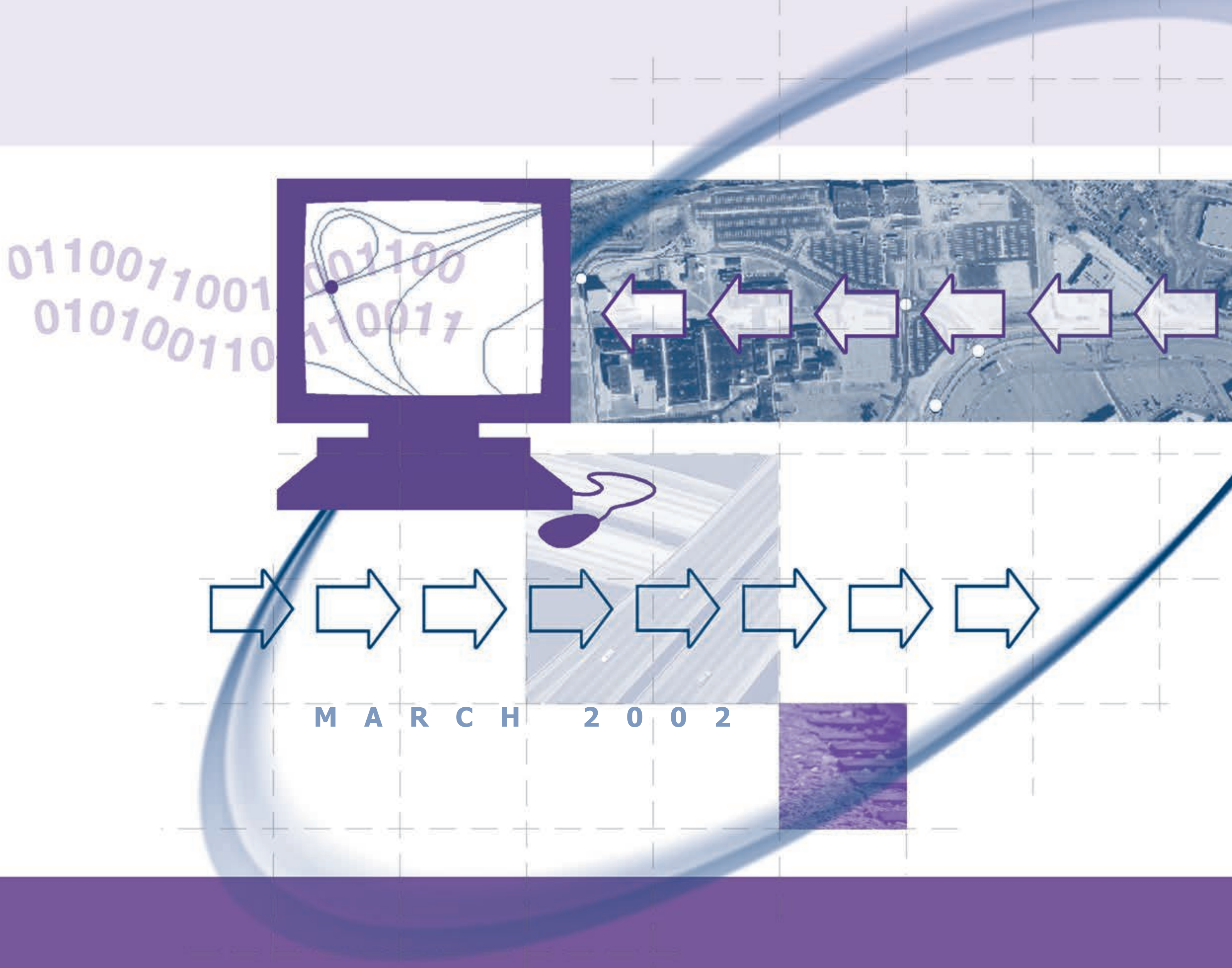
VOLUME III - TECHNICAL RECOMMENDATIONS

# Region-wide Transportation GIS Project Design and File Architecture



MARCH 2002

*Prepared For*
*The Delaware Valley Regional Planning Commission By:*

*Johnson, Mirmiran & Thompson/*
*Enterprise Information Solutions*
*A Joint Venture*

*In Association With: TransDecisions, Inc.*

DELAWARE VALLEY REGIONAL PLANNING COMMISSION

Created in 1965, the Delaware Valley Regional Planning Commission (DVRPC) is an interstate, intercounty and intercity agency that provides continuing, comprehensive and coordinated planning to shape a vision for the future growth of the Delaware Valley region. The region includes Bucks, Chester, Delaware, and Montgomery counties, as well as the City of Philadelphia, in Pennsylvania; and Burlington, Camden, Gloucester and Mercer counties in New Jersey. DVRPC provides technical assistance and services; conducts high priority studies that respond to the requests and demands of member state and local governments; fosters cooperation among various constituents to forge a consensus on diverse regional issues; determines and meets the needs of the private sector; and practices public outreach efforts to promote two-way communication and public awareness of regional issues and the Commission.

Our logo is adapted from the official DVRPC seal, and is designed as a stylized image of the Delaware Valley. The outer ring symbolizes the region as a whole, while the diagonal bar signifies the Delaware River. The two adjoining crescents represent the Commonwealth of Pennsylvania and the State of New Jersey.

DVRPC is funded by a variety of funding sources including federal grants from the U.S. Department of Transportation's Federal Highway Administration (FHWA) and Federal Transit Administration (FTA), the Pennsylvania and New Jersey departments of transportation, as well as by DVRPC's state and local member governments. The preparation of this document was funded by grants from the Federal Highway Administration and the Pennsylvania Department of Transportation. The authors, however, are solely responsible for its findings and conclusions, which may not represent the official views or policies of the funding agencies.

# Table of Contents

# List of Figures

# List of Tables

# Preface

## Project Purpose

As stated in the initial volume of this study, the primary purpose of this project is to "assure that each of these entities (DVRPC, its member city and county governments, and transportation operating agencies) has a GIS and data files that can be developed and seamlessly shared with each other to facilitate better transportation planning analysis and decision-making among the counties, the regions, and the states." It is the anticipation of DVRPC that accomplishing this purpose will require efforts on the part of all participants to supplement whatever data systems may be available from state, regional or federal sources. The purpose of the work to be performed for this project is to lay the foundation and establish the basic systemic and operational framework for these efforts.

Lying at the heart of this foundation and framework are a number of technical requirements and standards that must be established and observed in order for DVRPC and its member agencies to move forward in this endeavor. As presented throughout the previous volumes, these standards fall within the following major categories:

- Street address data
- Symbology
- Linework
- Database design
- Metadata
- Data dictionary

The purpose of this volume is to provide a series of specific recommendations for standards in each of these categories. The technical bases for these recommendations are the results of the needs assessment and research and prototyping efforts as detailed in Volumes I and II.

## Project Goals

Before discussing the specific recommendations, it is helpful to revisit the four major project goals and present the need for technical recommendations and standards within the context of these goals.

*Goal No.1: "Expand the use of GIS among all transportation planning partners and assist all members to improve their capacity as needed to reach a common operational level."*

Expanded usage of GIS for transportation planning among all of DVRPC's members and planning partners will not become a reality without effective and well thought-out plans for implementation. The implementation plans must address a number of issues at a variety of levels, including organizational and technical. The bulk of the organizational issues are addressed through the individual implementation plans presented in Volume IV. And, while these individual plans will also address technical issues, it is critical that these issues be addressed in a broader way through the recommendations the follow in this volume. On an organizational level, the recommendations for each organization are typically somewhat unique. Conversely, the technical aspects of the expanded use of GIS among the member organizations are less unique to the members themselves and are best addressed in a more uniform, comprehensive manner. Hence, the need for this volume.

*Goal No. 2: "Evaluate the transportation GIS files developed and maintained by federal and state agencies, DVRPC member governments and transit operators to determine how they can be used in an accurate and regionally consistent manner."*

The needs assessment process addressed this project goal. Through that process, evaluations were made and their results presented regarding the current status of GIS databases being developed and maintained by the member organizations. The results reflect a broad spectrum, extending from those organizations with virtually no GIS capabilities, through those with sophisticated, highly operational and productive systems. The likelihood of using the spatial and attribute databases from these systems within a regional system can only be measured through careful consideration of numerous technical and organizational issues. It is these technical issues that this volume of the study is intended to address.

*Goal No.3: "Provide for the seamless exchange of GIS data files and the integration of planning infrastructure among all member governments and operating agencies."*

This project goal can only be met through the acceptance and adoption of technical standards for database design and data communications. Prior information has indicated that the likelihood of developing and maintaining one regional centerline file is nonexistent. Therefore, the design of the attribute database is the critical element in this process. Designing a database structure that will facilitate the exchange and mapping of attribute and event data independent of the underlying graphic representation of the centerline is paramount. Furthermore, the development and implementation of metadata standards and a data dictionary structure that clearly and concisely conveys information sufficient to support complete understanding of the data is equally critical. Both the overall design of the database and standards for metadata and a data dictionary are addressed in Appendix A. Other factors that contribute to the achievement of this goal include street address data formats and the conveyance of information cartographically through the use of standard map symbology. These are addressed in Chapter II and Appendix A, respectively.

*Goal No. 4: "Structure the region-wide GIS design so that it can be expanded and enhanced by individual partners, while maintaining its consistency and exchangeability."*

This goal is inherent to any database design process. Designing a database that is extensible while maintaining consistency and normalization is certainly a goal for this project. This applies not only to the attribute databases, but to the linework data, as well. This is especially true for those organizations that have little or no GIS capabilities and data at this time. These criteria have been incorporated into the technical standards for database design that are spelled out in Chapters I and II.

## Common Linear Referencing System (LRS) Approach

Based on the detailed assessment performed throughout the DVRPC region as described in Volume I, it has been concluded that the Common Linear Referencing System (LRS) Approach makes the most practical sense at this point in time. Theoretical approaches have been reviewed that are technically sound, but lead to high costs for implementation as well as requiring a high degree of centralized data management. PennDOT and NJDOT have vast amounts of data tied directly to their respective LRS with associated business processes for collecting, storing, utilizing and maintaining this data. The quandary faced is that the majority of the local agencies in the region have no LRS, but do or will have very accurate centerlines that cover their entire road network, while the DOT's have LRS but do not cover the entire road network. The long-term solution needs to address the sharing of Transportation Planning data from both a "Top-Down / DOT to Local" as well as a "Bottom-Up / Local to DOT" way. This means that there needs to be a way of getting this event data *down* to the counties' centerlines and the more accurate centerline data and local event data back *up* to the DOT's. The DVRPC region needs a common cost-effective approach that allows them to share data between agencies, while maintaining current business processes.

The Common LRS Approach is a way of establishing a mechanism that will allow entities to share event data easily back and forth regardless of centerline accuracies or GIS packages being used. The commonality is that each LRS, at its core, references attributes against a linear element that has an agreed upon unique route identifier and an agreed upon set of distance measures.

The difference between New Jersey and Pennsylvania lies in the specific linear referencing method each DOT has chosen to deploy. NJDOT uses a Route/Milepoint method while PennDOT utilizes a County, State Route and Cumulative Offset method. By establishing these linear elements to each other, referenced attributes can be correlated. The Common LRS approach offers a de-centralized way to share the linear referenced data across agencies. And since the Common LRS Approach supports data, as it currently exists, it minimizes data and business process conversion and thus reduces the total cost of ownership.

## Street Address Geocoding

Within the realm of transportation planning at the local government level, there are a number of types of data that are collected that may have a street address as their primary locational reference. In order to use this data effectively, there must be a means available to reference this data to the centerline data. The process that facilitates this is called geocoding. Geocoding is the process of transforming a standard street address to a measured location along a street centerline segment. Most GIS software packages include tools for performing geocoding operations.

## Data Sharing is the Goal

The goal of the following sections is to outline the recommended approach for modifying/developing centerline datasets to allow region-wide data sharing across agencies. Initial data sharing will only occur for roads that are maintained in the state DOT GIS, due to the lack of a unique identifier for the other, local roads. Nevertheless, the architecture will be in place for complete data sharing. The unique identifier is the key to sharing data throughout the region and thus is the only limiting factor; if the road has a unique identifier it can and will be shared. It is recommended that the LRS of the State DOTs act as the standard that each local agency must meet. Therefore, each of the following sections represents an overview of how the DOT conducts business as well as a recommendation for the local agencies to follow.

# Chapter I - Linework Standards

## I-1 Current Status of Linework

In order to offer recommendations regarding future linework development and future linework maintenance or manipulation we must first revisit the assessment of the current linework. Currently there is no single regional methodology or approach used when developing a new centerline, nor is there a regional set of standards for uniquely identifying the segments or referencing events along that centerline. Generally speaking, these two deficiencies represent the most significant deficiencies in the region to establishing a methodology that will enable DVRPC to meet the goal of this project, which is to be able to share GIS data for Transportation Planning. As mentioned in previous reports, the DVRPC service area currently has considerable variance among the member agencies, with regards to their geography files that are used for GIS-T applications. This situation does not foster significant levels of regional linework development or regional GIS-T data sharing. For reference purposes, a brief overview of the current linework deficiencies is listed below.

### I-1.2 Extent, Roadway Systems & Scale

The map scales and extents of the centerline datasets that encompass the DVRPC region are adequate for Transportation Planning for each of the agencies that have current centerline files. The only major agency that has centerlines that work at the scale required by them but does not really fit into a scale appropriate for a more regional perspective is PennDOT's centerlines that are 1:24,000 scale. The key area of deficiency in this breakout is in the level of coverage of the road network. Once again each organization has a specific job to do with their road centerlines but from a regional perspective both PennDOT's and NJDOT's "GIS/LRS enabled" road networks only cover a small portion of the overall lane miles within the DVRPC service area. While PennDOT stops at the state maintained highway network, NJDOT has extended their GIS/LRS enabled road network further down to the county roads there still exists a significant number of roads outside of their GIS/LRS network. This gap in GIS/LRS road coverage poses a number of issues when planning for county and municipal transportation scenarios or extending a bottom up data transfer scenario.

### I-1.3 Centerline Representations

Most agencies throughout the DVRPC region represent the physical world logically, storing centerlines that are significantly separated by a median as two distinct records, and storing centerlines that are not separated as one record. Difficulties are presented with regional interoperability for unique road scenarios, Roosevelt Blvd in Philadelphia or road interchanges and ramps. Since there are no defined rules yet for dealing with these issues, interoperability issues may arise. Below is an outline of the current best practices for centerline representation as defined by the Federal Geographic Data Committee (FGDC).

#### I-1.3.1 Best Practices for Centerline Representation

In developing recommendations for designing, creating and maintaining the topological structure of the centerline data, the research effort has focused heavily on the ***National Spatial Data Infrastructure (NSDI) Framework Transportation Identification Standard (Public Review Draft)*** (FGDC-STD-999.1-2000) prepared by the Ground Transportation Subcommittee of the Federal Geographic Data Committee (FGDC) and distributed in December, 2000. As stated in this document, "(t)he importance of geo-spatial data depicting transportation features – especially road networks – extends well beyond their cartographic value. Road networks provide the basis for several indirect location referencing systems, including street addresses and various linear referencing methods commonly used to locate features like bridges, signs, pavement conditions, and traffic incidents. Geo-spatial transportation segments can be connected to form topological networks, which can be used to more accurately measure over-the-road travel distances between geographic locations." For each DVRPC member organization, those departments or individuals with authority to manage centerline data need to coordinate the development of a road database with all appropriate stakeholders, particularly with respect to the manner in which the various types of roads are modeled in the local database. The following include recommendations from the NSDI Standard for representing various types of roads in a transportation GIS database.

### I-1.3.2 NSDI Terminology

In NSDI parlance, a Framework Transportation Segment Reference Point (node) is defined as "the specified location of a (required) endpoint of a Framework Transportation Segment (road segment), or an (optional) reference point offset along the length of the road segment, on a physical transportation system." A road segment is defined as "(a) specified directed path between two Framework Transportation Segment Reference Points (nodes) along a physical transportation system that identifies a unique segment of that system.

### I-1.3.3 Road Types

According to the NSDI Standard, "the decision to represent a particular road by a single logical road segment or two or more parallel physical road segment should be based on scale, accuracy, cartographic and network application requirements. In general, network applications are facilitated where road segment and nodes can be directly replaced by network links and nodes."

### I-1.3.4 Roads with no Access Restrictions or Medians

The NSDI Standard further specifies that "one-way and two-way roads with no significant access restrictions or physical median separating directional roadways should be represented by a single road segment. Most local streets, connectors, and minor arterials fall into this category."

### I-1.3.5 Roads with Center Medians but no Access Restrictions

Again, according to the NSDI Standard, "some major urban and rural arterials have a center median which divides the travel lanes in each direction (e.g., Commonwealth Avenue in Boston). However, intersecting streets can access either direction of travel lanes via short transportation segments crossing the median at each intersection. These roads may be represented either by a single road segment which ignores the center median, or by two parallel road segment depicting directional roadways on either side of the median."

### I-1.3.6 Limited-Access Divided Highways

For these types of highways, the NSDI standard states, "most Interstate Highways and major, high speed expressways can only be entered or exited via specifically designated ramps. These roads almost always have some median strip or other physical barrier that prohibits vehicles from reversing direction without first exiting the highway at a designated ramp. These roads should always be represented by two road segment regardless of the actual physical separation between the lanes (e.g., even roads that are separated by a concrete "Jersey Barrier" should be represented by two road segment if each direction is served by its own entrance and exit ramps."

### I-1.3.7 Physically Separated, Limited-Access Parallel Lanes

From the NSDI Standard, "some high volume roads, particularly in urban areas, may designate certain lanes for high occupancy vehicles (HOV) or auto-only, and physically separate these lanes from the main travel lanes (e.g., I-395 in northern Virginia, or the New Jersey Turnpike outside New York City). If these physically separated lanes are served by their own entrance and exit ramps, they should be represented by their own Road segment. Furthermore, if the priority lanes are also separated directionally, each direction should be represented by its own road segment. Example: The northern end of the New Jersey Turnpike includes physically separated auto-only lanes, running parallel to the main traffic lanes in both directions. Both the main lanes and the auto-only lanes have their own entrance and exit ramps. This facility should be represented by four parallel road segment – one for each direction of the main lanes and one for each direction of the auto-only lanes."

### I-1.3.8 Entrance and Exit Ramps

Again citing the NSDI Standard, "entrance and exit ramps are one-way or two-way roads that provide general vehicle access to limited-access highways. Each entrance or exit ram should be represented by a road segment."

### I-1.3-9 Frontage Roads

The NSDI Standard further states that "a frontage or access road is a one- or two way, unlimited-access street that parallels but is physically separated from a more limited-access major arterial. Its main purpose is to provide access to establishments along the major arterial corridor while preventing access traffic from disrupting the flow of through traffic on the major arterial. Access from the frontage road to the major arterial is typically limited to intersections of cross-streets and/or specifically designated "gaps" in the median or physical barrier. Frontage roads should be represented by their own Road segment. Entrance "gaps" between the frontage road and the main arterial should be treated similar to an entrance or exit ramp."

The foregoing information has been presented as general recommendations for modeling various types of roads within a transportation GIS database. For further information, the actual NSDI Standard document can be found at:

http://www.fgdc.gov/standards/documents/standards/fr_trans_id/NSDI-Trans-Public_Review.pdf

### I-1.4 Map Resolution & Accuracy

The accuracy level throughout the DVRPC region is considered adequate for Transportation Planning as most of the road centerlines that exist are collected at an accuracy level that is more accurate than the collection of the GIS Transportation event data, which is to be mapped to the centerline. Although, local use of a PennDOT centerline may produce some inconsistencies when overlaying that layer with other more accurate layers in a local entities GIS.

### I-1.5 Maintenance

A majority of the agencies in the DVRPC region do not have a maintenance plan in place. Without a maintenance plan, the centerline data is obsolete almost immediately. The issues surrounding a strong regional maintenance plan for the sharing of road centerline changes or additions is a challenging task.

### I-1.6 LRS Methods

For the most part, there are no true Linear Referencing Systems being used by the member organizations of DVRPC with the exception of the DOT's and the Pennsylvania Turnpike Commission. Some form or type of LRS is critical to the successful use of GIS for Transportation Planning.

### I-1.7 Utilization

Currently, most agencies are not using their centerline for transportation planning. These agencies represent a deficiency in the regional model and are also less likely to maintain or produce a centerline that is usable for regional transportation planning.

### I-1.8 Route Network Topology

As a basis for implementing the Common LRS Approach – Route Network Topology is required. Route Network Topology is a means for allowing the GIS software to be able to build a route even if the route is composed of numerous graphic elements. The GIS software needs to know not only which graphic elements comprise a unique route but the exact order in which these graphic elements are joined to "build" or comprise the route. The GIS software handles this function as long as each graphic segment has the unique route identifier and the correct milepost or measurement value tied to each element. By having Route Network Topology the GIS software is able to place linear events along a route even if the event crosses several individual graphic elements. Route Network Topology should **not** be confused with a Routable Network that allows for things such as the routing of vehicles from one point to another within a road network. A Routable Network is not necessary for most transportation planning activities and hence is outside the scope of this project.

## I-2 Characteristics of Street Centerline Standards

It is of the utmost importance to define a standard for all of the aforementioned characteristics that will minimize the deficiencies of regional centerline datasets. As new centerlines are developed or existing centerlines are maintained or measured distances manipulated these standards must be strictly observed, allowing the region to maintain a centerline dataset that is highly accurate and usable for Transportation Planning. The most fundamental issues faced by DVRPC for Regional Transportation Planning is the establishment of a road centerline that has the following characteristics:

1. Owned and maintained by each local entity
2. Maintenance is performed on an on-going basis
3. Data is freely available to all DVRPC member agencies
4. Basic agreement on how centerlines representations are to be collected
5. Basic fundamental attribute information tied to each centerline
6. The geographic coverage is complete for the organizational needs.

# I-3 Critical Feature Attributes for Common LRS Approach and Address Geocoding

In order to implement the Common LRS Approach, there needs to be a consistent framework for incorporating LRS within the network data structure. Therefore, we recommend that a standard LRS data structure be strictly observed, throughout the region, respective to each State DOT. Establishment of this data structure will require careful consideration of the specific software requirements of key GIS software used in the region and the use of a software neutral data format for transferring the event data. The critical attributes for standardization of this structure are outlined in the following sections. It is recommended that all entities adhere to this structure and maintain these minimum attributes on their centerlines. Maintaining the minimum attributes on the centerlines will allow all agencies to share data with each other in the most efficient manner possible.

## I-3.1 Linear Referencing System (LRS)

In an LRS, features are, by definition, dynamically placed along the segments that make up the base network. In order to calculate the location of an event along a segment, the base network must include the following attributes.

| LRS Data Elements |
|---|
| Unique Route Identifier |
| Beginning Reference Point |
| Ending Reference Point |
| Direction* |

*Table I-1 LRS Data Elements*

1. Unique Route Identifier – The unique code that represents the route. A route may be comprised of multiple segments but all segments along a specific route will have the same unique route identifier.
2. Beginning Reference Point – The distance measure for the beginning of an individual segment of a route.

3. Ending Reference Point - The distance measure for the end of an individual segment of a route
4. Direction* – Population of this field is optional and only required in cases where the Unique Route Identifier does not include directionality as part of its definition

## I-3.2 Unique Numbering Systems

First and foremost, there needs to be a scheme in place in order to uniquely identify each road throughout the region. In some cases, it may be necessary to provide more than one attribute field in order to unambiguously identify a route. Typically, a second attribute field is used to differentiate among two or more subroutes, which share the same primary route number, but do not satisfy the topological requirements of a route spatial object. For example, a divided highway may be represented by two parallel lines in a map database, and for inventory purposes, both lines would have the same route number, but a different value in the direction field. It is recommended that the local agencies follow the DOT's scheme for uniquely identifying roads. New Jersey and Pennsylvania each have different methodologies for accomplishing this, which are described below.

### I-3.2.1 New Jersey

NJDOT employs the Standard Route Identifier scheme for uniquely identifying roads in New Jersey. It is strongly recommended that the agencies in the state follow this scheme, as well. The SRI scheme covers all roads in the State of New Jersey and currently NJDOT has enabled their GIS to handle the State Maintained Road Network and county roads down to the 500, 600 and 700 level. The unique identifier in the SRI scheme is an information-bearing identifier, meaning the field is coded with other useful information.

The coding follows the following scheme:

**CC|MM|RRRR|S|D**

CC = County
MM = Municipality
RRRR = Route Number
S = Suffix
D = Direction

### I-3.2.2 Pennsylvania

Currently, PennDOT uses NLF-ID to uniquely identify routes on the State Maintained Highway System. This unique identifier is not coded and is sequenced as new routes are added. In contrast to NJDOT, we must look at the data attributes in order to identify information about the route. This methodology for uniquely identifying routes works well for PennDOT, but it is arbitrary and offers no solution for identifying local roads. It is recommended that PennDOT adopt a numbering scheme similar to NJDOT, which will allow the local agencies to take part in the scheme and thus share data with the DOT.

Since the NLF-ID unique identifier is not reliable in uniquely identifying local roads, we recommend that the counties and the DOT use the following scheme for sharing data until PennDOT establishes a more defined scheme.

**CO|SR|Direction**

CO – County
SR – State Route
Direction

## I-3.3 Measurements

An LRS has, as its foundation, has a distance measurement along the centerline, from some beginning reference point to some ending reference point. The distance can be any unit of measure, including an address. The reference points can be at intersections, mileposts, stations, county boundaries or some other arbitrary location. New Jersey and Pennsylvania each have different methodologies for measuring distance, which are described below.

### I-3.3.1 New Jersey

In New Jersey, measurements, in miles, are made from south to north and from west to east. The measures go from the beginning of the route to the end and are not reset at county boundaries. This methodology is the same for both sides of divided highways as well. The centerlines are broken at arbitrary points as well as at county boundaries. The arbitrary break points are due to the graphic structure within their MicroStation CAD package that only allows for 99 vertices on an individual line string. Hence when 99 vertices are reached a new line is begun. The centerlines are not broken at intersections, making it difficult to ascertain mileages between intersections and an efficient manner. This is an important piece of information since the local agencies will need to break their centerlines at intersections in order to utilize street name/address range information for locating address events. This may force the agencies to add data to their centerline in two stages First, they will have to add the unique identifier through conflation or some other similar procedure. Then the agencies will have to add the measure values through a second, different procedure. One way to deal with this situation and not have to go through a two-step procedure is to use the Straight Line Diagrams to obtain the intersection mileages and attach that to the agency centerlines, a very tedious-time consuming process. Another solution involves a conversion between 2-dimensional distances, which is inherent in the geometry of the centerlines, and 3-dimensional real-world distances. The process for this will be discussed in later sections.

### I-3.3.2 Pennsylvania

In Pennsylvania, measurements are made in cumulative offset feet, from the beginning of the County to the end of the County measurements from south to north and from west to east. PennDOT has the capability to generate a file that will give the cumulative distance along a route within a county from one intersection to another intersection (local roads included). Again, this is an important piece of information since the local agencies will need to break their centerlines at intersections in order to utilize street name/address range information for locating address events. Since PennDOT does break the measurements at intersections the local agencies, the process for adding the unique identifier and the measure values is one step through conflation or some other similar process. The process specifics will be discussed in later sections.

## I-3.4 Street Address Geocoding

Address geocoding is today and will be in the near future the primary way event information is located by local entities along their centerline network. Address geocoding using road centerlines requires specific data elements for accuracy and consistency. Geocoding functions require that certain attributes be directly tied to the centerline. In order to reduce spatial errors and ensure geocoding accuracy among all centerline datasets, a standard must be in place that fosters consistency and accuracy throughout the region. The standard must cover the attributes that are included as well as the consistency of the data stored in those fields. The following table represents the attributes that must be tied directly to the centerline and are thus deemed critical for effective geocoding.

### I-3.4.1 Database Considerations For Street Centerline Geocoding

In order to promote consistency, certain field values should be obtained from a domain table, minimizing typographical errors and other user or operator created inconsistencies. These fields are marked with an asterisk (*). Additionally, to increase geocoding match rates the address event tables should have the following format.

1. <u>Prefix Direction</u> – The direction should be obtained from a domain table that contains all of the different variations for a direction (i.e. N, S, E, W, NW, etc…) The location of a direction designation may vary within an address. In some cases the direction field is stored in the prefix position and in others in the suffix position. As long as the storage is consistent for the event tables and the base geocoding table, the match rate will be high.
2. <u>Street Name</u> – Care should be taken to ensure that street names are not abbreviated or misspelled. In some cases streets may be known by more than one name. In these cases an alias or cross-reference table is needed. Streets with numeric names should be entered as $1^{ST}$ ST rather than First St.
3. <u>Street Type</u> – Street Types should also be obtained from a domain table that contains all of the United States Postal Service's recommended abbreviations.
4. <u>Suffix Direction</u> – see prefix direction
5. <u>Full Name</u> – This is the concatenated full name for a street segment, used for annotation.
6. <u>Address</u> – The street number is typically an integer, but it may also include alpha characters, (i.e. 142a or 216 ½). The left and right values determine the side of the street that the address occurs along. (FromAddressLeft, FromAddressRight, ToAddressLeft, and ToAddressRight)
7. <u>Zip Code</u> – The zip code is typically an integer. The left and right values again represent the side of the street that the zip code occurs along. The zip code value is important in distinguishing duplicate street names from each other.

| Address Geocoding Data Elements |
|---|
| PrefixDirection * |
| PrefixType * |
| StreetName |
| SuffixDirection * |
| SuffixType * |
| FullName |
| FromAddressLeft |
| FromAddressRight |
| ToAddressLeft |
| ToAddressRight |
| ZipCodeLeft * |
| ZipCodeRight * |

*Table I-1 Address Geocoding Data Elements*

On the basis of the foregoing information, the recommendations being suggested for inclusion of street addressing and address geocoding within the transportation GIS database design are as follows:

1. Each of the DVRPC member organizations should include street addressing as part of their transportation database design.

2. Certain attributes need to be included in the database design to effectively support the location of various point events along the street centerline using geocoding. These attributes are described in the foregoing sections.

3. The maintenance of point event data must be accomplished in a manner that will support location of the events along the centerline using geocoding. Specific attribute recommendations are provided in Table I-2 above.

### I-3.5 Data Sources for Common LRS and Street Address Data

There are several different data sources that are available for populating the various entities in the proposed data model. The Linear Referencing System data, namely the unique identifier and the from and to measure values should be obtained from the DOT's. This process is outlined in a following section and detailed in each agency's specific implementation plan. The Address data is a little more complex. There are several possible sources for this data including, E911, TIGER, any number of data vendors and possibly many others. Our recommendation is that the agencies use the most accurate and complete data that is available in their area. The decision of which source of the address data to use should be left up to the specific agency.

## I-4 Conclusions For Street Centerline Standards

On the basis of the foregoing information, the recommendations being suggested for establishing street centerline standards:

Based on the detailed assessment performed throughout the DVRPC region, it has been concluded that the Common Linear Referencing System Approach makes the most practical sense at this point in time.

1. The long-term solution needs to address the sharing of Transportation Planning data from both a “Top-Down / DOT to Local” as well as a “Bottom-Up / Local to DOT” way.
2. The Common LRS approach offers a de-centralized way to share the linear referenced data across agencies. And since the Common LRS Approach supports data, as it currently exists, it minimizes data and business process conversion and thus reduces the total cost of ownership.
3. There needs to be a scheme in place in order to uniquely identify each road throughout the region. In some cases, it may be necessary to provide more than one attribute field in order to unambiguously identify a route.
4. Information provided through the ***National Spatial Data Infrastructure (NSDI) Framework Transportation Identification Standard (Public Review Draft)*** (FGDC-STD-999.1-2000) prepared by the Ground Transportation Subcommittee of the Federal Geographic Data Committee (FGDC) and distributed in December, 2000 should be used to define the topological structure of the various road types represented through the centerline data.

# Chapter II - Database Design Standards

In order to achieve project goal number 3, which is to provide for the seamless exchange of GIS data files and the integration of planning infrastructure among all member governments and operating agencies, the design of a database structure that will facilitate the exchange and mapping of attribute and event data that is independent of the underlying graphic representation of the centerline is paramount. The previous section addressed the information that constitutes the base linework attribution that is required to support the recommend Common LRS Approach and Street Address geocoding. This section will deal with the information required to relate business data (ie. signs, traffic counts, accident locations, etc.) to the road centerlines.

## II-1 Event (Business) Data

Event data are typically stored and maintained in database tables that are not tied directly to the feature attribution of the road centerline. For example, a State DOT stores accident information in a separate event database table which is then related to the road centerline through the use an LRS. Event data can take the form of point events or linear events. Point events represent data that is placed at a distance along a centerline, including signs, accidents, signals, etc. Linear events represent data that placed between two distances along a centerline, including speed limit, pavement type, functional class, traffic volume, number of lanes, jurisdictions, etc. Again, Pennsylvania and New Jersey have different methodologies for locating event data, which are describe below.

### II-1.1 Linear/Point Event Tables

The formats for the Linear/Point Event Tables are similar. In both cases, all that is needed is the Unique Route Identifier and the measure information; two measures for a linear event a single measure for a point event. The location of the event along the road centerline can then be generated through the process of dynamic segmentation. Again, the user may want to include other information to identify the event more specifically. A value associated with the event, like a date for the accident or a numeric value for the traffic count. The recommended format for the event tables are shown below with a description of the fields.

1. <u>EventID</u> – The unique identifier for the event
2. <u>Unique Route Identifier</u> – The unique identifier for the route that the event occurs along.
3. <u>FromMeasure</u> – The beginning measurement for the event.
4. <u>ToMeasure</u> – The ending measurement for the event.
5. <u>Direction</u> – The direction of the event along the route. This is only populated in cases where there is a single centerline and the user wants to signify which side to place the event along
6. <u>User Defined Attribution</u> – This field(s) should contain the specific values for the event.

| Linear Events |
|---|
| EventID |
| Unique Route Identifier |
| FromMeasure |
| ToMeasure |
| Direction* |
| User Defined Attribution |

*Table II-1 Linear Event Attributes*

| Point Events |
|---|
| EventID |
| Unique Route Identifier |
| PointMeasure |
| Direction |
| User Defined Attribution |

*Table II-2 Point Event Attributes*

1. EventID – The unique identifier for the event
2. Unique Route Identifier – The unique identifier for the route that the event occurs along.
3. FromMeasure – The measurement for the event.
4. Direction – The direction of the event along the route. This is only populated in cases where there is a single centerline and the user wants to signify which side to place the event along.
5. User Defined Attribution – This field(s) should contain the specific values for the event.

## II-1.2 Address Event Table

The format for the Address Event Table is fairly simple. Theoretically, all that is needed is an address, in one field or parsed. This address can then be matched to the centerlines through the geocoding process. In reality, you may want to include other information to identify the event more specifically. For instance, there are also may be values associated with the event, like a date for the building permit or a numeric value for the traffic count. The recommended format for the event table is shown below with a description of the fields.

| Address Events |
| --- |
| StreetAddress |
| ZipCode |
| User Specified Attribution |

*Table II-3 Address Events*

1. Address – This field should maintain the complete street address for the location (i.e. 100 Main St). This field may be parsed into the components in the same manner as the Address Element Table.
2. ZipCode – This field should contain the zip code for the associated street address.
3. User Specified Attribution – This field(s) should contain data specific information on the event when appropriate.

## II-2 Data Sources For Event Data

In Volume I, Needs Assessment Overview, Chapter IV contains information regarding the various sources of event data that exist throughout the region. As might be expected, the primary sources currently are the State DOTs. However, both DVRPC and other region entities such as the New Jersey Transit Corporation, maintain databases that include transportation planning data that is of significant interest to the member organizations. At the local level, a few of the more advanced GIS users are actively developing and maintaining event databases that support transportation planning. It is apparent that initial sharing of data for transportation planning will be predominantly “top-down” with the State and regional agencies providing the major share of the data. However, as the DVRPC member organizations move forward with their implementation programs, the ability to share data from the local entities up to the State and regional organizations will become more feasible.

### II-2.1 New Jersey

New Jersey DOT will supply all of their event data in SRI/Milepost format. A single milepost for point data and a begin/end milepost for linear events will be supplied. All linear events occur along a single SRI number. Any event that traverses multiple SRI’s will be broken into multiple database records.

### II-2.2 Pennsylvania

PennDOT will supply all of their event data in County/State Route and Cumulative Offset format. A single offset will be provided for point events and a begin/end cumulative offset will be provided for linear events. All linear events occur along a single SRI number. Any event that traverses multiple County/SR’s will be broken into multiple database records.

## II-3 Recommended Data Model

The recommended GIS-T data model is structured to support the Unique Identifier, Address Information and Linear Referencing System information on all of the centerlines. Based on the experience gained during the Centerline Development Options Demonstration, the following data model is recommended.

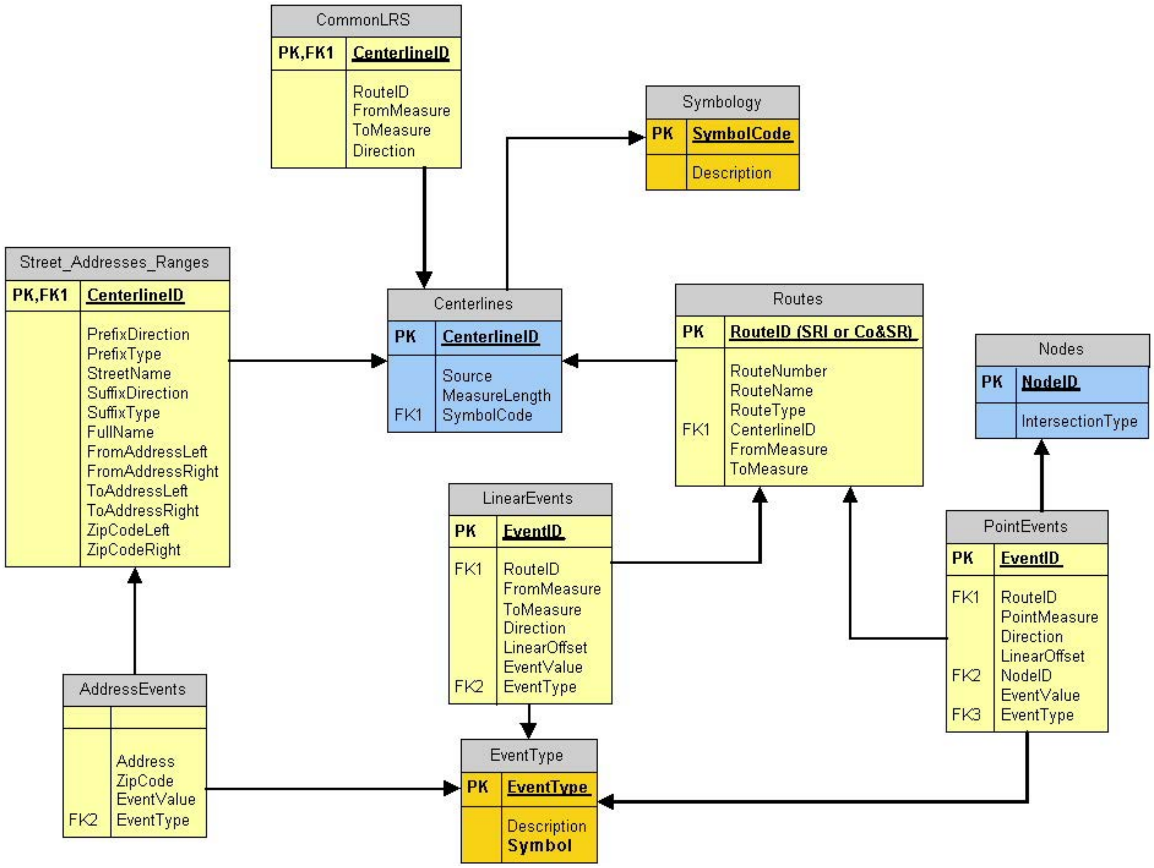



*Figure II-1 Recommended Data Model*

Blue entities are those tied directly to the geometry. The “Centerlines” entity is a table that represents each individual instance of a centerline segment. The “Nodes” table represents each individual instance of a node.

Yellow entities are those tables that represent Linear Referencing Methods. The “CommonLRS” entity represents data that must be included for the Common LRS. Although the “CommonLRS” entities are shown as separate from the “Centerlines” entity, they could be attributes of the “Centerlines” entity, since there is a direct one-to-one relationship between these. The attributes in each of those tables must be tied directly to each centerline segment from the appropriate source, NJDOT or PennDOT. These attributes depend on the state in which the agency resides. The “Routes” entity represents groups of centerlines that together, form a route. The “Routes” entity depends on attributes from the “CommonLRS” entity in order to build the Route-Milepoint type LRM, similar to that of NJDOT or PennDOT. The “RouteID” is the unique identifier and also depends on the state in which the agency resides. (SRI in New Jersey and County SR in Pennsylvania) The “FromMeasure” and “ToMeasure” fields represent the Begin Offset/Milepost and End Offset/Milepost, respectively. The “Street_Address_Ranges” entity represents unique address ranges for each individual centerline segment and uses the Street Name–Address Range type LRM. The “LinearEvent” or “PointEvent” entities represent events that occur along a centerline either at a discrete location or from one location to another. These entities are based on the "Routes" entity and the associated LRM in this model. The "AddressEvent" entity represents an event that occurs at a discrete location or from one location to another and is based on the "Street_Address_Ranges" entity and the Street Name-Address Range LRM.

Orange Entities are those tables that represent ways of implementing standard symbology. The "Symbology" table is an entity utilized for standard symbology for the road centerlines. The Symbol Code may correspond to the Census Feature Class Code outlined in the Assessment portion of this project. The "EventType" table is an entity that describes the various event data as well as offering a column that can be populated with a standard symbol type or style. This table may include further columns to account for font files, size, color and other symbology related identification.

## II-4 Conclusions For Database Design

On the basis of the foregoing information, the recommendations being suggested for designing a transportation GIS database are as follows:

**1.** In order to achieve project goal number 3, which is to provide for the seamless exchange of GIS data files and the integration of planning infrastructure among all member governments and operating agencies, the design of a database structure that will facilitate the exchange and mapping of attribute and event data that is independent of the underlying graphic representation of the centerline is paramount.

**2.** It is apparent that initial sharing of data for transportation planning will be predominantly "top-down" with the State and regional agencies providing the major share of the data. However, as the DVRPC member organizations move forward with their implementation programs, the ability to share data from the local entities up to the State and regional organizations will become more feasible.

3. As the DVRPC member organizations move forward with their implementation programs, the ability to share data from the local entities up to the State and regional organizations will become more feasible.

4. The recommended GIS-T data model is structured to support the Unique Identifier, Address Information and Linear Referencing System information on all of the centerlines.

(This page left intentionally blank)

# Chapter III - Common LRS/Street Address Geocoding Implementation

## III-1  Implementation Plan Outline

The general process for implementation of the Common LRS/Street Address Geocoding. Approach is shown below.  Again specific agency requirements will be outlined in their individual implementation plan.

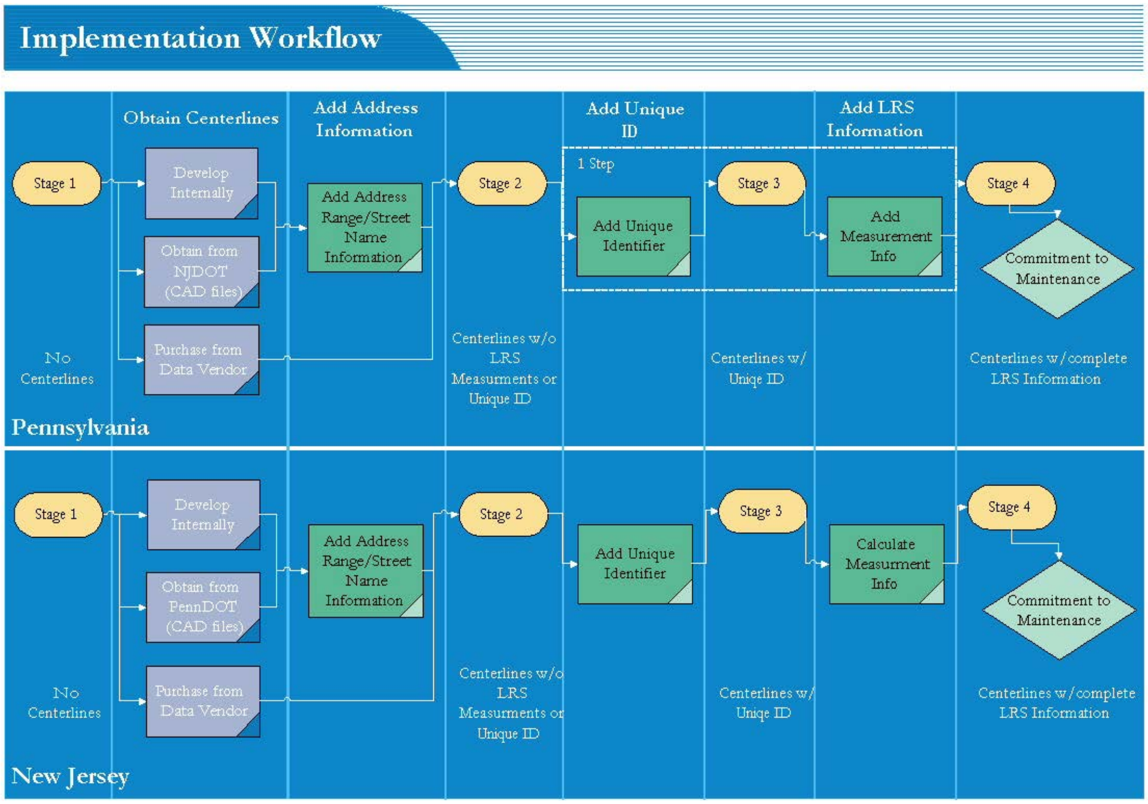



*Figure III-2  Implementation Workflow*

## III-2  Use of Conflation to Populate Databases

The primary goal of the implementation plan is to define a process that correctly builds and populates the necessary components of the recommended data model.  A major process of the workflow required to accomplish to populate base feature attributes is conflation.  Conflation is the term for the process of merging two different data sets. It is useful to refer to one set of data as the *source data* (data with valid attribute data), and the other data as the *target data* (data with accurate geometry). The process often involves moving attribute information from a less spatially accurate dataset called the source, to a more spatially accurate dataset, called the target that may have fewer if any usable attributes.

A normal conflation workflow is comprised of two major components: (1.) data preparation and (2.) the actual conflation process.

## III-2.1 Data Preparation

Most of the processes required to perform conflation are supported by software tools included in most GIS software packages. Before initiating the conflation process, it may be necessary to perform some data preparation.

### III-2.1.1 Update Centerline Geometry

Prior to commencing the conflation process it is highly recommended that the base centerline geometry is as current and accurate as possible. Ideally, the responsible agency will already have data maintenance policies and procedures in place that will help to ensure that the centerline data meets all accuracy and currency requirements.

### III-2.1.2 Arc-node topology

Both the target and the source need to have arc-node topology established, if it is not already. Upon conversion of the source from the existing format to coverage format, such topology is not established automatically. Most GIS software products provide tools for "cleaning" linework, a process that includes rectification, verification of topological integrity and consistency, and validation of projection and units. Cleaning linework typically establishes nodes at every point where two arcs meet in a 2D representation. This means that everywhere that two or more arcs representing line segments cross, (whether at the same grade or not); a node will be created. Other tools can be used to update feature attribute tables.

### III-2.1.3 Verify Address Consistency

Prior to commencing the actual conflation process, the responsible agency needs to identify discrepancies in the source file. Errors such as overlap in address ranges; address range reversal and arcs that may be flipped will be reported with this routine. All address errors reported via this process must be corrected in order for accurate geocoding to occur.

### III-2.1.4 Pseudonode Removal

Data preparation will also involve removing all pseudo nodes from the source coverage. A pseudonode is described as a node where two, and only two, arcs intersect, or a single arc that connects with itself. Pseudonodes can be eliminated first identifying the arcs that conjoin at a pseudonode, and then eliminating that node while distributing the address information specific to both arcs.

### III-2.2 Establish Rule-Base

The first step for conflation is to establish a Rule-Base that can be utilized during the conflation procedures. This Rule-Base will ensure that all parties working on the project will handle issues in the same manner. The recommended Rule-Base for dealing with situations illustrated in the diagram below follows.

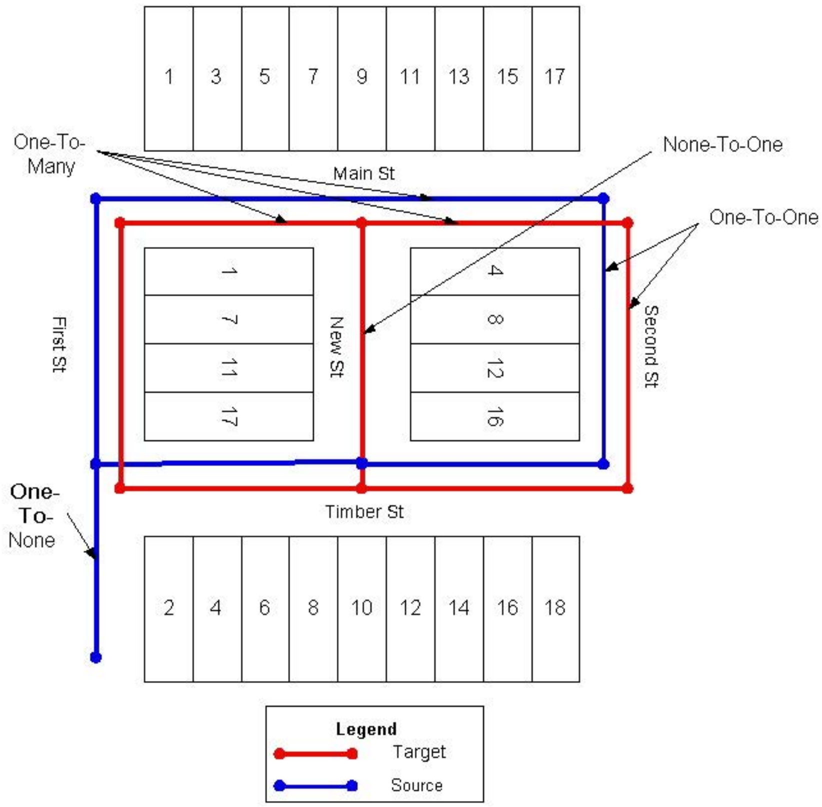



*Figure III-1 Street Centerline Conflation Rule-Base*

**One-to-One**: Conflate normally.
**One-to-Many**: (One source segment for many target segments) This is a frequent problem when conflating attributes from a less spatially accurate dataset to one that is more spatially accurate. In this situation the responsible agency should seek other sources to verify address range information on these segments. When all else fails it may be necessary to interpolate the address range for these segments based on segment length.
**Many-to-One:** (Many source segments for one target segment) Many to one relationships will require that the target dataset be split and then conflated as a one-to-one relationship.
**None-to-One:** (No source segments) The responsible agency should employ the same rule applied to the One-to-Many case whereby an additional source will be sought to verify the address range. Otherwise, the agency should leave this segment blank until a reliable source can be found.
**One-to-None**: (One source segment, no target segments) In this case the responsible agency should add all line segments that are in the source data set and not in the target dataset by sketching in the new segment using the source data with orthophotos as a backdrop and conflating all associated attributes. Additionally, the agency should snap the added arc to the existing geometry to preserve arc topology.
**Target Arc is shorter than the Source Arc** – When the target arc is shorter than the source arc, the full source address range should be applied to the target arc.
**Arc Directionality** —Arc directions should follow the address ranges as attributed in the address source file. Typically the arcs “from” node will be located at the end of

the street that contains the lowest address and the “to” node at the highest. If necessary, target arcs should be flipped to follow this rule. Major roads and ramps will be flipped to match the directionality of the source file.

### III-2.2.1 Execute Conflation Process

The general process for conflation is as follows:

1. If any automated conflation tools are available, run the automated conflation process. This typically will process approximately 50% of the segments in the target dataset.
2. Check the match rate and assess the geometry tolerances until optimal matching occurs.
3. Perform interactive conflation for the remaining un-conflated segments. This is done for the remaining unmatched segments based on the type of relationship that exists between the source and the target data.
    a. One-to-one
    b. One-to-many
    c. Many-to-one
    d. One-to-none
    e. None-to-one
4. QA/QC.
    a. Check the flow of the address ranges based on the road name.
    b. Fix any address anomalies.

## III-3 Implementation Outline

The Implementation Process for region-wide data sharing will take on a phased approach:

- Phase I – Only roads maintained in the state DOT GIS database will be shared. Address ranges should be added to all roads, but LRS information will only be added to the roads maintained in the DOT database. This is due to the lack of unique identifier on all roads.
- Phase II – LRS information will be added to all roads. Once the unique identifier is added to all roads Phase II can occur.

Again, with the experience gained during the Centerline Development Options demonstration the following methodology is recommended for implementation. There are several decisions that the agencies must make when implementing their specific solution. Associated with these decisions are benefits and consequences that are discussed in general terms in the following sections. The benefits and consequences will be discussed further in each agency’s Implementation Plan. Since PennDOT and NJDOT conduct business in slightly different manners, the implementation plans will be constructed with that in mind.

I. Obtain Centerlines (Internal, Borrow or Purchase from Data Vendor)

    a. Internal Development

        i. Professional training may be necessary in order to facilitate centerline collection.

ii. Develop Rule Base for photo-interpretation and data collection

1. This should help minimize the errors during development due to misinterpretation. Develop definitions for centerlines. This depends how the centerlines are to be used. (i.e. a centerline may be defined as 1 centerline per road, 1 centerline per direction per road, 1 centerline per lane per direction per road, etc…)

2. Other Rule Bases may include digitization direction, intersections, ramps and various other complex features. Standardization is the key.

iii. Collect the centerlines and any important attributes based on the pre-defined Rule Bases.

iv. Quality Assurance/Quality Control

1. Maintain a defined plan for QA/QC throughout the development process.

2. Develop Standards for maintaining topologic integrity and other rules that were defined during development.

3. The QA/QC plan can eventually become part of your Agency's maintenance plan.

b. Borrow from the DOT

i. The DOTs maintain Computer Aided Design (CAD) files by county for all of the roads that exist. These centerline datasets have no attribution but the geometry is clean and accurate and can easily be used as a foundation for the specific agencies centerline dataset.

c. Purchase from Data Vendor

i. Purchasing centerlines from a data vendor alleviates the necessary step of adding address information to the dataset. However, the overall ownership of the centerlines still belongs to the Data Vendor, making the dataset unavailable for data sharing and difficult to maintain. If the agency can come to an agreement with the data vendor to add and maintain the LRS information on the centerlines then this would be a viable option, otherwise, this option may not be worthwhile.

II. Add Address Information

a. Professional training may be necessary in order to facilitate centerline attribution

b. Depending on the availability of data, attributes can be obtained in a either of the following methods:

i. Conflate from other datasets

1. Develop Rule Base for conflating the attribute data

a. This should help minimize the errors due to misinterpretation.

b. Establishes rules for dealing with conflating attributes where the datasets do not have a one-to-one match.

ii. Manual Input

1. Field collection

a. Develop methodology and standards for field collection.

2. From Other datasets

a. Manual input from other datasets.

c. Quality Assurance/Quality Control

i. Verify the attributes for correctness through verification with other sources and field verifications where necessary.

III. Add Unique Identifier

a. The same procedures will be followed here as in the steps for adding the address information. It again, requires conflation of the unique identifier for the centerlines.

IV. Add LRS Information

a. Measure Values

i. Pennsylvania – Measure values will be added as part of the conflation process in Step III, above.

ii. New Jersey

1. Since NJDOT does not break the measure values (3-dimensional distances) at intersections of local roads, there needs to be a work around for determining the 3D distances between intersections. However, NJDOT does break the measure values at county boundaries, which can be used in the work around solution. Inherent in the centerline dataset is the 2-dimensional distance measured by the GIS software. The 3D distances can be estimated through a ratio. This process requires the calculation of a 2D/3D ratio in order to automate the addition of the measure values.

a. The process for converting the 2D distance to the 3D estimate is as follows:

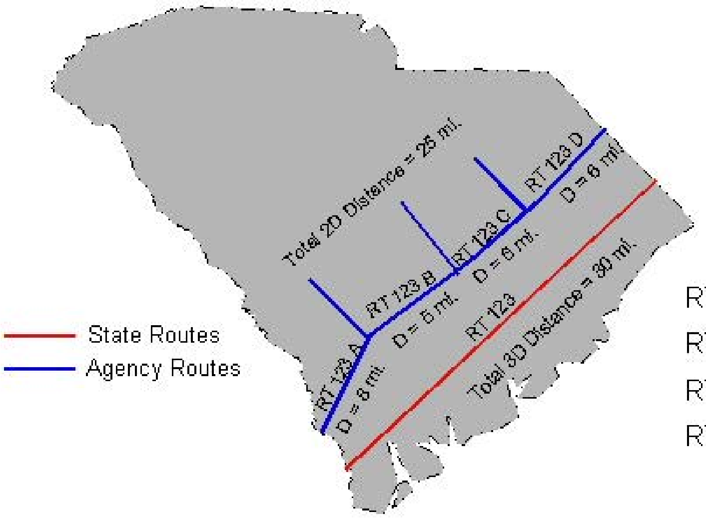
i. Determine the 3D distance for the length of the route that is inside of County A

ii. Determine the 2D distance for the length of the route that is inside County A.

iii. Calculate the ratio for 3D/2D

iv. Multiply this ratio value by the individual, intersection-to-intersection 2D lengths in order to get the estimated 3D length for each individual segment.

v. The following diagram outlines the process.




$$\frac{3D}{2D} = \frac{30}{25} = 1.2$$

RT 123 A = 8 X 1.2 = **9.6 mi**
RT 123 B = 5 X 1.2 = **6 mi**
RT 123 C = 6 X 1.2 = **7.2 mi**
RT 123 D = 6 X 1.2 = **7.2 mi**

| Routes By County | | |
|---|---|---|
| RouteID | FromMeasure | ToMeasure |
| 123 | 20.0 | 50.0 |

| Agency Centerlines | | |
|---|---|---|
| RouteID | FromMeasure | ToMeasure |
| 123 A | 20.0 | 29.6 |
| 123 B | 29.6 | 35.6 |
| 123 C | 35.6 | 42.8 |
| 123 D | 42.8 | 50.0 |

*Figure III-3  Converting 2D Distance to 3D Distance*

*Note*:  The “RouteID” includes a letter designation in the Agency sample database.  This is for illustrative purposes only.  The actual “RouteID” column will not contain values such as this, it will correspond to the state’s unique numbering scheme.

# III-4   Data Maintenance

Maintenance is the process of preserving the centerline dataset as changes are made in the real world.  These changes must be tracked and incorporated into the production centerline dataset within a pre-determined period of time.  In order for maintenance to function efficiently, the process and parameters must be strictly defined and observed.  A maintenance plan defines a context or methodology, including formal procedures for requesting, evaluating, and implementing changes to a database.  The objective is to identify and document the data that must be modified, the likely impact of the change on normal operations, and the time, cost and other resources required to implement the change.  Next, the change is analyzed, designed and tested.  It is important that all changes be made in a consistent manner.  Some maintenance problems require an immediate response.  Situations that threaten major integrity, release of new government regulations or others require a quick fix, and formal procedures must wait.  However, these emergency data patches should be formally incorporated into the database.  Maintenance Plans take many forms: corrective, adaptive, perfective and preventative.

## III-4.1 Corrective Maintenance

Corrective maintenance activities include both emergency corrections and preventative repairs.      Theses maintenance procedures include improving the integrity and the reliability of the data, streamlining and tightening data validation routines and correcting invalid data. This maintenance procedure involves cleaning and quality checking the data prior to inclusion in the overall database.

### III-4.2 Adaptive Maintenance

Adaptive maintenance refers to the process of enhancing the system by adding data features in response to new requirements or new problems. Adaptive maintenance is reactive; fix the data only when it is necessary. This type of maintenance should be completed when additional fields need to be included in the database, or changing the data to reflect a change in focus of the project or the organization.

### III-4.2 Perfective Maintenance

Perfective maintenance is the process of enhancing the data by improving efficiency, reliability or maintainability often in response to user requests. Here the idea is to fix the data before it breaks. Restructuring efforts are aimed at enhancing performance without changing how the database works or what it does. Perfective maintenance may include such things as: Data Normalization (if the data is not currently normalized), on-going fixes and updates that allow the data and associated systems to operate more efficiently.

### III-4.3 Ongoing Preventative Maintenance

Ongoing preventative maintenance is an important part of any database's standard operating procedure. The objective of preventative maintenance is to anticipate problems and correct them before they occur. Files and databases must be updated, periodically reorganized and regularly backed up.

### III-4.4 Data Maintenance Requirements

With the above information in mind, there are requirements for maintaining the data that will be shared throughout the region. These maintenance requirements go above and beyond the normal maintenance procedures for centerline datasets. These additional procedures require careful coordination between the agencies to ensure that the agencies are always on the same page. Although this sounds technically difficult, it is not. The data that needs to be maintained for effective data sharing is the attribute data only, and more specifically the attribute data that define the Linear Referencing System. Route identifications or measure values are the items that need to be maintained across agencies. Alignments, and other geometric changes are not necessary for the effective sharing of data. In the scenario we have recommended that data that will be transferred is the event data, and as long as the underlying LRS information is up to date, the event data will be placed in the appropriate location.

Although, the maintenance requirements will be slightly different is each state, there are some similarities. For instance, in both cases, attributes can be maintained through a change file that designates which routes have changed. The change file should include the unique identifier, the old value and the new value. This file can then be utilized to update the agency's centerlines through a simple database procedure. The specific procedures for maintenance for each state and are discussed below.

### III-4.2 New Jersey

The attributes for New Jersey agency centerlines can be updated through the change file mentioned above. Since NJDOT centerlines are not broken at intersections, when new roads are added locally the measure values will need to be updated by hand through estimations or interpolations. The methodology used will also utilize the 2D/3D ratio described above.

### III-4.3 Pennsylvania

The attributes for Pennsylvania agency centerlines can be updated through the change file mentioned above, as well. Since PennDOT centerlines are broken at intersections, when new roads are added locally the measure values can be obtained from PennDOT and updated on the local centerlines, automatically.

## III-5 Conclusions for Implementation Planning

1. The primary goal of the implementation plan is to define a process that correctly builds and populates the necessary components of the recommended data model.
2. A major process of the workflow required to accomplish to populate base feature attributes is conflation which is the term for the process of merging two different data sets. The process often involves moving attribute information from a less spatially accurate dataset called the source, to a more spatially accurate dataset, called the target that may have fewer if any usable attributes.
3. The Implementation Process for region-wide data sharing will take on a phased approach:
    - Phase I – Only roads maintained in the state DOT GIS database will be shared. Address ranges should be added to all roads, but LRS information will only be added to the roads maintained in the DOT database. This is due to the lack of unique identifier on all roads.
    - Phase II – LRS information will be added to all roads. Once the unique identifier is added to all roads Phase II can occur.
4. It is essential that each DVRPC member organization develop data maintenance programs and procedures to ensure the long-term validity and viability of their transportation GIS databases.

(This page left intentionally blank)

# *Appendix A - Metadata, Data Dictionary and Symbology*

# A-1 Metadata

As discussed previously in Chapter V of Volume I, "Needs Assessment Overview", the term "metadata" is commonly used to describe "data about data". Metadata typically describe the origin, content, quality, condition, availability and other characteristics of a data set. Put simply, metadata helps a user to locate and understand data.

The use of metadata is not unique to GIS data. Most information systems that store and manage large volumes of diverse types of data utilize some form of metadata to provide maintainers and users with the descriptive information that they need to track and understand the data sets.

The primary uses of metadata are:

- To organize and maintain an organization's investment in its databases
- To provide information to data catalogs and clearinghouses
- To provide information to aid in the transfer of data
- To improve the users understanding concerning the available data sets

The prior reference to Metadata also cited the dynamic nature of GIS data and the fact that mechanisms, procedures and standards for maintaining current records of changes to spatial databases are vital to ensure their usefulness. As defined previously, mechanisms are the software tools that are required to integrate the maintenance of metadata into an overall program of database maintenance and management. Procedures refer to the prescribed operational steps that form an orderly process for employing the mechanisms to create and maintain the metadata. Standards refer to the common terminology and definitions that are required to ensure that the metadata is useful and meaningful to a widespread, diverse user community. These standards also typically include the format and content requirements for metadata documentation.

## A-1.1 Recommendations for Standards

The most common standard in use is the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). This standard is quite comprehensive and is designed to be applied to a broad range of GIS data types. For this reason, it is often regarded as being too complex for compliance by most organizations. Nonetheless, the primary objective of the standard is to provide a common set of terminology and definitions for the documentation of digital geospatial data. The information included in the standard was selected based on four roles that metadata play:

- availability -- data needed to determine the sets of data that exist for a geographic location.
- fitness for use -- data needed to determine if a set of data meets a specific need.
- access -- data needed to acquire an identified set of data.
- transfer -- data needed to process and use a set of data.

The metadata research process has identified subsets of the FGDC CSDGM that meet the objective mentioned above for the DVRPC member organizations. These subsets could serve as the standard metadata elements for the DVRPC member organizations. The recommended items should be populated and kept current as spatial data is created and updated.

#### A-1.1.1 New Jersey

For member organizations in New Jersey, the recommended metadata solution is the standard promoted through NJMetaLite. This is a software application and corresponding metadata specification, which runs on Windows 95/98 and Windows NT (Service Pack 5 or above. It is a modification of MetaLite, a metadata tool developed by the USGS and the United Nations Environmental Program. The data format is a modified version of the FGDC standard developed by the New Jersey Department of Environmental Protection include data fields and features relevant to GIS work being done in New Jersey. Samples of this metadata format are depicted in Appendix A. For more information, refer to the following Internet address: *http://www.state.nj.us/dep/gis/endex/njml.htm*.

#### A-1.1.2 Pennsylvania

One option for Pennsylvania member organizations is the metadata standard promoted through the Pennsylvania Spatial Data Access (PASDA) Data Interview Form. This is an online tool for creating metadata by completing a relatively easy to use interactive form. This form is accessible through the following URL:

http://www.pasda.psu.edu/metadata/metamaker-ez.shtml.

### A-1.2 Recommendations for Mechanisms

At the present time, there exist a number of mechanisms, or tools for creating and maintaining metadata for GIS databases. There are several software tools that have been developed and are available free of charge. Most of these tools support the Federal Geographic Data Committee (FGDC) Metadata Standard that is described briefly below, or some form of modified version of this standard. Chapter V of Volume I lists several examples of tools for metadata maintenance and creation. That list is repeated below for reference.

- *Pennsylvania Spatial Data Access (PASDA) Data Interview Form* – (See description in section A-1.1.2, above.)

- *NJMetaLite* – (See description in section A-1.1.1, above)

- *ESRI® ArcInfo® ArcCatalog* - ArcInfo 8.1 GIS software comes bundled with the FGDC Metadata Editor. This tool, accessible through ArcCatalog™ facilitates the creation of FGDC-compliant metadata for the selected item in the ArcInfo Catalog tree. The use of this tool will support the creation and maintenance of the various FGDC subsets mentioned previously.

- *ESRI® ArcView® Metadata Collector* – Developed by NOAA, this is an easy-to- use application that can be utilized by any ArcView 3.x user for creating Federal Geographic Data Committee (FGDC)–compliant metadata. The tool can create metadata for any data type supported by ArcView, including ARC/INFO coverages, ArcView shapefiles, as well as any supported image formats. (http://www.csc.noaa.gov/ metadata/text/download.html)

- *CORPSMET95* – Developed by the U.S. Army Corps of Engineers, this is a Windows-based desktop application that leads the end user through the process of developing FGDC compliant metadata. (*http://corpsgeo1.usace.army.mil/*)

- *Metamaker 2.30* - produced to support the National Biological Information Infrastructure (NBII), this product supports the FGDC Content Standard for Geospatial Metadata (CSDGM) as well as additional fields recording biological related information as required by the NBII. This is a standalone database based on Microsoft Access Version 2.0 for Windows 3.1. It is a 16-bit program that stores metadata elements in a relational database. Output can be produced using cns & mp which are integrated into the program. Refer to the following URL for further information: *http://www.umesc.usgs.gov/metamaker/nbiimker.html*.

#### A-1.2.1 New Jersey

It is recommended that New Jersey member organizations adopt the NJMetaLite software as the mechanism for creating and maintaining metadata for their transportation related GIS data. This recommendation is consistent with the prior recommendation to adopt the NJMetaLite metadata standard.

#### A-1.2.2 Pennsylvania

It is recommended that Pennsylvania member organizations utilize the tools available through the PASDA website to create and maintain their transportation GIS metadata.

### A-1.3 Recommendations for Procedures

The adoption of a standard and a mechanism alone is not sufficient to ensure that a complete, accurate and sustainable set of metadata will become a reality. The DVRPC member organizations are encouraged to incorporate the practice of creating and maintaining metadata into the standard operating procedures or maintaining all of their spatial data. Whatever metadata editor software tool is employed, its use must be a step in the overall data maintenance process. It is suggested that each agency responsible for spatial data creation and maintenance also hold responsibility for the upkeep of the corresponding metadata document. It is also suggested that the individual metadata documents be assembled into a single document containing information about the complete record of the enterprise GIS data sets. This complete document should be made available to both internal and external users of the accessible spatial data.

### A-1.4 DVRPC Responsibility

To assist its member organizations in the implementation and use of the recommended metadata standards, DVRPC should strive to promote the inclusion of a metadata creation or update task as a part of all GIS data development and maintenance programs. Part of this promotional effort may include the implementation of a GIS-T Metadata page on the DVRPC website. By serving as a clearinghouse agency for transportation GIS metadata, DVRPC would be providing a valuable service to its member agencies while actively promoting a vital component of a regional GIS framework. Putting this into practice would require the adoption of standards, procedures and mechanisms that are consistent throughout the region and will readily support the exchange of metadata between DVRPC and its member agencies.

## A-2 Data Dictionary

In Chapter V of Volume I, Needs Assessment Overview, there is a discussion on the need for a data dictionary to support transportation GIS databases and the role that the data dictionary should play in providing this support. To reiterate, a data dictionary is a repository of information in a GIS in which information is stored on all the objects within the database and their relationships. The data dictionary serves as a reference source suited for many purposes in an operational GIS environment, including: training, metadata, quality control, and data development. By specifying the exact GIS data structure (layer names, table layouts, valid values, and annotation information), this information can be used by the GIS end users to better understand the data that they are using. If DVRPC and its member agencies are to establish any type of effective mechanism to share GIS data related to transportation planning, it will be absolutely essential that all data generators design, develop and deploy a data dictionary to support the use of their data on a region-wide basis.

### A-2.1 Data Dictionary Versus Metadata

Within the GIS industry and community of practitioners, there often exists some confusion and even disagreement over the distinction between metadata and data dictionaries. Quite often, these terms are used interchangeably. For the purposes of this project and its requirement for technical recommendations, the following distinctions are established between these two often-conflicting terms.

Metadata, in this case, comprises information about the dataset and is typically based upon some standard format, in most cases, the FGDC standard cited earlier. Metadata may include the data dictionary and typically provides more detail about the creation, maintenance processes. Metadata also describes limitations of use and access. The data dictionary may be included as a component of the overall metadata document. However, the term data dictionary does not necessarily include all the other facets typically included in the "standardized" metadata form. Metadata is normally used as a comprehensive, all-inclusive, source of information for a multitude of characteristics of the data to which it refers.

On the other hand, a data dictionary is itself a database about data and databases. It holds a variety of information, including the name, type, range of values, source, and authorization for access for each data element in the organization's files and databases. It also indicates which application programs use that data so that when a change in a data structure is contemplated, a list of affected programs can be generated. The data dictionary may be a stand-alone system or an integral part of the DBMS. Data integrity and accuracy is better ensured in the latter case. A data dictionary is typically more focused than the overall metadata, with most of this focused being placed on the actual structure of the database.

### A-2.2 Recommendations

There are a number of versions of both metadata and data dictionaries currently in use among DVRPC member organizations. As part of the earlier phases of this project, several sets of sample data were obtained from various organizations. Included were several samples of data dictionaries from the New Jersey and Pennsylvania Departments of Transportation, DVRPC, City of Philadelphia, Chester County, and others. While each of these differ somewhat, they, for the most part, include certain common elements which form the basis of the data dictionary structure and content recommendations.

#### A-2.2.1 Feature Definition

It is recommended that the transportation GIS data dictionary include elements that define the various features in the database. Typically, the components of feature definition include a standard term for the feature, a definition of the standard term, and the type of spatial representation used to define the spatial extent of the feature (point, line, polygon, etc.). Other components may include a description of the derivation of the feature's spatial representation and attribution and details about the specifications of the original data capture.

| Standard Term | Definition | Spatial Representation |
|---|---|---|
| Street segment | A section of roadway connecting two intersections | Line |
| Intersection | A point lying at the intersection of two or more street segments | Point |
| Accident site | A location along a street segment or within an intersection marking the occurrence of a traffic mishap or collision | Point |

*Table A-1 Feature Definition Example*

#### A-2.2.2 Feature Attribution

Along with the definition of the spatial features within the database, it is also recommended that the data dictionary include definitions of the various attributes associated with each feature. The format that is used in most cases, is a layout of the attribute data table structures. This format also includes the name of the attribute table, a description of the table and, for each attribute in the table, the components of a typical attribute definition include a standard term for the attribute, its data type (numeric, character, etc.) the length of the field or row in the database table and a description of the attribute. The definition sequence normally concludes with a list of "valid values" for each attribute, where appropriate. An example is shown in the following figure.

**Cover:** TRANS **Theme**: PLANIMETRIC & TOPO **Source:** AERIAL PHOTOGRAPHY (APRIL, 1993) **Describe:** Transportation Facilities other than roads, rails & bridges

This coverage contains the transportation and parking planimetric features. This includes roadways, parking lots and airports. The features are stored as both an Arc Attribute Table (AAT) and a Polygon Attribute Table (PAT). Annotation subclasses are defined for parking, airport and runway. Each uses ARC/INFO symbol 1.

**File size:** Approximately 495KB **Created:** ASI **Contact:** GIS Staff (301) 952-4779
**Data format:** Arc/Info Coverage 8.x Format
**Data structure:**

Polygon Attribute Table (PAT)        Table Name: trans.pat

| ITEM NAME | WIDTH | OUTPUT | TYPE | N.DEC | ALTERNATE NAME |
|---|---|---|---|---|---|
| FTR-CODE | 2 | 5 | B | - | - |

FTR-CODE - feature code

1201 - street
1202 - parking
1203 - airport
1204 - median
1205 - unpaved parking
1301 - non transportation feature

ARC Attribute Table (AAT)        Table Name: trans.aat

| ITEM NAME | WIDTH | OUTPUT | TYPE | N.DEC | ALTERNATE NAME |
|---|---|---|---|---|---|
| FTR-CODE | 2 | 5 | B | - | - |
| HID-SEG | 1 | 1 | C | - | - |

FTR-CODE - feature code

1101 - neatline
1102 - project boundary
1103 - paved road
1104 - unpaved road
1105 - paved parking
1106 - unpaved parking
1107 - airport
1108 – runway

*Figure A-1  Attribute Definition Example*

## A-3  Conclusion For Metadata and Data Dictionaries

The foregoing examples represent just a few of the many types of formats currently being employed to build and deploy GIS data dictionaries. Some of these formats include more detail, others less. It is important that DVRPC, perhaps through the project steering committee, more clearly define an appropriate data dictionary standard for its member organizations. The information conveyed through the prior sections is to serve as a basic foundation for this process.

# *Appendix B - Brief Metadata Elements List*

# B-1 Metadata Elements

The following list represents recommended items to be included in the standard metadata elements list for transportation GIS databases. Please keep in mind, the items are not presented in an order consistent with the suggested metadata content structure. *Note*: Information having no corresponding metadata element may be placed in the "Supplemental Information" field

The metadata content structure follows the section groupings listed below. These groupings are described in more detail in the Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata, Version 2 (revised June 1998) FGDC-STD-001-1998.

## B-1.1 Metadata Elements Sections

1. Identification Information
    - Citation Information
    - Description
    - Time Period Information

2. Data Quality Information
3. Spatial Data Organization Information
4. **Spatial Reference Information**
5. Entity and Attribute Information
6. Distribution Information
    - Contact Information

## B-1.2 Metadata Elements Fields

**Originator**
The name of the department, government agency, or other organization that developed the data set; this may be the same agency that is creating the metadata.

**Publication Date**
The date when the data set was published or otherwise made available for release.

**Title of Data Set**
A title, which describes the data set.

**Edition**
Indicate the version of the data set (it may be a date).

**Presentation Form**
The mode in which the geospatial data is represented.

**Publication Place**
The name of the city (place) where the data set was published or released.

**Publisher**
The name of the individual or organization that published the data set. (Often the same as the *Originator*.)

**Online Linkage**
If the data set is available online, provide the web, ftp or gopher address where the file can be found.

**Abstract**
A brief, narrative summary of the data set that describes the "what" aspects of it.

**Purpose**
A summary of the intentions with which the data set was developed. The Purpose describes the "why" aspects of the data set.  This field may also be used to indicate what the data set is **not** suitable for.

**Supplemental Information**
Any other important descriptive information about the data set.  This will include a brief summary about data quality, information pertaining to source documents used to create the data set, or additional information having no corresponding metadata element.

**Dates**

- **Beginning Date:** The first year (and optionally month, or month and day) of the event or period of coverage.
- **Ending Date:** The last year (and optionally month, or month and day) for the event or period of coverage.

**Currentness Reference**
Indicates the basis on which the time period of content information was determined. Most potential users are interested in a data set's currentness with regard to the "ground condition".

**Progress**
The state of the data set. Choices are: *Complete, In work, Planned*

**Intended Data Set Maintenance and Update Frequency**
The frequency with which changes and additions are made to the data set after the initial data set is completed.

**Bounding Coordinates**
Latitude and longitude values, expressed in decimal degrees, that form a bounding rectangle indicating the limits of coverage of the data set.

**Theme Keyword(s)**
Identifies one or more of the provided theme keywords.

**Other Theme Keyword(s)**
Other words or phrases used to describe the cultural and demographic content covered by the data set.

**Place Keyword(s)**
The geographic name(s) of the location covered by a data set.

**Attribute Accuracy Report**
A description of the accuracy of the information in the dataset, including the identification of any tests used, testing methodology and results obtained.

**Logical Consistency Report**
An explanation and assessment relative to the fidelity of the line work, attributes and/or relationships in the data set.

**Completeness Report**
Information about omissions, selection criteria, generalization, definitions used, and other rules used to derive the data set.

**Entity & Attribute Overview**
A brief summary of the information contained in the data set, such as the number of fields and records. This overview may also provide information related to the attribute collection and maintenance and overall quality.

**Entity & Attribute Details and Citations**
Reference to the complete description of the entity types, attributes and attribute values for the data set, such as database field names, characteristics and valid values.

**Horizontal Positional Accuracy**
An estimate of the accuracy of the horizontal positions of the spatial objects.

**Vertical Positional Accuracy**
An estimate of the accuracy of the vertical positions in the data set.

**Process Description**
An explanation of the event and related parameters or tolerances, which constructed the dataset.

**Process Date**
The date when the event was completed.

**Spatial Data Type**
The system of objects used to represent space 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.

**Horizontal Datum**
The identification given to the horizontal reference system used for defining the coordinates of points.

**Data Set Name as Known by Distributor**
The identifier by which the distribution organization knows the data set.

**Limits on Data Accessibility**
Restrictions on or legal prerequisites for accessing the data set. These may include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the data set.

***NOTE***: The types of constraints intended are those applied to ensure rights of privacy or intellectual property, and any other special restrictions, limitations, or warranties on obtaining the information resources, or its component products.

**Limits on Use of Data**
Restrictions and legal prerequisites for using the data set after access is granted. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the data set. Choices are:

***NOTE***: The types of constraints intended are those applied to ensure rights of privacy or intellectual property, and any other special restrictions, limitations, or warranties on using the information resources, or its component products.

**Liability Held by Distributor**
Statement of the liability (or lack thereof) assumed by the distribution agency. Include any disclaimers which may apply and which have not already been included under "Limits on data accessibility" or "Limits on use of data".

**Native Data Set Environment**
A description of the data set in the producer's processing environment, including items such as the name of the software (including version), the computer operating system, file name (including host-, path-, and filenames), and the data set size.

**Digital Transfer Format Name**
The name of the data transfer format.  The format in which the data is available.

**Distribution Organization**
Provide the name of the organization from which the data set may be obtained. This will generally be the same agency that is listed as the "Originator" of the data set.

**Distribution Contact Position/Person**
Provide the name and/or title of the individual to contact in order to obtain the data set.

**Address Type**
Indicate whether the address information that follows is the:

- mailing address,
- physical address, or
- mailing and physical address of the distribution contact person.

**Distribution Address**

**Address:** The street address for the distribution contact person.
**City:** The city of the distribution contact person.
**State or Province:** The state or province of the distribution contact person.
**Postal Code:** The ZIP code or other postal code of the distribution contact person.
**Country:** The country of the address for the distribution contact person.

**Phone**
The telephone number for the distribution contact person.

**Fax**
Fax number of the distribution contact person.

**E-mail**
E-mail address for the distribution contact person.

**Fees**
The fees and terms for retrieving the data set. If fee schedules are posted on the organization's web site, the URL may be included here.

**Region-wide Transportation GIS Project Design and File Architecture**
***Volume III – Technical Recommendations***

**Publication No.:** 02016

**Date Published:** March 2002

**Geographic Area Covered:** Delaware Valley Region

**Key Words:** Linear Referencing System (LRS), Transportation, Geographic Information System (GIS), DVRPC Member Governments, Operating Agencies, Database, Attributes, Data Dictionary, Metadata, Standards, Data Maintenance, Conflation, Address Geocoding,

**ABSTRACT**
The primary purpose of this project is to assure that DVRPC, it's member city and county governments, and transportation operating agencies have a GIS and data files that can be developed and shared with each other to facilitate better transportation planning analysis and decision-making. This report, divided into five volumes, serves as the foundation to establish the operational framework for these efforts.

Volume III – Technical Recommendations presents a series of specific recommendations for standards based upon the results of the needs assessment and research and prototyping efforts detailed in Volumes I and II. Standards are recommended for street address data, symbology, linework, database design, data dictionary, and metadata. The Common Linear Referencing System approach is recommended as the most efficient method of sharing transportation data and should be implemented in phases.

Delaware Valley Regional Planning Commission
8th Floor – The Bourse Building
111 South Independence Mall East
Philadelphia, PA 19102-2582

| | |
|---|---|
| Phone: | 215-592-1800 |
| Fax: | 215-592-9125 |
| Internet: | www.dvrpc.org |

| | |
|---|---|
| Staff contact: | Michael Ontko |
| Phone: | 215-238-2824 |
| Email: | montko@dvrpc.org |