FTA National Transit GIS:

Data Standards, Guidelines and Recommend Practices

DTRS57-95-P-80861

January 1996

prepared for:
Volpe National Transportation Systems Center
Cambridge, MA

prepared by:
VIGGEN Corporation
Boston, MA

Mission Statement

The mission of the National Transit GIS is to promote the collection, use, and sharing of public transit spatial data.

The goals pertinent to the development of the Standard, Guidelines and Recommended Practices for the National Transit GIS include:

1. improve the use of geospatial data in response to the ISTEA of 1991 by providing transit-domain standards.

2. provide standards, guidelines and recommended practices for transit data, specifically the National Transit GIS, compatible with the National Geospatial Data Clearinghouse,

3. provide standards, guidelines and recommended practices for transit data, specifically the National Transit GIS, to help the public collect, use, and share public transit information.

4. support the mission requirements of the Department of Transportation, particularly as established by the Intermodal Surface Transportation Efficiency Act of 1991. Most notably this includes the development of a GIS-based National Transportation System for transit routes as a major element of the National Spatial Database Infrastructure (NSDI).

Executive Summary

The Federal Transit Administration (FTA) National Transit Geographic Information System (GIS) is a representative inventory of the public transit assets of the country. Creation of this national system is an ongoing and collaborative effort on the part of many within the transportation industry. Use of these transit data will facilitate the exchange of information within the U.S. Department of Transportation and throughout the transit industry. This effort supports the mission requirements of the Department of Transportation, particularly as established by the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991. Most notably this includes the development of a GIS-based National Transportation System for transit routes as a major element of the National Spatial Database Infrastructure (NSDI).

This spatially referenced data base will provide transit planning and operations data such as population served, ridership, passenger miles and route/rail miles for all modes of public transit. The systems and facilities includes rural and urban bus systems, commuter rail, subways, light rail, people mover systems, high occupancy vehicle systems, ferry terminals and transitways.

The *Standards, Guidelines and Recommended Practices* establishes a framework for maintaining the National Transit GIS (NTG) database, ensuring data integrity, interoperability and consistency. The methods and quality control used in creating, storing, exchanging and documenting the data in the National Transit GIS is known by recommending feature type definitions, formats, file formats, update procedures and other standards. The document outlines Feature Type Definitions and Descriptions, Addressing and Street Naming Conventions, Feature Type Automation and Conversion Guidelines, Transfer Formats, and Update and Maintenance Procedures.

The use of the NTG will depend on the quality and content of the digital map product. Creation, maintenance and record keeping practices are important vehicles which inform data quality and appropriate use. Data automation and documentation procedures for data currency, positional accuracy, attribute accuracy, logical consistency and completeness are among the practices recommended. **Digital Map Products and Attribute Guidelines** and **Metadata standards** deal with documentation standards. The Federal Geographic Data Committee established the *Content Standards for Digital Geospatial Metadata* as a standard for providing the means of cataloging data sets in the National Spatial Data Infrastructure (NSDI). This document applies that standard to the NTG and includes definitions, descriptions and examples of the relevant sections.

The section on **Digital Map Products and Attribute Guidelines** includes definition and documentation (metadata) specifications for the base map product, including transit feature type definitions, data quality reporting issues, feature content and file types, and street naming and address conventions. These standards provide a framework for

describing the content and format of the digital data set accessed by NTG users or provided to the NTG by Transit Agencies.

Transit Feature Automation and Conversion Practices describe the underlying requirements for core transit feature types. These requirements include recommendations on identifying and describing core transit features and recommended practices for encoding each feature type on the digital base map. Specifically, data automation is the means used to input spatial data into a digital base map. For example, route databases may be created by tracing the road network from bus stop to bus stop, by selecting links from a digital base map and storing the links in sequential order in a file, or by maintaining a point database composed of bus stops (and time points). Data requirements for the core transit feature types were derived by the Advanced Public Transportation Systems (APTS) Map and Spatial Database Working Group of the Intelligent Transportation Society of America (ITS-A) as part of a functional decomposition of transit applications and functions. The APTS Map Database User Requirements Document (MDURD) identifies the uses and requirements of the data. The *Transit Feature Automation and Conversion Practices* section recommends encoding practices which meet the application requirements of the core set of features.

Data Exchange Specifications recommend database formats and a data dictionary based on the interim format published by the Bureau of Transportation Statistics. Any access to or update of the NTG requires a common format for storing the data files. This format will enable GIS vendors to develop translation software to import these databases directly into their own internal format until a formal Transportation Network Profile (TNP) is adopted as part of the Spatial Data Transfer Standard (SDTS). This format if accepted by transit application software vendors (e.g., scheduling, customer information, CAD) will enable exchange of transit data among applications which use map databases and related attribute data.

Maintenance and Update Recommendations include guidelines for data users and providers to ensure the integrity and quality control of data the National Transit GIS. These procedures recommendations for ensuring that records are neither missing or duplicated in the database, and that all the changes to the data sets are recorded. More detailed procedures will be developed when the NTG is set up in a location which is publicly accessible. These procedures will be similar to those of the National Transportation Atlas (NTA) maintained and disseminated by the Department of Transportation (DOT) Bureau of Transportation Statistics (BTS)

These standards, guidelines and recommended practices adapt existing, industry standards to address transit concerns. Implementation strategies and technological innovation change standards and practices over time. The users of the National Transit GIS and GIS-for-Transit must remain diligent to ensure the applicability of this document to industry practice. For that purpose, change forms for the document and feature type definitions are included in the appendix.

TABLE OF CONTENTS

1.0 Introduction

Background

The Federal Transit Administration National Transit Geographic Information System (GIS) is a representative inventory of the public transit assets of the country. Creation of this national system is an ongoing and collaborative effort on the part of many within the transportation industry. Use of these transit data will facilitate the exchange of information within the U.S. Department of Transportation and throughout the transit industry. Locally, transit managers will have access to information that will allow them to better utilize resources and make informed policy, operations, and planning decisions. At the national level, this information will represent the nation's public transportation infrastructure that is weaved throughout the country. It will also facilitate improved analysis of policy and planning decisions.

The implementation of GIS will enable the immediate display of inventory and selected data associated with fixed route public transit facilities in the United States. It will also provide for the display of spatial and attribute information on other transportation facilities within the country, including highways, airports, marine ports and passenger rail systems.

The Transit GIS will promote the exchange of information among the modal administrations and the transit industry to improve transit access to transit infrastructure data sed to make transportation decisions. Transportation management and planning can be based on immediate access to current transit facility and property inventory data represented in relation to the latest available geographic and census demographic data.

The National Transit GIS will include geographic information based on state, county, city and town, urbanized areas, and other political boundaries. Streets, municipal buildings, hospitals, schools, etc., will be represented as well as rivers, streams, lakes, and parks.

This spatially referenced data base will provide transit planning and operations data such as population served, ridership, passenger miles and route/rail miles for all modes of public transit. The systems and facilities includes rural and urban bus systems, commuter rail, subways, light rail, people mover systems, high occupancy vehicle systems, ferry terminals and transitways.

This effort supports the mission requirements of the Department of Transportation, particularly as established by the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991. Most notably this includes the development of a GIS-based National Transportation System for transit routes as a major element of the National Spatial Database Infrastructure (NSDI).

Scope

This document establishes a framework for maintaining the National Transit GIS (NTG) database, ensuring data integrity, interoperability and consistency. Though, this document was written specifically for the upkeep of the National Transit GIS, many transit agencies requested that the guidelines pertain to local transit provider and planner implementation issues. As such, this document outlines Feature Type Definitions and Descriptions, Addressing and Street Naming Conventions, Feature Type Automation and Conversion Guidelines, Transfer Formats, and Update and Maintenance Procedures.

Contents

The document is divided up into eight different sections (including the Introduction). The chapters deal with technical documentation guidelines and recommended practices, including:

Chapter 2 -- Digital Map Products and Attribute Guidelines includes definition and documentation (metadata) specifications for the base map product, including data quality reporting issues, feature content and file types, and street naming and address conventions.

Chapter 3 -- Transit Feature Automation and Conversion Practices describes the underlying requirements for core transit feature types. These requirements include recommendations on identifying and describing core transit features and recommended practices for encoding each feature type on the digital base map. Specifically, data automation is the means used to input spatial data into a digital base map. For example, route databases may be created by tracing the road network from bus stop to bus stop, by selecting links from a digital base map and storing the links in sequential order in a file, or by maintaining a point database composed of bus stops (and time points).

Chapter 4 -- Data Exchange Specifications contains recommendations for database formats and data dictionary based on the interim format published by the BTS. This format will enable GIS vendors to develop translation software to import these databases directly into their own internal format until a formal Transportation Network Profile (TNP) is adopted as part of the Spatial Data Transfer Standard (SDTS).

Chapter 5 -- Maintenance and Update Recommendations include guidelines for data users and providers to ensure the integrity and quality control of data the National Transit GIS.

Appendix A -- Glossary of Terms defines the terms used throughout this document.

Appendix B -- Transit Feature Types define and describe feature types related to transit functions. These feature types provide the foundation for the National Transit GIS data dictionary and data exchange feature codes.

Appendix C -- Document Change Forms contains forms for the public to submit recommendations for changes or enhancements to this document.

Appendix D -- Adaptation of the FGDC *Content Standard for Digital Geospatial Metadata* for the National Transit GIS details the FGDC Metadata and identifies default information needed or provided by the NTG.

Appendix E -- *National Transportation Atlas Data Dictionary and Database Formats* [BTS, September 1995], developed by the Bureau of Transportation Statistics, describes geospatial features type files and formats to access National Transportation Atlas (NTA) data. The NTG is part of the NTA.

Appendix E -- Guidelines for Use of a Route Object for Transit Features Types describes methods to transfer transit feature types which inherit route object characteristics. The data structure described in this white paper supports access and exchange of data from the NTG.

2.0 Digital Map Product, Feature and Attribute Guidelines

The use of the NTG will depend on the quality and content of the digital map product. Creation, maintenance and record keeping practices are important vehicles which inform data quality and appropriate use. Data automation and documentation procedures for data currency, positional accuracy, attribute accuracy, logical consistency and completeness are among the practices recommended. This chapter (and Appendix D) deal with documentation standards, and Chapter 3.0 discusses recommended practices for transit feature automation and conversion. Specifications for base map feature automation (links, nodes, topology, adjacency) are not detailed in this document, rather verification and reporting practices are suggested in the metadata documentation (Appendix D).

Base map feature documentation need not be provided if published elsewhere. For example, the NTG uses the TIGER/Line files as its base map; a reference to the version number will provide sufficient information to a user to know how to access the map data product information. However, transit features and attributes which add value and designate that data set as part of the NTG need documentation on quality, currentness, content, and other characteristics. This documentation is typically referred to as metadata.

Metadata

Metadata describe the content, quality, contacts, conditions and other characteristics of a data set. Many information specialist say that in the computer age, data is not power, "metadata" is. More than any other standard, the metadata provides a "roadmap" to information in a data set. The metadata provides information on the organization of, maintenance of and investment in data, data catalogs, access paths, and data transfer.

A metadata document on geospatial data helps people who use geospatial data find the data they need and determine how best to use that data. To this end, the Federal Geographic Data Committee (FGDC) developed a Content Standard for Digital Geospatial Metadata to facilitate access to data inventoried in the National Geospatial Data Clearinghouse. This standard provides a format to catalog information about geospatial data sets. Appendix D contains an adaptation of the Content Standard for Digital Transit Geospatial Metadata reflecting transit domain considerations. The appendix describes and explains elements of the FGDC Metadata standard, and augments the FGDC version with guidelines for transit features, themes and databases. Metadata documents created for each data set in the NTG will advance appropriate use of the data, and those created for data sets submitted to the NTG will improve the data's integration with the NTG.

Because the National Transit GIS uses TIGER '92 as the basis for describing transit-related features, the metadata should reflect information related to the base map and the NTG. TIGER documentation describes feature quality issues related to Census' obligations and glances over issues not relevant to their mission such as positional accuracy, lineage and completeness. Nevertheless, public domain data such as TIGER and USGS Digital Line Graphs (DLG) easily acqired need only be referenced. Metadata documentation on changes made to public domain data such as corrected TIGER or attributed DLGs should be submitted on the sources and changes transacted. All commercial or custome-built base map (e.g., MPO developed base maps) should include metadata information.

For many years, the FTA has been collecting "mass transportation financial and operating information...to help any level of government make a public sector investment decision" [Title 49 U.S.C. 5335(a)]. The National Transit *Database* (NTD), formerly known as Section 15 data, and NTG should link the NTD aggregate data to the NTG physical infrastructure to support FTA decision making.

Standard Addressing Conventions

Addressing conventions provide a standard format for automated processing of data files containing address information. Standard formats, abbreviations and conventions exist for address components, city (place names), state and zip codes. Other Federal agencies develop and maintain these standards; many of them are already incorporated into the TIGER database. For example, Federal Information Processing Standard (FIPS) Codes are used for place and state names. Also, Census is currently working with the U.S. Postal Service (USPS) to adopt addressing conventions, and abbreviations for address components, states, and place names.

The conventions used for the National Transit GIS should adopt standards promulgated by the USPS and adopted by the Census Bureau. These conventions, described in the *Postal Addressing Standard* [Publication 28, August 1995], standardize the delivery address line, city, state and zip+4. In particular, the delivery address line identifies five relevant components: primary address number, predirectional, street name, suffix, and postdirectional. The document describes the format, use of abbreviations, acceptable abbreviations, address exceptions and approved abbreviations for juxtaposition of duplicate values[1].

[1] The publication is available from USPS, National Customer Support Center, Memphis TN 38188-0001.

3.0 Transit Feature Automation and Conversion Practices

Core Data Requirements

This section describes the underlying requirements for data that supports GIS for Transit. These recommended practices include approaches to uniquely identify core transit features and to unambiguously reference features on a digital base map. Also, the "Data Automation and Conversion" paragraphs recommend practices for encoding each core feature or data set described herein. Specifically, data automation is the means used to input spatial data into a digital base map. For example, route databases may be created by tracing the road network from bus stop to bus stop, or it may be created by selecting links from a digital base map and storing the links in sequential order in a file.

These discussions deal with issues to ensure data fitness for different applications. In the transit route example, if the route overlaps with the street network for traffic congestion analysis, then the two networks should be identical. Yet, in the case of a subway there is no direct overlap.

As described above, the means by which spatial features are coded depend on their use. Different GIS functions may be performed only on certain types of data. For example, a map projection translation is not effective on data not referenced by spatial coordinates. The *Uses* section briefly describes transit applications and GIS functions which help to determine data automation techniques which facilitate use of the data. Some of these GIS functions may be applied specifically to point, line and area data. Certain operations require that the data be referenced in the same manner prior to operation. For example, if the line-of-sight distance between two objects is required, both objects need to be defined as coordinate pairs (or triplets). This is not complicated because the GIS automatically converts the data to its own Cartesian frame to calculate the distance. A line or polygon will be associated with its centroid for this distance operation. However, if the travel distance along a road network is requested, the position on and relationship to the network is needed. In general, any application associated with networks require information about the network topology.

Generally, each data set requires at least two primary unique references. The first is a table reference to uniquely identify an instance of a feature. For example, each bus shelter owned by a property should be assigned an identifier for inventory purposes. Most agencies use an index as the primary key; this index may or may not have embedded intelligence. For example, the first three digits of the feature's name concatenated with a four digit number. A primary key should be assigned the most stable identifier possible. Thus, a number based on a sequence of locations may not be appropriate as a primary key.

Also, each spatial feature requires at least one location identifier, that is, a geographic reference which corresponds to a position on a digital base map. This reference ensures a mapping between the tabular data and digital base map. If the tabular data are used by multiple applications which use different base maps, the common identifier must be invariant to the characteristics of the different base maps, including resolution, positional accuracy, link-node structure and extent. With increased deployment of advanced technologies and computerized systems in transit agencies, and between transit agencies and other transportation/planning organizations, this issue will emerge as the significant technical obstacle to integrating spatial data.

Furthermore, in many cases different location referencing methods are required within a single agency. Translation among these methods (e.g., coordinate snapping and projecting distance, etc.) may propagate errors that affect analysis results. Some standards and procedures may be applied to mitigate or eliminate these errors. These issues are not discussed in this document.

Access Point Inventory

Uses

The Access Point Inventory or Bus Stop database is used by almost every department in the transit agency. Typically, created by Service Planning the inventory is used for designing the routes, and analyzing route design, level of service, and run schedules. The locations of the bus stops are needed for designing the instructions for operator training, maintenance of facilities, and directing and planning customer trip itineraries for both fixed route and demand-responsive customers.

Data Automation and Conversion

The Access Point Inventory is typically stored as a point database. Many times, agencies tag these points with a coordinate pair (i.e., latitude and longitude) for ease of display. This kind of representation enables visualization through overlay and point-in-polygon analysis. Some transit agencies select a point directly from a digital base map. This method, highly dependent on positional accuracy of the selected data set, provides accurate location references relative to a single digital base map.

Another technique gaining in popularity is collecting data using satellite positioning technology -- Global Positioning System (GPS). Depending on receiver and antenna accuracy, and applied corrections, these techniques will provide accurate spatial coordinates in World Geodetic System 1984 (WGS-84). The GPS method provides good positional accuracy when differential GPS corrections are applied. Yet, significant errors may appear when this data is overlaid onto representations of the road network which possess different non-linear errors and resolution. For example, the geometric accuracy of existing TIGER/Line files is accurate to about 250 feet, making a bus stop location -- derived using GPS -- appear inside a building.

Spatial coordinates do not provide accurate point locations on the road network (point-on-line analysis) or path finding algorithms. For example, to find a path to the nearest bus stop, or to find all the bus stops along a transit route, the bus stops must be attached to the road network. Applications which reference the road network require use of a linear referencing method. Typical techniques use address matching, node-offset, milepoint, position relative to an intersection (e.g., near side, far side, mid block) or relation to landmark. Most of these techniques are invariant to resolution, positional accuracy, and extent; some are also invariant to link-node structure.

Each of these linear referencing methods are described in the paragraphs below.

Address Matching is a procedure to match street addresses associated with entities (e.g. roads, tracks, facilities, houses) stored in the digital base map. In most cases, an entity's street number, street name, street type, and other prefixes and suffixes must produce an exact match of the spelling and format of the digital base map's tabular information. Very few systems have achieved better than 90 to 95% automation. Also, no addresses exist for most rural routes, interstates and many state highways. Tiger/Line files specify standard naming conventions for street address numbering, indexing, and abbreviations. Future Census products will use *Postal Addressing Standards* (see Chapter 2 on Standard Addressing Conventions).

Node-offset (or intersection) is a procedure that identifies a node by index, name, or intersecting street names, and the linear distance from that point. For example, if a bus stop is located 20 feet from the intersection of Main St and South St, it may be defined in a database as:

Run Street	Cross Street	Offset
Main St	South St	20

This type of referencing requires street naming and offset direction conventions. Typically, transportation agencies assume orientation from south to north, west to east. This assumes that if Main St runs southeast and northwest, the bus stop is on Main St, north of South St. An additional field or fields may be required to designate orientation such as

Orientation: *side of street or intersection*
Domain Values: north (N), south (S), east (E), west (W), northeast (NE), northwest (NW), southeast (SE), southwest (SW).

Relative Position from intersection is a procedure similar to node-offset because it uses a node, indexed location (e.g., landmark), address, or intersection, but it does not use a distance measure from that point. Instead, this technique uses a *relative* position. The typical terms used are near-side (NS), far-side (FS), mid-block (MB), opposite (OPP) and at (@). Near side indicates a location prior to the node; far side -- after the node; mid block -- middle of the block after the node; opposite -- on the other side of the street; and @ -- at that location, on the same side of the street. This procedure assumes that a direction or orientation is specified. Similar to the node-offset procedure if this technique is used for a bus stop inventory, conventions for directionality require standardization.

Milepoint is typically used by transportation agencies to reference a location along a route. For example, mile marker signs along the interstate are references of distance from the start of a route. This approach is a technique used to precisely map linear attribute data to the road network.

Composition
The access point inventory catalogs all the locations where transit patrons board or alight a transit vehicle in revenue service. This data set includes at least the following fields:

<u>Unique Bus Stop/Access Point ID</u>. Each access point should contain a unique identifier to which every department has access.

<u>Location Reference</u>: Each access point requires at least one type of location identifier in order to locate it on a digital base map. Selection of a location reference depends on the use of the data.

Transit Route Database

Uses
The route database may be used to generate route descriptions for training drivers, input for runcutting, generating on/off survey forms, customer information services and schedule graphics. These applications also use other information such as landmarks, "car time" or "impedance" between stops, wait time for travelers, actual run time between stops, time transfers, origin/destination information, route alignment, ADA ride eligibility, among other applications.

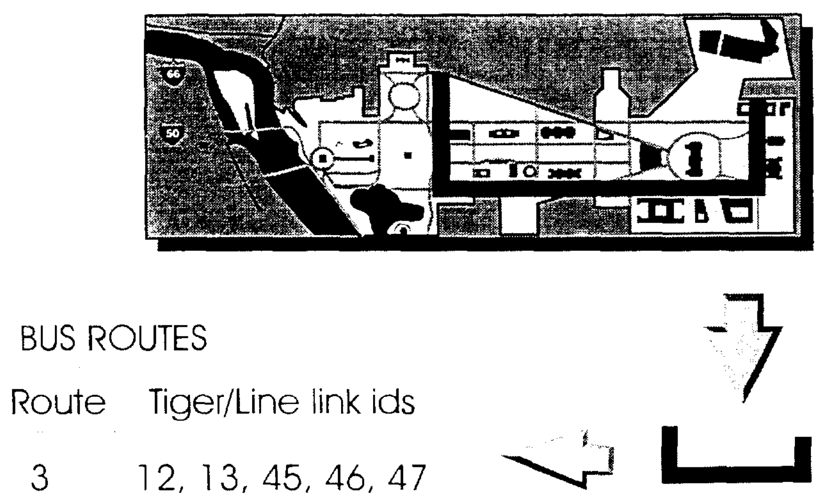
Overlaying of economic growth, demographics, land use and other data with a transit route database enables regional analysis applications such as Title 6 compliance, level of service analysis and other planning functions.

Data Automation and Conversion
Many transit organizations represent their time points coincident with bus stops. In those cases, time points coincide with bus stops, and time segments coincide with segments between bus stops. In cases not coincident, a separate network for time points and time segments may be needed. The keys used to join those two point databases are the route number and pattern number.

In order to link the route database with the digital base map, a common reference needs to be established. Also, most agencies have several data sets to which they wish to compare their route database, e.g., demographics and economic growth measures.

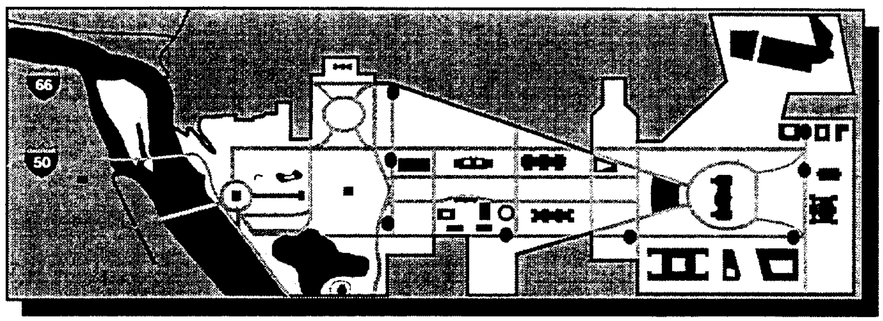
The most efficient ways of ensuring that the route database coincides with the base map is to cross reference the route segments with the base map links.

Figure 1 Bus Route Coincident with Links in Digital Base Map



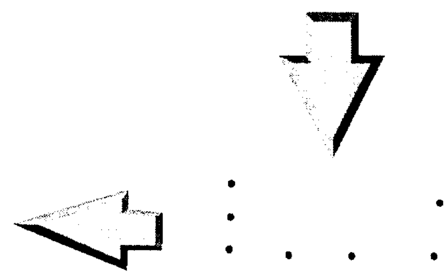
This approach may create a non-normalized table, and consequently many agencies forego attaching segments between bus stops along a route, preferring to rely on the topological relationship and a shortest path algorithm to connect bus stops.

Figure 2 Bus Stops Defining a Bus Route in a Digital Base Map



BUS ROUTES

Route	Pat #	Seq #	Run	Cross	Ref
3	1	1	1st	Main	fs
3	1	2	1st	Mall	ns
...					



A more sophisticated method using a function called dynamic segmentation currently solves this problem. This tool builds a table between two paths designating a virtual link-node structure. This function is not found in most desktop GIS tools, only in the more expensive GIS products. An approach to applying this method is described in more detail in [Peng, GIS-T '95].

If a route is designated without the bus stops (i.e., bus stops are represented in a separate data table) as in the case of the National Transit GIS, then the route may be represented by using the link node structure of the base map or a sufficient number of nodes to convey the topological relationship between the points. Again, to ensure integration with the base map, the segments comprising the route should be coincident with the base map. Coincident features are those that are common between data layer and base map. Examples include transit routes overlaid on the road network or traffic analysis zones that are bounded by road segments. Coincident features must be digitized only once to avoid slight differences such as "slivers" and to ensure network integrity when features are topologically joined with the base map. This clearly demonstrates the need for selecting segments and points which compose the transit feature *directly* from the base map.

Composition

Depending on the approach used to represent different routes either (1) a point database for each pattern, (2) a segment table for each route or (3) a network database for the route network will be created. The National Transit GIS is an example of a route network database. This type of database catalogs the corridors (route alignment) supported by transit services.

A route database required for most transit functions should include information on all the patterns in the route, the sequence of access points, access point inventory information, distance/travel time between access points, and time point information. Information such as distance between access points, route direction, and other information may be derived from an accurate digital base map. The GIS could even determine the path between access points if certain assumptions were encoded in the path finding algorithm[2].

Different agencies may implement different data models for transit routes. Whether a transit route database is implemented as a single or multiple tables, the combination of these tables require the following information:

Route Number. A unique number given to a Transit Route, a collection of paths or traversals in revenue service.

Pattern Identifier. A unique identifier (alphanumeric or symbol) which references a series of contiguous segments linking an origin to a destination. Most transit agencies designate patterns as Inbound, Outbound and circular. Typically, any branching or scheduled route deviation is considered a separate pattern.

Access Point Identifiers. The Unique Bus Stop/Access Point ID from the Access Point Inventory. Including this term will ensure coupling between the route, and access point location and inventory information.

Access Point Sequence Number. A reference which identifies the order of access point for each *pattern* in every route. This information is important to include for level of service studies, customer service trip planning, scheduling timed transfers, estimating time of arrival information, and other applications.

Time Point. A time point is a point where time is measured. This data is used for generating operators' pieces of work, monitoring headways, constructing passenger schedules, and coordinating transfers.

Derived Data
In addition, the following may be derived from the digital base map, but is useful information:

Line Identifier. An identifier which references the coincidence of the route pattern and digital base map one dimensional object.

Line Sequence Number. An ordered number identifying the sequence of line identifiers for each pattern in every route.

---

[2] For example, buses could only travel on certain classes of roads.

For large agencies who manage a thousand or more bus stops, this database contains thousands of entries, particularly, since transfer points may be listed multiple times. (See Chapter 4, Data Exchange Specifications, for file formats for data submitted to the NTG.)

Run Database

Uses
In practice, two run databases may exist, one used for the Pick, the other derived from operator sheets or starter reports, and generated on a daily basis. This latter run database goes to the administration services/payroll.

Many transit agencies are tied to particular report forms for displaying run sheets due to past practices or union rules. This discussion does not recommend a format for these documents, it merely comments on database design issues related to their conceptual representation.

Data Automation and Conversion
The same data conversion techniques used by the transit route database apply to the run database. In fact, the run database references routes, patterns and trips.

Composition
The run database, similar to the route database, is composed of a combination of spatial features, primarily segments (not necessarily adjacent segments or consecutive time periods). There are different types of segments represented, e.g., deadhead, pull out, pull in, layover, revenue service, swing. The run database describes an operator's piece of work on a given day or time period (e.g., week day, Sat, Sun, Holiday). Also, it contains a cost measure based on duration of each trip or run.

Run Identifier. An identifier which references the coincidence of the run pattern and digital base map one dimensional object.

Sequence of paths A list of identifiers related to the transit route number, pattern and trip in sequential order of work. The path features may not be geographically sequential, that is adjacent or connected (e.g., for an operator swing).

Sequence of time points Typically, these time points refer to the start and end points of a type of path within the run, particularly, when the work is performed on multiple transit routes, as in the case of interlining.

Duration (cost) hrs/min This value is derived from the links which compose the run. Typically, each link is assigned an "impedance" which is the "cost" to traverse or turn unto or from that link. The unit may be in miles per hour or dollars per mile per hour, and may depend on other factors such as time of day, day of week, travel mode, seniority, or other condition. Duration is the sum of all these costs for each run.

Incident Database

Uses
The incident database contains the location of incidents and accidents that occur on transit property or involve a transit vehicle. The incident database is used to track the data related to

each incident. Dispatchers and operators input information related to the incident, including location, and weather and road conditions. Safety inspectors compile their reports, track liability claims, and analyze safety statistics using this database.

Data Automation and Conversion
Similar to the Access Point Inventory, the locations of these data may be geocoded as a latitude and longitude pair, offset of a milepost, location within a facility, intersection, landmark or address. Again, the criteria for selection depends on the use, but the reference method should be consistent with the database. Only one method should be used to store the data (other methods may be stored for other purposes). When one location identifier type is more appropriate for a particular application, the data may be translated through build-in or automated translation routines. Also, the location reference must be appropriate for the "scale" of the base map. Resolution of incidents within a facility are not feasible on a base map derived from DLGs of 1:100,000 scale (80 meter accuracy).

For an incident involving a transit vehicle in revenue service, reference to the transit route and access point may be significant information to include.

Composition
Incident ID. Each incident should contain a unique identifier which is assigned when an incident is first recorded.

Location Reference: Each incident requires at least one type of location identifier in order to locate it on a digital base map. Selection of a location reference depends on the use, resolution and accuracy of the data.

Facility/Vehicle identifier. The Facility/Vehicle identifier refers to the reference for a transit facility or vehicle wherein the incident occurred. This ID links the safety database to the transit vehicle and facilities databases.

If necessary:

Operator ID. This ID links the safety database to the operator.

Transit Route ID. This ID links the safety database to the transit route, pattern, and specific trip.

Access Point ID. This ID links the safety database to the bus stop or access point inventory database.

4.0 Data Exchange Specifications

Purpose

A spatial data transfer standard is a common set of formats for geospatial data usually available in the public domain. These formats are developed to enable access and extraction of geospatial data sets from/to any platform, system, or tool. For example, many transit software developers use the "TIGER" file format to transfer a base map into applications like an Automated Vehicle Location (AVL) systems, and scheduling, runcutting or customer pre-trip planning software.

The NTG must employ data exchange standards for acquiring and updating transit route information, maintaining the currentness of the digital base map, integrating the NTG with the National Transportation Atlas, making the database available on the National Information Infrastructure (NII) and allowing access to the transit infrastructure information by researchers, policy-makers and other users.

Current data exchange standards can facilitate the transfer of TIGER/Line data, augmented and corrected TIGER/Line data. However, transit route data are geospatial features not currently supported by most transfer standards. The features require data structures that support automated translation of base map data between systems, software and tools, and also, allows for manual data input by transit operators. *Appendix F -- Guidelines for Use of a Route Object for Transit Feature Types* explores the issues related to reference methods which describe route objects. Eventually, the NTG will use a representation supported by industry geospatial exchange standards. This discussion establishes requirements for transit, and highlights the advantages and disadvantages of each method.

National Transportation Atlas Exchange Specification

The fixed guideway facilities of the NTG are currently maintained by the Bureau of Transportation Statistics (BTS) as part of the National Transportation Atlas (NTA). The infrastructure files are available from BTS in a format described in the *National Transportation Atlas Data Dictionary and Database Formats* [September, 1995] (see Appendix E). The NTA Data Dictionary defines six distinct record types: Link, Node, Point, Area, Geography and Attribute. Furthermore, feature types are represented by multiple record types, grouped into three categories: Transportation Networks, Transportation Point Facilities and Areas.

Transportation Networks are composed of four related record types: Link, Node, Geography, and Attribute. Examples include highways, railroads and waterways.

Transportation Point Facilities require only two related record types: Point and Attribute. Examples include airports and multimode terminals

Areas are made up of three related record types: Areas, Geography and Attribute. Examples include Congressional Districts, States and National Parks.

The NTA will add a seventh record type for linear reference. The linear reference file type should accommodate a route feature. *A route feature is a contiguous, ordered set of 1-dimensional objects.* Transit routes, blocks, runs, fixed guideway facilities and attributes associated with contiguous links may be described by a linear reference file type. The Linear Reference file is conceptually similar to the Geography file in that a route ID replaces the line ID, and is followed by a variable number of location identifiers "that define the shape of the linear feature" [BTS, p. 8].

Any submissions to the National Transit GIS should conform to the formats identified in the NTA Data Dictionary (including the route feature type/linear reference file format when available).

Route Data Exchange Specifications

As mentioned above, route data is not easily transferred using current file formats and data models. Though routes may be ascribed to a link or node, when multiple routes overlap on a link or partial link, both the flat file and relational model become cumbersome, and many times the relational model becomes non-normalized. Using the six file formats defined in the NTA file formats and the linear reference/route object file format described in Appendix F, a transit route feature should be represented as follows:

Transit Route Alignment (<u>without</u> bus stop or time point information) is composed of the Linear Reference and Attribute record types. The Link, Node and Geography record types should be included to describe the base map upon which the transit route is overlaid.

Transit Route (<u>with</u> bus stop or time point information) is composed of the Linear Reference, Geography, Link, Node, Point, and Attribute record types. The Point record type will describe the bus stop or time point information. Attribute record types will describe the relational link to the tabular data tables. Each attribute record type possess its own data description. A subset of transit feature type attributes are listed in *Appendix B -- Transit Feature Types*. Optional attributes are described in *Chapter 3.0 -- Transit Feature Automation and Conversion Practices*.

Transit Route (with level of service information) is composed of the Linear Reference and Attribute record types. The Link, Node and Geography record types should be included to describe the base map upon which the transit route is overlaid. A standard format should be developed to capture level of service and/or timetable information. This tabular data set should be consistent and integratable with scheduling and customer information system tools currently supported by transit properties.

5.0 Maintenance and Update Recommendations

Purpose

The road network and transit routes change over time. This necessitates adding, editing and deleting information from the data set. The FTA should appoint a data custodian to monitor and record all changes to the NTG and to ensure data quality and control. This information will be available to users who access the data.

Quality Control/Quality Assurance

Referential Integrity

Definition

Referential Integrity refers to the accuracy, validity or correctness of the data in meeting the constraints and rules defined by the internal structure and content of the data. These rules apply to multiple levels of data and data base management systems. Tabular data are constrained by data domain and type rules, and data base management system rules. The relational database model must be normalized, and referential integrity constraints ensure that a unique primary key exists for each table and no foreign key is unmatched to a primary key.

General integrity constraints associated with geographical data relate to data quality, such as ensuring fundamental relationships in the graph structure. For example, the sum of the degrees of the graph vertices equals twice the number of edges, where the vertex is a node and the edge is the line between two nodes.

Maintenance Functions

The tabular data maintenance function may be almost totally automated because most DBMS provide functions to check constraints and flag violations including data type, domain, entity and referential. The data custodian needs to monitor changes to the database to verify updates and insertions to the database.

Updates and insertions of geographic data require more extensive training. Some of the techniques used to verify topologic integrity are described in the section on Data Quality in the metadata specifications (see Appendix D). Some GIS provide functions to check feature constraints. These should be used and documented in the transaction log as part of every transaction.

Transaction Processing

Definition

The process and act of changing a data set should be fully documented. The action is called a transaction. Three major functions occur during a transaction: insertion (I), update (U) or deletion (D).

Maintenance Functions

Most GIS and DBMS support insert, update and delete functions, as well as rollback (undo) and commit actions. This document identifies a format for two data sets which should be created and updated after every transaction. The two data sets -- including a ***transaction log*** that records information about the change in the data set; and a ***transaction list*** that records the new information and the old information that was replaced -- may be created by the GIS and/or DBMS. For submissions to the NTG, these *Standards* recommends a format and content to these files. The content is more important than the format. The transaction list should be available with each new version of the NTG.

The transaction process is composed of four steps: check out, change, verify, check in. A data set being changed should undergo a formal check-out and check-in procedure. Typically, an electronic library allows concurrent use of data sets for different purposes. Moreover, only a data custodian is allowed to change information in a data set. The data custodian will sign out the data for maintenance, including insertion, update or deletion. The data will be verified for accuracy, completion, and consistency, and then the new data will be signed into the system for use.

The information required as part of the transaction log includes: Transaction Number, Transaction Type, File Name, Record Type, Date, Time, User, Comment, Layer, Action(s) including verification technique. The transaction list contains the transaction number, transaction type, new information, old information, data, file name, version and revision, and record type.

(Note: these file types are consistent with the NTA Data Dictionary and Database Format record types.)

Transaction Log Format:

TRANSACTION	This unique index identifies the transaction number.
TRANSTYPE	The Transaction Type is associated with the change to the data set. These include: I insertion U update D deletion
FILNAME	This string refers to the data set file from which the data was changed.
RECTYPE	This one character field identifies the record type associated with the file. These include: L Link N Node P Point A Area G Geography T Attribute R Route (Linear Reference)
DATE	The date on which the transaction process was completed and committed. The date is entered in the format 'mmddyyyy'.
TIME	The time at which the transaction process was completed and committed. The time is entered in the format 'hh:mm' where hh is a value between 00 and 23.
USER	User is a string which refers to the person who made the changes to the data set.

COMMENT	The comment field may include any free text which describes the reason for the changes.
LAYER/TILE	This string refers to the theme of information or region contained in the file/record type. For example, bus stops in TGR20025 (Suffolk County).
ACTION(S)	This free text refers to the changes performed by the user, including verification methods, sources and other participants in the process.

Transaction List Format

TRANSACTION	This unique index identifies the transaction number.
TRANSTYPE	The Transaction Type is associated with the change to the data set. These include: I insertion U update D deletion
DATE	The date on which the transaction process was completed and committed. The date is entered in the format 'mmddyyyy'.
FILNAME	This string refers to the data set file from which the data was changed
OLD_VERSION	The version number is a 2-digit number that is incremented for all records in the database whenever a new release is distributed.
OLD_REVISION	The revision number is a 2-digit number that is incremented individually for each record whenever a change is made to one of its fields.
NEW_VERSION	The version number assigned to the database with the changed field.
NEW_REVISION	The revision number assigned to the changed field.
RECTYPE	This one character field identifies the record type associated with the file. These include: L Link N Node P Point A Area G Geography T Attribute R Route (Linear Reference)
NEW_ENTRY	This free text lists the changes to the old version/revision.
OLD_ENTRY	This free text lists the original entry to the old version/revision.

Appendix A: Glossary of Terms

ACRONYMS

APTS	Advanced Public Transportation System
BTS	Bureau of Transportation Statistics (part of the U.S. DOT)
DLG	Digital Line Graph
FGDC	Federal Geographic Data Committee
FHWA	Federal Highway Administration
FTA	Federal Transit Administration
GIS	Geographic Information System
ITS	Intelligent Transportation Systems
ITS-A	Intelligent Transportation Society of America
NTG	National Transit GIS
TIGER	Topologically Integrated Geographic Encoding and Referencing
SDTS	Spatial Data Transfer Standard
US DOT	United States Department of Transportation
USGS	United States Geological Survey
USPS	United States Postal Service
VNTSC	Volpe National Transportation Systems Center

DEFINITIONS

Much of the terminology used to represent a map database and its fundamental elements come from geography, graph theory and data processing fields. Many of these terms were adapted from the Federal Information Processing Standard (FIPS) 173, *Spatial Data Transfer Standard*[3], Chapter 1 list of definitions.

Accuracy. The closeness of results of observations, computations, or estimates to the true values or the values accepted as being true.

Arc. A locus of points that forms a curve defined by a mathematical expression.

Area. A bounded, continuous, two-dimensional object that may or may not include its boundary.

Attribute. Non-graphical data that describes or is associated with geospatial entities or spatial features. For example, schedule information and ridership are associated with bus routes.

Attribute Accuracy. Refers to accuracy related to an attribute.

Cadastral Feature. Elements created from legal descriptions of property boundaries and other land rights.

Chain. A directed nonbranching sequence of nonintersecting line segments and (or) arcs bounded by nodes, not necessarily distinct, at each end.

Completeness. Measure related to omissions, selection criteria, generalization, definitions used, and other relevant mapping rules used to derive the data set.

Digital Base Map. A spatial representation, usually graphic on a flat surface, of spatial phenomena.

Digital Image. A two-dimensional array of regularly spaced picture elements (pixels) constituting a picture.

Geocode. The process of identifying and assigning a spatial index code to an attribute which associates it to unique points, lines, or areas stored in a map database.

Graph. A set of topologically interrelated zero-dimensional (node), one-dimensional (link or chain), and sometimes two-dimensional objects that conform to a set of defined constraint rules...Three such types [are] planar graph, network, and two-dimensional manifold. All three share the following rules: each link or chain is bounded by an ordered pair of nodes, not necessarily distinct; a node may bound one or more links or chains; and links or chains may only intersect at nodes.

---

[3] SDTS, 4/21/92

Geographic Information System. Software tools for collecting, storing, retrieving, analyzing and displaying spatial data.

Grid. A set of grid cells forming a regular, or nearly regular, tesselation of a surface.

Grid Cell. A two-dimensional object that represents the smallest nondivisible element of a grid.

Layer. An areally distributed set of spatial data representing entity instances within one theme, or having one common attribute or attribute value in an association of spatial objects. In the context of raster data, a layer is specifically a two-dimensional array of attribute values associated with all or part of a grid or image.

Lineage. Information about the events, parameters, and source data which were used to construct the data set, and information on source data publication dates, documents and "responsible party" contacts.

Line Segment. A direct [straight] line between two points.

Link. A topological connection between two nodes. A link may be directed by ordering its nodes.

Logical Consistency. The measure of valid or permissible relationships encoded in the data structure of the digital spatial data.

Map Database. A digital base map and related subject information stored as a volume set, volume, file set or file.

Network. A graph which consists of a series of links and nodes and their relationships. The network may be a planar or non-planar graph. If non-planar and projected onto a two-dimensional surface, a network can have either more than one node at a point and (or) intersecting links or chains without corresponding nodes.

Node. A zero-dimensional object that is a topological junction of two or more links or chains, or an end point of a link or chain.

Pixel. A two-dimensional picture element that is the smallest nondivisible element of a digital image.

Planar Graph. The node and link or chain objects of the graph occur or can be represented as though they occur upon a planar surface. Not more than one node may exist at any given point on the surface. Links and chains may only intersect at nodes.

Planimetric Feature. A natural or cultural physical entity referenced by its horizontal position on the earth's surface. Examples include: roads, buildings and water bodies that are visible and identifiable from aerial photography.

Point. A zero-dimensional object that specifies geometric location. One coordinate pair or triplet specifies the location.

Positional Accuracy. The assessment of the closeness of a map feature location to the actual position in the universe.

Quality. An essential or distinguishing characteristic necessary for spatial data to be fit for use.

Raster. A format for storing and displaying graphic data in which image values are stored as uniform grid cells or picture elements (pixels).

Resolution. The minimum difference between two independently measured or computed values that can be distinguished by the measurement or analytical method being considered or used.

Spatial Data. Representation of facts or ideas which possess a dimension in space.

Spatial Object. Representation of dimensional objects. Elemental spatial objects include point, line and polygon.

String. A connected nonbranching sequence of line segments specified as the ordered sequence of points between those line segments.

Topology. A branch of mathematics dealing with graph theory. The spatial relationship among connecting and adjacent spatial objects. Properties include orientation, containment and proximity.

Transversal. A directed, nonbranching sequence of chains.

Two-Dimensional Manifold. A planar graph and its associated two dimensional objects. Each chain bounds two and only two, not necessarily distinct, atomic two-dimensional components. The areas are mutually exclusive and completely exhaust the surface.

Appendix B: Transit Feature Types

(based on APTS Map Database User Requirements Document)

The Transit Feature Types are mapped to a set of primitive and complex spatial objects. Each object is associated with principal attributes which describe the object. The definition of each object is printed in italics; the object's principal attributes are listed below the definition.

Spatial Object Set

POINT : 0-Dimension Object

a unit of location

location[4]

PIECE : 1-Dimension Object

a combination of start, end and shape points

Start Point
End Point
Shape Point(s)
Linear Reference[5]

POLYGON : 2-Dimension Object

two or more contiguous pieces with a common start and end point

Collection of pieces
Adjacent Polygons

PATH : Composite 1-Dimension Objects

one or more contiguous pieces

Sequence of pieces

PLANE : Composite Spatial Objects

topologically interrelated spatial elements

Collection of component spatial elements
Relationship of component spatial elements

---

[4] Representation of a point on a map, including geopolitical, address, intersection, coordinates, milepost, offset from start of Piece, etc.

[5] Representation of an address range, milepost, point or marker, link id or standard location reference id.

Spatial Transit Feature Types

Adapted from the APTS Map Database User Requirements Document, Vs. 1.0 (Oct. 21, 1995)

Map database spatial features are abstract categories which are used to specify types of services and facilities. The transit feature type focuses on common functionalities of an entity. At transit term in one agency does not necessarily mean the same thing in another. Transit professionals, even from the same organization, disagree on names and definitions of terms. Yet, all agencies providing public transportation services employ similar functionality in their systems. We chose terms that are generally understood by transit professionals and are defined such that they are not ambiguous. Feature definitions were chosen from standard reference documents and refined to capture the essence of transit functions.

ACCESS POINT (transfer point) : POINT

A point where passengers board or alight a vehicle.

ID
Name
Descriptor
Location
Type: Bus stop, park and ride, tracks, platform, pickup/dropoff.

ACCESS ZONE : POLYGON

A buffer surrounding a transit route or access point.

ID
Name
Descriptor
Collection of pieces
Adjacent Polygons
Width/Radius
Type: pedestrian, vehicle, proximity zone

ADMINISTRATIVE/POLITICAL/STATISTICAL REGION : POLYGON

(Defined by type)

ID
Name
Descriptor
Collection of pieces
Adjacent Polygons
Type: census tracts, zip code areas, area code regions

BARRIER : POINT

Any object that precludes or prevents movement through a part of the transportation network.

ID
Name
Descriptor
Location
Type: curb

BARRIER : PIECE

Any object that precludes or prevents movement through a part of the transportation network.

ID
Name
Descriptor
Start Point
End Point
Shape Point(s)
Linear Reference
Type

BLOCK : PATH

The sequence of trips made by a transit vehicle (also called a vehicle run)

ID
Name
Descriptor
Sequence of pieces
Type

CENTROID : POINT

A representative point for a polygon

ID
Name
Descriptor
Location
Type

DETOUR : PATH

A temporary series of (geographic) sequential, contiguous segments

ID
Name
Descriptor
Sequence of pieces
Duration[6]

---

6 Duration: *Period during which the detour lasts. Values include short/long term, start/end time, start/end date.*

Type

EVENT : POINT

A spatial-temporal entity that occurs over a specified period of time

ID
Name
Descriptor
Location
Duration
Type: accident or incident

EVENT : POINT

A spatial-temporal entity that occurs over a specified period of time

ID
Name
Descriptor
Location
Duration
Type: accident or incident

EVENT : PIECE

A spatial-temporal entity that occurs over a specified period of time

ID
Name
Descriptor
Location
Duration
Type: special event

EVENT : POLYGON

A spatial-temporal entity that occurs over a specified period of time

ID
Name
Descriptor
Location
Duration
Type: special event, sports venues, planning areas, construction sites

EVENT : PATH

A spatial-temporal entity that occurs over a specified period of time

ID
Name
Descriptor

Location
Duration
Type: parade, marathon, special event

FACILITIES/ ASSETS : POINT

Buildings and structures that support transit infrastructure.

ID
Name
Descriptor
Location
Type: bus stop shelter, sign

FACILITIES/ ASSETS : PIECE

Network structures or goods that support transit operations.

ID
Name
Descriptor
Start Point
End Point
Shape Point(s)
Linear Reference
Type

FACILITIES/ ASSETS : POLYGON

Buildings and structures that support transit infrastructure.

ID
Name
Descriptor
Collection of pieces
Adjacent Polygons
Type: real estate, station

FARE ZONE : POLYGON

An area partitioned by a transit agency for fare pricing

ID
Name
Descriptor
Collection of pieces
Adjacent Polygons
Type

MAP : PLANE

A graphical representation on a plane of certain selected features of a part or the whole of the surface of the Earth or any other entity.

ID
Name
Descriptor
Collection of component spatial elements
Relationship of component spatial elements
Type: Bicycle paths, Transit Routes

MODE PATH : PATH

A collection of pieces with restricted mode use.

ID
Name
Descriptor
Sequence of pieces
Type: bike trail/path, pedestrian way, HOV lanes, bus only lanes

NETWORK : PLANE

A collection of points and pieces that have a relationship to eachother

ID
Name
Descriptor
Collection of component spatial elements
Relationship of component spatial elements
Type: road, transit, transportation, bicycle, pedestrian, communication

NODE : POINT

A (topological) connection between pieces, or the starting or end point of a segment.

ID
Name
Descriptor
Location
Type: Intersection, time point[7], (see note [8])

[7]Time Point: *A location at which time is measured.*
[8] Dependent on function of segment.

ORIGIN/ DESTINATION : POINT

The beginning or end of a trip-maker's trip.

ID
Name
Descriptor
Location
Type: Origin, Destination

PARCEL : POINT

A package or group of packages or things

ID
Name
Descriptor
Location
Type: wheelchair, bicycle, luggage

PERSON : POINT

A human being

ID
Name
Descriptor
Location
Type: tripmaker, operator, customer

POINT OF INTEREST (POI) : POINT

A point of interest

ID
Name
Descriptor
Location
Type: landmark, observation point

POI : POLYGON

A polygon of interest

ID
Name
Descriptor
Collection of pieces
Adjacent Polygons
Type: mall, hospital, landmark

RUN : PATH

A transit operator's assignment of trips per day.

ID
Name
Descriptor
Sequence of pieces
Cost[9]
Service[10]
Type: straight, swing, part-time, tripper

SEGMENT : PIECE

A piece in a network.

ID
Name
Descriptor
Start Node
End Node
Shape Point(s)
Linear Reference
Type: road, bicycle, rail, communication, foot, "transit", "transportation"

SERVICE AREA : POLYGON

A legal, jurisdictional or functional area in which a transit agency provides its service.

ID
Name
Descriptor
Collection of pieces
Adjacent Polygons
Type: full, commuter, ADA, access

STANDBY, HOLDOVER LOCATION : POINT

A point where a non operating vehicle waits for an assignment (other than its storage facility)

ID
Name
Descriptor
Location
Type

---

[9] Cost: *Outlay of monetary value to operate a run.*

[10] Service: *Refers to day of week to which run applies: 1 = weekday, 2 = Saturday, 3 = Sunday, 4 = Holiday.*

STREET : PATH

A collection of road pieces that have a common name.

ID
Name
Descriptor
Sequence of pieces
Type

TOLL : POINT

A point at which a toll is collected or accumulated

ID
Name
Descriptor
Location
Type: plaza, smart card reader

TRANSIT ROUTE : PATH

A collection of patterns[11] *("paths") in a revenue service*

ID
Name
Route number[12]
Descriptor
Sequence of pieces
Collection of Time Points
Collection of Access Points
Type: Bus, Light Rail, Commuter Rail, Heavy Rail, emergency, fixed, variable, express, limited, supplemental service

TRANSIT VEHICLE : POINT

A motorized conveyance owned by a transit agency.

ID
Name
Descriptor
Location
Type: bus, articulated vehicle, auto, van

---

[11] Pattern: *A series of contiguous segments linking origins to destinations.*

[12] Route number: *A unique identifier assigned to a transit route.*

TRANSIT VEHICLE : PIECE

A motorized conveyance owned by a transit agency.

ID
Name
Descriptor
Start Point
End Point
Shape Point(s)
Linear Reference
Type: train

TRIP : PATH

A one-way movement of a person or vehicle between two points.

ID
Name
Descriptor
Sequence of pieces
Type: person, vehicle

Person Trip : PATH

(includes part of a trip that is not part of transit service)

Type (linked, unlinked)
Collection of origin(s) and destination(s)

Vehicle Trip : PATH

(includes deadhead, pull in, pull out)

Type (revenue, non-revenue)
Collection of time points
Collection of access points

Appendix C: *Standards* Change Forms

The following forms are included as a means to submit changes, corrections, and enhancements to the *FTA National Transit GIS: Standards, Guidelines and Recommended Practices*. Suggestions and recommendations are welcomed and encouraged, particularly by NTG users and transit operators. These *Standards* are intended to meet the needs of the many constituent groups using the NTG. The following forms are provided for your convenience in submitting revisions and comments:

The **Standards Correction Form** should be used to identify errors in this document.

The **Standards Review Form** should be used to submit changes to the content version, or recommend enhancements or additional guidelines to forthcoming *Standards*.

The **Transit Feature Review** should be used to comment on the Transit Feature Types defined in Appendix B. The Advanced Public Transportation Systems (APTS) Map and Spatial Database Working Group of ITS-America is the forum which reviews the nomenclature and specifies the attributes for those terms. These review forms will be examined and forwarded to that group for consideration.

Standards Correction Form

The National Transit GIS: Standards, Guidelines and Recommended Practices contains errors on the following pages:

____

____

____

____

A copy of the page(s) referenced with the error(s) noted, are attached.

Contributed by:

Name ______________________________

Organization ______________________________

Address ______________________________

____

City __________________ State ________ Zip ________

Phone __________________ Fax __________________

Date Submitted ______________________________

Please mail or fax to: NTG Guidelines c/o VIGGEN Corporation
21 Union St., 2nd Floor
Boston, MA 02108
ph: (617) 742-5114 fx: (617) 742-5124
email: okunieff@world.std.com or pokunieff@viggen.com

Standards Review Form for the *National Transit GIS: Standards, Guidelines and Recommended Practices*

(Please submit a separate form for each comment, recommendation or enhancement.)

❑ Comment on Existing Std ❑ Revision of Existing Std ❑ New Std

Standard, Guideline or Recommended Practice: ____________________

____

____

____

Page No. (if applies): __________

Comment (include justification and sources):
Attach additional comments on separate page.

____

____

____

____

____

Contributed by:

Name ____________________

Organization ____________________

Address ____________________

____

City ____________________ State __________ Zip __________

Phone ____________________ Fax ____________________

Date Submitted ____________________

Please mail or fax to: NTG Guidelines c/o VIGGEN Corporation
21 Union St., 2nd Floor
Boston, MA 02108
ph: (617) 742-5114 fx: (617) 742-5124
email: okunieff@world.std.com or pokunieff@viggen.com

Transit Feature Review Form

Please check the items which apply:

❑ Recommend Changes to Transit Feature Type(s)

❑ Recommend New Transit Feature Type(s)

Change to Existing Transit Feature Type

Current Feature Type	Page No(s)

Recommended Revision/New Specification:

____

____

____

Contributed by:

Name ______________________________

Organization ______________________________

Address ______________________________

____

City ______________ State ________ Zip ________

Phone ______________ Fax ______________

Date Submitted ______________________________

Please mail or fax to: NTG Guidelines c/o VIGGEN Corporation
21 Union St., 2nd Floor
Boston, MA 02108
ph: (617) 742-5114 fx: (617) 742-5124
email: okunieff@world.std.com or pokunieff@viggen.com

Appendix D Adaptation of the FGDC *Content Standard for Digital Geospatial Metadata* for the National Transit GIS

(Note: Most of the definitions come from the Content Standards for Digital Geospatial Metadata Workbook, FGDC, March 24, 1995)

Identification Information

Identification Information is the basic information about the data set.

Contact Name -- the name of an organization or individual that developed the data set.
Publication Date -- date of publication or release.
Title of Data Set -- name by which data set is known
Edition -- version number
Publication Information
    Publication Place --name of city and state where the data set was published
    Publisher -- name of organization that published data set
On-Line Linkages -- name of on-line computer resource that contains the data set. Entries should follow the Uniform Resource Locator conventions of the Internet.
Larger Work Citation -- the information identifies the larger work in which this data set is included. Use same fields as documented in the section Identification Information (above).

Example
Identification Information
    Contact Name: William Wiggens, FTA, (compiler)
    Publication Date: January 1996
    Title of Data Set: National Transit GIS.AnyTown_MTA
    Edition: 1996
    Publication Information
        Publication Place: Bridgewater, MA
        Publisher: Federal Transit Administration
    On-Line Linkages: ftp.bts.gov
    On-Line Linkages: http:\\www.bts.gov
    On-Line Linkages: gopher.bts.gov

    Larger Work Citation:
        Contact Name: Geographic Information Services, Bureau of Transportation Statistics, Bruce Spear (comp.)
        Publication Date: 1996
        Title of Data Set: National Transportation Atlas Databases
        Edition: 1996
        Publication Information
            Publication Place: Washington, D.C.
            Publisher: Bureau of Transportation Statistics
        On-Line Linkages: ftp.bts.gov
        On-Line Linkages: http:\\www.bts.gov
        On-Line Linkages: gopher.bts.gov

Description. Description of data set included intended use.

Abstract -- a brief narrative summary of the data set.
Purpose -- a brief narrative on why the data was developed.
Supplemental Information -- other descriptive information about the data source.

Example
Description
Abstract: The NTG.AnyTown_MTA depicts the 15 square mile service area and 27 route patterns of AnyTown MTA. The nominal scale is 1:100,000 with an error of ±80 meters.
Purpose: The provides information about coverage of public transit and level of service available in AnyTown.
Supplemental Information: The base map data source is from the Census TIGER/Line files (1992).

Time Period of Content. Time period for which the data set is valid.

Currentness Reference -- The basis on which the time period of content information is valid. The acceptable reply for the NTG should be based on the transit feature and attribute information.
Range of Dates -- the beginning and ending date for the time period of content
Beginning Date -- mm/dd/yy
Ending Date -- mm/dd/yy

Example
Time Period of Content
Currentness Reference AnyTown MTA System Map
Range of Dates
Beginning Date -- 03/15/95
Ending Date -- 06/30/95

Status. The state of and maintenance information for the data set.
Progress -- the state of the data set (e.g., complete, in-work, planned)
Maintenance and Update Frequency -- the frequency of changes and additions made to the data set after the initial set is completed (*e.g.*, continually, monthly, annually, as needed, daily, weekly, none planned, unknown).

Example
Status
Progress: In-work
Maintenance and Update Frequency As needed

Spatial Domain. The geographic extent of the data set.

Bounding Coordinates -- the limits of coverage of a data set expressed by latitude and longitude values.

West Bounding Coordinate -- west most coordinate of the limit of coverage expressed in longitude.
East Bounding Coordinate -- east most coordinate of the limit of coverage expressed in longitude.
North Bounding Coordinate -- north most coordinate of the limit of coverage expressed in latitude.
South Bounding Coordinate -- south most coordinate of the limit of coverage expressed in latitude.

Example
Spatial Domain
Bounding Coordinates
West Bounding Coordinate: -xxx.xxxxxx
East Bounding Coordinate: -xxx.xxxxxx
North Bounding Coordinate: xx.xxxxxx
South Bounding Coordinate: xx.xxxxxx

Keywords. Words or phrases summarizing an aspect of the data set.
Theme -- subjects covered by the data set.
Theme Keyword Thesaurus -- reference to a formally registered thesaurus or a similar authoritative source of theme keywords. Possible transit sources include: *Guidelines*, Chapter 3.0: 1996, *Reporting Manual*: 1995 National Transit Database Report Year; TRANSMODEL.
Theme Keyword -- common-use or thesaurus word or phrase used to describe the subject of the data set.

Place -- geographic locations characterized by the data set.
Place Keyword Thesaurus -- reference to a formally registered thesaurus or a similar authoritative source of place keywords. (e.g., Geographic Names Information System, APTA Directory, National Transit Database (Section 15) Profiles).
Place Keyword -- the geographic name of a location covered by a data set which uses a common name or one from the thesaurus.

Stratum -- layered, vertical locations characterized by the data set.
Stratum Keyword Thesaurus -- reference to a formally registered thesaurus of a similar authoritative source of stratum keywords.
Stratum Keyword -- the name of a vertical location used to describe the locations covered by a data set.

Temporal -- time period(s) characterized by the data set.
Temporal Keyword Thesaurus -- reference to a formally registered thesaurus of a similar authoritative source of temporal keywords. This could include patterns and level of service during "a.m. peak" or "p.m. peak."
Temporal Keyword -- the name of a time period covered by a data set

Example
Keywords
Theme

Theme Keyword Thesaurus: Standards, Guidelines and Recommended Practices for the National Transit GIS.
Theme Keyword: Route Database

Theme Keyword Thesaurus: None
Theme Keyword: Network
Theme Keyword: Transportation
Theme Keyword: Transit
Theme Keyword: National Transit GIS
Theme Keyword: TIGER/Line

Place
Place Keyword Thesaurus: None
Place Keyword AnyTown

Place Keyword Thesaurus: National Transit Database Profiles
Place Keyword: AnyTown MTA

Temporal
Temporal Keyword Thesaurus: AnyTown MTA Timetable
Temporal Keyword: Spring

Access Constraints. Restrictions and legal prerequisites for accessing the data set. These include any access constraints applied to assure the protection or privacy or intellectual property, and any special restrictions or limitations on obtaining the data set.

Example
Access Constraints: None

Use Constraints. Restrictions and legal prerequisites for using the data set. These include any use constraints applied to assure the protection or privacy or intellectual property, and any special restrictions or limitations on using the data set.

Example
Use Constraints: Should not be used for schedule information.

Point of Contact. Contact information for an individual or organization that is knowledgeable about the data set.
Primary Contact Person -- primary individual to contact about the data set.
Contact Person -- name of contact person
Contact Organization -- organization of contact person
Contact Address -- postal address of contact
Address
City
State
Postal Code

Country
Contact Voice Number
Contact Facsimile Number
Contact Electronic Mail Address

Example
Point of Contact
Primary Contact Person
Contact Person: Larry Harman
Contact Organization: Bridgewater State College
Contact Address
City: Bridgewater
State: MA
Postal Code: 02325
Country: USA
Contact Voice Number: 508-279-6144
Contact Electronic Mail Address: lharman@bridgew.edu

Data Set Credit. Recognition of those who contributed to the data set.

Example
Data Set Credit: The summer interns at Bridgewater State College. This work was supported by the Federal Transit Administration and the Volpe National Transit Systems Center.

Native Data Set Environment. A description of the data set in the producer's processing environment.
Software Tool (version):
Operating System:
File Name:
Size:

Example
Native Data Set Environment
Software Tool (version): Caliper's TransCAD (2.1 and 3.0)
Operating System: Windows 3.1
File Name: AnyTown.xxx
Size: ???

Data Quality

Data users use quality information to help evaluate the adequacy and applicability of geographic data sources for a particular use. In particular, the documentation related to the data should provide answers to the following questions:

"How good are the data? Is information available that allows a user to decide if the data are suitable for his or her purpose? What is the positional and attribute

> accuracy? Are the data complete? Were the consistency of the data verified? What data were used to create the data set, and what processes were applied to these sources?" [from *Content Standards for Digital Geospatial Metadata Workbook*, Workbook Version 1.0; Federal Geographic Data Committee, March 24, 1995.]

This section describes categories in which an assessment of the quality of the data set is documented. The descriptions used were extracted (without explicit reference) from the Spatial Data Transfer Standard (FIPS 173) and Digital Geospatial Metadata Workbook which describes the June 8, 1994 version of the FGDC Metadata standard. Data quality information pertaining to general TIGER/Line data quality can be found in Census TIGER/Line file documentation (i.e., chapter 5 and other referenced data files).

Attribute Accuracy

Attribute accuracy is the association and assessment of the accuracy of feature characteristics in the data set, and the assignment of values to those characteristics. A report describing attribute validation methods should accompany the data set. Only those features populated by the data custodian should be documented. The attribute accuracy report for the underlying TIGER/Line data, described in TIGER documentation, need not be documented unless updated or augmented by others.

All attribute data should be verified using an accuracy test. The report of a test of attribute accuracy should include the date of the test and the dates of the materials used. The report should make reference to the map scale.

Recommended accuracy tests for attributes include:

Deductive Estimate. These tests are based on assessing potential errors that may occur in each production step and the propagation of those error. Any estimate, even a guess based on experience, is permitted. The basis for the deduction shall be explained. Statements such as "good" or "poor" should be explained in as quantitative a manner as possible.

Tests based on Comparison with Independent Sample. The most common technique involves comparing the source data with another source of equal or higher quality data (including ground truth) for a statistically significant sample region. In the case of bus stops, a list of physical bus stops may be chosen at random to compare to the source data. The source data is compared against the independent sample. Variances in measurements are noted and a misclassification matrix should be reported as counts of sample units crosstabulated by the categories of the sample and tested material. The sampling procedure and the location of the sample units should be described.

Tests based on Polygon Overlay. Similar to the comparison of independent samples, this test compares the source data with another source of equal or higher quality data. This test involves comparing two data sources using a set of rules such as the U.S. National Map Accuracy Standards [US Bureau of the Budget, revised June 17, 1947]. A misclassification matrix shall be reported as areas. The relationship between the two maps shall be explained; as far as possible, the two sources should be independent and one should have higher accuracy.

Logical Consistency

Logical consistency is the measure of valid or permissible relationships encoded in the data structure of the digital spatial data. A report on logical consistency should detail the tests performed and the results of the tests. Recommended tests include tests for valid values, general tests for graphic data, and specific topological tests. The tests on graphic data and topology apply to the underlying base map. The Census Bureau subjects TIGER to general and specific automated testing to validate feature and attribute values and ensure node-line-area relationships satisfy topological requirements. The node-line-area requirements include:

1. Complete chains must begin and end at nodes;
2. Complete chains must connect to each other at nodes;
3. Complete chains do not extend through nodes;
4. Left and right polygons are defined for each complete chain element and are consistent for complete chains connecting at nodes;
5. Complete chains representing the limits of a file are free from gaps.

Transit feature and attribute data (specifically route structures), and geospatial data added to the TIGER data should undergo the following tests:

Test for Valid Values. Tests for permissible values may be applied to any data structure. Use of a database management system (DBMS) allows application of range checking and referential integrity triggers and procedures during data input. Such a test can detect major errors, but may not ensure all aspects of logical consistency.

General Tests for Graphic Data. A data source containing lines may be subjected to the following general questions:

-- Do lines intersect only where intended?
-- Are any lines entered twice?
-- Are there any overshoots or undershoots?
-- Are any polygons too small, or any lines too close?

Different tests may be applied to address these questions, but the quality report should contain a description of the tests applied or a reference to documentation of the software used.

Specific Topological Tests. Topologic tests should be performed to ensure that chains intersect at nodes and complete closure of chains bounding polygons. Conditions may be verified by automated procedures and reports generated. The quality report should identify the software (name and version) used to verify these conditions.

Date of Test. The report shall include the date on which the tests were applied. Test report dates of subsequent tests performed after update should be included in this field.

Completeness

The assessment of completeness measures information about omissions, selection criteria, generalization, definitions used, and other relevant mapping rules used to derive the data set. For example, TIGER/Line documentation identify four areas in their completeness report: attribute codes of original source data, Census Bureau feature and attribute internal codes, feature network, and landmark features. [From TIGER/Line Documentation, Chapter 5]:

> "TIGER/Line contains at least the same level of content and detail as shown on the source[;]
>
> ...Information used to create the [internal code] file is as complete as possible[;]
>
> ...[and] In some areas, local officials reviewed the census maps and identified new features and feature changes.
>
> The TIGER/Line files contain limited point and area landmark data."

This report is particularly important for the NTG. The transit infrastructure, e.g., fixed facilities and route miles, provides valuable information for policy decision makers. The completeness report should evaluate the existence of all graphic and non-graphic data in the data set. The graphic data may be tested by exercising random network paths (linked trip planning) and non-graphic data may be compared against a master list of the universe of data and tested for all necessary data fields in the data file.

The procedures used for testing and the results should be described in the quality report. For example, a test of the completeness of bus stops in a transit property's service area should be compared with the agency's master bus stop data list. All the bus stop data file attributes including amenities, location references, etc. should be examined for entries.

Positional Accuracy

Positional accuracy is the assessment of the closeness of a map feature location to the actual position in the universe. Horizontal and vertical positions should be evaluated separately. The TIGER/Line is a planimetric map and has no information on elevation. Other digital base maps store the vertical or "z" position in an attribute field.

The Census Bureau's mission does not compel it to be concerned about positional accuracy. The features based on the Digital Line Graph files inherit the positional accuracy of their source. So, the positional accuracy of the USGS DLG who possess an accuracy of approximately $\pm$ 167 feet (1:100,000-scale maps) are specified as meeting the National Map Accuracy Standards for 1:100,000-scale maps. The Census Bureau cannot specify the accuracy of updated features and GBF/DIME-File features. Visual comparison tests against source materials were made with check plots. [from TIGER Documentation, Chapter 5]

Positional Accuracy

The quality report portion on positional accuracy should include the degree of compliance to a spatial registration standard. More information on spatial registration standards can be found in Section 4.1.3.5 of SDTS.

Suggested methods to obtain measures of positional accuracy include similar methods described for attribute accuracy [from Workbook].

Deductive Estimate. Any deductive statement based on knowledge of errors in each production step shall include reference to complete calibration tests and shall also describe assumptions concerning error propagation. Results from deductive estimates shall be distinguished from results of other tests.

Internal Evidence. Federal Geodetic Control Committee procedures will be used for tests based on repeated measurement and redundancy such as closure of traverse or residuals from an adjustment.

Comparison to Source. When using graphic inspection of results ("check plots"), the geometric tolerances applied shall be reported and the method of registration shall also be described. Use of check plots shall be included in the lineage portion.

Independent Source of Higher Accuracy. The preferred test for positional accuracy is a comparison to an independent source of higher accuracy. The test shall be conducted using the rules prescribed in the "ASPRS Accuracy Standards for Large Scale Maps" [see Section 1.3.3 of SDTS]. When the dates of testing and source material differ, the report shall describe procedures used to ensure that the results relate to positional error and not to temporal effects. The numerical results in ground units, as well as the number and location of the test points, shall be reported. A statement of compliance to a particular threshold is not adequate in itself. This test may only be applicable to well-defined points.

Lineage

Lineage is the information about the events, parameters, and source data which were used to construct the data set, and information on source data publication dates, documents and "responsible party" contacts. The FGDC Metadata standard has more detailed requirements for lineage including references to specific control information and transformation algorithms.

The NTG metadata will not explicitly cite the TIGER/Line lineage since it is well documented elsewhere.

The lineage fields relevant to the National Transit GIS are described below:

Source Information. A list of sources and a discussion of the information extracted from each.

Source Citation
- Originator:
- Publication Date:
- Title:

Publication Information:
    Publication Place:
    Publisher:
Source Scale Denomination
Source Time Period of Content

Process Step. A description of the feature automation or conversion processes.

    Process Description
    Process Date
    Source Produced Citation Name
    Process Contact

Update Procedures. A description of the update procedures.

    Update Procedures
    Update Date
    Update Contact

Example

Source Citation
    Originator: Census Bureau
    Publication Date: 1992
    Title: TIGER/Line files
Publication Information
    Publication Place: Washington, DC
    Publisher: Bureau of the Census
    Publication Metadata/Documentation Citation: TIGER/Line® Files, 1992 Technical Documentation prepared by the Bureau of the Census. Washington, DC: The Bureau, 1992.

Source Citation
    Originator: AnyTown MTA
    Publication Date: Spring 1995
    Title: AnyTown MTA System Map
Publication Information:
    Publication Place: AnyTown
    Publisher: AnyTown MTA

Source Scale Denomination Not Applicable
Source Time Period of Content
    Start Date: 03/15/95
    End Date: 06/30/95
    Publication Metadata/Documentation Citation:

Process Step
    Process Description: Route patterns were extracted from system maps (for small transit agencies), route schedules (for larger agencies), operator

directions, and other GIS data sources supplied by transit agencies. The TRANSCAD 2.1 and 3.0 software tool route function was used to code the networks, and route data bases were stored in TRANSCAD route files.

Process Date: June - August 1995
Source Produced Citation Name: NTG Route Patterns xxx_xx.xx
Process Contact: Madu Rao, Moakley Center, Bridgewater State College, Bridgewater, MA. 02325 (508) 697-1390 x2120, email: mrao@bridgew.edu

Update Procedures
Update Procedures: See Chapter 5, National Transit GIS Standards, Guidelines and Recommended Practices. FTA, Washington, D.C. Draft, December 1995.
Update Contact: Paula Okunieff, Viggen Corporation, 21 Union Street, Boston, MA 02108. (617) 742-5114; Fax (617) 742-5124.

Spatial Data Organization Information

Spatial Data Organization Information is the mechanism used to represent spatial information in the data set. The data may be represented using an "indirect" or "direct spatial reference. The direct spatial reference uses spherical or Cartesian coordinates (e.g., latitude and longitude) to describe point and vector objects. The indirect spatial reference uses any other method to describe the location of a spatial object. The reference may use a name (e.g., Cook County, I-94), code (e.g., FIPS code) or linear identifier (including street address, milepoint, or distance along a linear feature).

Indirect Spatial Reference. Name of types of geographic features, addressing schemes, or other means through which locations are referenced in the data set.

Reference Method
Labeled Point ID: (St Name, x, y) ! see *Address and Street Naming Conventions*
Intersection: (On St Name, At St Name)
Node ID: (Node #) ! *e.g.*, node no., bus stop no., time point no.
Milepost: (Section No., Offset) ! offset = Milepost + 00.nnn
Reference: (Node ID, Distance Offset) ! or
(On St, At St, Distance offset) ! direction based on conventions ! or
(On St, At St, Offset Code) !offset code=[NS, FS, MB, @, OPP]
Address: (Primary Address No., Predirectional, Street Name, Suffix, Postdirectional)

Direct Spatial Reference Method. The system of objects used to represent space in the data set.

Point and Vector Object Information -- the types and numbers of vector spatial objects in the data set.

SDTS Point and Vector Object Type -- name of point and vector spatial objects used to locate zero-, one-, and two-dimensional spatial locations in the data set.

"Node, network"
"Link"
"Complete chain"
"GT-polygon composed of chains"
"Route"

Point and Vector Object Count -- the total number of the point or vector object type occurring in the data set.

Example
Indirect Spatial Reference.
Reference Method: Address

Direct Spatial Reference Method
Point and Vector Object Information
SDTS Point and Vector Object Type: Route (Nt'l Transit GIS), Node (TIGER), Complete chain (TIGER), GT-polygon (TIGER)
Point and Vector Object Count: [no. of links]

Spatial Reference Information

Spatial reference information refers to the description of the reference frame for, and the means to encode, coordinates in the data set.

Horizontal Coordinate System Definition. The reference frame or system from which linear or angular quantities are measured and assigned to the position that a point occupies.

Geographic -- the quantities of latitude and longitude which define the position of a point on the Earth's surface with respect to a reference spheroid.

Latitude Resolution -- the minimum difference between two adjacent latitude values expressed in Geographic Coordinate Units of measure.

Longitude Resolution -- the minimum difference between two adjacent longitude values expressed in Geographic Coordinate Units of measure.

Geographic Coordinate Units -- units of measure used for the latitude and longitude values. Units may be expressed in the following formats: "Decimal degrees" "Decimal minutes" "Decimal seconds" "Degrees and decimal minutes" "Degrees, minutes, and decimal seconds" "Radians" "Grads"

Example

Horizontal Coordinate System Definition. The reference frame or system from which linear or angular quantities are measured and assigned to the position that a point occupies.

Geographic

Latitude Resolution: 0.xxxxx

Longitude Resolution: 0.xxxxx

Geographic Coordinate Units: Degrees, minutes, and decimal seconds

Entity and Attribute Information

The entity and attribute information section contains information about the content of the data set, including the entities types, their attributes, and the domains from which attribute values may be assigned. This section is particularly important for core transit data described in Chapter 3.0 and Transit Feature Types specified in Appendix B. This section provides a way to describe the meaning of transit feature data, attributes, and attribute value information so users understand the information content of the National Transit GIS (or any transit related data set) and use the data appropriately.

The SDTS has provided a convenient mechanism to map specific glossaries to a standard glossary. The feature type is an abstract data type; an "included terms" as defined by the SDTS is an instance of a feature type or the particular term used by a transit property to describe that feature. For example, a transit route may be referred to as a line, journey, pattern or route. The *APTS Map Database User Requirements Document* includes an appendix of included terms (Appendix 1 Data Template).

The Overview Description provides the elements needed to give users a sense of the information content and a reference to the data dictionaries or source(s) of the complete description. The Detailed Description provides the elements needed to describe these meanings. Appendix B of this document may be substituted for the detailed description section. The two approaches can be used together to document a data set.

Detailed Description. Description of the entities, attributes, attribute values, and related characteristics encoded in the data set.

Entity Type -- the definition and description of a set into which similar entity instances are classified.

Entity Type Label -- the name of the entity type.
Entity Type Definition -- the description of the entity type.
Entity Type Definition Source -- the authority of the definition.

Attribute -- A defined characteristic of an entity.

Attribute Label -- the name of the attribute.
Attribute Definition -- the description of the attribute.
Attribute Definition Source -- the authority of the definition.
Attribute Domain Values -- the valid values that can be assigned for an attribute.

The domain values list acceptable values for an attribute. For example, the domain for the attribute "pad" consists of concrete, asphalt, and dirt. The Metadata standard contains four types of domain value types: enumerated, range, codeset and unrepresentable. The enumerated consists of a list of values, similar to the example above. The range domain is comprised of a sequence, series or scale of values, typically numeric values. The codeset domain is defined as a set of codes, for example the Federal Information Processing Standards contains numeric codes for nations, States and counties. An unrepresentable domain is one for which the set of data values cannot be represented. The Workbook states:

> "Reasons include attributes whose values do not exist in a known, predefined set (for example, the values for an attribute of people's names), or attributes whose values cannot be depicted using the forms of representation (available character set, etc.) used for the metadata. In these cases, the information content of the set of values should be provided." [Workbook, 5-2]

The following are fields related to attribute domain values:

Enumerated Domain -- the members of an established set of valid values.

> Enumerated Domain Value -- the name or label of a member of the set.
> Enumerated Domain Value Definition -- the description of the value.
> Enumerated Domain Value Definition Source -- the authority of the definition.

Range Domain -- the minimum and maximum values of a continuum of valid values.

> Range Domain Minimum -- the least value that the attribute can be assigned.
> Range Domain Maximum -- the greatest value that the attribute can be assigned.

Codeset Domain -- reference to a standard or list which contains the members of an established set of valid values.

> Codeset Name -- the title of the codeset.
> Codeset Source -- the authority for the codeset.

Unrepresentable Domain -- description of the values and reasons why they cannot be represented.

The following is additional information which describes the attributes:

Attribute Units of Measure -- the standard of measurement for an attribute value.

Attribute Measurement Resolution -- the smallest unit increment to which an attribute value is measured.

Beginning Date of Attribute Values -- earliest or only date for which the attribute values are current. In cases when a range of dates are provided, this is the earliest date for which the information are valid.

Ending Date of Attribute Values -- latest date for which the information are current. Used in cases when a range of dates are provided.

Attribute Value Accuracy Information -- an assessment of the accuracy of the assignment of attribute values. This is an individual accuracy measure field assigned to each attribute

Attribute Value Accuracy -- an estimate of the accuracy of the assignment of attribute values.
Attribute Value Accuracy Explanation -- the definition of the Attribute Value Accuracy measure and units, and a description of how the estimate was derived.
Attribute Measurement Frequency -- the frequency with which attribute values are added.

Example
Detailed Description.
Entity Type
Entity Type Label: Transit Route
Entity Type Definition : A collection of patterns in a revenue service
Entity Type Definition Source: *APTS Map Database User Requirements Document*, Version 1, October 21, 1994.

Attribute
Attribute Label: ID
Attribute Definition: Unique sequential number assigned to each database entry
Attribute Domain Values: Integer

Attribute Label: Name
Attribute Definition: An alphanumeric identifier
Attribute Domain Values: Free Text

Attribute Label: Route number
Attribute Definition: A unique identifier assigned to a transit route
Attribute Domain Values: Integer

Attribute Label: Descriptor
Attribute Definition: A free text comment on the entity
Attribute Domain Values: Free Text

Attribute Label: Sequence of Pieces
Attribute Definition: A link to the geographic data file which defines the route geospatial description

Attribute Label: Collection of Time Points
Attribute Definition: A link to a point data file which defines the time point locations associated with the transit route.

Attribute Label: Collection of Access Points
Attribute Definition: A link to the point data file which defines the access point

locations associated with the transit route.

Attribute Label: Type
Attribute Definition: A name given to the transit route that describes the service provided.
Attribute Domain Values: "Bus", "Light Rail", "Commuter Rail", "Heavy Rail", "emergency", "fixed", "variable", "express", "limited", "supplemental service", free text.

Overview Description. -- summary of, and citation to detailed description of, the information content of the data set.

Entity and Attribute Overview -- detailed summary of the information contained in a data set.

Entity and Attribute Detail Citation -- reference to the complete description of the entity types, attributes, and attribute values for the data set.

Example
Entity and Attribute Overview
The National Transit GIS includes rural and urban bus systems, commuter rail, subways, light rail, people mover systems, high occupancy vehicle systems, ferry terminals and transitways, and also data on population served, ridership, level of service, passenger miles and route/rail miles for these modes of public transit.

Entity and Attribute Detail Citation: Fixed guideway systems: Spear, Bruce, (1995) National Transportation Atlas schema: 1996. Bus route database: Harman, Larry, (1995) National Transit GIS data dictionary.

Distribution Information

Distribution Information -- information about the distributor of and options for obtaining the data set.

Distributor -- the party from whom the data set may be obtained.
- Contact Person Primary
  - Contact Person
  - Contact Organization
- Contact Organization Primary
  - Contact Position
  - Contact Address
    - Address
    - City
    - State or Province
    - Postal Code
    - Country
  - Contact Voice Telephone
  - Contact TDD/TTY Telephone
  - Contact Facsimile Telephone

Contact Electronic Mail Address
Hours of Service
Contact Instructions

Distribution Liability -- statement of the liability assumed by the distributor.

Standard Order Process -- the common ways in which the data set may be obtained or received, and related instructions and fee information.

Non-digital Form -- the description of options for obtaining the data set on non-computer-compatible media.

Digital Form -- the description of options for obtaining the data set on computer-compatible media.

Digital Transfer Information - description of the form of the data to be distributed.

Format Name -- the name of the data transfer format. Codes used for different formats include:

"ARCE" ARC/INFO Export format
"ARCG" ARC/INFO Generate format
"ASCII" ASCII file, formatted for text attributes, declared format
"COORD" User-created coordinate file, declared format
"DGN" Microstation format (Intergraph Corporation)
"DIGEST" Digital Geographic Information Exchange Standard
"DLG" Digital Line Graph (U.S. Geological Survey)
"DWG" AutoCAD Drawing format
"DX90" Data Exchange '90
"DXF" AutoCAD Drawing Exchange Format
"IGDS" Interactive Graphic Design System format (Intergraph Corporation)
"IGES" Initial Graphics Exchange Standard
"MIF" MapInfo Interchange Format (MapInfo Corporation)
"NTG" Route Data Structure (common ASCII format)
"SDTS" Spatial Data Transfer Standard (Federal Information Processing Standard 173)
"RTS" Route Data Format (Caliper Corporation)
"TIFF" Tagged Image File Format
"TGRLN" Topologically Integrated Geographic Encoding and Referencing (TIGER) Line format (Bureau of the Census)

Format Version Number -- version number of the format.

Format Version Date -- date of the version of the format.

Format Specification -- name of a subset, profile, or product specification of the format.

File Decompression Technique -- recommendations of algorithms or processes (including means of obtaining these algorithms or processes) that can be applied to read or expand data sets to which data compression techniques have been applied.

Transfer Size -- the size, or estimated size, of the transferred data set in megabytes.

Digital Transfer Option -- the means and media by which a data set is obtained from the distributor.

Online Option -- information required to directly obtain the data set electronically.

Computer Contact Information -- instructions for establishing communications with the distribution computer.

Network Address -- the electronic address from which the data set can be obtained from the distribution computer.

Network Resource Name -- the name of the file or service from which the data set can be obtained. For example: ftp:ftp.bts.gov/pub/

Offline Option -- information about media-specific options for receiving the data set.

Offline Media -- name of the media on which the data set can be received. This field contains formats such as: "CD-ROM" "3-1/2 inch floppy disk" "5-1/4 inch floppy disk" "9-track tape" "4 mm cartridge tape" "8 mm cartridge tape" "1/4-inch cartridge tape"

Recording Capacity -- the density of information to which data are written. Used in cases where different recording capacities are possible.

Recording Density -- the density in which the data set can be recorded.

Recording Density Units -- the units of measure for the recording density.

Recording Format -- the options available or method used to write the data set to the medium. Examples include: "zip" "tar" "High Sierra" "ISO 9660".

Compatibility Information --- description of other limitations or requirements for using the medium.

Fees -- the fees and terms for retrieving the data set.

Custom Order Process -- description of custom distribution services available, and the terms and conditions for obtaining these services.

Available Time Period -- the time period when the data set will be available from the distributor.

Example

Distributor
- Contact Person Primary
  - Contact Person: William Wiggins
  - Contact Organization: Federal Transit Administration
- Contact Organization Primary
  - Contact Position:
  - Contact Address
    - Address: 400 Seventh Street, S.W.
    - City: Washington
    - State or Province: DC
    - Postal Code: 20590
    - Country: USA
  - Contact Voice Telephone: (202) 366-0255
  - Contact Facsimile Telephone: (202) 366-3765
  - Contact Electronic Mail Address: wiggins@tts.dot.gov
  - Hours of Service: weekdays, 9 am - 5 pm
  - Contact Instructions: Call for further instructions.

Distribution Liability. Although these data have been processed successfully on a computer system at Bridgewater State College, no warranty expressed or implied is made by the FTA regarding the utility of the data on any other system, nor shall the act of distribution constitute any such warranty.

Standard Order Process
- Non-digital Form: available as map sheets for each Transit District.

- Digital Form.
  - Digital Transfer Information
  - Format Name: TGRLN
  - Format Version Date: 1992
  - Format Specification: TIGER/LINE
- Digital Form.
  - Digital Transfer Information
  - Format Name: RTS
  - Format Version Number: TransCAD Version 3
  - Format Version Date: 1995

Digital Transfer Option

Offline Option
- Offline Media: 3-1/2 inch floppy disk
- Recording Format: ASCII

Ordering Instructions
- Available Time Period
  - Calendar Date: 1996

Metadata Reference Information.

The Metadata Reference Information includes data on the currentness of the metadata information, and the responsible party.

Metadata Date -- the date that the metadata were created or last updated.

Metadata Review Date -- the date of the latest review of the metadata entry.

Metadata Future Review Date -- the date by which the metadata entry should be reviewed.

Metadata Contact -- the party responsible for the metadata information.

Metadata Standard Name -- the name of the metadata standard used to document the data set.

Metadata Standard Version -- identification of the version of the metadata standard used to document the data set.

Metadata Time Convention -- form used to convey time of day information in the metadata entry. Used if time of day information is included in the metadata for a data set.

Metadata Access Constraints -- restrictions and legal prerequisites for accessing the metadata. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the metadata.

Metadata Use Constraints -- restrictions and legal prerequisites for using the metadata after access is granted. These include any use constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on using the metadata.

Example
Metadata Date: 19951209

Metadata Contact
- Contact Person Primary
  - Contact Person: Paula Okunieff
  - Contact Organization: Viggen Corporation
- Contact Organization Primary
  - Contact Position: FTA NTG Standards Program Manager
  - Contact Address
    - Address: 21 Union Street, 2nd Floor
    - City: Boston
    - State or Province: MA
    - Postal Code: 02108
    - Country: USA
  - Contact Voice Telephone: (617) 742-5114
  - Contact Facsimile Telephone: (617) 742-5124
  - Contact Electronic Mail Address: okunieff@world.std.com
  - Hours of Service: weekdays, 9 am - 5 pm

Contact Instructions: Call for further instructions.

Metadata Standard Name: Content Standards for the National Transit GIS Metadata
Metadata Standard Version: 1995
Metadata Access Constraints: None.
Metadata Use Constraints: None.

Appendix E -- *National Transportation Atlas Data Dictionary and Database Formats* [BTS, September 1995]

September 1995

National Transportation Atlas
Data Dictionary and Database Formats

The Bureau of Transportation Statistics (BTS) has created a standard set of file formats for geospatial databases included in the National Transportation Atlas (NTA). These formats were developed to make it easier to view and extract NTA data by establishing a common structure for each of the three geospatial feature types depicted in the NTA: points, networks, and areas. These formats are also being provided to GIS software vendors so that they can develop translation software to import the databases directly into their own internal formats. The BTS will distribute all of its geospatial data using these formats as an interim standard until a formal Transportation Network Profile (TNP) is adopted as part of the Spatial Data Transfer Standard (SDTS).

The BTS geospatial file formats are based on a set of standard record types. Six distinct record types are currently defined: **Link, Node, Point, Area, Geography,** and **Attribute.** [1] Each of the three spatial feature types included in the NTA consists of an interrelated combination of these record types defining the geometry, topology, and attributes associated with a specific transportation or background feature. Feature types and their composite record types are described below:

Transportation Networks are composed of four related record types: Link, Node, Geography, and Attribute. Examples of transportation networks are highways, railroads, and waterways.

Transportation Point Facilities such as airports, water ports, and truck terminals require only two related record types: Point and Attribute.

Areas are made up of three related record types: Area, Geography, and Attribute. Features such as Congressional Districts, States, and National Parks are examples of areas.

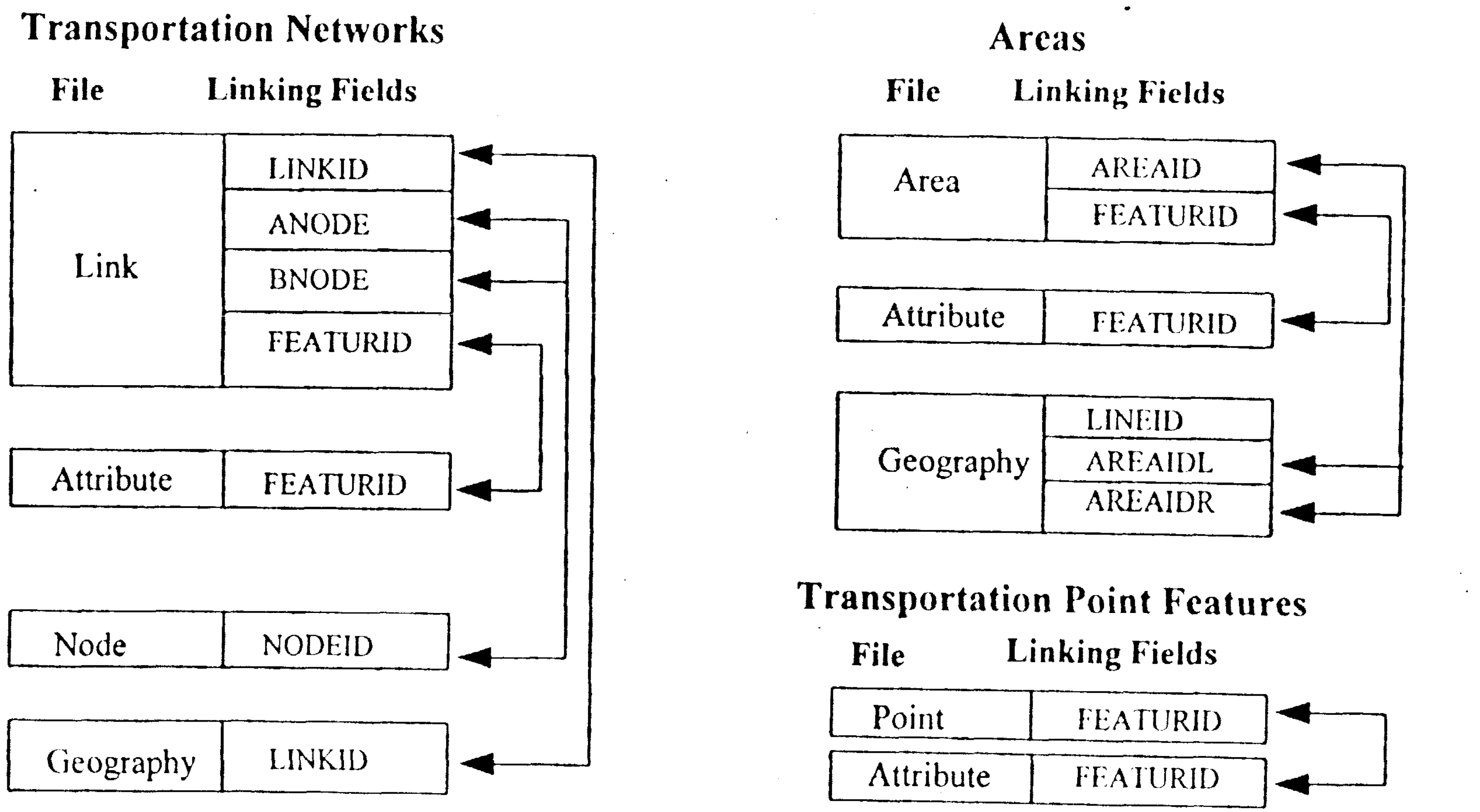
Figure 1 illustrates the relationships among the record types for each feature.

Each geospatial database distributed by the BTS will consist of a set of files sharing a common file name and file extensions identifying the corresponding record type. File names will be consistent with DOS file naming conventions (e.g., eight characters or less, no spaces, etc.). The file extensions for each record type are given below:

---

1 A seventh record type, **Linear Reference**, is still under development, but will eventually be added to this set.

File Linkage Relationships in the Databases Contained within the National Transportation Atlas



June 1995

Figure 1

.lnk	link record type
.nod	node record type
.pnt	point record type
.are	area record type
.geo	geography record type
.t__	attribute record type[2]

Each record type employs a standard ASCII character set and fixed length records with fixed length attribute fields. The first four fields are the same for every record type. They include the record type identifier, version number, revision number, and modification date as described below:

1 RECTYPE

The RECTYPE identifies the record type associated with the file, where:

L	Link
N	Node
P	Point
A	Area
G	Geography
T	Attribute

This field will have the same value for each record in the file.

2 VERSION

The version number is a 2-digit number that will be incremented for all records in the database whenever a new release is distributed.

The BTS plans to issue a new release of each NTA database on an annual basis, incorporating updates of attribute fields as well as all corrections and enhancements made to the geography and/or topology since the previous release.

Between each version release, the BTS will post interim updates on its Internet site, incorporating corrections identified by users and enhancement work in progress. Interim updates will include both a transaction database, containing only those records which have been updated since the last major release, and a transaction log file, identifying what specific changes were made to each updated record.

3 REVISION

The revison number is a 2-digit number that will be incremented individually for each record whenever a change is made to one of its fields. Revision numbers are also included in the transaction log file so

[2] A geospatial database may have more than one attribute or table file. The two blank characters allow the BTS or users to establish a sequential numbering scheme where multiple attribute files are linked to the same geospatial feature.

that database users can identify what specific change was made to the record. Revision numbers will not be reset with each new release

4 MODDATE — The modification date indicates when each record in the database was last changed. For release 0100 this field is blank. As changes are made to a record, the modification date is entered in the format 'mmddyyyy'.

Following these first four fields, each record type has its own characteristic set of fields. These fields are described below:

LINK FILE

The link file contains basic information related to transportation network links. Each record contains the four standard fields (defined above) plus seven additional fields.

1 RECTYPE
2 VERSION
3 REVISION
4 MODDATE

5 LINKID — The LINKID is a unique sequential number assigned to each record in the link file. It is used internally by most GIS software to index records and to establish topological relationships between spatial objects. It is not permanent and may change between different versions of the file.

6 FEATURID — The FEATURID is a unique character string or numeric value associated with the spatial feature. The metadata associated with each network database describes the significance of and method for assigning this identifier.

While not necessarily a permanent identifier, the FEATURID is more stable than the LINKID. Any changes made in the FEATURID will be recorded in a transaction file maintained by the BTS and distributed with each new version of the database.

7 ANODE — The ANODE identifies a record (NODEID) in the node file that corresponds to the starting position of the link.[3]

8 BNODE — The BNODE identifies a record (NODEID) in the node file that corresponds to the ending position of the link.

---

[3] Starting and ending positions for links are generally determined by the direction in which the link was digitized.

9 DESCRIPT — The DESCRIPT field can contain any character string describing the link, such as a river name, route id name, or number. These descriptions need not be unique for each record.

10 STFIPS1 — The STFIPS1 field contains the FIPS code for the State in which the link is located. If the link borders two States, STFIPS1 contains the lower numeric value FIPS code.

11 STFIPS2 — The STFIPS2 field contains the higher numeric value FIPS code for a second State in those cases where the link borders two States. If the link is located completely within a State, this field contains 00. In no case should a link border three or more States because a node should exist at the point where the States intersect.

NODE FILE

The node file contains basic information related to transportation network nodes. Each record contains the four standard fields (defined above) plus six additional fields.

1 RECTYPE
2 VERSION
3 REVISON
4 MODDATE

5 NODEID — The NODEID is a unique sequential number assigned to each record in the node file. It is used internally by most GIS software to index records and to establish topological relationships between spatial objects. It is not permanent and may change between different versions of the file.

6 FEATURID — The FEATURID is a unique character string or numeric value associated with the spatial feature. The metadata associated with each network database describes the significance of and method for assigning this identifier.

While not necessarily a permanent identifier, the FEATURID is more stable than the NODEID. Any changes made in the FEATURID will be recorded in a transaction file maintained by the BTS and distributed with each new version of the database.

7 LONGITUD — The longitude of the node expressed as a signed integer with six (6) implied decimal places.

8 LATITUDE — The latitude of the node expressed as a signed integer with six (6) implied decimal places.

9 DESCRIPT — The DESCRIPT field can contain any character string describing the node, such as a town name, landmark, or border crossing point. These descriptions need not be unique for each record.

10 STFIPS — This field contains the FIPS code for the State in which the node feature is located. If the node is located on the border of two or more States, this field contains 00. If the node is located outside the borders of the United States, the field contains 99.

POINT FILE

The point file contains basic information related to transportation point features and is similar to the network node file. Each record contains the four standard fields (defined above) plus six additional fields.

1 RECTYPE
2 VERSION
3 REVISION
4 MODDATE

5 POINTID — The POINTID is a unique sequential number assigned to each record in the point feature file. It is used internally by most GIS software to index records. It is not permanent and may change between different versions of the file.

6 FEATURID — The FEATURID is a unique character string or numeric value associated with the point feature. The metadata associated with each point database describes the significance of and method for assigning this identifier.

While not necessarily a permanent identifier, the FEATURID is more stable than the POINTID. Any changes made in the FEATURID will be recorded in a transaction file maintained by the BTS and distributed with each new version of the database.

7 LONGITUD — The longitude of the point feature expressed as a signed integer with six (6) implied decimal places.

8 LATITUDE — The latitude of the point feature expressed as a signed integer with six (6) implied decimal places.

9 DESCRIPT — The DESCRIPT field can contain any character string describing the point feature, such as a town name or airport. These descriptions need not be unique for each record.

10 STFIPS — This field contains the FIPS code for the State in which the point feature is located. If the point feature is located on the border of two or more States, this field contains 00. If the point feature is located outside the borders of the United States, the field contains 99.

AREA FILE

The area file contains basic information related to areal features. Each record contains the four standard fields (defined above) plus six additional fields.

1 RECTYPE
2 VERSION
3 REVISION
4 MODDATE

5 POLYID — The POLYID is a unique sequential number assigned to each record in the area file. It is used internally by most GIS software to index records and to establish topological relationships between spatial objects. It is not permanent and may change between different versions of the file.

6 FEATURID — The FEATURID is a unique character string or numeric value associated with the area feature. Where an area feature crosses a State boundary, separate records exist for the portions of the feature located within each State. For example, Yellowstone National Park exists as three records, one each for Wyoming, Montana, and Idaho. Each of those three records will have a unique POLYID but the same FEATURID. The metadata associated with each area database describes the significance of and method for assigning this identifier.

While not necessarily a permanent identifier, the FEATURID is more stable than the POLYID. Any changes made in the FEATURID will be recorded in a transaction file maintained by the BTS and distributed with each new version of the database.

7 CNTRLONG — The longitude of the area feature record centroid expressed as a signed integer with six (6) implied decimal places.[4]

8 CNTRLAT — The latitude of the area feature record centroid expressed as a signed integer with six (6) implied decimal places.

[4] In those cases where an area feature is composed of multiple polygons, the CNTRLONG and CNTRLAT fields pertain to the centroid of each polygon.

9 DESCRIPT — The DESCRIPT field can contain any character string describing the area feature, such as a national park or military installation name. These descriptions need not be unique for each record.

10 STFIPS — The STFIPS field contains the FIPS code for the State in which the area feature is located. For those features which cross State boundaries, this field contains the FIPS code for the State in which the individual polygon is located.

GEOGRAPHY FILE

The geography file contains the shape information for network links or area boundaries. The file structure includes a header record followed by a variable number of shape point records. The header record contains the four standard fields (defined above) plus four additional fields.

1 RECTYPE
2 VERSION
3 REVISION
4 MODDATE

5 LINEID (LINKID) — The LINEID is a unique sequential number assigned to each record in the geography file. It is used internally by most GIS software to index records and to establish topological relationships between spatial objects. In transportation networks the field name is LINKID, and matches the LINKID field in the Link File. It is not permanent and may change between different versions of the file.

6 POLYIDL — The polygon identification number (POLYID) from the area file for the polygon record located on the left side of the line. For transportation networks, this field is blank.

7 POLYIDR — The polygon identification number (POLYID) from the area file for the polygon record located on the right side of the line. For transportation networks, this field is blank.

8 NPOINT — The number of coordinate pairs (longitude and latitude) that define the shape of the linear feature and follow the header record. For transportation networks, the geography file includes the coordinates of both the start and end nodes of the link as well as all intermediate shape points. Therefore every linear feature will have at least two coordinate pairs.

ATTRIBUTE FILE

The attribute file(s) contains additional information about network links or nodes, or point or area features. There may be multiple attribute files associated with any geospatial feature database, and each attribute file may have a different number of attributes and different attribute formats. However, each specific attribute file is characterized by a fixed format and fixed record lengths.

Each record in an attribute file contains the four standard fields (defined above), one feature identification field, and a variable number of attribute fields.

1 RECTYPE
2 VERSION
3 REVISION
4 MODDATE

5	FEATURID	The FEATURID is a unique character string or numeric value associated with the geospatial feature, and matches the FEATURID field in the corresponding Link, Node, Point, or Area files. This field provides the relational link between the spatial object which is used to represent a feature and the various attributes associated with it.

The FEATURID field is followed by a variable number of attribute fields pertaining to the feature. Data descriptions and formats for each attribute field are defined in the metadata for the geospatial database.

In addition to fixed format, fixed record length ASCII files, the BTS plans to distribute attribute files in dBase format. Many existing GIS software package can read dBase files directly, and this will obviate the need for users of those packages to build import templates for each attribute file.

Spatial File Formats

TRANSPORTATION NETWORKS

Link File

Field Num	Field Name	Field Type	Field Length	Beg Pos	End Pos	Field Description
1	RECTYPE	char	1	1	1	Record type: always 'L'
2	VERSION	char	2	2	3	Version number
3	REVISION	char	2	4	5	Revision number
4	MODDATE	num	8	6	13	Modification date
5	LINKID	num	10	14	23	Link ID
6	FEATURID	char	10	24	33	Feature ID
7	ANODE	num	10	34	43	Node ID for beginning node
8	BNODE	num	10	44	53	Node ID for ending node
9	DESCRIPT	char	35	54	88	Name/Identification
10	STFIPS1	num	2	89	90	Primary State FIPS Code
11	STFIPS2	num	2	91	92	Secondary State FIPS Code

Node File

Field Num	Field Name	Field Type	Field Length	Beg Pos	End Pos	Field Description
1	RECTYPE	char	1	1	1	Record type: always 'N'
2	VERSION	char	2	2	3	Version number
3	REVISION	char	2	4	5	Revision number
4	MODDATE	num	8	6	13	Modification date
5	NODEID	num	10	14	23	Node ID
6	FEATURID	char	10	24	33	Feature ID
7	LONGITUD	num	10	34	43	Longitude (6 implied dec)
8	LATITUDE	num	10	44	53	Latitude (6 implied dec)
9	DESCRIPT	char	35	54	88	Name/identification
10	STFIPS	num	2	89	90	State FIPS code 00 if in multiple states 99 if not in the United States

Geography (shape point) File
Header record

Field Num	Field Name	Field Type	Field Length	Beg Pos	End Pos	Field Description
1	RECTYPE	char	1	1	1	Record type: always 'G'
2	VERSION	char	2	2	3	Version number
3	REVISION	char	2	4	5	Revision number
4	MODDATE	num	8	6	13	Modification date
5	LINKID	num	10	14	23	Link ID
6	blank	num	10	24	33	blank
7	blank	num	10	34	43	blank
8	NPOINT	num	3	44	46	Number of coordinate pairs

Shape records

The header record is followed by INT(NPOINT + 1 / 4) fixed length shape records. Each shape record contains up to 4 coordinate pairs and is formatted as numeric with 6 implied decimal places.

TRANSPORTATION POINT FACILITIES

Point File

Field Num	Field Name	Field Type	Field Length	Beg Pos	End Pos	Field Description
1	RECTYPE	char	1	1	1	Record type: always 'P'
2	VERSION	char	2	2	3	Version number
3	REVISION	char	2	4	5	Revision number
4	MODDATE	num	8	6	13	Modification date
5	POINTID	num	10	14	23	Point ID
6	FEATURID	char	10	24	33	Feature ID
7	LONGITUD	num	10	34	43	Longitude (6 implied dec)
8	LATITUDE	num	10	44	53	Latitude (6 implied dec)
9	DESCRIP	char	35	54	88	Name/identification
10	STFIPS	num	2	89	90	State FIPS code 00 if in multiple States 99 if not in the United States

AREA FILES

Area File

Field Num	Field Name	Field Type	Field Length	Beg Pos	End Pos	Field Description
1	RECTYPE	char	1	1	1	Record type: always 'A'
2	VERSION	char	2	2	3	Version number
3	REVISION	char	2	4	5	Revision number
4	MODDATE	num	8	6	13	Modification date
5	POLYID	num	10	14	23	Polygon ID
6	FEATURID	char	10	24	33	Feature ID
7	CNTRLONG	num	10	34	43	Longitude of feature centroid (6 implied dec)
8	CNTRLAT	num	10	44	53	Latitude of feature centroid (6 implied decimal)
9	DESCRIPT	char	35	54	88	Name/identification
10	STFIPS	num	2	89	90	State FIPS code

Geography (shape point) File
Header record

Field Num	Field Name	Field Type	Field Length	Beg Pos	End Pos	Field Description
1	RECTYPE	char	1	1	1	Record type: always 'G'
2	VERSION	char	2	2	3	Version number
3	REVISON	char	2	4	5	Revision number
4	MODDATE	num	8	6	13	Modification date
5	LINEID	num	10	14	23	Line ID
6	POLYIDL	num	10	24	33	Polygon ID on left
7	POLYIDR	num	10	34	43	Polygon ID on right
8	NPOINT	num	3	44	46	Number of coordinate pairs

Shape records

The header record is followed by INT(NPOINT + 1 / 4) fixed length shape records. Each shape record contains up to 4 coordinate pairs and is formatted as numeric with 6 implied decimal places.

Appendix F -- Guidelines for Use of a Route Object for Transit Features Types

Objectives

Much of the transit infrastructure is represented as route features, e.g., bus routes, subway lines, and people mover systems. Managing, disseminating and updating this information requires the development of a robust "route" data model to support the transfer and conversion of the route features included in the GIS. Yet, no common route model exists which supports both geographic representation and attribute requirements for transit feature types. This appendix presents some observations and recommendations for a common data structure or file format to store route information for transfer and conversion.

Development of a route data structure is essential since:

- Map databases are translated from one format to another, and
- Feature and attribute data are transferred between significantly different digital base maps.

Translation

The translation function, typified by standards such as the Spatial Data Transfer Standard, DIGEST, or specialized translators supported by vendor products (e.g., ArcInfo to MGE format), simply translates one set of symbols and file formats to another. The accuracy, completeness and lineage of the base map does not change, so feature type data is translated to a common format, and attribute files remain linked to the geospatial data through the feature codes.

Transfer

Transfer between dissimilar base maps must integrate feature data from one base map into another. Depending on the dissimilarity of the conversion, this transfer may involve functions such as "join," "geocoding," "conflation," or "feature automation." Within the next decade, if not sooner, the process of integrating or dynamically conflating data will be standard practice, quickly and inexpensively executed. Attribute data may require translation to accomodate to differences in the positional accuracy, link-node structure, or attribute resolution of the target database. For example, a link travel time on a link with an effective distance of 35 meters must be translated when transferred to the same link on another database with an effective distance of 44 meters.

The management, update and dissemination of the National Transit GIS requires that both translation and transfer be supported. Though different in application, both require a common data structure for storing and disseminating transit route feature types. The data structure must possess the following characteristics (in order of priority):

Flexibility: Data structures must be able to represent location a transit route in a variety of ways including descriptively, and with GIS tools which *do* and *do not* support dynamic segmentation..

Robust: Route features must be unambiguously and accurately referenced.

Efficiecy: A minimal set and length codes and conventions to reference features must be used.

The FTA presently collects most of its data from transit properties that do not use GIS tools or full-function GIS tools. Therefore, the data structure must have the flexibility to descriptively identify route features. Data collection efforts of this sort fall into the transfer between dissimilar databases category. Adding to the need for flexibility is the fact that many transit agencies who possess GIS tools use TIGER Line Files as their base map update and improve their TIGER base maps (which the FTA does not do). This will require that the FTA incoporate those changes into the NTG base map. Others purchase data sets from commercial sources or share base maps created by their Metropolitan Planning Organizations (MPO), State DOT or other regional/local transportation agency. This requires that route data submitted from those transit agencies be integrated with the NTG TIGER Line base map.

Moreover, no mandate or regulation requires transit agencies to contribute their data to the National Transit GIS, nor does FTA have funding to undertake a data collection effort similar to that which allowed the creation of the transit route database. This situation compels the route data structure to be simple and useful to both transit agencies, transit application vendors and the FTA.

Use of Route *Object Data Structure*

Major transit feature types possess route characteristics. The National Transit GIS is composed of directional route miles for bus routes and fixed guideway systems. Over 530 transit properties and 3,000 routes belonging to the Fixed Guideway Systems, Fixed Bus Route Systems, Major Urban Transit Providers, and Rural and Paratransit Grantees of the U.S. and its territories are defined. The route object data structure will be used to store and disseminate these routes. Also, transit agencies maintain other transit "route" features including the "block", "run", and "trip" for other applications and obligations. These may also be stored and transferred in this data structure.

Route Object Data Structure Description

The route data structure is defined as a string representing the Route Identifier (Route_ID) and an ordered set of location identifiers (Route_Path). The ordered set of location identifiers may take many forms. When a route database is transferred between users, a data dictionary or alternate module should describe the type of location identifier used. A single location identifier should be used for each database. The Route data structure and location identifiers are described below. The location identifiers are grouped according to transfer functionality (see Purpose section), and the description contains recommendations on the use and density, exceptions and example of each alternative.

ROUTE DEFINITION

A route is a contiguous, ordered set of 1-dimensional objects.

ROUTE DATA STRUCTURE

```
ROUTE   :   Record Of
        {     Route_ID      :     String,
              Route_Path    :     Ordered_Set of Location Identifiers
        }
```

! Location Identifiers

Point ID :	(x, y)	!longitude/easting, latitude/northing; in decimal form

! Alternative Location Identifiers:

Labeled Point ID:	(St Name, x, y)	! see *Address and Street Naming Conventions*
Intersection:	(On St Name, At St Name)	
Node ID:	(Node #)	! *e.g.*, node no., bus stop no., time point no.
Milepost:	(Section No., Offset)	! offset = Milepost + 00.nnn
Reference:	(On St. At St. Distance offset)	! or (direction based on conventions)
	(Node ID, Distance Offset)	! or
	(On St. At St. Offset Code)	! or offset code = [NS, FS, MB, @, OPP]
Address:	(Primary Address No., Predirectional, Street Name, Suffix, Postdirectional)	

The elements of the data structure may be stored in ASCII or binary format, each row on a separate line, and each element of the record separated by a comma, tab or space. The record should be framed by less than (<) and greater than (>) signs. For example,

---

Record of Bus Route 101

<101, long1, lat1, long2, lat2, long3, lat3, long4, lat4, long5, lat5, long6, lat6, long7, lat7, long8, lat8, long9, lat9, long10, lat10, long11, lat11, long12, lat12, long12, lat12, long14, lat14, long15, lat15, long16, lat16>

Methods Described

Translation of Route Features between Base Map Formats

Point ID

Geographic coordinate quality is preserved, without loss of resolution or increased ambiguity when translated between computer systems or file formats. A data set which uses the same version of TIGER Line Data as the National Transit GIS possesses the same geographic coordinates. Thus, a string of geographic coordinates will unambiguously describe the segments of the route. Furthermore, a string of geographic coordinates defines arcs and chains, so that "sections" and mileposts can be referenced to ground-truth using the ordered set of point identifiers; this record type is easily incorporated into a feature based model of a route object.

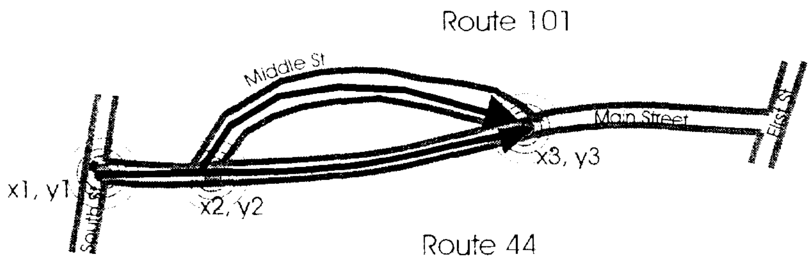
Density of Geographic Coordinates in Ordered Set

The string of Position IDs should ensure that all the arcs, chains and/or lines in the route are identified. The start and end points of the route feature should be included in the string.

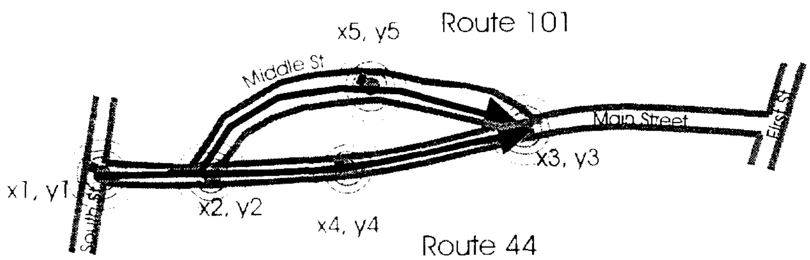
Limitations

Split roads may require additional references (i.e., shape points) of a chain. Figure 3 illustrates a split route for which link or chain nodes will not be sufficient to unambiguously identify the route path. In this case, the two entries <101, x1, y1, x2, y2, x3, y3> and <44, x1, y1, x2, y2, x3, y3> are equivalent. An additional point from the chain representing Middle Street and Main Street should be added to each data structure representing routes 101 and 44, respectively.

Figure 3 Split Route



Limited number of Point IDs may cause ambiguity in route description



Increased density of Point IDs clearly identify route path.

Example

Many GIS vendors store and export line features similar to this representation. An instance of a transit route would appear as in Figure 4 and Table 1.

Figure 4 Transit Route

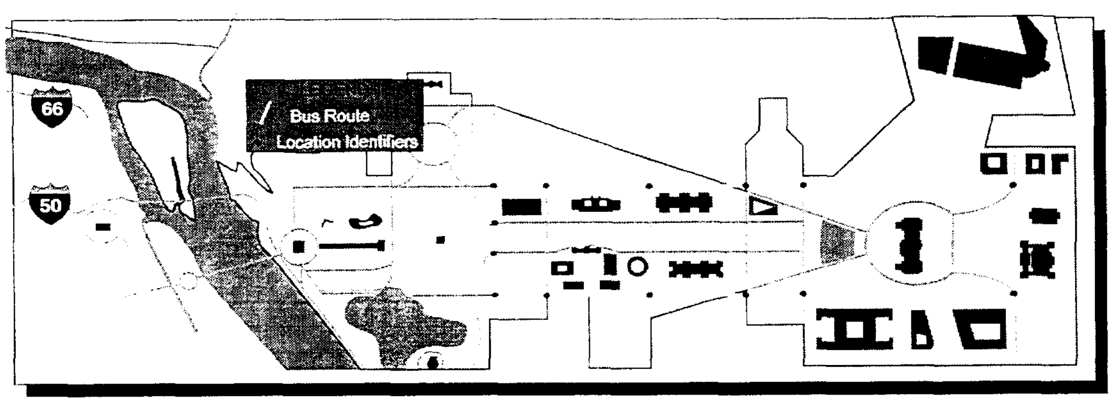


Table 1 Instance of Transit Route Record

<101,
long1, lat1,
long2, lat2,
long3, lat3,
long4, lat4,
long5, lat5,
long6, lat6,
long7, lat7,
long8, lat8,
long9, lat9,
long10, lat10,
long11, lat11,
long12, lat12,
long12, lat12,
long14, lat14,
long15, lat15
>

Transfer of Route Objects Between Dissimilar Data Sets

Differences in positional accuracy and attribute completeness between data sets may be significant. Moreover, GIS tool functionality may not allow derivation of the certain location references described above. The ability to choose the linear reference best suited for both data sets mitigates any ambiguity and error in the placement of the feature. For example, organizations with data sets whose positional accuracies are closely matched may prefer the Labeled Point ID; or organizations with data sets populated with complete road features, addresses and labels may use the Intersection or Addressing method.

Only one method should be used in each data file. Header information in the file and entries in the metadata document should describe the extension or format used

<u>Labeled Point ID</u>

The labeled point ID method is similar to the Point ID method in that it relies primarily on the spatial coordinate for identifying the route. The label or name is used to resolve any ambiguity. For example, the road name (*i.e.*, Middle or Main Street) resolves the ambiguity between Route 101 and Route 44 in Figure 3. The two entries would appear as <101, Main, x1, y1, Middle, x2, y2, Main, x3, y3> and <44, Main, x1, y1, Main, x2, y2, Main, x3, y3>. The full street name with all its component parts (*e.g.*, pre and post directionals, and type) need not be specified because the coordinate pair typically resolves the ambiguity[13]. This method is useful when the feature data is transferred to a map database of similar resolution and accuracy as the source data set. For example, two digital databases derived from the same source, such as commercial products

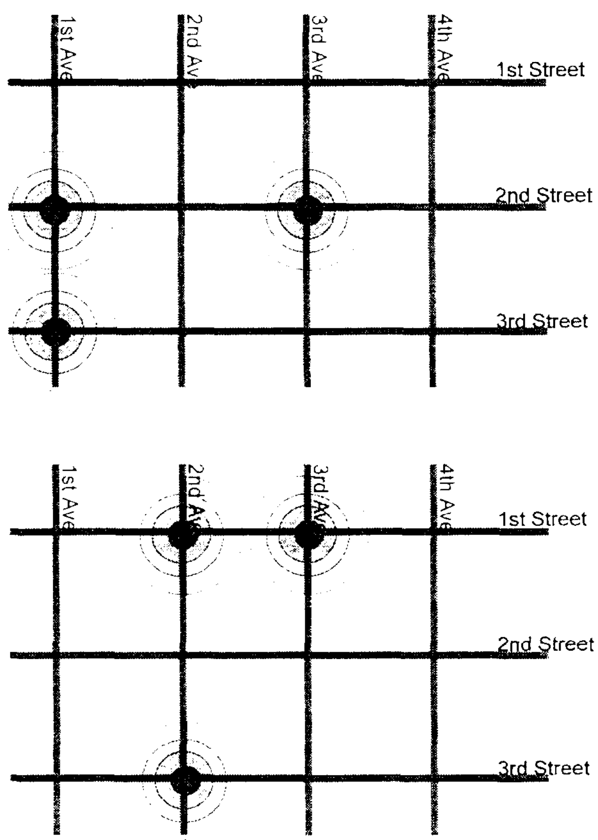
[13] Except in the case of the intersection between 1st Street and 1st Avenue, or streets with similar names within proximity of the database accuracy.

derived from Digital Line Graph (DLG) Data and completely attributed with common street names should be closely matched.

Density of Geographic Coordinates in Ordered Set
The string of Labeled Position IDs should ensure that the pattern of arcs, chains and/or lines in the route are identified. An arc, chain or line need be not be explicitly identified if it is contained within two endpoints of a chain or contiguous chains associated with the same label. The start and end points of the route feature should be included in the string. For example, Route 39 buses travel about 3 miles on Huntington Avenue between South Huntington Ave. (x9, y9) and Dartmouth Street (x10, y10). This could be represented as <39, ..., Huntington, x9, y9, Dartmouth, x10, y10, ...>.

Limitations
An exception as described in Footnote 2 where two streets with similar names in close proximity, may produce an ambiguous reference. The ambiguity may not appear if the direction of travel is not ambiguous. For example, the route data structure assumes directionality due to the ordered nature of the location identifiers. The route is also assumed to be contiguous. If First Street is assumed instead of First Avenue, then the next segment in the record will typically parallel First Street and not intersect it as illustrated in Figure 5.

Figure 5 Labeled Point ID with Similarly Named Streets



Examples

A Labeled Point ID record format would appear as follows:

Labeled Point ID Record Format

<101,
14th, long1, lat1,
14th, long2, lat2,
Independence, long3, lat3,
Independence, long4, lat4,
Independence, long5, lat5,
Independence, long6, lat6,
Independence, long7, lat7,
1st, long8, lat8,
Constitution, long9, lat9,
Constitution, long10, lat10,
Constitution, long11, lat11,
Constitution, long12, lat12,
Constitution, long13, lat13,
14th, long14, lat14,
14th, long15, lat15
>

Condensed Form

<101,
14th, long1, lat1,
14th, long2, lat2,
Independence, long7, lat7,
1st, long8, lat8,
Constitution, long13, lat13,
14th, long15, lat15
>

Intersection

The Intersection location identifier is principally for identifying the route descriptively. This technique is used when the base maps for which the features are exchanged maintain complete and standard addresses and street names. Advanced algorithms for address matching and geocoding improve matching rates, though this technique is almost never fully automated. The intersection referencing method is useful for descriptively identifying the path of the route, particularly, for agencies that do not possess GIS tools.

A comma or tab should separate each item in the record, and a space is used to delimit the constituent parts of the street name. Street naming standards used by the U.S. Post Office should be employed as the conventions used to identify the On and At Streets.

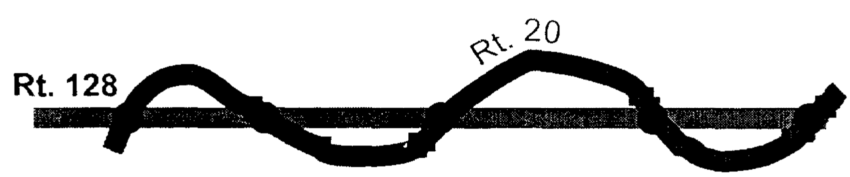
Density of Geographic Coordinates in Ordered Set

The string of Intersections should ensure that the pattern of arcs, chains and/or lines in the route are identified. The start and end points of the route feature should be included in the string. An arc, chain or line need be not be explicitly identified if the contiguous chains are associated with the same On Street between two intersections. For example, Route 39 buses travel about 3 miles on Huntington Avenue between South Huntington Ave. and Dartmouth Street. This could be represented as <39, ..., S. Huntington St, Huntington St, Huntington St, Dartmouth St, ...>.

Limitations

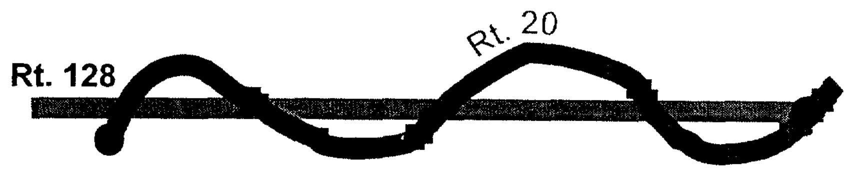
A major problem may occur with a route that weaves through roads that intersect in more than one place; multiple instances of the intersection exist. The path of the route may be ambiguous, as illustrated in Figure 6. In this case, the path of the route will be clarified by including additional intersections.

Figure 6 Weaving Roads

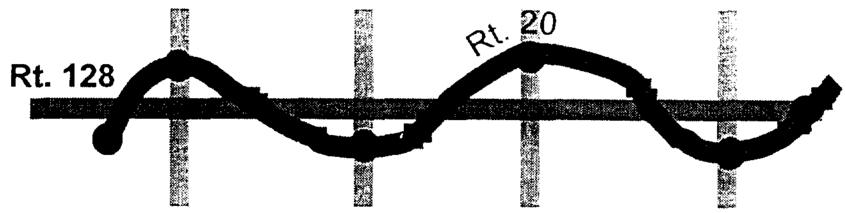


Other problems will occur when a route record is transferred from a richly attributed source to a sparsely populated source. With missing cross streets the sparsely attributed source will not be capable of defining the route structure (see Figure 7).

Figure 7 Transfer of Intersections Between Differently Attributed Data Sources



Sparsely populated road network



Richly populated road network

This method is limited to locations which are designated by intersections. A precise position cannot be specified for routes that start, end or change at a location other than an intersection.

Examples
A common practice used by many transit agencies to identify their time points and route descriptions for operator training, the Intersection Method provides a means of building route descriptions without the use of a GIS, survey or positioning/navigation device.

Record Structure
<101,
14th St. Jefferson Dr.
14th St, Independence Av,
Independence Av, 12th St,
Independence Av, 9th St,
Independence Av, 7th St.
Independence Av, 4th St,
Independence Av, 3rd St,
Independence Av, 1th St,
1st St, Constitution Av,
Constitution Av, 3rd St,
Constitution Av, 4th St,
Constitution Av, 7th St,
Constitution Av, 9th St,
Constitution Av, 12th St.
Constitution Av, 14th St.
14th St, Madison Dr,
14th St, Jefferson Dr
>

Condensed Form

<101,
14th St, Madison Dr,
14th St, Independence Av,
Independence Av, 1st St,
1st St, Constitution Av,
Constitution Av, 14th St,
14th St, Madison Dr
>

Node ID

In most cases, node identifiers differ between different data sources. However, in many transit agencies, bus stops, time points and other node types are assigned standard codes. These codes are inserted into a number of digital products. This enables the translation and conversion of route feature types between different systems and data sources.

Density of Geographic Coordinates in Ordered Set

Node IDs are subject to the same limitations as the Position IDs. The string of Nodes IDs should ensure that all the arcs, chains and/or lines in the route are identified. The start and end points of the route feature should be included in the string.

Limitations

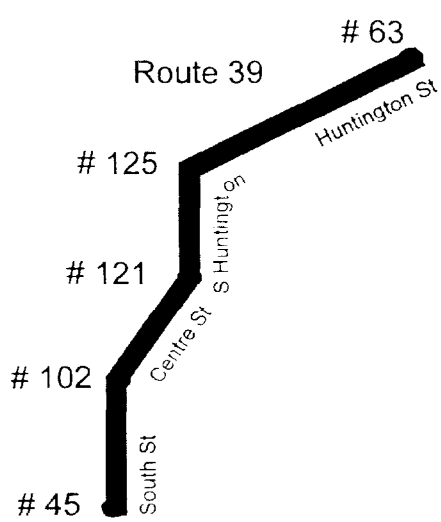
Most data sources do not share node identifier codes. TIGER Line Files do not explicitly index and maintain nodes and node identifiers.

Examples

Perhaps, the most concise of the location identifier formats, the record appears as a route id and series of node ids.

Route Instance using Node IDs: <39, 45, 102, 103, 104, 105, 107, 121, 122, 123 124, 125, 126, 65, 64, 63>

Figure 8 Use of Node Identifiers



Node	Name
45	Forest Hills Station
102	Monument
121	Hyde Sq
125	Brigham Cir
63	Copley Sq

Milepost

Typically, state transportation agencies use mileposts to reference roadway characteristics and equipment. Milepost markers and section references are described using the Position ID method: the Milepost method is used to identify attributes along a mileposted road. For example, pavement type and accident site are referenced to the milepost. The resolution of the milepost reference depends on the accuracy of the milepost location. Also, mileposts, though identified as marking units in "miles", may actually be kilometers or short/long miles. The physical marker is merely a reference to a physical location.

Density of Geographic Coordinates in Ordered Set

The string of Mileposts should ensure that the pattern of arcs, chains and/or lines in the route are identified. The start and end points of the route feature should be included in the string. An arc, chain or line need be not be explicitly identified if the contiguous chains are associated with the same Section Number, Road Name or Road Label. This method follows the same rules as the intersections and labeled position IDs.

Limitations

This method requires a well maintained national index of road names, and establishment of mileposts along urban roadways. This practice has limited application for metropolitan areas. Additional information should be included in the metadata or file header detailing the region or jurisdiction of the road name since highway names and mileposts may be duplicated in adjacent regions.

Examples

Route 39 buses travel about 3 miles on Huntington Avenue between South Huntington Ave. and Dartmouth Street. If these streets were indexed, such as in Table 1, this could be represented as:

```
<39,
....
432, 2.885,
431, 35.559,
57, 1.009,
...>
```

Table 1 - Section/Street Name Lookup

Section No/ Index	Street Name
57	Dartmouth St
431	Huntington Av
432	S Huntington Av

Reference

The reference point method is similar to the milepost in that it references a position from a known point. In the milepost method, the point is a physical marker; in the reference point method, the position is based on attributes of the road network, e.g., intersection (on and at street names) or node identifier. The offset element is similar to the milepost offset, though, the same street name. This method follows the same rules as the intersections, and labeled position IDs.

Limitations
This method works well in urban areas where the addresses are densely and sequentially organized. However, in rural areas with few addresses and few rural route numbers, this method propagates large distance errors between address locations. Also, because highways and interstates are typically not attributed with addresses, the use of addresses is inappropriate.

Additional information should be included in the metadata or file header detailing the region or jurisdiction of the address since street names recur in adjacent regions.

Examples

Record Structure
<39
1 South St
779 Centre St
461 Centre St
351 S Huntington Av
551 Huntington Av
10 Huntington Av>

Conclusion

These methods for encoding a route may not be efficient for all applications or data base management systems; however, the structures enable the transfer of "route" features between different platforms, GIS application tools, information systems applications and databases. Attribute features associated with the route features should continue to be transferred using a route feature code.

The route data structure used for the National Transit GIS should provide a robust, flexible and efficient definition. Not all methods or coding structures fill these characteristics in every situation. This paper identified methods which best fit alternative environments. In summary, Table 2 identifies the route location identifiers which support different alternative environments.

Table 2 Summary of Recommended Uses for Route Location Identifiers

Alternatives	Pt ID	LPID	Inters	Node	MPost	Ref/off	Ref/code	Ref/node	Add
1	✓								
2			✓			✓	✓		✓
3		✓							
4					✓	✓			
5		✓	✓				✓		
6			✓				✓		✓
7					✓	✓	✓	✓	✓

where *Pt ID = Point ID*
LPID = Labelled Point ID