given junction may have many attributes, and there will be one record for each intersecting transportation feature in the Junction/Event Table. This design leads to the use of a partial primary key for the Junction Table of junction ID to identify all the attribute records for a given junction. An alternative would be to put all junction attributes in a single record, thereby eliminating the need for attribute to be part of the Junction Table's primary key.

One should note that the Jurisdiction, Event, and Junction Tables are highly normalized in that they will provide storage space for any number of attributes without the need to deal with null values. An alternative would be to use a separate row (data element) for each attribute.

The Event Table's design requires that event ID be different for each attribute. This may be undesirable in the case of an event described by multiple attributes. In this case, more attribute columns (data fields) could be added, or the attribute field could be added to the primary key. The full primary key for the Event Table must be included in the Junction/Event Table, so any changes to the Event Table's primary key must be reflected in changes to the Junction/Event Table.

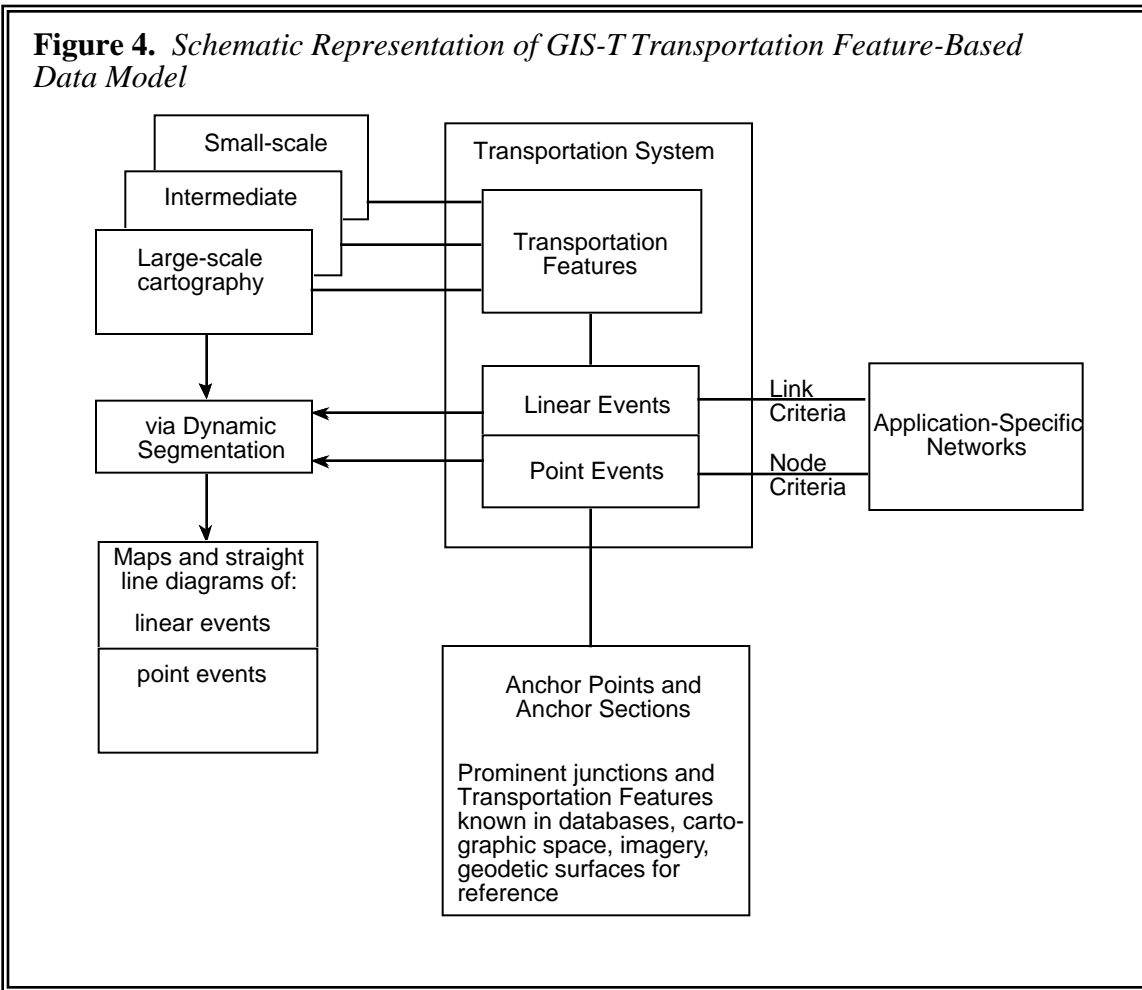
Note that an end-to-end junction is required at the jurisdiction boundary where a transportation feature is subdivided. This means that a given junction ID can be applicable to multiple jurisdictions; therefore, junction ID must be unique in the entire database. Some people may question the efficiency of creating junctions where intersections do not occur except relative to a boundary. However, this approach can increase the efficiency of compartmentalizing the database into manageable pieces, such as for base map maintenance.

Similarly, some Transportation Features, such as bridges, may cross streams that are jurisdiction boundaries. Since this happens frequently, bridge identifiers might be made unique within states rather than counties or are assigned to a specific county for identification.

Note also that unless there is some mechanism for relating linear and non-linear LRSs, non-road transportation features may not be positioned relative to the described roads and event points. A data model for airports and similar non-road transportation features may consist of only Jurisdiction and Transportation Feature. Attributes that apply to the entire feature should be included in the Transportation Feature entity of the model.

Note that events and transportation features are independent of cartographic and topologic representations. These connections are needed for use in a GIS and are made in the GIS-T data model proposed by Dueker and Butler (1997).

Figure 4 is a schematic of the Dueker-Butler GIS-T data model. It is a feature-based model rather than link-based or geometric-based. Consequently, it can handle multiple cartographic representations and can generate application-specific networks.



Implementation and testing of the GIS-T feature-based model is difficult for two reasons. First, many organizations are working with legacy data that is link- and cartographic-based, but are deficient in some ways, either out of date or geographically inaccurate. The legacy files are structured wherein the transportation system attributes, the cartography, and the network links are integrated into a single record for each network chain. Changing one part requires changing all parts and competing legacy files do not have matching network chains. They partition the transportation system into chains using different criteria.

Secondly, other organizations, seeing this mismatch, may choose to build digital street and road centerline files from scratch and avoid geometric conflation of inconsistent networks. This choice of an integrated data model is facilitated by technology wherein a GPS-based van can collect transportation attributes and cartography for network chains very efficiently by using the real world as a large digitizing tablet. The result is another unique legacy database, albeit more spatially accurate and more detailed in attributes and links, but equally difficult for sharing of data among organizations for updating purposes.

The solution inherent in the Dueker-Butler GIS-T enterprise data model is to convert legacy database network chains to linear events of Transportation Features, and then assign Transportation Feature IDs and linear referencing (milepoint values) to each of the network chains

in the legacy files. This enables link-based applications, but relies on using Transportation Features for data sharing.

### **The Problem with Link-based Approaches**

The illustrated data model does not require any transportation feature segmentation except at the jurisdiction boundary (end-to-end). Most digital transportation map databases are link based, which poses a problem for data sharing. Parties must agree on a base network and external identifiers for links to assure trouble-free data exchange. Yet, it is difficult to agree on a common base network. A more fundamental data model is needed to facilitate data sharing. Interoperability is best addressed by adopting an approach that is non-linked based and allows multiple cartographic representations.

Transportation systems are currently difficult to digitally represent and uniquely identify because of the large number of different strategies by which they can be subdivided and represented. Bridges may be point objects in one database and linear objects in another. Airports can be represented by point symbols at small scales or by runways (lines) and terminal buildings (areas) at large scales. The most difficult transportation features are highways since they are usually multiple-element objects. Cartographers segment highway features for ease of digitizing or drawing, while pavement managers segment them by type of pavement, construction engineers by project limits, and traffic engineers at intersections--but at which intersections? Usually not intersections with driveways and often not with local roads. As this illustrates, segmenting of roads is not clear cut. Consequently, road segments are not unambiguous geographic features that can be uniquely identified for purposes of maintaining interoperable digital transportation map databases. Similarly, intersections of transportation features are not easily defined and uniquely identified. These issues argue against a link-based and fixed cartography approach.

Nevertheless, attempts to share digital transportation map databases have tended to concentrate on finding a single representation of the transportation system that can be agreed to and adopted. However, this quest for a single detailed base network requires a permanent identifier for each minute link. This strategy has been unilaterally implemented by data developer organizations, such as the U.S. Bureau of the Census in developing TIGER and by Etak corp. in building MapBase 2.0. However, sharing transportation data using a TIGER or Etak involves a difficult cartographic conflation process unless link identifiers are compatible and networks are consistent. Yet most arrangements have data that are not consistent. It is doubtful that the transportation community would agree to a single highway network representation or that the National Spatial Data Infrastructure (NSDI) would adopt any single organization's representation. Other solutions are needed. The solution proposed by the authors (Dueker and Butler, 1997) is to permanently identify the more global Transportation Features and to specify types of junctions as nodes or line events as links, and then build networks from the tables of line events and junctions.

### **The Problem with Using Actual Names**

More user friendly external identifiers, such as road names with defined endpoints, have been proposed (Dueker, 1995). Common names of transportation features can be used as one criterion for identification, but transportation features need more permanent identifiers. Road names and route numbers, for example, need to be handled as aliases and/or traversals for external access to data. Such external identification of transportation features is necessary to compare and share data, irrespective of the cartographic or network representation.

However useful such an approach may be, it fails to provide a universally applicable solution. Many jurisdictions may include the same road name or slight variations in spelling may exist within a single jurisdiction. The problem with using road naming jurisdictions as a basis for road segmentation and identification is the fact that the boundaries of the municipalities often responsible for declaring official road names frequently move. Zip code boundaries, within which each street name must be unique, are also subject to realignment as development occurs. The only sub-state political jurisdictions that seem to be stable are counties. This jurisdictional level is

therefore proposed to serve as the most detailed road identifier reference. It also has the benefit of being a reasonable scale for base map maintenance, supports the form of linear LRS now used by many states, allows sub-state regions to be created using multiple counties, and is consistent with local mapping applications such as property records.

## **TRANSPORTATION FEATURE NAMING STANDARD**

Before a standard identifier naming convention can be proposed, a formal definition of transportation feature is needed. Our working definition is "a portion of any transportation system that is referenced by a unique identifier." The more formal definition is:

**Transportation Feature.** *An element of a transportation system that may be uniquely identified in the real world and for which attributes are provided.*

Transportation features are confined to the limits of a jurisdiction, which forms the basic unit for providing feature names. Transportation features serve to organize the entire database, with jurisdiction being the highest level of the organizational hierarchy. Each transportation feature name, or reference, need be unique only within the context of a given jurisdiction. Under no circumstances should the name or identifier of a transportation feature be something that may change, such as its route number or facility name. It is the physical entity that is referred to by a transportation feature identifier. The total unique transportation feature identifier could be the concatenation of a jurisdiction identifier and transportation feature identifier (preferably a numeric identifier.) Other jurisdictions and area-specific data are tied to an area event or polygon and do not control the road naming process. One or more alias names may be used for each transportation feature; these alternative names need not be unique as they can be stored as linear or point events (attributes).

External references (foreign keys) may be used to extend the new GIS-T database to legacy databases. For example, one could use bridge numbers stored as a point event attribute to access a bridge inventory. Airports could be referenced by site number, and railroads by name and milepost (railroads typically have their own linear LRSs based on milepost). Transit services do not have a uniform LRS statewide except for service provider identity; each provider uses its own route naming convention. These local linear LRSs could be readily overlaid on the anchor section and transportation feature systems. Relating route identifiers used in legacy databases should be done by defining them as linear events of transportation features in the GIS-T database, whereas more ephemeral things, like delivery routes, should be defined as traversals.

The solution to the road naming problem has two organizational variations. One is to name a single organization as the czar for assignment of unique roadway and street ID numbers (transportation feature names) and agree that all other organizations will follow their lead. The second is a decentralized approach of adopting street and road naming standards. In this case, the standard name is used as the unique external identifier for data sharing, allowing each organization to employ their own internal identifiers for database management. Alias names may also be offered.

The shortcoming of the first approach is that the lead organization, say the state DOT, would have to become responsible for managing the assignment of unique identifier numbers including those of local streets, or delegating to local governments procedures by which to do it in a consistent manner. A potential shortcoming of the second approach is the need to define transportation feature beginning and ending points, and to choose which of two or more overlapping routes is primary for use as an external identifier. Resolving such issues may require the formation and operation of an inter-organizational standards committee.

In spite of the problems with the second approach, it seems the preferable way to foster data sharing in a decentralized environment. Naming rules could be designated by many relevant agencies (e.g., USGS, U.S. DOT, and/or AASHTO) or through an SDTS Transportation Profile. A proposed draft set of naming rules is the subject of the balance of this paper.

## **Overview of Proposal**

The basic naming system is a character string composed of substrings with specific meanings separated by periods. The substrings identify transportation feature type, jurisdiction, feature sequence number, and feature subdivision sequence number. The first component allows the ready identification of what type of transportation feature is being described, with implications for both how the transportation feature ID is determined and what type of attribute data may be available. The last two components form the transportation feature ID within the specified jurisdiction.

The transportation feature type codes use two-letter mnemonics to represent the various possible types. Jurisdiction, since it is defined uniformly at the county level for roads, usually consists of the combined federal standard codes for states and counties. The feature sequence number is a randomly assigned number in the case of roads, or a specified identifier for other transportation features. A feature sequence subdivision number is used to identify portions of multiple-part features. An example would be entrance and exit ramps for an Interstate highway or a realigned section of a road. The following illustrates the overall naming standard:

TF.aa.nn.nnn.aa.nnnnn.nnnn

*where*

TF = the data item as being a transportation feature identifier;  
a = an alphabetic character; and  
n = a number.

An illustrative example of how a transportation feature ID might look is provided by a federal-aid eligible road located in Kansas:

TF.RD.KS.031.ST.00294.000

where

- RD = Transportation Feature Type of "Road;"
- KS = State of Kansas zip code abbreviation;
- 031 = County FIPS code;
- ST = Transportation Feature Sequence Number is assigned by state DOT;
- 00294 = Transportation Feature Sequence Number; and
- 000 = Transportation Feature Subdivision (original mainline).

The initial identifier of 'TF' is provided as a means of differentiating this set of objects from others that may be included; e.g., city polygons, census tracts, and water bodies. The order is not precisely reflective of the jurisdiction ID/transportation feature ID format used in the earlier discussion of the data model and sample database. In that discussion, transportation feature type would be an attribute of the Transportation Feature entity. However, the example database design addressed only one type of transportation feature (roads), so there was no need to explicitly include that attribute in the database. Here, however, a single identification method is proposed for all types of transportation features, so transportation feature type must be explicitly stated. It is listed first in the naming hierarchy as it is expected to be one of the primary selection criteria for database queries.

The separating periods need not be included in the actual data, but can be generated when displaying the information as a useful way of improving visibility of specific components. It is expected that separate storage of each ID component is the most efficient for some uses.

Previous work by the authors opposed the use of identifiers that included intelligence, as is proposed here, presenting an apparent conflict. (Dueker and Butler, 1997) There is actually no conflict since the proposal made here is for an external, or foreign, key to access data. The internal, or primary, key of the database being accessed would remain exclusive to the subject database and should not be tied to any specific attribute.

### **Transportation Feature Types**

The first element of the naming system is transportation feature type. The following types are suggested as a starting point:

<u>Transportation Feature Type</u>	<u>Coded Value</u>
Airport	AP
Air Route	AR
Bridge	BG
Bus Route	BR
Bus Stop	BS
Bus Terminal	BT
Dock	DK
Ferry Route	FR
Intermodal Terminal	IT
Overhead Sign	OS
Post-mounted Sign	PS
Road Interchange	RI
Railroad Junction	RJ
Railroad Track	RT
Railroad Station	RS
Road	RD
Junction	JN

<u>Transportation Feature Type</u>	<u>Coded Value</u>
Shipyard	SY
Traffic Signal	TS
Waterway	WW

The list of defined transportation features includes many that may be primarily recognized as point events occurring on a linear transportation feature, or may be separate point and area features. Examples of point events include Junction, Traffic Signal, and Overhead Sign. Features that could be either point or area features include Harbor Dock, Railroad Station, and Shipyard. A highway agency may have data using only the Road type of transportation feature. Road junctions may be included in such a database as point events without using the separately listed Road Junction type.

The provided list is illustrative only. It is intended to show how a standard may be constructed. A potential issue for resolution is whether such objects as Overhead Sign are really transportation features. Since the list is illustrative, there has been no attempt to unambiguously define the various terms presented. Also, a further typology of junctions may be defined.

A useful starting point for making final selections may be the work of the Ground Transportation Subcommittee of the Federal Geographic Data Committee, which is working on a standard GIS-T dictionary of feature terms.

### **Jurisdiction Codes**

The jurisdiction component of the identifier will vary according to the transportation feature type. The jurisdiction portion of a road transportation feature identifier consists of two elements. The first is the two-letter Zip code abbreviation for each state. The second is the standard four-digit code for each county, as given in FIPS Publication 6-4. This list must be extended to use highway district codes for Alaska, parish codes for Louisiana, and municipio district codes for Puerto Rico. These extensions are the same as presently used for Highway Performance Monitoring System (HPMS) reporting. The state and county substrings are separated by a period. A third component of the jurisdiction substring could be used to designate country for worldwide implementation. Such a component would physically be the first part of the jurisdiction substring.

In the case of airports, which are assigned identifying site numbers at the national level, the state code could be the state in which the airport is located, or 'US' to represent a national jurisdiction. The county portion of the code could be the actual county in which the facility is located, or it could be '000'. Features, such as bridges, that can occur in multiple counties (cross a boundary) could also show a county code of '000' and a national or state jurisdiction. A similar approach could be used for waterways, air routes, transit service routes, and other transportation features which have national identifiers.

### **Transportation Feature Sequence Number**

Airports and transit service providers are uniquely identified at the national level so their feature sequence number should be the nationally assigned site number in the case of an airport, or the provider identifier used for Section 15 reports in the case of transit service providers. Roads would receive a randomly assigned sequence number. The issue is who would make the assignment.

Roads eligible for federal aid must be assigned a unique identifier by the state DOT that provides HPMS data to the Federal Highway Administration, so a mechanism already exists to provide the needed identifier. Other roads do not have a universally established ID creation mechanism. This situation may not represent a real problem since the reason for having a sequence number is to provide a foreign key to attribute data records, not locate roads on the surface of the earth. Any database provider would have to utilize some mechanism for identifying the features being described. Nevertheless, a requirement of the sequence numbering system is to distinguish between those road identifiers assigned by the state DOT to federal-aid eligible roads and those assigned by other entities or for other purposes.

Roads are not the only linear transportation features that need to be defined at this level. Transit service routes must also be identified, and airport runways, railroad yard tracks, and similar objects may need to be separately identified in a database.

To address these needs, it is proposed that the first component of the transportation feature sequence number identify the entity that made the numerical assignment. The proposed list is, again, only illustrative:

<u>Designating Agency</u>	<u>Coded Value</u>
Airport authority	AA
County	CO
Harbor authority	HA
Municipality	MU
Railroad company	RC
State transportation agency	ST
State water resources agency	SW
Transit service provider	TS
U.S. Bureau of Land Management	LM
U.S. Dept. of Transportation	US
U.S. Forest Service	FS
U.S. Geodetic Survey	GS

To this agency component would be added a five-digit sequence number, separated by a period. The resulting substring would consist of a two-character designating agency code, a period, and a five-digit sequence number.

Bridges would be identified using a state-designated structure number. Railroad grade crossings would be identified using the existing crossing number. Bus stops, railroad yards, and similar private facilities would be designated by their owning entity. Note that a transit route following a roadway has its own transportation feature identifier separate from that of the road and the bus stops that it may connect.

The five digits provided should be sufficient to address all numbering schemas given the ability to separate features by designating agency and/or jurisdiction. Using this approach, each designating agency needs to keep track of only the sequence numbers it assigns. Sequence numbers may be duplicated for different designating agencies.

### **Transportation Feature Subdivision Number**

A single number may not be adequate to handle all identification needs. A primary exception is the need to identify Interstate and other limited-access highway ramps. Another use would be to identify links of a traversal for those organizations that must use a link-based approach. Another example is the need to separately identify and store attributes for a realigned section of road, or to identify various branches of a railroad passenger service route. These various components of the road network may be best identified in the context of the primary route of which they are a part. A four-digit feature subdivision number is proposed to accommodate these diverse needs.

## **CONCLUSION**

This paper has shown that a common approach to defining transportation features can be developed using a simple, decentralized approach. It is necessary to first establish an overall data model and database design approach before a transportation feature identification method can be properly stated. The proposed data model is essentially that already being used for numerous legacy highway databases, with the addition of the concept of junction to expand the "one road at a time" approach of straight-line diagrams to a full 2-dimensional GIS-T product. A sample relational database design shows how the data model can be implemented.

The suggested transportation feature naming standard operates in a manner similar to that used for internet addresses. Separate components of the character string used as a transportation feature identifier provide information on where the feature is located, the type of feature it is, and who designated the identifier. Rules for determining the end points of linear transportation features are simple. Most non-linear transportation features are discrete points or areas and are simply identified by a single identifier. Roads and railroads are linear transportation features that may need to be subdivided to make data management more reasonable. The proposed level of subdivision is the county for roads; railroads and transit service routes may be subdivided in other manners as determined by their owners.

For more information on the larger concepts of data modeling extensions to handle multiple cartographic representations and network generation, and the development of data standards for transportation applications, the reader is invited to visit the Portland State University web site to download the authors' paper, "GIS-T Enterprise Data Model with Implementation Choices." The address is <http://www.upa.pdx.edu/CUS/>.

## REFERENCES

- American Association of State Highway and Transportation Officials, *AASHTO Linear Referencing Workshop Technical Notes*, 1995 GIS-T Conference, Reno NV, April 1995.
- Dueker, K. *National Framework Data: Models for Ground Transportation Data*, Discussion Paper 95-4, Center for Urban Studies, Portland State University, 1995.
- Dueker, K. and J.A. Butler, *GIS-T Enterprise Data Model with Implementation Choices*, Center for Urban Studies, Portland State University, 1997.
- Dueker, K. and R. Vrana, "Dynamic Segmentation Revisited: A Milepoint Linear Data Model," *Journal of the Urban and Regional Information Systems Association* 4, No. 2, pp. 94-105, 1992.
- Fletcher, D., Espinoza, R., Gordon, S., Spear, B. and A. Vonderohe, *The Case for a Unified Linear Reference System*, Presented at a Workshop on Enterprise Location Referencing Systems: Policies, Procedures and Standards for Implementation, Department of Geography, University of Utah, July 26-27, 1996.
- Fletcher, D., et al., *The Case for a Unified Linear Reference System*. Alliance for Transportation Research, New Mexico, 1996.
- Gordon, S., *Overview of ITS Location Referencing Issues*, presented at a Workshop on Enterprise Location Referencing Systems: Policies, Procedures and Standards for Implementation, Department of Geography, University of Utah, July 26-27, 1996.
- Liu, W., *Sharing Transportation Data*, Discussion Paper 96-5, Center for Urban Studies, Portland State University, forthcoming.
- Okunieff, P., D. Siegel, Q. Miao, and S. Gordon (1995) "Location Referencing Methods for Intelligent Transportation Systems (ITS) User Services: Recommended Approach," *Proceedings*, 1995 GIS-T Symposium, AASHTO, pp. 57-75.
- Viggen Corp. , *ITS Location Referencing System: Dynamic Assignment of Network Attributes*, Draft, 21 Union St. Boston MA 02108, 1994.
- Vonderohe, A., et al., National Cooperative Highway Research Program (NCHRP) Project 20-27 reports published by National Academy Press, Washington, DC
- a. "Implementation of Geographic Information Systems (GIS) in State DOTs," *Research Results Digest 180*, Aug. 1991;
  - b. "Management Guide for Implementation of Geographic Information Systems (GIS) in State DOTs," *Research Results Digest 191*, Oct. 1993;

c. *Report 359: Adaptation of Geographic Information Systems for Transportation*, 1993.

Vonderohe, A., C.L. Chou, F. Sun, and T. Adams, "Results of a Workshop on a Generic Data Model for Linear Referencing Systems," *Proceedings of Geographic Information Systems for Transportation Symposium*, April 1995, Sparks, NV, pp. 23-56.

Vonderohe, A., and T. Hepworth, *A Methodology for Design of a Linear Referencing System for Surface Transportation*, prepared for Sandia National Laboratories, Project AT-4567, Dec. 1996, University of Wisconsin, Madison, WI.