GIS-T Data Sharing Issues

Kenneth Dueker
Portland State University

J. Allison Butler

Let us know how access to this document benefits you.

Follow this and additional works at: http://pdxscholar.library.pdx.edu/cus_pubs

Part of the Transportation Commons, and the Urban Studies and Planning Commons

Citation Details
http://pdxscholar.library.pdx.edu/cus_pubs/15

This Working Paper is brought to you for free and open access. It has been accepted for inclusion in Center for Urban Studies Publications and Reports by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
GIS-T Data Sharing Issues

Kenneth J. Dueker
Portland State University

J. Allison Butler
Hamilton County
Tennessee

May 1999

Discussion Paper 99-02

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

www.upa.pdx.edu/cus/
Introduction

Current data models used to represent streets and roads, e.g., TIGER, Arc, 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.

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 linear transportation. 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. 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 that cannot be navigated. Yet, public organizations responsible for road maintenance follow different rules; when they omit private roads and include paper streets. 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 data sharing 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

¹ Sperling and Sharp (forthcoming) 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 geospatial database into another.”
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.
ramps and lanes. Further, differing levels of real-time use of systems dictate response requirements of databases and interfaces.

The purpose of this paper is to develop a framework and principles for sharing of transportation data. The framework is intended to clarify roles of the various participants and the principles are intended to provide guidance for the participants. Both the framework and the principles are based on a GIS-T data model that defines relations among transportation data elements. (See Dueker and Butler (1998) for a detailed description of the data model. A simplified version is provided in the next section.) The data model guards against ambiguities and provides a basis for the development of the framework and principles for sharing of transportation data.

A Brief Description of the Dueker-Butler GIS-T Data Model

- Jurisdiction defines context for one or more Transportation Feature
- Transportation Feature must be defined within the context of one Jurisdiction

The Transportation Feature entity is the central object, so we will start out tour there. We are speaking here only of linear transportation features. The full Dueker-Butler Model support other structural types. Transportation features are facilities that comprise part of a transportation systems, such as roads comprise a highway network. A jurisdiction sets the context for defining the extent and identifier 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.
• Transportation Feature may have one or more Event Point
• Event Point must relate to one Transportation Feature

Characteristics of a linear transportation feature may be applicable to a single point or to a section of the feature. These characteristics are called “events,” and may thus be classified as linear events and point events. A linear event would be something like a section of roadway with a constant speed limit or pavement condition. A point event might be something like a sign or a traffic count site. Some characteristics might be both; e.g., a bridge could be a linear event because it has length or it could be a point event because it is viewed as being located at one place.

• Event Point may begin or end Linear Event
• Event Point may locate Point Event
• Linear Event must begin at one Event Point
• Linear Event must end at one Event Point

Point Event must be located by one Event Point

The location of a point event and the endpoints of a linear event, are defined by measures stated in the linear LRM. These defining locations, including their linear LRM measure values, are called “event points.” Therefore, linear events are defined by a pair of event points, while point events are defined by a single event point. This association causes what may appear to be an unusual set of relationships in that there is no direct connection from Transportation Feature to Linear Event and Point Event. This is due to the use of event points to locate the position of linear and point events. It is the set of event points that is “owned” by the transportation feature, while event points define the position of events on that transportation feature.

• Intersection is a type of Point Event
• Point Event may be an Intersection
• Intersection must be part of two or more Transportation Feature
• Transportation Feature may intersect at one or more Intersection

One type of point event is the intersection, which, by definition, represents the junction between two or more linear transportation features. The location of the intersection, from the point of view of each of the intersecting transportation features, must be defined within the context of each feature’s linear LRM. Thus, if two roadways intersect at a single intersection, the intersection will be represented by two point events—one for each of the roadways. [Note: The ERD looks similar to the detailed model but is really very different. In this simplified version, we are showing a true Intersection entity. In the detailed model, we show a resolution (associative) entity.]

• Anchor Section may establish linear LRM datum for one Transportation Feature
Transportation Feature may have one or more Anchor Section
Anchor Section begins at Anchor Point
Anchor Section ends at Anchor Point
Anchor Point may begin Anchor Section
Anchor Point may end Anchor Section

There has been substantial interest in providing a formal geographic datum to establish a linear location referencing method (linear LRM) framework that is utilized to describe locations on a linear transportation feature with a given degree of accuracy. The basic datum component for the linear LRM is the anchor section, which primarily defines the official length of all or part of a transportation feature. Anchor sections begin and end at anchor points. Since anchor points are not allocated to a specific transportation feature—but anchor sections must be—then there is no explicit relationship between Anchor Points and Transportation Feature except that conveyed through Anchor Section.

Reference Object may be used to locate one or more Anchor Point
Anchor Point may be located using one or more Reference Object
Reference Object is located on the Earth using one Geographic Point

Conflation, or the combination of multiple data sets, frequently requires a user to match locations on one source with the same place on another source. Anchor points may be defined using many location methods as a means of registering the transportation feature on the surface of the Earth. In order to function most effectively, an anchor point is typically defined as a logical position, such as the center of the intersection of two streets. Intersections are easy to find on maps. While logically understandable, it is difficult to precisely define what each user might unambiguously interpret to be the exact center.

We have provided the concept of a reference object that can be used to precisely locate the position of an anchor point. A good reference object would be something that can be found on a map and in the field, such as traffic signal pole or right of way marker. The position of a reference object is defined by a geographic point. A single reference object could be used to define the position of one or more anchor points, or several reference objects could be used to define the location of a single anchor point.

Geographic Point may be transformed to one or more Cartographic Point
Cartographic Point must represent one Geographic Point
Cartographic Point may shape one or more Line Segment
Line Segment is shaped by one or more Cartographic Point
Cartographic Point may locate the position for one Point Symbol

Mapping software has a means to convert geographic locations to an internal mapping framework. This allows each geographic point to be defined by one or more cartographic points, depending on the selected datum transformation. Cartographic points begin, end and otherwise shape line segments, or may be used to state the position for displaying a point symbol.

Transportation Feature may be illustrated by one Base Map String
• Base Map String may represent one Transportation Feature
• Base Map String is composed of one or more Line Segment
• Linear Event String is composed of one or more Line Segment
• Base Map String may be subdivided into one or more Linear Event String through dynamic segmentation
• Linear Event may define one Linear Event String
• Linear Event String must be defined by one Linear Event
• Point Event may select one Point Symbol
• Point Symbol may be selected by Point Event

Base map strings should be constructed that correspond to the extent of each transportation feature. The dynamic segment function will subdivide (segment) the base map string to correspond with the extent of linear events on the feature, or to properly place point event symbols.

• Point Event may be Node
• Node may be Point Event
• Node may begin one or more Link
• Node may end one or more Link
• Link must begin at one Node
• Link must end at one Node
• Link may define the extent of one or more Traversal Segment
• Traversal Segment must be defined by one Link
• Linear Event may provide attribute for one or more Traversal Segment
• Traversal Segment must receive attribute from one or more Linear Event
• Point Event may provide attribute for one or more Traversal Segment
• Traversal Segment may receive attribute from one or more Point Event
• Traversal Segment may be part of one or more Traversal
• Traversal must contain one or more Traversal Segment

The logical transportation feature network is defined by links and nodes that form the explicit topology of the network. Links must begin and end at nodes. Nodes may be equivalent to point events, which are expected to be physically real things, or they may represent “virtual” events, such as traffic analysis zone centroids. When attributes are placed on a link, it becomes a traversal segment. One or more traversal segments may be assembled to construct traversals through the network.

For pathfinding applications, a traversal segment might contain attributes associated with minimum carrying loads of bridges, highest traffic volume, or other characteristics that might be used to determine an optimum path through the network. We use the terms “static traversal” to mean a persistent traversal, such as the entire path of a numbered state route across all the component transportation features, and “dynamic traversal” to mean such things as a path to get a fire truck to a fire.

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.

GIS-T Data Modeling Issues

A National Cooperative Research Program project, 20-27(3), is in the process of trying to
bring consensus in the area of GIS-T data models. There are two competing approaches.
The Dueker-Butler GIS-T data model represents one approach, which can be
characterized as a feature (object) database approach. It is best suited for a federated
systems environment with legacy data of varying spatial accuracy.

An alternative approach can be characterized as a location (geometry) approach (Sutton,
1999). This alternative approach would work well in a centralized environment wherein
the location of transportation features would be redigitized with high precision 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, as issues, such as repeatability of GPS positioning, how to
abstract networks, how to relate to other location referencing systems, and representation
at smaller scales have not been adequately developed.

In addition there is a missing element in both approaches, which is a basic problem with
GIS data models in general that become paramount in GIS-T. In 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 a GIS. There needs to be a new
Dynamic 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 objec.

Transportation Data Participants

Organizations that have ownership and maintenance responsibilities for transportation
infrastructure are simultaneously data producers, data integrators, and data users.
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) has 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 that a disabled vehicle is located is a particularly demanding task. Similarly, ITS tolling applications may require tracking vehicles by lane of multiple-lane facilities.

**Transportation Data Sharing Principles**

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 network primitives or 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 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 county boundaries or major intersections if consistency with assignment networks is desired.
• Collect minor road facilities by street name to minimize the number of unique identifiers.

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 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 the use of scripts to bulk-assign state and county codes.

Issues of Definition and Identification of Transportation Features

Dueker and Butler (1998) defined a Transportation Feature as “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).” This definition is encompassing while some specific applications may restrict their databases to roads, pedestrian paths, or waterways. The important point is to code the type of Transportation Feature so that in sharing of data the type can be used to select those features of common interest.

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 Road Data Model 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 coordinate or name matching.

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 were developed and procedures were developed

---

4 Refer to http://www.fgdc.gov for links to the NSDI Framework Road Data Model.
for bulk assignment of higher level codes to sequenced numbered roadway features. 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. We 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. Arterial road identifiers in metropolitan 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 DOTor county DOT road ID. 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 within the concatenated state and county FIPS codes. Some judgement 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 IDs are assigned. Also, there are situations where name changes occur arbitrarily, such as at municipal boundaries where a different ID may not be needed.

**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 accuracy of the cartography ranges from 1:24K to 1:100K.

Similarly, the spatial accuracy 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 digital orthophoto quarter-quadrangle program offers the opportunity for states to update their road vectors and register them using 1:12K imagery. 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).
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)

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 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 ID. 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.

Network Issues

Again, 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.

**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)

**In Conclusion**

Sharing of GIS-T data poses challenges. This paper identifies 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.

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.
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 the USGS digital orthophoto quarter-quadrangles. A second stage would by to update the vectors using replications of vehicle tracking data.

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 E9-1-1.

Acknowledgements

The contributions of Mark Bosworth, Senior GIS Analyst, Metro, Portland and Paul Bender, Graduate Research Assistant, Center for Urban Studies, Portland State University for the case study materials is gratefully acknowledged.

References


