VOLUME II - TECHNOLOGY VISION AND IMPLEMENTATION OPTIONS

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

(This page is intentionally blank)

Preface

Basis For Research

Up until this point, the needs assessment process has been working towards an understanding of the basic status of GIS for transportation planning throughout the region served by DVRPC. This process has not only assessed the status, accuracy and maintenance of the centerline files but also the types of transportation data that are being collected against the centerline.

Detailed information has been gathered pertaining to the methods being employed to place or reference that transportation data along the centerline. The methods of how that data is being placed along the centerline are termed a linear referencing method. Route/Milepost, Street Addressing, Offset from Intersections and Route/x,y are all examples of linear referencing methods.

Basic Questions

Now come the most basic questions faced in this project:

- How can the organizations that have been identified as being in the lower stages of GIS-T development best work their way up to the higher stages and play a more active role in the regional GIS-T environment?
- How can addressed-based data and data models that best support the geocoding of these data be effectively and consistently exploited for GIS-T applications?
- How can the different agencies that collect transportation data share that data amongst themselves and others without forcing a single regional linear referencing method or a single, regional centerline file?
- And, by what means can such a concept be proven beyond the realm of merely writing about it?

Demonstrations and Prototypes

The answers to these questions can best be found through the use of a prototype. By using actual datasets acquired from member organizations through the interview process, it will be possible to develop prototypical, or test, databases to evaluate and validate various scenarios. At this time, the following scenarios or models are anticipated:

1. ***Centerline Development Options-*** These are various centerline development scenarios for organizations that do not currently actively use and maintain a centerline. The process that has been employed examines a variety of alternatives that may be available and evaluates how each of these alternatives will aid these organizations in moving towards a higher stage of GIS development to support transportation planning.

2. ***The Coordinate, Route Model*** – This model will test the concept of establishing a unique identifier for each road segment and using this unique identifier as a key for sharing data among various

organizations. This scenario will also test the validity of using the unique identifier along with x,y coordinate information in defining the location and spatial extent of various phenomena.

3. ***The Common LRS Model*** – This model will test the concept of establishing a unique identifier for each road segment and using this unique identifier along with milepost measurements along the segment for sharing data among various organizations.

4. ***The National Spatial Data Infrastructure (NSDI) Transportation Framework Model*** – This model will apply the new NSDI transportation framework and will perform the same functions as SRI prototype except the data to be loaded will have native LRS references.

Conclusion

The building and execution of these various demonstrations and prototypes will serve a number of valuable purposes for this project:

- They will provide a solid, independent and practical foundation for the various technical recommendations that are required.
- They will provide a solid, independent and practical foundation for the implementation plans that will be developed for each member organization.
- They will help to establish reasonable estimates of implementation costs, resource requirements and schedules—important factors for the member organizations and DVRPC to consider as they move forward.
- They will help to enable DVRPC and its member organizations to move to the next levels of GIS-T and its related technologies.

The ability to prove concepts through the use of demonstrations and prototypes is a proven technique. By applying this approach to this project, a wealth of information can be generated and the project goals can be realized.

Chapter I -The Vision

I-1.0 Introduction

This document describes an approach that Delaware Valley Regional Planning Commission (DVRPC) can take as they move towards a collaborative data-sharing environment for its member agencies. This approach is designed to provide a range of solutions that represent different types of data transformation methodologies, and different levels of technological and human resource investments. Individual agencies can elect to adopt a methodology and a level of technology investment appropriate to their needs, and are provided a series of technology migration paths that they can follow over time to best suit their internal and DVRPC's needs. To provide context, the technical approaches of two prototype applications (Coordinate Route and Extended NSDI) and a demonstration (Common LRS) that map along the range of outlined solutions will also be described in detail in the latter sections of this document.
The adopted approach is based on the following assumptions:

- Different member agencies will elect to invest differing levels of technology and human resources.
- Each agency, either on their own or facilitated by DVRPC, will want to achieve the highest levels of data accuracy, stability, and systems integration to meet their needs. However, they will want to migrate to higher-order technologies over time to increase some measure of effectiveness.
- Agencies will adopt newer technologies at different rates. However, some agencies may not migrate at all, or may migrate at a much slower rate than others due to internal constraints.
- The number and size of the data sets that are shared among member agencies will increase over time.
- DVRPC can provide a technology transfer service role for its member agencies.
- This approach facilitates the maintenance of data.

These assumptions are important for a number of reasons. The first assumption requires that any feasible solution must be able to accommodate member agencies at different levels of technology. In addition, each agency within DVRPC could be at a different level of technology at any point in time. Further, any solution cannot make assumptions about the geographical size of the agency and associated level of technology or sophistication. In fact, it is likely that the largest agencies in terms of geography will have the highest cost of adoption and, therefore, could be expected to lag behind agencies that have less data to manage.

The second assumption predicates that all agencies would like to achieve a technological level that provides the highest degree of data stability, accuracy, and systems integration that meets their immediate needs and level of investment. However, it is also safe to assume that some, or all, of the member agencies will not want to incur additional costs associated with achieving a higher level of technology than meets their current needs. As their missions change, and as the benefits of data collaboration begin to be realized, the assumption is made that agencies will want to participate at a higher level and will want to migrate to a more sophisticated technology.

The third assumption reflects the notion that different agencies have different internal missions, and that their internal political process will affect the rate at which they adopt changing technologies.

The fourth assumption is a general statement about GIS and data environments as a whole. The importance of this assumption is that any solution must be able to scale over time, and effectively meet the challenges associated with high rates of growth in the amount of data that is shared.

The last assumption reflects DVRPC's role as an umbrella organization, and assumes that DVRPC is willing to mitigate the costs incurred by a member agency adopting a technology solution. DVRPC would do this by providing technology transfer services, and facilitating the data investments that are necessary to achieve a basic level of collaboration technology across all of its member agencies.

These assumptions can be reflected as a series of requirements that the proposed solution should address:

- Agencies must be able to participate at any supported level of technology.
- Technology solutions must remain viable over time.
- Agencies should be able to select a transformation methodology that meets their immediate needs, and stay with that methodology until they choose to migrate to another.
- Agencies should be able to migrate to more sophisticated levels of technology and different transformation methodologies over time.
- DVRPC should be able to provide technology transfer services at any technology solution that they choose to support for their member agencies.
- Data will be maintained in an accurate, complete and timely manner.

To these ends, a technology ramp model was adopted whereby member agencies can elect to participate in the collaborative environment using a transformation methodology at a level of technology that meets their immediate needs, but allows them to achieve higher levels of integration over time at a pace that suits their collaborative goals.

I-2.0 Technology Ramp

The technology ramp describes a matrix of data transformation technologies that can be used to perform data-sharing workflows between DVRPC member agencies. The matrix is composed of four rows, which represent four discrete data transformation methods: geometric, LRS, NSDI, and NCHRP 2027. The columns of the matrix represent seven levels of investment in the network data that are required to support a particular transformation solution. The cells of the matrix that have entries represent a valid combination (point solution) of transformation methodology (row) against a level of investment in the data environment (columns), and together represent a data sharing technology. The "sweet-spot" represents those combinations of technologies that provide a reasonable solution for agencies at different times and technology investment.

If an agency were to adopt a technology, the horizontal line extending to the right from that technology illustrates the notion that the selected solution is still usable through time and as the investment in the network data changes. For example, an organization electing to participate using a geometric transformation and a topological network could continue to use that level of technology even if other organizations, or DVRPC as a whole, invest in an NSDI or Extended NSDI framework model.

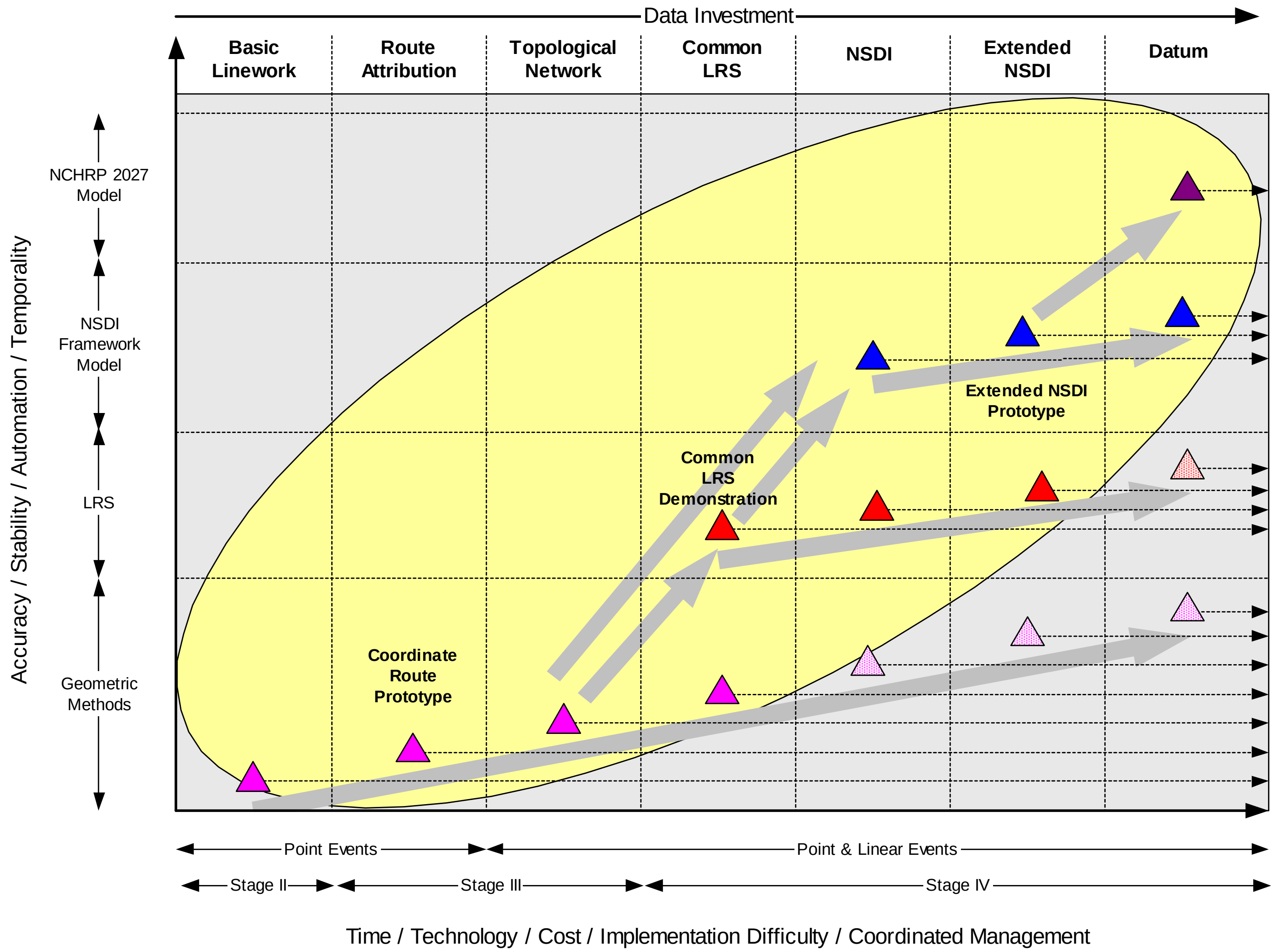

Figure I-1 – The Technology Ramp

In Figure I-1, the larger gray lines between matrix elements represent technology migration paths (refer to the figure on the previous page). Those within the same transformation methodology (row) indicate incremental migrations associated with investing more in the data and software associated with transformations. Those migration paths between rows are associated with changing the underlying transformation method that is used by an organization, and represent a more significant investment in terms of data, software, human resources, and participation framework.

Incremental changes provide a positive benefit for an agency by increasing the accuracy and/or stability of the underlying transformation method. Changes in the transformation models provide a basis for much increased accuracy and stability, but also allow higher degrees of automation and handling of temporal data to be achieved.

I-2.1 Technology Ramp Axes

The axes of the technology ramp depict the increasing levels of relative complexity and investment for each of a variety of components of the transformation methodologies (vertical) and the levels of data investment (horizontal).

I-2.1.1 Horizontal Axis

The horizontal axis reflects the expected increase in some time, technology, cost, engineering difficulty, and collaboration required for either a participating agency or for DVRPC as a centralizing agency to adopt a technology. These factors constitute the components of the data investment that is required in order to implement a data transformation solution.

I-2.1.1.1 Time

This is the relative amount of time that is required to implement a technology. This would be the time invested by an organization before a data transformation solution was available. On the left, the Basic Line Work solution could be implemented quickly, whereas a NSDI or NCHRP 2027 model would take much longer to create and manage the spatial and other associated data structures necessary to support the models.

I-2.1.1.2 Technology

This is the relative technological sophistication (software, hardware, know-how, show-how) of a solution. On the left are relatively low technological solutions that require minimal hardware and software and use basic spatial data. The solutions increase in technological sophistication to the right. The technology dimension in some ways reflects the “total cost of ownership” of the underlying data requirements.

I-2.1.1.3 Cost

This is the expected cost of implementing a system that can sustain the utility of the technology. This notion has several interpretations that can be made:

- The cost of preparing a data set to accommodate the requirements of the transformation method,
- The software and hardware costs associated with a transformation method,
- The investment costs associated with training users and administrator personnel,
- The management costs of the solution, or
- An overall system cost that accommodates any or all of the above.

I-2.1.1.4 Implementation Difficulty

This dimension reflects the difficulty in engineering a solution for a single participant agency (legacy systems, workflows, etc.), or the difficulty in engineering a framework for DVRPC to accommodate an organization-wide transformation solution. The least difficult methods are geometric based transformations and can be performed in an ad-hoc manner on simplistic clients, while the most difficult transformation methodologies require sophisticated multi-tier systems and a large amount of supporting business rules and maintenance workflows.

I-2.1.5 Coordinated Management

This dimension reflects the degree of centralization necessary to develop and administer a solution across all the DVRPC participating agencies. The low-technology solutions use ad-hoc workflows and can exist as decentralized (participating agency-centric) solutions. In contrast, the higher-order transformation methodologies of Extended NSDI and NCHRP 2027 are organization-centric and require a strongly centralized management with rigorous change, update, and versioning of the base network. The 2027 model requires total centralized management of the system.

I-2.1.2 Vertical Axis

The vertical axis reflects the expected increase in the quality of solution based on the accuracy, stability, degree of automation, or ability to manage temporally indexed data for a technology. These factors constitute the components of the various transformation methods that may be applied by a particular agency.

I-2.1.2.1 Accuracy

As the type of transformation model changes, the accuracy of the results can be controlled. Geometric transformation are the least accurate, LRS transformations are more accurate. NCHRP 2027 provides the most accurate transformation methodology, as all distances in linear space are controlled in the system. Within a transformation methodology, investments in the quality of the data support marginal increases in accuracy and performance. For example, adding route identifiers to basic line work is an investment that allows higher degrees of accuracy to be realized by removing some of the ambiguity inherent in the basic geometric algorithm. Similarly, building a topological model allows linear events to be transformed.

I-2.1.2.2 Stability

The stability of a solution is important since projecting a data set from one network to another and back again should, in theory, provide the original data set. If the system has instabilities and error associated with its methods, then the results of a transformation would not be expected to be symmetrical. This can lead to issues in accuracy, etc. Geometric transformations are the least stable because they are based on different software algorithms (i.e., ArcView™, ArcInfo™, Oracle®, and GeoMedia® all have different snapping algorithms), and are greatly affected by the precision, scale, and projection of the data. Performing a snap of a point in one GIS versus another would be expected to provide different results, as they use different snapping algorithms and tolerances. Transformations in linear space are more stable, but as the degree of measurement control increases the stability of a solution also increases. Thus, it would be expected that a NCHRP 2027 model would provide the most stable solution as all distance is carefully controlled. In fact, the 2027 model is predicted on the notion that the stability of data transformations is of paramount importance.

I-2.1.2.3 Automation

This dimension captures the degree of automation that can be used to perform transformations of data sets between disparate agencies. The more rigorous transformation models that are based on well-defined data base schemas and less on ad-hoc user-workflows allow a greater degree of automation. This is important if a lot of data is to be shared, or if data is shared repeatedly (e.g., monthly or quarterly reporting requirements). Automating transformation procedures reduces the likelihood of human-induced error and increases the repeatability of a transformation process.

I-2.1.2.4 Temporality

Temporality captures the ability of a transformation model to manage temporal data. The higher-order transformation methods (NSDI and 20-27) provide different levels of support for temporally indexed data. The NCHRP 20-27 model has temporality as an organizing principle, while NSDI provides a basic state-model for organizing objects in time and network-space.

I-3.0 Transformation Methods

The term “transformation methods” refers to the ordered processes and procedures that are applied to convert the locational reference of a feature or event from one system to another. Examples include the calculation of an x,y coordinate pair in State Plane coordinates from a street address or the conversion of a location defined by a route-milepost-offset measurement to one defined by latitude-longitude. The adoption of a transformation method is critical to the success of any system designed to support the exchange of transportation-related spatial data. The transformation methods that have been applied to the development of the technology ramp are described as follows.

I-3.1 Geometric Transformation Methods

Geometric methods transform from one network to another using a common geometric reference. An event is transformed from its reference system (LRS, Street Address, GPS, etc.) to a coordinate location on a network. The coordinate is then projected to the “to” network using basic GIS “snapping” software. The transformation method requires that the “from” and “to” networks have a similar scale and projection system, and that they have a common “model” of the physical world.

The use of sophisticated GIS software can negate the need for a common projection system across all member agencies, as the event data could be re-projected either as part of the export process or as part of the import process. This re-projection would require the definition of metadata statements that describe the measurement units, projection systems, datum, and spheroid for the exported data. Using a sophisticated GIS, however, still would not alleviate the problems associated with network representations at different scales.

Adding route identifiers to the network would allow increased accuracy and remove the reliance on scale to a certain degree, as pre-selecting a set of features to snap against can control the snapping algorithms.

Adding network connectivity (topology) to the data would allow linear events to be moved between agencies.

I-3.2 Transformation Based Upon a Linear Referencing System (LRS)

Dynamic segmentation software would be used to map events specified as linear references between networks. All events would be referenced using a standard linear reference (e.g., route/offset), and every location on the network would need to have a unique unambiguous linear reference. Input events using non-linear referencing methods, such as GPS and street addresses, must be mapped to a linear reference before they can be transformed to another network. This is because the linear reference is the only common referencing scheme between the two disparate networks.

The networks that are sharing data should have similar scale so that the measurement errors (i.e., errors associated with interpolating an offset) are minimized. The “from” and “to” networks would need to be concurrent with respect to the attribution of the LRS, otherwise transformation errors would occur.

The LRS approach can make good use of commercial LRS/GIS software and, once the networks are attributed, would provide a stable and accurate transformation methodology.

I-3.3 Transformation Based Upon the NSDI Transportation Framework

The NSDI framework model provides a transformation methodology, which is stable and has some basic support for temporality. The transformation model is based on locations that are parametrically defined over transportation features and an equivalence of transportation features between disparate networks. Data is transformed between networks by converting the event into an NSDI (feature) reference, and then converting the NSDI reference through the equivalence table to create an NSDI reference for the “to” network. The resulting NSDI reference would then get mapped to a linear or address reference in the “to” network so that it can be displayed. The approach obviates the issues associated with scale and representation, as the NSDI framework provides facilities for equivalencing single and dual-aligned representations, and is fundamentally a topological model (not a spatial model).

The basic NSDI model assumes that a simple bi-directional relationship is made between two disparate networks. Using this model, four networks wanting to share data would have to develop and maintain six equivalence tables (assuming equivalence tables are symmetrical; if they are not, then twelve equivalence tables are required). The extended NSDI model provides a single master network against which all participating agencies develop a single symmetrical equivalence table.

Using this approach, the four agencies would have to maintain four, and not six (or twelve), equivalence tables. The master network would have to be a superset of all participating network representations but could be modeled easily using a commercial base map (GDT, NavTech for example).

Any topological network can be converted into an NSDI network very easily using customized computer programs. However, there is a cost associated with creating and maintaining a mapping (equivalence table) between the "from" and the "to" network. If all of the networks have the same LRS associated with them (i.e., upgraded from the common LRS), then the creation of the equivalence tables could be fully automated.

I-3.4 Transformation Based Upon NCHRP 20-27

The NCHRP 20-27(3) model is the cumulative result of ten years of research and design in Linear Referencing Systems, sponsored by the National Cooperative Highway Research Program and state programs. The model represents state-of-the-art network data management. The 20-27 model is discussed in a variety of reports and workshops by prominent researchers and practitioners

The 20-27 model provides a panacea solution for agencies with multiple networks and cartographic representations. It provides a stable framework from which data can be converted from one referencing system on one network to a referencing system in another disparate network. Central to the 20-27 solution is the definition of a stable linear datum, which forms the basis of the transformation model. The 20-27 conceptual framework is illustrated in the Figure I-2 diagram, which was taken from NCHRP Research Results Digest No. 218, September 1997.

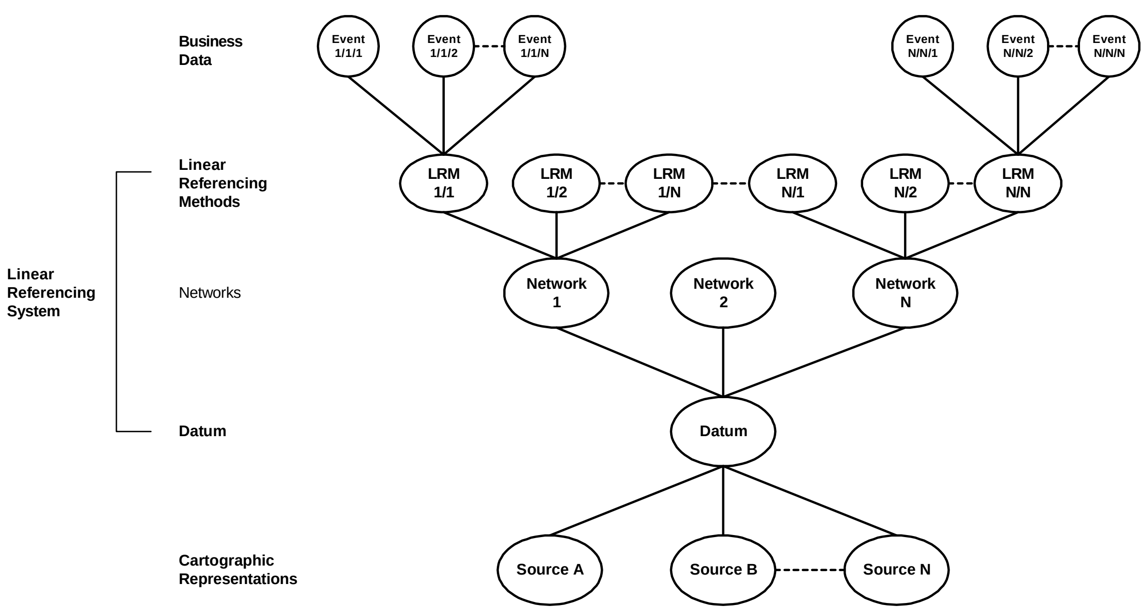

Figure I-2 – 20-27 Conceptual Framework

The datum is the organizing principle for the model. There is a single datum that is identified that would cover the entire DVRPC area. This datum would change over time, as changes to the transportation infrastructure are made. The changes are stored so that events located on the network can be located in the correct temporal context by "rolling back" the datum and associated data layers to reflect the time when the event occurred.

The datum represents the theoretical extent of linear space, and can be represented using a cartographic representation that meets the needs of the application. The 20-27 model provides for several cartographic representations, 1:24K, 1:100K, 1:1mill for example, which would be defined by the application context that the model is implemented within. In practice, however, the overhead of maintaining multiple cartographic representations outweighs the benefits, and so a single representation is provided.

Once a datum is defined, one or more networks can be identified over it. A network is defined in context of an application of the datum. A state DOT and a county street network would be expected to have different representations over the same datum, as the referent parts of the transportation system differ between the agencies. On top of the network layers, are layers of data that define the linear referencing systems including the definition of routes (traversals) and offsets that form a third layer of data abstraction. This is illustrated in the following diagram taken from the NCHRP 20-27(3) report.

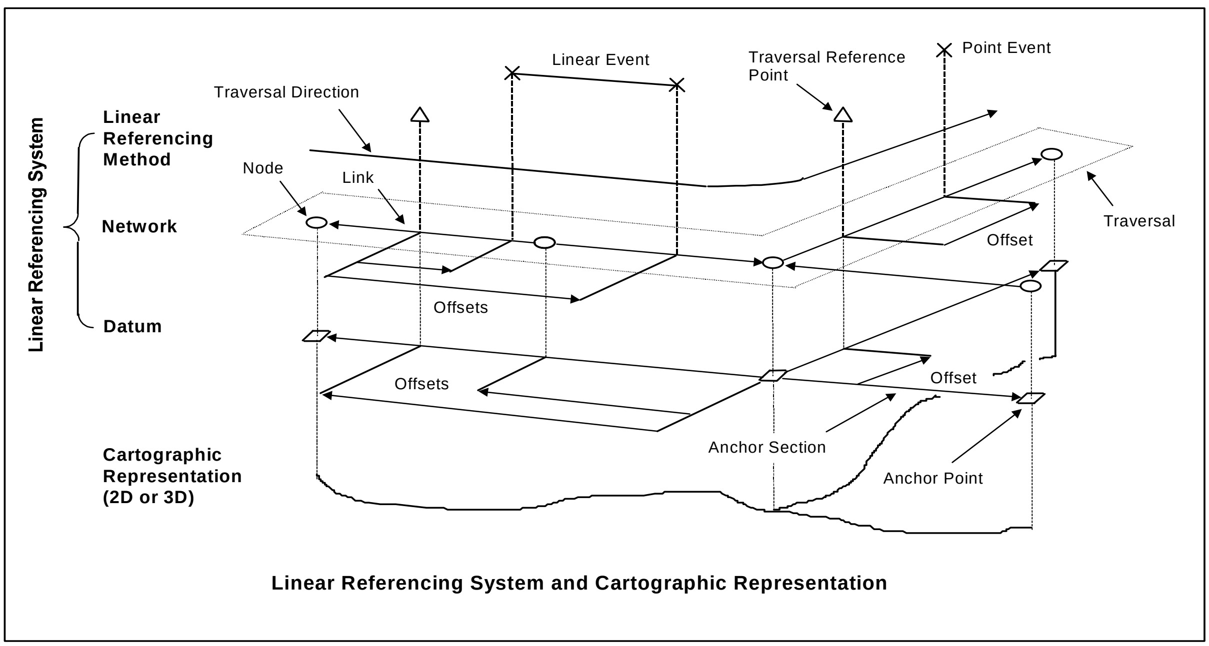

Figure I-3 – Linear Referencing System and Cartographic Representation

The 20-27(3) transformation model requires that all events be projected from one organization's LRS to the network, and then from the network to the datum. To place the event on another network, the datum reference is projected to the other network and then defined as an LRS in the other network. To maximize the stability of data over time, all events should be stored using datum references.

There are currently several implementations of the 20-27 model in development. Wisconsin DOT has been developing a modified model on top of an Arc/Info framework for a number of years.

Iowa DOT has a working pilot model developed on top of an Oracle database, and is currently starting a full implementation of the pilot across the entire state. These implementations are costly to develop and maintain. For example, the Iowa DOT 20-27 model requires a database schema with over 100 tables to support the run-time transformation engine.

I-4.0 Data Investments

As mentioned previously, the term "data investments" refers to the total investment of resources that would be required to implement a particular transformation method. This investment is measured in time, technology, cost implementation difficulty and coordinated management. The following sections describe relative data investment levels associated with the various types of GIS data models that could be used, on an increasing scale of complexity.

I-4.1 Basic Line Work

This requires all participants to be in the same coordinate system and have a relatively consistent network model. Disparate networks need to have the same scale and structure (ramps, treatment of dual alignments, etc.), but the data could potentially exist in any software system including CAD, GIS, or Oracle Spatial. Note that this will not work for linear events, as there is no underlying topological model.

Use of a sophisticated GIS negates the requirement of consistent projection systems because the data can be re-projected, as it is either imported to a network or exported for some other network. The projection system into which the data was exported would need to be known (maybe use FGDC Metadata Standard).

I-4.2 Route Attribution

A basic model utilizes a controlled snapping methodology whereby a point in one network space is projected to a selected set of features in another network. The use of a common route attribute in the two networks removes the dependency on scale as a control. In its basic form, it requires that both networks have a common projection system.

This will not work for linear events, as there is no underlying topological model.
Use of a sophisticated GIS negates the requirement of consistent projection systems because the data can be re-projected, as it is either imported to a network or exported for some other network. The projection system into which the data was exported would need to be known (maybe use FGDC Metadata Standard).

I-4.3 Topological Network

The topological network transformation model is built on top of the coordinate route transformation, and provides a connectivity model that allows linear features to be transformed between networks. The network model would have to be built for each participant network that wants to import linear features from other sources. Arc/Info, ArcView Network Analyst, GeoMedia Transportation Manger, and MGE/Network are the COTS products that could perform this functionality. MGE/SM could be used if there were NLFs created in the "to" network. All but MGE/SM require a coded algorithm to perform the work, and the algorithm would be different for each host environment. Custom code can be developed for these environments (including

Oracle Spatial), and a projection module can be created that would not require a COTS network manager product.

I-4.4 Common LRS

All networks participating in data sharing have to have a common LRS composed of common route identification and linear measurements.

I-4.5 NSDI

The NDSI methods focus on the adoption of a NSDI Framework model. This model provides equivalence between two disparate networks at a feature level (i.e., rows in a database). The basic NSDI model allows any network to be equivalenced to any other network that represents components (links and nodes) of the same physical system.

If there were three participating networks, each network would have to construct and maintain two equivalence tables (one for each other participating network), for a total of six equivalence tables. The NSDI model provides a control network termed the Master Network, which provides a centralized reference through which all transformations are made. Each participating network would have to develop a single equivalence table, for a total of three equivalence tables for the three participating networks.

To perform a transformation, there must be some method available that will take a linear reference and calculate a feature-based reference (what is termed a network reference). This network reference can then be used to derive the NSDI reference for an event. The events would be transmitted to another agency as a set of NSDI references. The receiving “to” agency would need the NSDI equivalence table that maps the “from” network to the “to” network, and can then translate the NSDI references from the events into new NSDI references associated with the “to” network. Once this conversion is performed, the NSDI references can then be mapped to a local feature-based (network) reference, and eventually to a geometric reference (for mapping) or a linear reference.

I-4.6 Extended NSDI

This is an extension of the basic NSDI model that provides a master network against which all transformations are performed. It provides a very stable and version-able system.

The extension to the basic NSDI is the definition of the master network, and the extension of the equivalence table to support partial-to-partial mappings of linear features.

The Master Network would have to be an investment by the agency charged with its maintenance, such as a state DOT, but the investment reduces the cost associated with each participating agency because then need only construct and maintain a single equivalence table.

Equivalence tables could be managed by the responsible agency for the participating agencies, or managed independently by each participating agency and then posted to the responsible agency.

I-4.7 Datum

A datum is composed of anchor sections and anchor points that together provide an exhaustive representation of the transportation networks. The master network from an Extended NSDI model could be used to provide the basic datum, but DVRPC would have to provide significant resources to validate and maintain the system over time, as it becomes the centralized component against which all transformations are developed.

Participating agencies would have to maintain a transport layer and associated location network structures and referencing methods that reflect their needs. A single centralized database is the most reasonable model for storing the large numbers of tables and volume of data that would have to be collected and managed, but a distributed database could work.

I-5.0 Prototypes and Demonstration Overview

Three prototype applications and a demonstration technology were created by JMT/EnterInfo/TransDecisions. for DVRPC to investigate and illustrate three different data sharing methodologies for member agencies. The prototypes and demonstration include:

- Stage I/II Centerline Development Options
- Coordinate Route Prototype
- Common LRS Demonstration
- Extended NSDI Prototype

Each of these are described in detail in the subsequent chapters.

Chapter II - Centerline Development Options

This demonstration focused on establishing a street centerline and related attribute database, for the Stage I & II agencies in the DVRPC region, which can be utilized for transportation planning. These agencies either do not currently own a street centerline dataset, or the current dataset does not include the attributes necessary for transportation planning. This prototype will investigate and quantify available centerline and attribution solutions, estimating implementation costs and emphasizing a reasonable return on investment. The investigation will include internal development, region-wide data sharing or possibly even purchase from a data vendor. Following the investigation, the JMT/EnterInfo Team will recommend the best solutions based on cost (dollars & resources), accuracy (spatial & attribute), compatibility and usability.

II-1.0 Investigation Summaries

Based on the specific needs of the agencies, the JMT/EnterInfo Team has investigated the following possible solutions.

II-1.1 In-House Development

This solution involves each specific agency developing their own centerlines based on a set of region-wide standards for centerline development. The solution is dependent upon the availability of the proper resources, including: personnel, hardware, software and time. Although it may appear to be more inexpensive to complete the development "in-house," the required resources may not be available rendering this solution unattainable. Our investigation assumes the availability of the required resources and the man-hour estimates are based on that fact. This estimate is based on the development of 25,000 segments. Estimates for any participating agency will be scaled from this value.

Other issues involved with this solution include the complexity involved with discerning street centerlines from an aerial photograph as well as the complexities involved with conflation. Interpreting aerial photography and determining centerlines requires formal training in order to obtain high accuracy, which may not be available. Conflation is an extremely tedious task. Although there are methods for automating this procedure, those methods usually result in less than half of the segments being conflated. This forces the use of manual conflation. Manual conflation requires thorough knowledge of the centerlines in question. The user must have previous knowledge of the centerlines, or several different source materials in order to accurately determine which source centerline matches with which target centerline.

Positive aspects of this solution include: minimal cost and maximum previous knowledge. The minimum cost is due solely to the "in-house" development. Again this assumes that the personnel, hardware and software are already in place and all have the ability to perform the required tasks. If the agency does not have the resources in place the cost will rise exponentially. In addition, since the line work is being developed and conflated by "in-house" personnel, we can assume they have a wealth of knowledge about the centerlines. This will make development and conflation much easier and more accurate, unless these personnel are interns or other temporary employees. (Rating: Cost = 5 – Excellent; Accuracy = 5 – Excellent)

Since the data is developed completely "in-house," the Compatibility with other agency datasets is very high for this solution. The Agency can decide on the data model to employ, the accuracy and extent of the data that will be the best fit with their current datasets. Although, the Agency should decide on the most accurate and extensive coverage and dataset available, this may not be the best solution with regards to fitting their datasets. For example, the centerline data that is developed may be more accurate then their existing data causing a problem during overlay analysis and cartographic production. (Rating: Compatibility = 5 – Excellent)

Again, since the data is developed completely "in-house," the Usability for Transportation Planning factor is very high for this solution. The data model we have proposed includes tables and columns that can be directly utilized for Transportation Planning. If the agency follows this model, they can design their specific database with the ability to include the attributes as well and easily use their new centerline dataset for Transportation Planning efforts. (Rating: Usability = 5 – Excellent)

The Maintenance Requirements for this solution are stringent. Unless the data is maintained carefully and meticulously it will become a snapshot in time and obsolete rather quickly. Since the data was developed "in-house" it requires that "in-house" personnel maintain it. This also requires that a strict maintenance policy is in place and is adhered to in a timely manner. This policy must include all departments that are involved during each maintenance step throughout the life cycle of a road. In addition, trained personnel must be assigned to this process. Since we are assuming that the Internal Development project will include provisions for maintenance this parameter can be rated very highly. (Rating: Maintenance = 5 – Excellent)

II-1.2 Contract Development

This solution is very similar to internal development in terms of personnel requirements. Hardware and Software requirements will not be an additional cost for this solution since the responsibility lies with the consultant. Again, costs in this estimate are only concerned with man-hours and is based on 25,000 segments. This solution involves each agency hiring contractors/consultants to develop the centerlines.

Issues involved with this solution are similar to those involved with "in-house" development. They include the complexity involved with discerning street centerlines from an aerial photograph as well as the complexities involved with conflation. In this case, the consultants must be experts in aerial photograph interpretation and development of planimetrics. Therefore, the data that is developed in this solution will have the highest level of accuracy. Regardless of the amount of training one possesses, conflation is still an extremely tedious task. Since the consultant will also be required to perform manual conflation procedures, the accuracy will suffer due to an overall lack of thorough knowledge of the centerlines in question. In this case having several different source materials available will be a viable substitute in order to accurately determine which source centerline matches with which target centerline. (Rating: Accuracy = 4 – Good)

The man-hour cost for this solution should be similar to that of the "in-house" cost. The difference in these two solutions, with respect to cost, will be the rates that are charged for consultants versus those charged for "in-house" personnel. In addition to the contract cost, "in-house" staff may require training and will need to monitor the progress and direction of the project. They will also need to provide quality control and other on-going maintenance costs. With this in mind, the cost for contract development will be much higher than the cost for internal development. (Rating: Cost = 1 – Very Poor)

Although a consultant develops the data, it is still under the direction of "in-house" personnel. Thus, the Compatibility factor is very high for this solution. The Agency can decide on the data model to employ, the accuracy and extent of the data that will be the best fit with their current datasets. The same pitfalls are associated with Contract development as Agency development as far as compatibility. (Rating: Compatibility = 5 – Excellent)

Again, since the data is developed under the direction of "in-house" personnel, the Usability for Transportation Planning factor is very high for this solution. The data model is the foundation for the newly developed dataset and thus has the ability to support transportation planning. (Rating: Usability = 5 – Excellent)

As with In House development, the Maintenance Requirements for this solution are stringent. Again, unless the data is maintained carefully and meticulously it will become a snapshot in time and obsolete rather quickly. This solution requires that "in-house" personnel maintain it. Therefore, this solution has the same Maintenance Requirements as the Agency Development solution. Since we are assuming that the Contract Development project will include provisions for maintenance this parameter can be rated very highly. (Rating: Maintenance = 5 – Excellent)

II-1.3 Centerline Borrowing

Through the Linework Assessment the JMT/EnterInfo Team has established stages for centerline development for all of the member agencies. This is very useful in deciding which agencies to borrow data from. It is logical to attempt to establish a sharing agreement with the agency that has obtained the highest rating through our analysis. Although, in some instances this rating may be put aside due to some factors or characteristics carrying more weight than others. For example, an Agency may have the most accurate centerline database that is maintained aggressively but it may not have the local roads that are important for the borrowing agency. With that in mind, we see centerline borrowing as a viable solution for centerline development. Which agency is picked to borrow from depends directly on the needs and wants of the agency attempting to "borrow" the data. Therefore, we cannot make a general recommendation regarding who should borrow from whom.

The cost for this solution is minimal, most likely only including the media costs for reproducing the data. (Rating: Cost = 5 – Excellent)

The accuracy of the data depends on the agency from which you would like to borrow. Since the lending agencies are essentially limited to PennDOT or NJDOT, we know the accuracy of the dataset is relatively high, although, the coverage may not be as complete as the borrowing agency would like. Due to this fact, accuracy for this solution does not get the highest possible rating. (Rating: Accuracy = 4 – Good)

Again, the Compatibility Factor for this solution depends on the agency from which the data is borrowed. The accuracy of the PennDOT and NJDOT data is known; therefore, the compatibility will depend directly on the agency's datasets. Since we can assume that borrowing would not be a viable solution unless the compatibility is reasonably high, we have rated this solution as average. (Rating: Compatibility = 3 – Average)

The Usability Factor for this solution depends on the agency from which you would like to borrow. PennDOT and NJDOT use their data for Transportation Planning efforts on a daily basis; therefore, the usability will be very high for any agency that would like to obtain their data from either DOT. Since, as was mentioned before, the coverage may not be as complete as the agency would like, this solution does not receive the highest rating for usability. (Rating: Usability = 4 – Good)

Maintenance may be an issue. If maintenance/edits are performed in-house following a one-time only borrowing of the centerline, then the effort is identical to that under the In-House option. However, if the centerline maintenance is left to the lending agency, then there will be an arduous maintenance procedure. The maintenance procedure will require obtaining new "cuts" of the data on a regular basis. Any edits the agency makes will have to be recreated every time a new version of the data is received. This data must then be reloaded into your system. This process will be time-consuming and totally reliant on the lending agency's maintenance schedule. Since we have assumed lending agency maintenance, we have given it a very poor rating. (Rating: Maintenance = 1 – Very Poor)

II-1.4 Purchase from Data Vendor

Although the specifics of this solution are directly dependent on the vendor (GDT, Navtech, or Tele-Atlas), we have grouped the ratings in our Comparison Matrix. In most cases this is adequate, but we have detailed the specific requirements and issues for each of three data vendors in Appendix A.

Depending on the number of users that are required, the purchase price can vary within a wide range of values as detailed in Appendix A. In general, for a small number of users, the cost for this solution is relatively low. As users are added the cost increases greatly and that is why we have given cost an overall poor rating for this solution. However, GDT offers a free dataset with its Community Update Program, which is described in Appendix A. (Rating: Cost = 3 – Average)

Accuracy is reasonable for all of the data vendors. They all utilize several different source datasets. However, the accuracy level that the data vendors deem as high is not nearly accurate enough for the local agencies. Due to this fact we have given Accuracy a low rating. (Rating: Accuracy = 2 – Poor)

Since compatibility is dependent directly on accuracy, it is an issue for this solution. Depending on the accuracy of the agency's current data, the purchased data most likely will not overlay properly and cause issues during analysis and cartographic production. (Rating: Compatibility = 2 – Poor)

If the agency intends to perform transportation planning efforts using Street Name and Address Range as the Linear Referencing Method then the Usability factor is very high. These vendor-supplied centerlines do not contain attribution that equates to a true linear referencing system. Therefore, if the agency would like to perform transportation planning using linearly referenced data (i.e. DOT datasets), then the usability is very low. Therefore we have given usability an average rating for this solution. (Rating: Usability = 3 – Average)

Maintenance is very poor for this solution. Although, all of the data vendors have a maintenance policy in place and do so in a set time frame, they will not maintain information and attributes that the agency adds to the data, including attributes like the Unique Identifier. This also raises the issue of life-cycle cost for the data. With this in mind we have given maintenance for this solution the lowest rating. (Rating: Maintenance = 1 – Very Poor)

II-2.0 Solution Comparison Matrix

The results of these investigations are tabulated the Comparison Matrix shown below, in Figure II-1.

II-2.1 Costs – The cost associated with developing centerline and attributes using each solution.

II-2.2 Accuracy – The accuracy associated with centerlines developed using each solution. The relative accuracy standards depend on quality of the data (accuracy of geometry as well as attributes) and extent of the data (area covered by the dataset).

II-2.3 Compatibility – The compatibility factor is based on spatial overlay conflicts with other agency datasets.

II-2.4 Usability – The usability factor is based on how usable the dataset will be for transportation planning efforts.

II-2.5 Maintenance – The maintenance factor depends on how difficult it will be to establish a maintenance plan and maintain the data and its relationship to any event datasets.

II-3.0 Data Model

Regardless of its source, the centerline will not be useable for transportation planning applications unless it is integrated into a logical GIS data model that is characterized by components that support the management and analysis of transportation planning data. Typically, these components include a linear referencing scheme, a route structure that supports dynamic segmentation based on point and linear event data, and the event data used as input to the dynamic segmentation process. An example of such a model is described in the following section.

II-3.1 Logical Data Model

The data model depicted in Figure II-2 includes each of the components described above.

Available Solutions / Solution Characteristics	Developed Internally	Developed Externally (By Consultant)	Centerline Borrowing	Purchased From Data Vendor
Cost	**5**	**1**	**5**	**3**
Accuracy	**5**	**4**	**4**	**2**
Compatibility	**5**	**5**	**3**	**2**
Usability	**5**	**5**	**4**	**3**
Maintenance	**5**	**5**	**1**	**1**
Total	**25**	**20**	**17**	**11**

Correlation	
5	Excellent
4	Good
3	Average
2	Poor
1	Very Poor

Figure II-1 Solution Comparison Matrix

II-3.1.1 Geometry

At the base of the model are the entities that represent the underlying network geometry of the GIS. Entities tied directly to the geometry are indicated by the blue boxes in the diagram. 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. In this model, nodes typically occur at the intersection of two or more centerline segments.

II-3.2 Linear Referencing

The entities represented by the yellow boxes are those tables that define 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

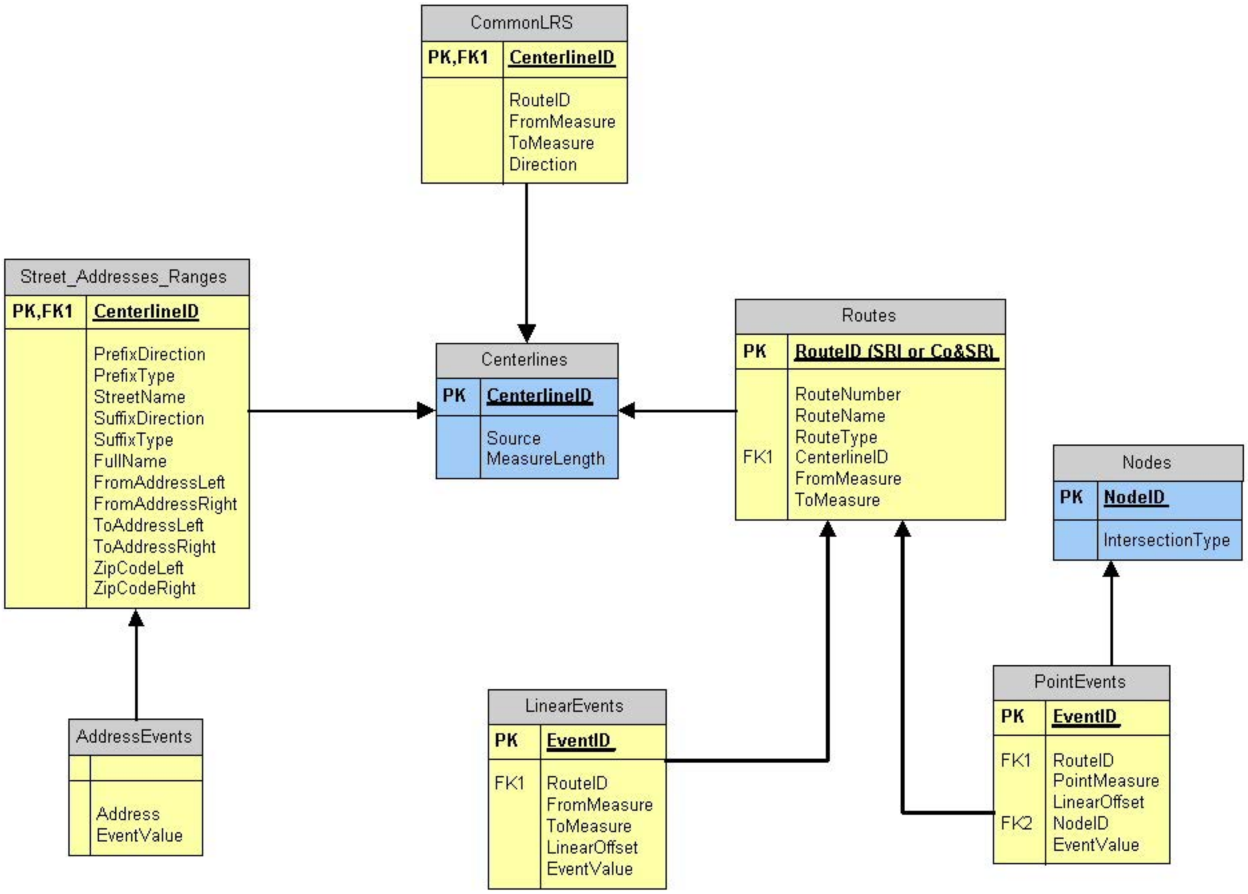

Figure II-2 Logical Data Model

" 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.

II-4.0 Demonstration Methodology

In order to test the aforementioned possible solutions a demonstration exercise was performed that tested the possibility of developing a centerline and adding attributes to the centerline. For this demonstration, a study area was defined consisting of the area surrounding City Line Avenue in Philadelphia. This study area was chosen due to the fact that the boundaries of various member agencies converge in the area, and there are several ramps and other complex features in the area making it an excellent sample set for the region. In order to develop the demonstration, the following procedures were employed:

1. Develop Centerlines from aerial photography interpretation
2. Conflate Street Address Info from the available Tele-Atlas data
3. Develop a Data Entry Tool Template for maintaining the centerline database.

Street Name and Address Range was chosen as the Linear Referencing Method for several reasons. Using this type of LRM facilitated the population of the centerline database with all of the necessary attributes to perform linear referencing procedure. In addition, choosing Street Name and Address Range provided the advantage of using the source data that was available.

Tele-Atlas was chosen due to its availability and cooperation. Tele-Atlas agreed to allow the use of their data for the demonstration. In the demonstration, the Tele-Atlas data emulates the type of data an agency might receive from their E-911 system and it is being used with that assumption. Although E-911 data is not always the most geographically accurate data, it usually contains the most accurate attributes for street names and address ranges.

II-4.1 Demonstration Procedures

In order to develop the prototype we went through the following procedures.

1. ***Create new street centerline feature class in a Personal Geodatabase***

    a. The new ArcGIS 8.1 tools were used for the initial creation of the new centerline file.

    b. Using ArcCatalog, we defined a personal geodatabase with a feature dataset to store the new centerline feature class.

    c. In ArcMap, we used the aerial photographs provided by DVRPC and the inherent digitizing tools to create the new centerlines. This process involved interpreting the aerial photos to determine where to correctly place the centerlines. For the prototype, only one rule was enforced during the process. The rule stated that, if the road was obviously divided by a large median then two centerline features were digitized to represent both sides of the road. No attributes are populated during this process. Other rules will definitely be necessary for the complete project, such as: collection of highway ramps, intersections and any other complications.

2. ***Convert street centerline feature class and source data to coverages***

    a. The conflation tools that were utilized in our prototype were developed in ArcInfo workstation. For that reason, it is necessary to convert both the new street centerline feature class and the source data to ArcInfo coverages.

3. ***Prepare new street centerline and source coverages for conflation***

    a. Both coverages need several steps of preparation before continuing to the conflation process. The following steps are crucial to the overall process.

        i. The source coverage will probably have pseudonodes that need to be removed. A pseudonode is a node where two, and only two, arcs intersect or a single arc that connects with itself. Pseudonodes can be caused from other data sets such as: police sectors, fire districts, county boundaries and so on. . A custom AML is used to remove these pseudonodes. The AML will replace the attributes in the to and from address fields with the minimum from address and the maximum to address between the two arcs separated by a pseudonode as well as delete all the pseudonodes.

ii. The target coverage might also have some pseudonodes (extraneous nodes) that need to be removed. This step may only require a simple deletion of the pseudonode if the dataset lacks any important attributes as in our case. In some cases, you may want to keep some important attributes, which means that the pseudonodes need to be removed with same procedure as the source data.

4. ***Conflating street centerline coverage (target) with the source coverage***

a. There are two major steps in the conflation process, automated conflation and manual conflation.

i. Automatic Conflation **-** The target coverage should first be conflated with the source by using the matchcover command in ArcInfo Workstation. Ideally, this automated process will conflate about 50 percent of the target centerlines. During the prototype, this process converted 11 percent of the centerlines. The process should be run several times with different match tolerance values until the best results are achieved. The best results would be achieved when the minimum number of multi matches and maximum number of actual matches are obtained.

ii. Manual Conflation **-** The next step is to manually conflate the centerlines. A custom menu tool for ArcEdit was used to complete this task. The tool allows you to manually select the source centerline and then select the appropriate matching target centerline. This will automatically transfer the attribute values to the target coverage. This process was repeated until all the matching centerlines were conflated. The process of transferring attribute values requires a good knowledge of the datasets due to user interpretation.

b. At this point, all the matching target centerlines have been completely conflated with the source attributes.

5. ***Convert target data back to a feature class with a personal geodatabase or to your target format***

a. The conflated target coverage is now converted back to a geodatabase feature class to take full advantage of the new ArcGIS model.

6. ***Set up domain values for specific fields in the geodatabase***

   a. Certain fields in the database may only require certain values. These values can be limited by setting up domains. Domains are a set of allowed values for a table column. Domains can be manually set up from the properties window of the geodatabase using ArcCatalog. These domain values make the maintenance\update tool a lot easier to use. They will provide a drop down window for all valid values for each field specified. For the prototype we created domain values for the following fields: [street type], [direction], [county], [zip code]. Domain tables can be created for any field in the database.

II-5.0 Recommended Solution

Based on above investigations and prototyping, the Team has determined that the most effective solution for developing a centerline dataset with attributes necessary for transportation planning for those agencies currently classified as Stage I or II is some combination of Agency Development, Contract Development and Data Vendor purchase.

In cases where the Agency has the resources and training necessary to develop and attribute the centerlines effectively, internal development is the best solution. This allows the Agency to take advantage of their inherent knowledge of the street network as stakeholders in the data. The quality of the data will be high and since the agency has developed the data internally the stake they hold is greater. Therefore the data will most likely be maintained more aggressively and most useful in the long run. Maintenance is the stumbling block for solutions other than the In-House development solution. Stakeholders are more likely to maintain a centerline that they have spent their time and money developing rather than a purchased centerline.

If the Agency does not have the “in-house” resources to develop and attribute the data effectively, then the next best solution would be to either hire a consultant or a team of consultants to perform the entire project or enter into an agreement with GDT through their community update program. Hiring a consultant will ensure the Agency owns the data at the completion of the project but it is expensive and it does not ensure maintenance. The Agency will be responsible for maintaining the data once it has been developed and this may become an issue, if the personnel is not trained and in place to perform this task. An agreement with GDT is free but it does not offer outright ownership nor does it ensure maintenance. Again, the Agency is responsible for maintenance through GDT and if the personnel is not in place maintenance will not occur.

II-6.0 Basic Implementation Plan

The basic components of a plan for implementing and maintaining the recommended solution are as follows: (All of these components may not be necessary depending on the specific agency)

II-6.1 Obtain Aerial Photography

Aerial Photography can be obtained through purchase, data sharing or hiring a contractor to develop from a new flight. The most viable option is sharing with DVRPC, since they have very accurate aerial photographs covering the entire region.

II-6.2 Develop Centerlines (Internal or Contract Development)

The following steps are recommended as a basic procedure for developing street centerlines.

1. Develop Rule Base for photo-interpretation and data collection. 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 defined as 1 centerline per road, 1 centerline per direction per road, 1 centerline per lane per direction per road, etc…) Other Rule Bases may include digitization direction, intersections, ramps and various other complex features.

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

3. Quality Assurance/Quality Control. A defined plan for QA/QC should be maintained throughout the development process. It is highly recommended that standards for maintaining topologic integrity be developed and adhered to. The QA/QC plan should eventually become part of your Agency's maintenance plan.

II-6.3 Attribute Centerlines (Internal or Contract Development)

The second step in the recommended centerline creation process includes the capture of attributes associated with each centerline segment. At a minimum, the appropriate Unique Route Identifier should be added to the centerlines. Depending on the availability of data, attributes can be obtained in either of the following methods:

II-6.3.1 Conflate from other datasets

This approach requires the development of a rule base for conflating the attribute data. This should help minimize the errors due to misinterpretation. It also establishes rules for dealing with special situations such as conflating attributes where the datasets do not have a one-to-one match.

II-6.3.2 Manual Input

This approach typically includes two principle methods, field collection, which requires the development of methodology and standards and the manual input of data from other data sources such as external databases, spreadsheets, and ASCII text files as well as paper documents.

II-6.3.3 Quality Assurance/Quality Control

The process of adding attributes to the captured centerlines also requires a program of quality assurance and control. It is necessary that the accuracy of the attributes is confirmed through verification with other data sources and also through field verifications where necessary.

II-6.4 Maintenance (Internal, Contract and Purchase from Data Vendor)

Once the centerlines and associated attributes have been captured and validated through the quality control process, it is essential that a program of ongoing data maintenance be put into place to ensure the continued validity and usefulness of the data.

II-6.4.1 Establish an Aggressive Maintenance Plan

This plan should detail which attributes are to be maintained, where the new data will come from and how often the data should be maintained. Wherever possible, it is important that the maintenance plan be integrated with other organizational business processes and work flows so that the program becomes an integral part of the day-to-day operations.

II-6.4.2 Establish the personnel that will be responsible for maintenance

Specific assignments of personnel to the data maintenance program should be made to ensure that the process is successful and experiences little or no interruptions.

Chapter III - Coordinate Route Model

III-1.0 Overview

This Coordinate Route technology provides agencies with the ability to use a combination of geometric location (x,y) and a unique road identifier to transfer information between multiple network representations. Using this technology, point event data can be “registered” to multiple agencies’ centerline data. The approach relies on having a unique identification for previously identified routes as an attribute of the network data which allows the same centerline to be found, even if that centerline was collected at a different map scale.

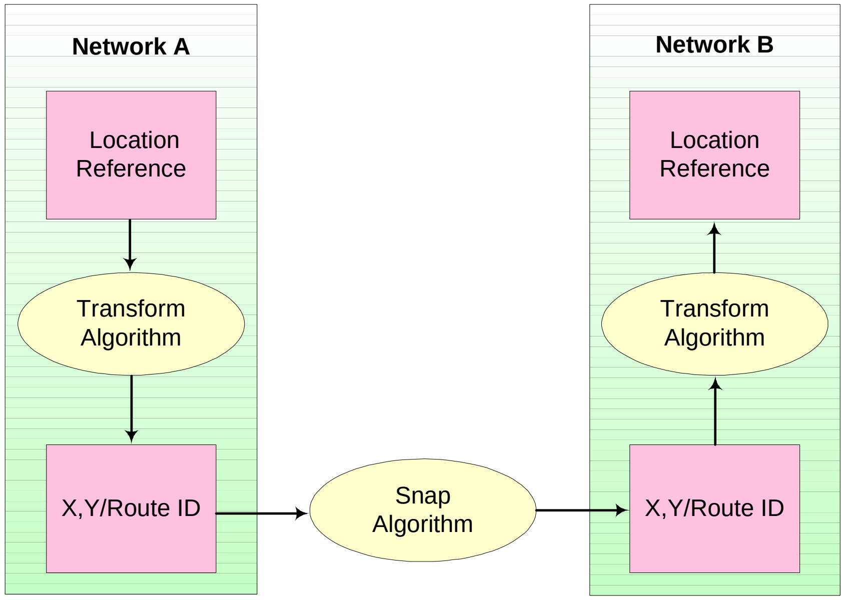

Figure III-1 - Overview of Coordinate Route Approach

The approach uses the following algorithm (refer to Figure III-1).

1. A location reference on Network A is transformed via a transformation algorithm to an (x,y) reference with an assigned route ID.
2. A snap algorithm is executed, which projects the coordinate on Network A to the closest matching route in Network B.
3. The resultant location in Network "B" is transformed via a transformation algorithm to a location reference.

These steps are illustrated in the following diagram.

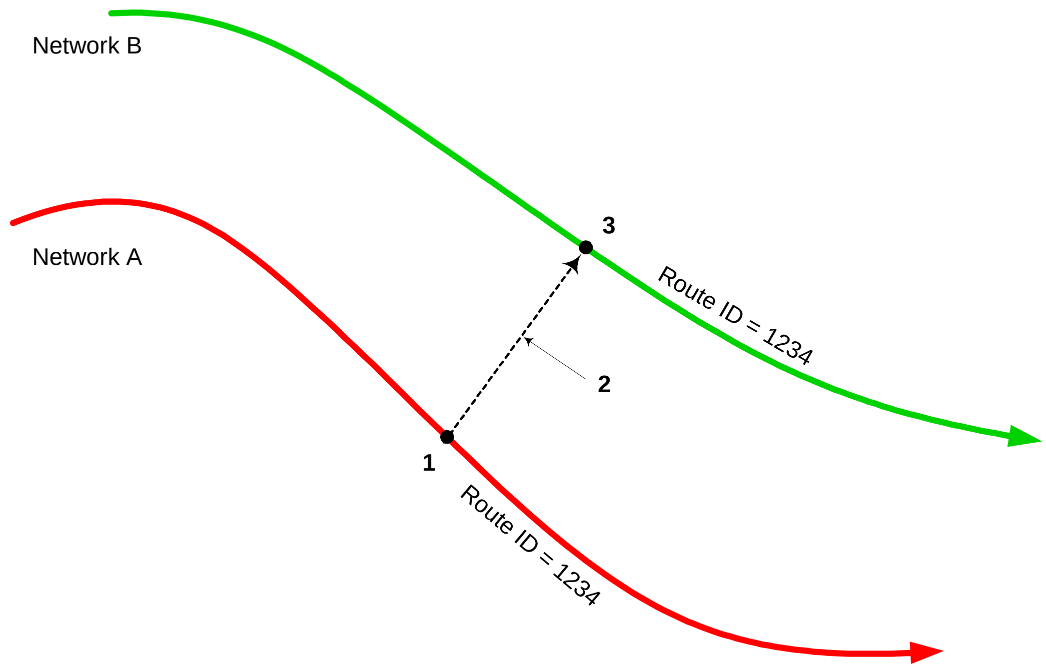

Figure III-2 - Coordinate Route Transformation Overview

III-2.0 Collaboration Architecture

The collaboration architecture relies on a distributed model that allows agencies to share event data with each other in an "ad-hoc" fashion (refer to Figure IV-3). The mechanisms used to share data between agencies are varied and may include: FTP, Web, Sneaker Net, or any other reasonable method of delivering a data set from one member agency to another. The data must include Route ID and X,Y. When the data is received at its destination, it is converted locally using the transformation algorithms to "map" the data onto the desired transportation network. Any additional workflows that are required to pre-format the data, such as projection conversion, must be determined and specified as a part of the collaboration process.

For data sets that are shared often or if formalized data sharing procedures are required, data clearing-houses can be set up in either a centralized or decentralized manner to allow member agencies to post data sets once and allow them to be used by any other member agency without burdening the provider agency. Data clearing-houses could be as simple as an FTP site or local web page.

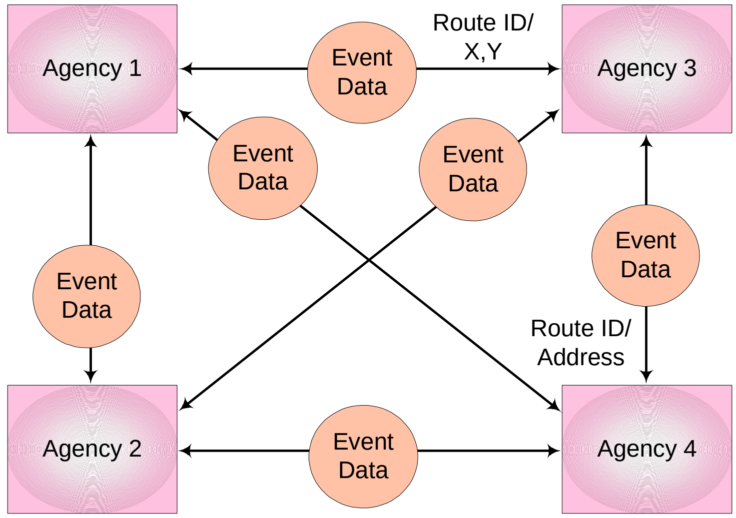

Figure III-3 - Coordinate Route Collaboration Architecture

III-3.0 Data Requirements

The data requirements for the Coordinate Route approach are:

- A source network.
- A set(s) of event data that is either linearly referenced or represented as x,y locations in a source network that are to be transformed to locations in a destination network.
- A destination network to transform and map the input event data against.
- A common set of route identifiers for the routes contained in the source and destination networks. In the prototype, TransDecisions used New Jersey DOT's 10-digit Standard Route Identifier (SRI) as a common numbering system. However, any common unique referencing scheme can be adopted.
- The source and destination networks typically require the use of a common geo-spatial reference framework. The sets of geometry for both networks should reflect the same scale factors, projection datum, and relevant attribution. Alternatively, appropriate GIS software packages can manage the conversion from one projection datum to another, which could be incorporated as a part of the workflow to manage cases where uncommon datums are present.

III-4.0 Prototype Implementation

The coverage area for this prototype is Burlington County, New Jersey as shown in Figure IV-4. County road data contains street address range data and the State roads contain New Jersey's Linear Referencing System (LRS). The prototype transforms point events defined by address information in the county network to common LRS-defined point events in the state route network, and vice versa.

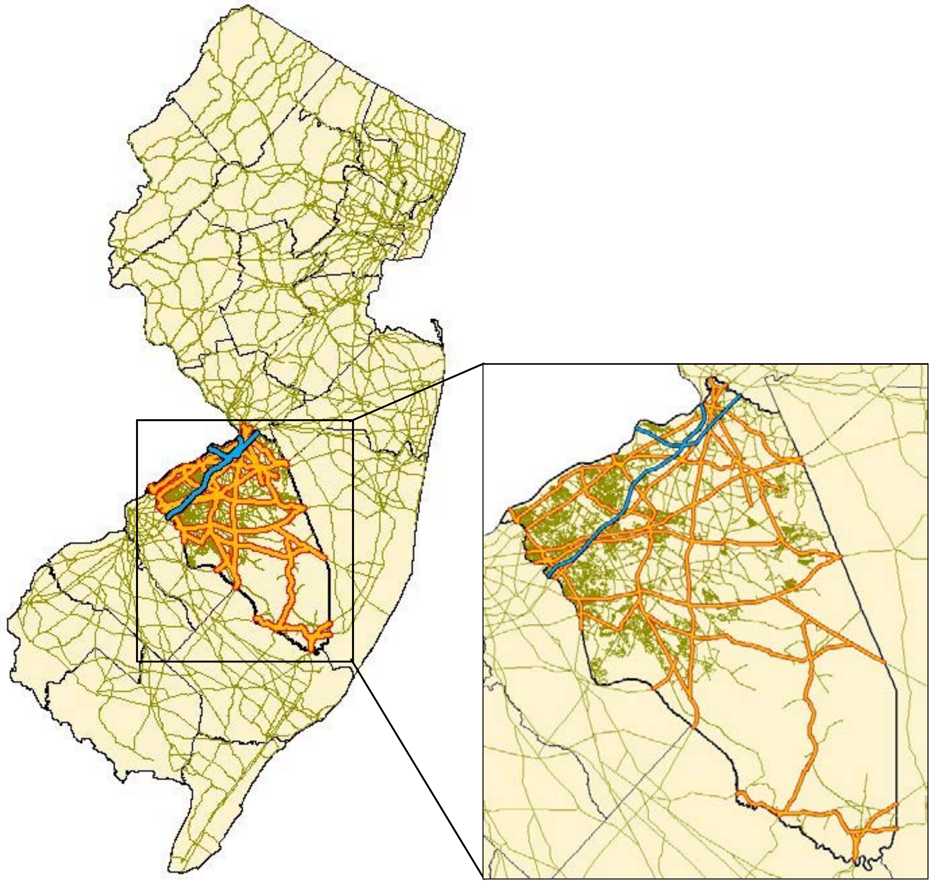

Figure III-4 - Area of Coverage (Burlington County, NJ)

III-5.0 Data Sets

The data sets selected for use in this prototype include:

- ***Burlington County Road Centerlines***
  This centerline data set was provided by Burlington County. It was hand collected by the county using a mobile GPS (approx 1:200 scale). This data is appropriate for use in the prototype because it contains road centerline data, street names, and SRI numbers.
- ***New Jersey DOT Road Centerline File***
  This centerline data file was provided by New Jersey DOT (NJDOT). This data is appropriate for use in the prototype because it includes SRI numbers and milepost information.
- ***NJDOT Burlington_Traffic_99 Event Table***
  This point event file was provided by NJDOT. It is used as an input data set to illustrate the state-to-local transformation of data.
- ***Burlington County Simulated Event File***
  This simulated point event file is used as an input data set to show the local-to-state transformation of data.

III-6.0 Transformation Overview

Geocoded location references on the Burlington County Street network are transformed to x,y coordinates/route ID and vice versa using a geocode algorithm. Linear Location References on the NJDOT Sate Route Network are transformed to x,y coordinate/route ID and vice versa using dynamic segmentation algorithms (see Figure III-5).

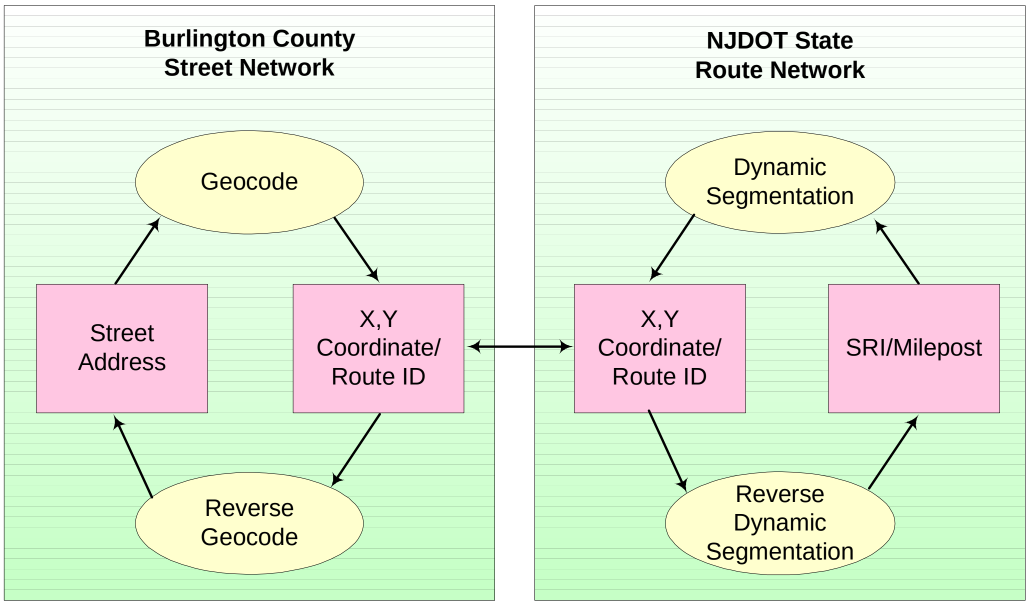

Figure III-5 – Transformations within a Network

III-6.1 Street Address to SRI/Milepost Transformation

Starting with an event table containing events referenced using street addresses on the Burlington County (source) network, the following steps are performed to transform the data to an linear refercnce defined using a State Route Identifier (SRI) /milepost on the New Jersey DOT (destination) network (see Figure III-6).

The following procedure is performed for each event that is mapped from the Burlington County network to the State DOT's network.

1. Using an address geocoding algorithm, an x,y location that represents the corresponding location of the street address is generated.
2. A query is preformed in the database to determine the appropriate route ID for that location.
3. The coordinate is projected to the NJ DOT centerlines using a snap algorithm. The snap process first determines the roadway segments in the destination network that have a corresponding route ID, and then determines the x,y point on that route that is closest to the x,y point on the source network.
4. A reverse dynamic segmentation is performed on the resulting route ID/x,y location on the destination network, which derives a corresponding SRI and milepost for that location.

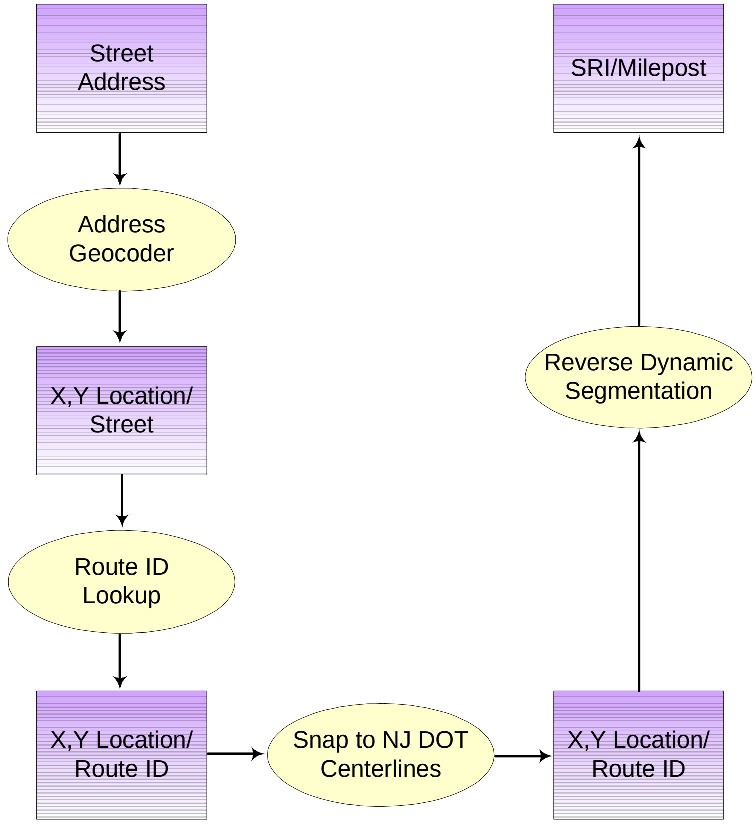

Figure III-6 - Street Address to SRI/Milepost Algorithm

III-6.2 SRI/Milepost to Street Address Transformation

Starting with an event table containing SRI/milepost linear event data on the NJ DOTs (source) network, the following steps are performed to transform the data to a street address reference on the Burlington County (destination) network (see Figure III-7).

The following procedure is performed for each event that is mapped from the state DOT's network to the Burlington County network.

1. Using a dynamic segmentation algorithm, the SRI/milepost is converted from a linear event to an x,y location. The SRI number is used as the route ID.

2. A snap algorithm is then performed. It first determines the roadway segments in the destination network that have a corresponding route ID, and then determines the x,y point on that route that is closest to the x,y point on the source network.

3. A spatial query is performed to determine the name of the street that corresponds to the route ID at the specified x,y location.

4. A reverse geocode algorithm is executed, which takes the street and the x,y location and returns the corresponding street address for that location.

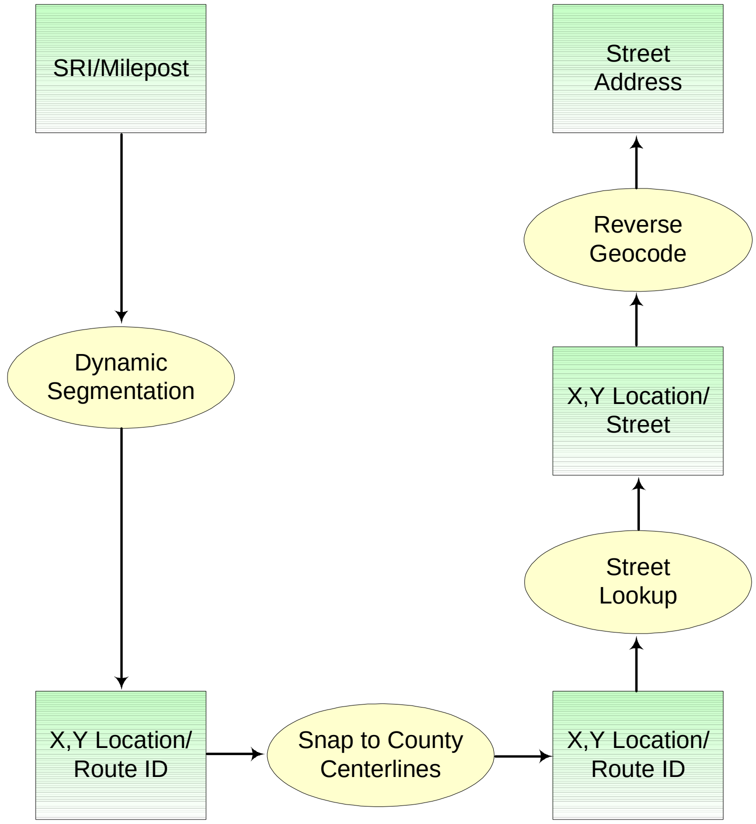

Figure III-7 - SRI/Milepost to Street Address Algorithm

III-7.0 Prototype System Overview

The Coordinate Route prototype was implemented using the following system architecture.

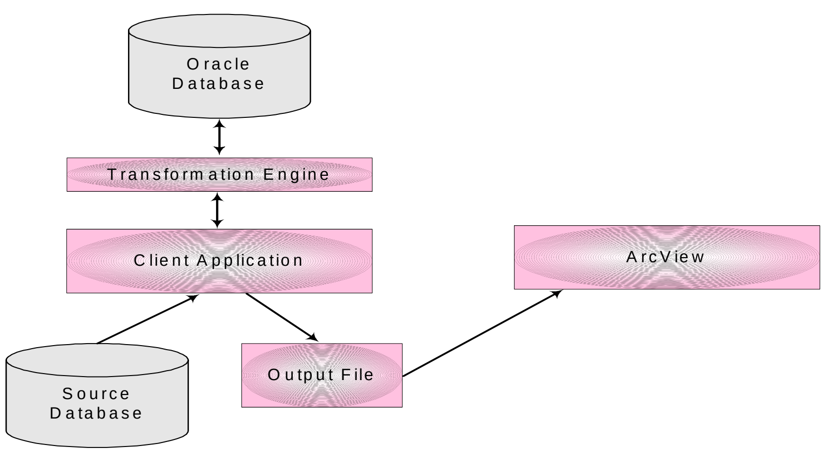

Figure III-8 - Coordinate Route Prototype Implementation

The following components are contained in the prototype:

- A Visual Basic® client program that imports/exports data and manages the transformation process.
- A Microsoft Access database with tables that contain the input data sets.
- An Oracle database that contains the road networks and the transformation algorithms.

III-8.0 Prototype Operation

The prototype operates as follows:

1. The user starts the client program and connects to the Oracle database.
2. Using the client, an input data set is selected from a list of tables retrieved from the Microsoft Access database.
3. The user selects the desired transformation (address to SRI/milepost, x,y) or (SRI/milepost to address, x,y) and presses the transform button.
4. The transformation algorithms execute (see Figure III-5 in the Transformation Overview) and the results are output to a comma separated value(CSV) file. The user can specify the name and location of the file prior to performing the transformation.
5. The user then starts ArcView and opens an ArcView project file that contains the appropriate base map and thematic layers.
6. The CSV file is imported into ArcView as a layer, allowing the results of the transformation to be viewed.
7. Additional CSV files can be imported, which allows comparison of the locations between the source network and the destination network to verify the transformation.

III-9.0 Data Preprocessing Steps

In order to prepare the data sets analysis over the area of coverage, the following preprocessing steps were preformed on the data:

III-9.1 County Road Preprocessing Steps

The County Road data that was supplied did not contain address range information. The data set was populated with address ranges that contained simulated addresses. To overcome this issue, the data was imported into an Oracle database and an SQL procedure was executed that applied simulated address block information to each county route segment. The SQL procedure applied odd addresses to the left side of each street, and even addresses to the right side. Left and right sides of the street were determined based on the noted primary direction of the roadway segments in the data set. 352 segments in the data set contained a NULL street name, and a NULL SRI number that is approximately 1.3% of the overall data set. When the prototype attempts geocoding, these data entries will report a “Not Found” error. An error is also reported when the prototype attempts to determine the SRI and milepost based on an address on the route where the SRI number is not present.

III-9.2 State Route Preprocessing Steps

The state route overlay was filtered to include only Burlington County, New Jersey (the area chosen to research). The highway segments that laid outside the county needed to be filtered. To do this TransDecisions imported the state transportation network into an Oracle database, and used Oracle Spatial to apply a county FIPS ID to each highway segment. Then an SQL query was executed to select only those highway segments that exist inside Burlington County, New Jersey.

III-9.3 Base Map Preprocessing Steps

To construct a spatially correct base map in ArcView for the presentation portion of the prototype, certain layers of the provided thematic layers had to be re-projected from within the NAD 83 state plane projection system for New Jersey with the underlying units converted from feet to meters. Layers were added to the base map consisting of county and state boundaries, state highways for the entire state of New Jersey, and county roads for Burlington County, New Jersey.

III-9.4 Input Data Sets

To properly demonstrate the prototype, two subsets of input data were derived from the provided source data sets. The first subset is an event file that lists the locations of traffic events on the county road network by address location. This data set is used to demonstrate the mapping of the equivalent locations onto the state highway network. Because TransDecisions did not receive any point event data for Burlington County, a simulated input data set that mirrors the access permits event table from NJDOT was built. The street addresses were taken from the ACCESS_PERMITS table to create the data set. The second data subset is a traffic event file that has locations linearly referenced by state route and milepost on the state highway network. This data set is used to demonstrate the mapping of the equivalent locations on the County Road Network, and was imported into Microsoft Access without any pre-processing. Both datasets were constrained into separate tables to a subset of the total event locations available to make the performance of the transformation algorithms reasonable for the prototype demonstration.

III-10.0 Benefits of the Coordinate Route Approach

Benefits of this approach include:

- Each entity can collect its transportation event information as it currently does, as long as it includes Route ID and coordinates.
- Departments of Transportation can collect information in Route/Milepost or County/SR/Segment and Offset format.
- Local governments can collect information as street addresses, known address points (for local governments that have points along each road that relate to the parcel addresses from their cadastral database), or x,y coordinates (Lat/Lng) from GPS systems.
- Compatible with many COTS packages including ArcView and GeoMedia.

III-11.0 Tradeoffs of the Coordinate Route Approach

This approach has some of the following tradeoffs:

- All participants must use common reference scheme. (Scale, Projection, Attribution, etc.)

- There is no support for linear events due to no underlying connectivity model in the system. (However, the endpoints of a linear event could be supported as point event data.)
- The ad-hoc approach used for data sharing could have additional data management costs.

(This page is intentionally blank.)

Chapter IV - Common LRS Model

IV-1.0 Overview

The Common LRS transformation method provides member agencies with the ability to use a common linear referencing system to transfer event information between multiple agencies using dynamic segmentation. Both point and linear event data can be "registered" to multiple agencies' centerline data using this approach. Data is transformed between different networks using a common linear reference composed of a route identifier and an offset for the event over the route. The offset could be defined in terms of miles, kilometers, feet, etc., but would have to be standardized across the DVRPC member agencies within each state. The transformation method is illustrated in Figure IV-1.

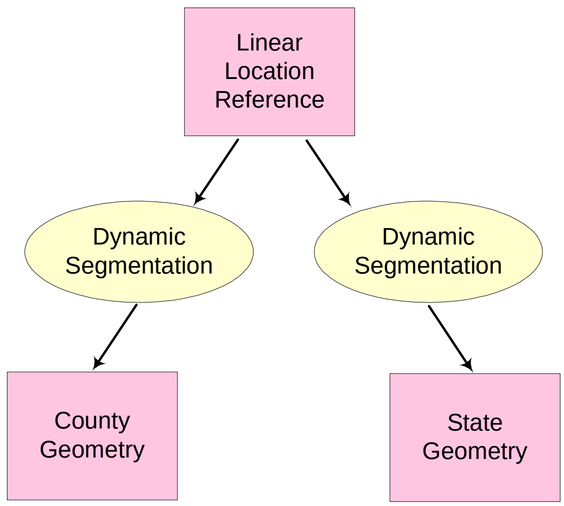

Figure IV-1 - Common LRS Transformation Overview

Note that, using this approach, all data is referenced using a common linear location reference and can be referenced to any participating network without requring intermediate transformation processes.

The following algorithm is used to transform data:

1. A location reference in Network A is placed on the network.
2. The linear reference for this location is established.
3. A dynamic segmentation algorithm finds the location corresponding to the linear reference on the opposing road network.
4. The location reference for the corresponding location is derived in Network B.

These steps are illustrated in Figure IV-2.

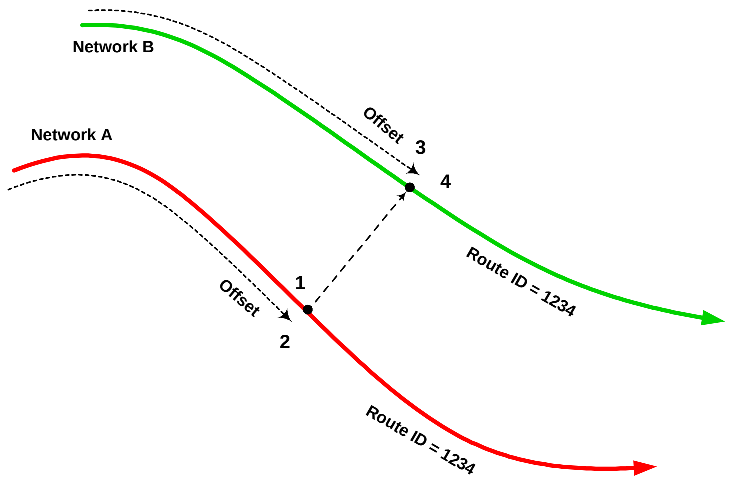

Figure IV-2 - Common LRS Transformation Overview

IV-2.0 Collaboration Architecture

The collaboration architecture relies on a distributed approach that allows agencies to share event data with each other in an ad-hoc fashion (refer to Figure V-3). The mechanisms used to share data between agencies are varied and may include FTP, Web, Sneaker Net, or any other reasonable method of delivering a data set from the source location to its destination. For point events, the data sent must contain Route ID and distance. For linear events, the data sent must contain Route ID, beginning, and ending distances. When the data has been received at its destination, it is loaded into a database and a commercial dynamic segmentation software program is used to place the events along the agencies road network. The dynamic segmentation process could be handled by a central authority for local agencies that do not have a dynamic segmentation tool available to them. This would require that those governments share their centerline files with the central agency.

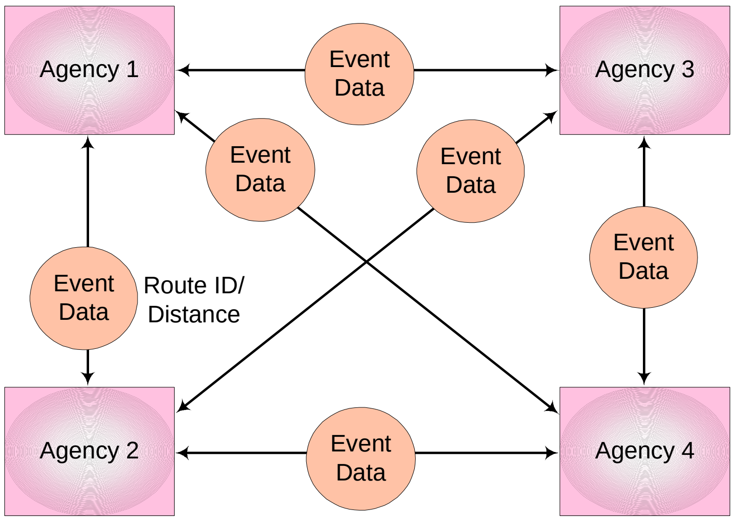

Figure IV-3 - Common LRS Collaboration Architecture

Similar to the Coordinate Route methodology, this approach allows for formal or semi-formal data clearinghouses to be set up in either a centralized or decentralized manner to allow member agencies to post data sets once and allow them to be used by any other member agency without burdening the central authority. Data clearinghouses could be as simple as an FTP site or local web page.

IV-3.0 Data Requirements

The data requirements for the Common LRS approach are:

- A destination/destination network attributed with the same linear referencing system composed of route identifiers and measured lengths for the routes.
- A set(s) of either point or line event data that contains both the unique route identifier and the offset information.
- Appropriate Dynamic Segmentation software that can take the event data and place the events at the correct locations along the network.

IV-4.0 Demonstration Implementation

The coverage area for this demonstration is in southeastern Burlington County, New Jersey as shown in Figure IV-4.

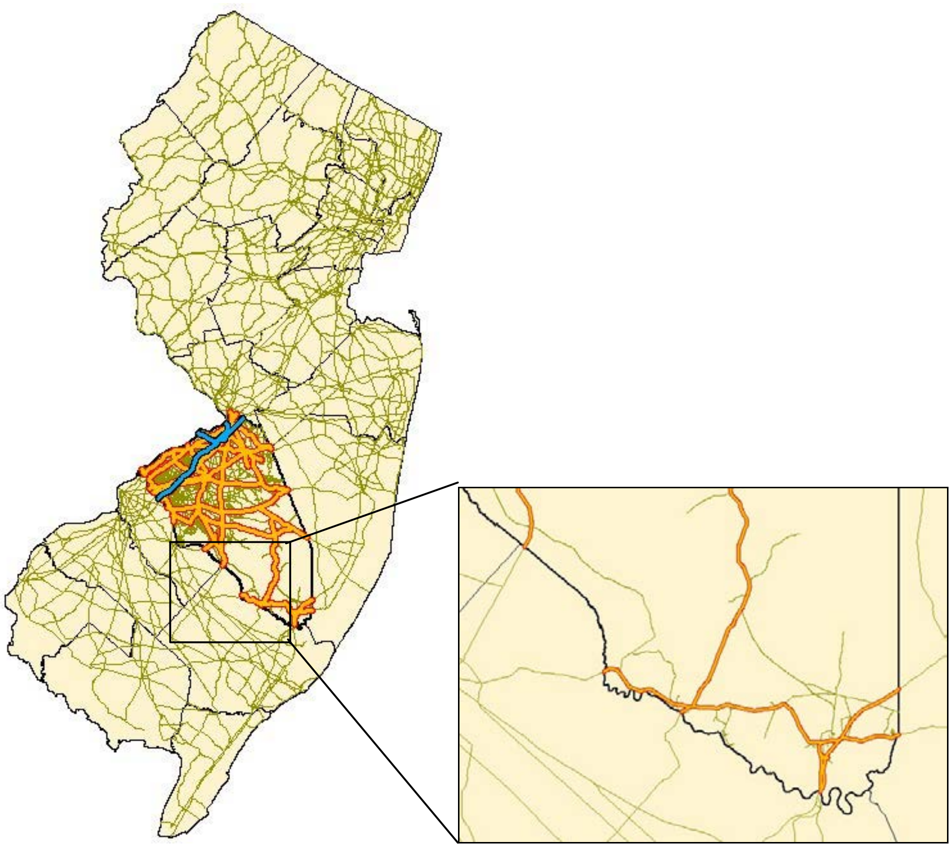

Figure IV-4 - Area of Coverage (Southeastern Burlington County, NJ)

The county road data and the New Jersey state road data contains the SRI numbers and LRS distance-milepost information. This demonstration transforms point and linear events from the county network to the state network, and from the state network to the county network using SRI and LRS milepost measures.

IV-5.0 Selected Data Sets

The data sets selected for use in the demonstration include:

- ***Burlington County Road Centerlines***
  This data set was provided by Burlington County. The centerline data was collected by the county using a mobile GPS device (approx 1:200 scale). The data is appropriate for use in this demonstration because it includes road centerline data with SRI numbers, mileposts, and segment lengths.
- ***New Jersey DOT Road Centerlines***
  This dataset was provided by NJDOT. This centerline data file is appropriate for use in the demonstration because it includes SRI numbers and milepost information.
- ***Event Data***
  Sample accident and AADT (Linear Traffic Counts) data was manually created for demonstration purposes.

IV-6.0 Implementation Examples

Example 1: Dynamic Segmentation along State Route
In this example, point and linear event data is placed along the state route network using dynamic segmentation. The combination of SRI number and milepost values is used.

Example 2: Dynamic Segmentation along County Route
In this example, point and linear event data is placed along the county road network using dynamic segmentation. The combination of SRI number and milepost values is used.

IV-7.0 Data Preprocessing Steps

In order to prepare the data sets we received for analysis over the area of coverage, the following preprocessing steps were performed on the data:

IV-7.1 County Road Preprocessing Steps

The county data supplied had the correct SRI numbers, but the milepost information was the same for each segment along a particular SRI. For example, SRI 00000542 was comprised of 30 individual road segments; however, each segment had the same begin/end milepost values. TransDecisions evaluated the begin and end milepost for each route, and looked at each of the individual segments lengths to determine the correct begin and end mileposts for each of the segments that comprised the route.

IV-7.2 State Road Preprocessing Steps

No preprocessing was required.

IV-7.3 Event Data Preprocessing Steps

Two database tables were generated for accident point data, and one for AADT linear event data. These tables were created and entered into the Oracle database.

IV-7.4 Linear Referencing Setup

Using TransDecisions' LRSx™ application (see the Components in the Demonstration section below for details), a route system was built for the county road network, and a route system was built for the state road network. Metadata was added to LRSx, which determines how the software reads and places the event data along each of the networks.

IV-8.0 Demonstration Operation Description

The Common LRS prototype works as follows:

1. The user opens an SQL prompt in Oracle and logs into the database.
2. Using a series of insert record input statements, accidents and traffic volume (AADT) records are entered into the Oracle database.
3. A database trigger is fired, and each of the new database records are dynamically segmented.
4. The user opens an ArcView 3.2 .apr file, which includes the New Jersey base-level GIS data.
5. Using Safe Software's FME ArcView extension (see the Components in the Demonstration section below for details), the user requests to view either the accident or traffic count data along either the state or the local road network.
6. FME reads the records from Oracle and converts it to a shape file, which is displayed in ArcView.

IV-8.1 Components in the Demonstration

This demonstration relies on the following software tools:

- ArcView 3.2
- Safe Software's FME ArcView Extension (www.safe.com). FME stands for (Feature Manipulation Engine). This third party extension for ArcView was used in the demonstration to facilitate reading geometry directly from Oracle Spatial for presentation in ArcView.
- Oracle 8.1.6 with Oracle Spatial installed
- TransDecisions' LRSx software. LRSx stands for Linear Referencing System Extensions. This product is Oracle based and provides an integrated set of functions and procedures that facilitate the organization, retrieval, and validation of linear referenced data.

IV-9.0 Benefits of this Approach

Benefits of this approach include:

- The ability to handle both point and linear event data.
- Centralized authority is not required but can be utilized for government agencies not wanting to purchase/implement dynamic segmentation software.
- Compatibility with higher-end COTS GIS packages.
- Stable and accurate transformation between networks.

IV-10.0 Tradeoffs of the Common LRS Approach

This approach has some of the following tradeoffs:

- Requires a common linear referencing system to be adopted by all constituents.
- To support linear events, the underlying transportation networks must have topological connectivity. This may require additional data investments to be made.
- The ad-hoc approach used for data sharing could have additional data management costs.

Chapter V - Extended NSDI Prototype

V-1.0 Overview

Extended NSDI provides a framework based on equivalenced network components that tie multiple networks together. Central to the approach is the definition and construction of a master network which represents a superset of all participant networks and provides a single-base reference system that can be used to map between disparate "local" reference systems using a "star" transformation model (see Figure V-1).

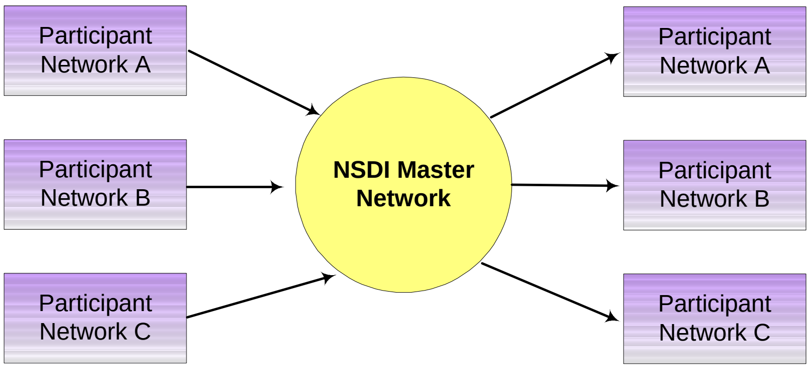

Figure V-1 - Extended NSDI Approach

Each local referencing system must have an equivalency table created that maps the local system to the global system in the master network. While the master network is centrally managed and maintained, individual data sets are locally managed and maintained within their respective agencies. Hence, the hybrid centralized/decentralized aspect of this approach.

Using a generic code framework and data schema naming standard, it is possible to provide a single system that delivers seamless transformations from any participant network to any other participant network (provided that they share a common master network, and that their equivalence table is mapped to that master network). Such a system could seamlessly take an NSDI reference, provide a forward transformation to the master network to generate a master NSDI reference, and then perform a reverse transformation to generate a new NSDI reference for a participant network.

This transformation is a mathematical process that can reference a location in one transportation network to another. The business problem associated with these transformations, however, is more complex. The issue being how to take a location reference in one network and state it in terms of a location reference on another network. The location references could be from differing LRMs, street address, and LRS in this case. The only assumption that can be made is that the location reference in the first network has an analogous location reference in the second network, however that location reference might be NULL.

To solve this problem there needs to be additional transformations that can take a location reference in a participant network and state it in terms of an NSDI reference, and vice versa. Any network can support multiple location references. So with NSDI, an arbitrary location reference in a participating network can map to any other arbitrary location reference in any other participating network.

V-1.1 Forward Transformation

The forward transformation takes an NSDI reference and projects it to the master network . This projection is performed by searching the equivalence table for the ID of the NSDI reference, and passing back the master NSDI reference ID and offset.

V-1.2 Reverse Transformation

The reverse transformation takes a master NSDI reference and projects it to a participant network. This projection is also performed by searching the equivalence table for the ID of the master NSDI reference, and returning the ID and offset values for the participant network.

V-2.0 Collaboration Architecture

The collaboration architecture relies on a hybrid centralized/distributed approach that allows agencies to share event data with each other in a controlled fashion (refer to Figure V-2). The data sharing mechanism relies on the master network, which acts as the hub of a star topology of particpant networks (A, B, and C). The data sharing capability is attained through a client/server methodology, which could have a varied number of underlying transport protocols including, TCP/IP, HTTP, etc. Event data from local participants would be "posted" to the master network server, which would process the request and send back the results. Once the data has been received locally, any additional workflows that are required to pre-format the data, such as attribution merge, must be determined and specified as a part of the collaboration process.

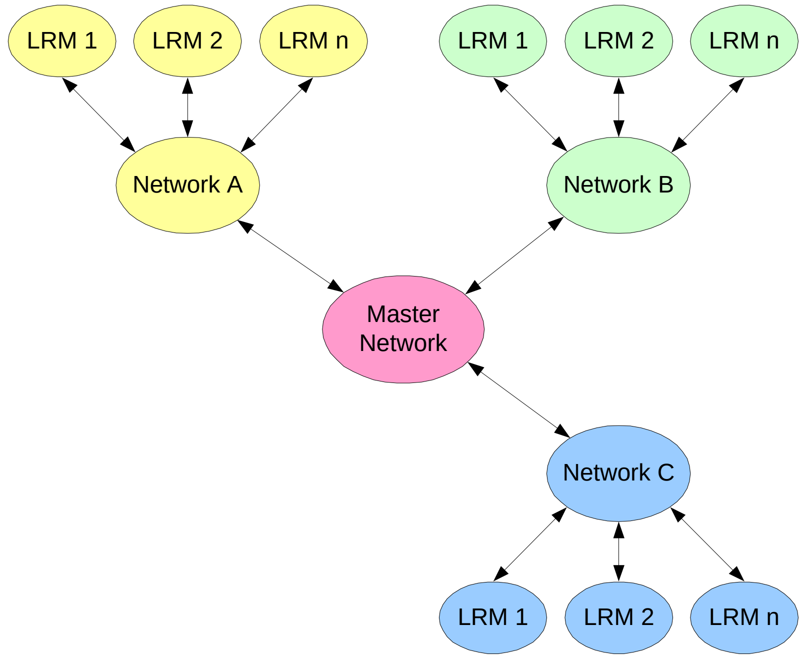

Figure V-2 - Extended NSDI Collaboration Architecture

This approach allows for formal or semi-formal data clearinghouses to be set up in either a centralized or decentralized manner to allow member agencies to post data sets once and allow them to be used by any other member agency without burdening the provider agency. Data clearing houses could be as simple as an FTP site or local web page.

V-3.0 Data Requirements

To use this approach, the following is required:

- A set of participant networks containing data that is linearly referenced.
- A master data representation to be used in the master network. Ideally, this data representation is a superset of the combined participant networks.
- To support linear events, the underlying transportation networks must have topological connectivity. This means that the linkages between the various road segments that are traversed from the starting point to the ending point of the linear event are known and can be traversed to facilitate plotting the relevant geometry.
- A defined equivalence table for each participant network that maps the data to the master network must be created and maintained.
- A centralized database must be provided to host the master network . In the prototype, an Oracle database is used. However, this approach is readily portable to any database architecture.
- A sponsor to host and manage the master network for all constituents must be established. This sponsor could be a vendor (i.e., NavTech® or GDT), or a planning agency (such as DVRPC).

V-4.0 Prototype Implementation

The coverage area for this prototype is Tredyfferin township, Chester County, Pennsylvania as shown in Figure V-3.

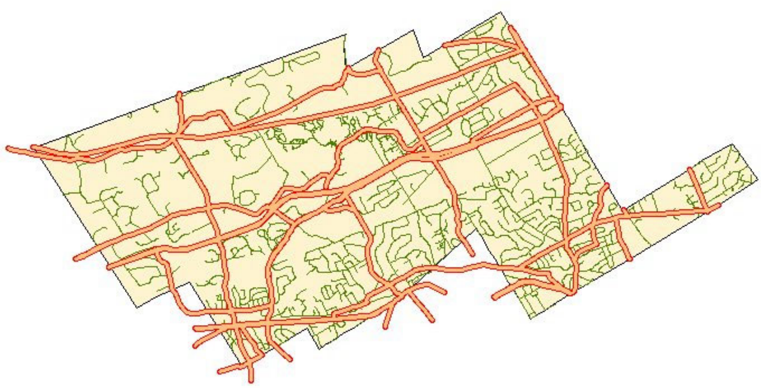

Figure V-3 - Area of Coverage (Tredyfferin Township, Chester County, Pennsylvania)

County road data contains street addresses, while state roads contain LRS information. The prototype transforms point events defined by address information in the county network to common LRS defined point events in the state route network, and vice versa.

V-5.0 Selected Data Sets

The data sets selected for use in this prototype include:

- ***Chester County Roads***

This data set was provided by Chester County, Pennsylvania. The data is appropriate for use in the prototype because it contains ArcInfo network topology that was used to create the FTRP table, and road segments that were designated as FTSegs. Each were required to create master network , which was designated to be based on the county road network (as it is a superset of both participant networks). In addition, this road network was used as the county road participant network.

- ***Pennsylvania State Routes.***
  This data set was provided by Pennsylvania DOT. This data set is appropriate for use in the prototype because it contained appropriate segments and network topology to be used as a participant network.
- ***Linear event tables***
  A set of simulated linear event data that is used as input data to illustrate the state-to-local and local-to-state transformation of linear event data.
- ***Point Event tables***
  A set of simulated point event data that is used as input data to show state-to-local and local-to-state transformation of point event data.

V-6.0 Transformation Overview

The general transformation model for participant networks is shown in Figure VI-4. A linear location for a participant network is mapped to a network location via a process. For the county network, this is an address geocoding process. For the state network, the process is dynamic segmentation, which results in a network location rather than a geometry. Once a network location has been determined, the network location can be looked up via cross referencing tables to determine an NSDI reference. The NSDI reference is then looked up in the master NSDI network to generate a master NSDI reference.

The second stage of this process cross references an NSDI reference in the master NSDI network to a participant network NSDI reference via a reverse lookup. Once this is completed, the NSDI reference can be cross referenced to a network location through a reverse lookup and then to the specific LRM using a reverse geocode or a reverse dynamic segmentation process.

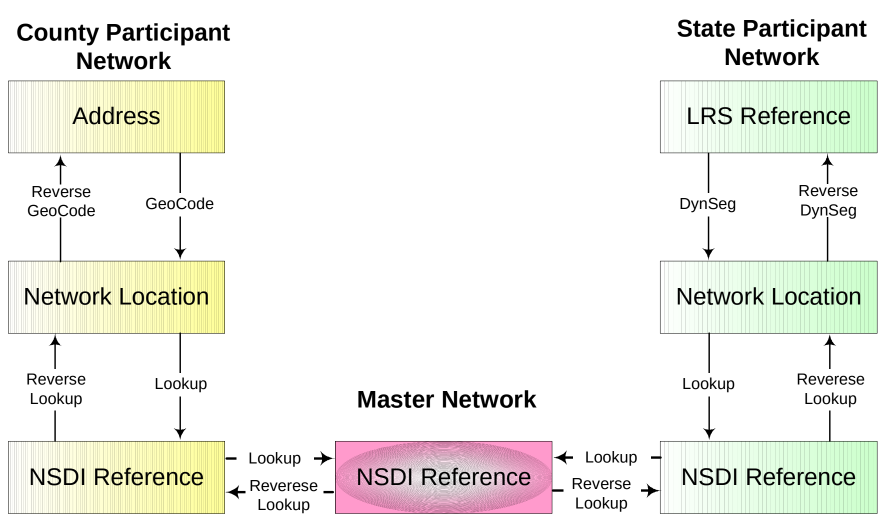

Figure V-4 - Extended NSDI Transformations

V-6.1 County Participant Network To Master Network Transformation

The county participant network uses a street address location reference method. A street address is located using an address matching algorithm and returns a network location. A network location is an unambiguous reference on a network, stated in terms of a primary key of a network segment and an offset over that segment in the direction of the segment. The network location is then used to generate a NSDI reference, which is an equivalent entity over a different set of primary keys. The primary key in the table for the county road network is used to generate the NSDI Segment ID. The network location primary key is transformed into a Segment ID using an SQL function. As there is a one-to-one correspondance between the network segments and Segment IDs, the offset of the network location is the same for the Segment ID.

V-6.2 State Participant Network To Master Network Transformation

The state participant network uses a linear reference, which uses a primary route key and offsets as milepoints over this system of linear features. A LRS, offset value can be converted to a network location, which can be used to generate a NSDI reference. There is no direct connection between a primary key for a network location and a NSDI Segment ID, a segment cross refernce table is used to obtain the mapping. The cross refernce table is used to lookup the cross reference between the network location and the Segment ID. The offset is the same.

V-7.0 Implementation Examples

Example 1: State Route to County Road Transformation

In this example, data is transformed from the state route to the county road network (refer to Figure V-5).

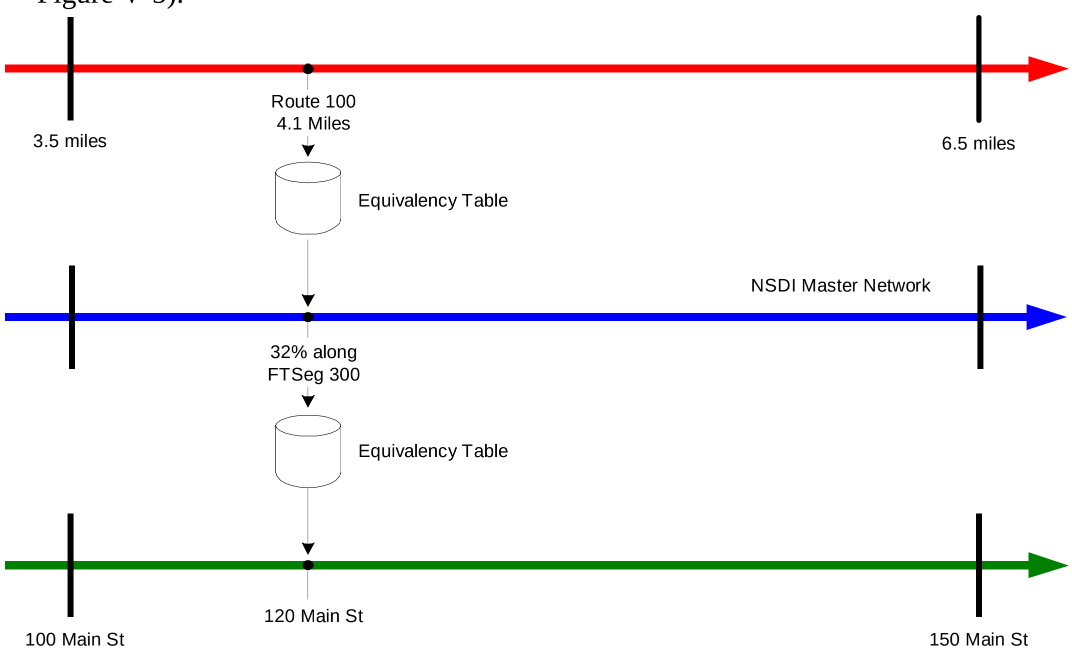

Figure V-5 - State Route to County Road Transformation

1. State DOT places an event (point or linear) onto their road centerline using their current linear referencing system. Route/milepost or Co/Sr/Seg.

2. Local data using the equilavency tables is referenced against the NSDI master network, and a Segment ID and % offset along the segment is determined.

3. Using the Segment ID and % offset along the segment, the NSDI data is mapped to the local road network using the NSDI master network and the information is returned to the local government in the method that they use to place information along their road network (address range, address point, x,y coordinate, intersection/offset).

Example 2: County Road To State Route Transformation
In this example, data is transformed from the county road to the state route network (refer to Figure V-6).

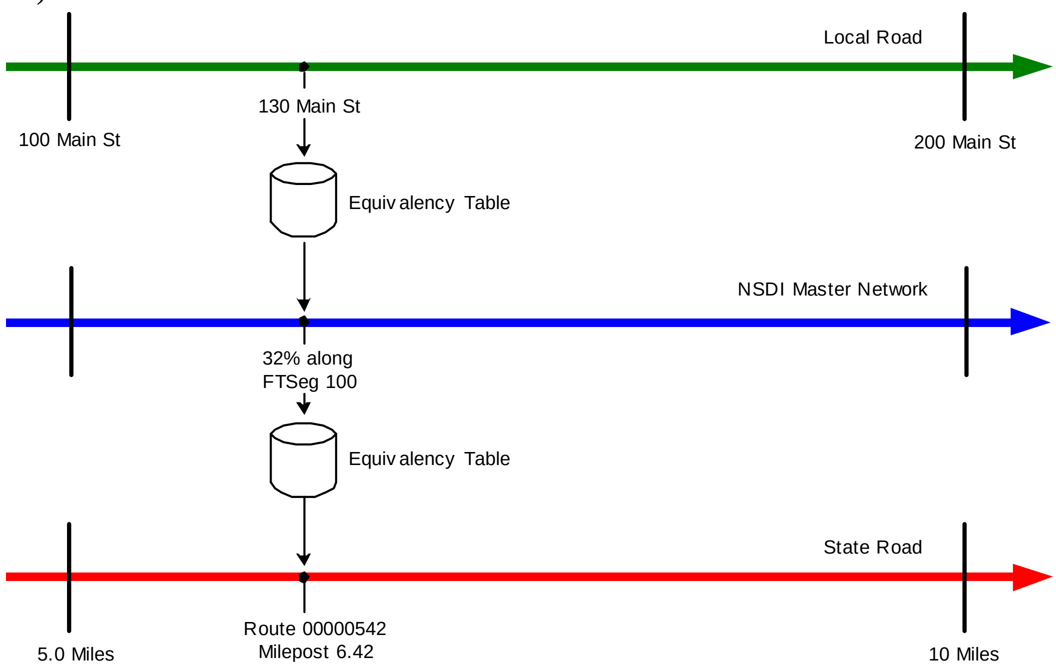

Figure V-6 - County Road to State Route Transformation

1. A local government places an event (point or linear) onto their road centerline using their standard method (address ranges, address points, x,y coordinates, offset from intersection).

2. Local data is referenced against the NSDI master network using the equivalency tables, and an Segment ID and % offset along the segment is determined.

3. Using the Segment ID and % offset along the segment, the NSDI data is mapped to the state road network using the NSDI master network and the information is returned using the state's LRS.

V-8.0 System Overview

Figure V-7 illustrates the conceptual system architecture for the prototype.

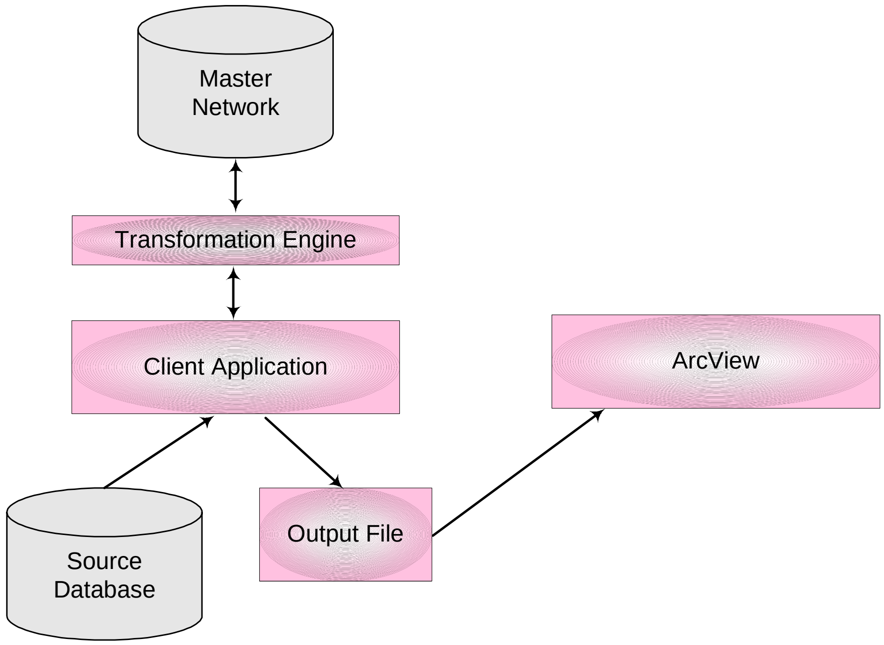

Figure V-7 - Extended NSDI Prototype Implementation

The developed Extended NSDI prototype was implemented as follows:

1. A VisualBasic client application was used to import a source data set containing event data from the source database, which for the prototype is a Microsoft Access database. In a production system any type of database that can be accessed an ODBC data source may be used.

2. The VB client passes the input event data to the Transformation Engine, which handles the transformation and passes back the results. The Transformation Engine is a set of functions contained in a database package within an Oracle 8*i* database with a set of PL/SQL™ wrappers that enable the software to be called from the VB client.

3. The Transformation Engine transforms data in one of two ways: street address to route/offset, or vice versa.

4. The attribution and geometry contained within the source and destination networks are preloaded and stored in the Oracle 8*i* database, and are accessed by the Transformation Engine as required to compute the results. The VB client calls the appropriate functions in order to complete the transformation process.

5. Once the transformation process is complete, the VB client outputs the resulting data set to a comma separated value file.

6. The resulting CSV file can then be imported into ArcView and overlaid as a thematic layer on top of the desired base map for results presentation.

V-9.0 Logical Data Model

The logical data model is designed to be symmetrical and replicatable for all participating networks. Additional participant networks can be added by simply constructing the requisite physical model, and populating the tables with the appropriate data. The diagram of the complete logical data model for Extended NSDI is shown in Figure V-8 on page 55. The entities represented by the yellow boxes are county data inputs. The green boxes represent state agency data inputs. The purple boxes in the middle of the diagram represent the master network components. The two salmon colored boxes in the lower left portion represent configuration and metadata components.

V-9.1 County Data Model Subset

The county data model subset shown in Figure V-9 on page 56 illustrates the relationships for a typical participant network. The color-coding of the boxes follows the same relevant representations as those depicted in Figure V-8 as described above. The most notable addition to the NSDI model is the inclusion of a cross-reference table (NSDI_COUNTY_ID_XREF). This table stores the relationships between features identified in the participant network, and their equivalent NSDI identifiers. This table is an important addition because it allows the NSDI transformation code to generically map between network references and NSDI references. It also becomes an efficient domain verification system for validating the existence of feature identifiers in both the NSDI and network domains.

V-9.2 State Data Model Subset

The state data model subset shown in Figure V-10 on page 57 reflects the county sub-model with the exception that the participating network is replaced with the state highway network. The color-coding of the boxes follows the same relevant representations as those depicted in Figure V-8 as described above.

V-9.3 Master Network Data Model Subset

The master network data model subset shown in Figure V-11, below, is essentially an unaltered NSDI model.

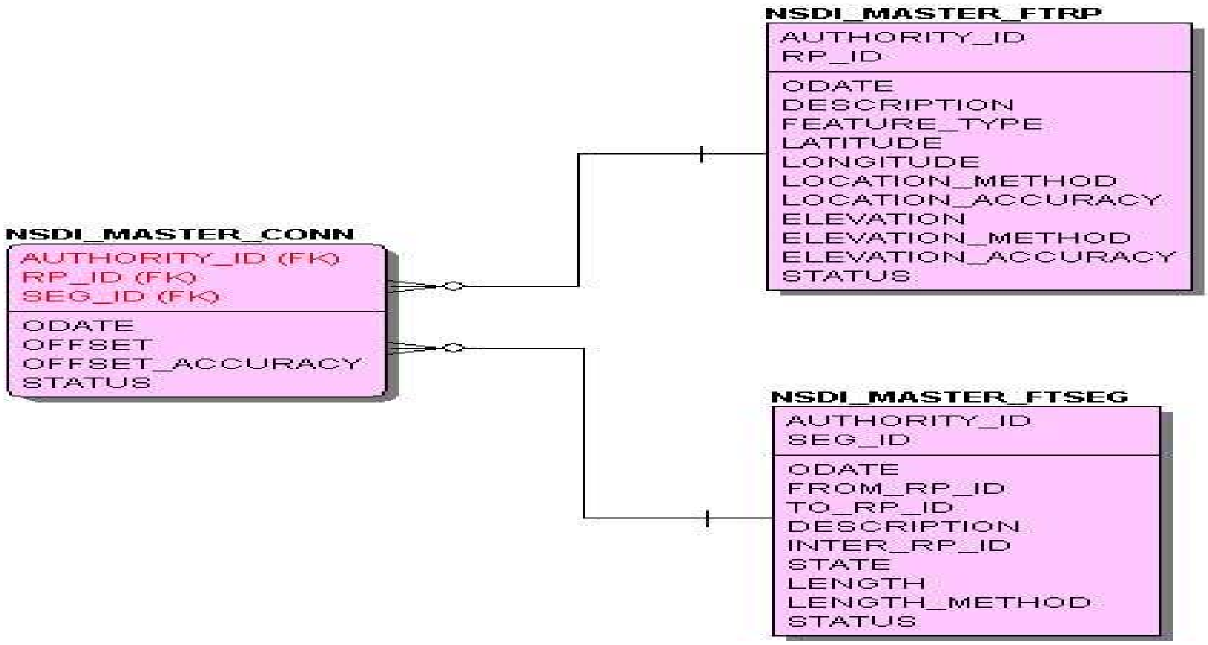

Figure V-11 - Extended NSDI Master Network Model Subset

V-10.0 Data Preprocessing Steps

To prepare the data for the prototype, TransDecisons began with three sets of road network data:

V-10.1 County Road Preprocessing Steps

The county road network for Chester County was imported into Oracle, and roadways outside Tredyfferin Township were filtered out of the dataset.

V-10.2 State Route Preprocessing Steps

In order to obtain a state route overlay just for Tredyfferin Township, all highway segments that were outside the township were filtered. The state transportation network was imported into an Oracle database, and roadways outside Tredyfferin Township were filtered out of the dataset using an Oracle Spatial query.

V-10.3 Master Network Preprocessing Steps

Because the County Road network represented a "superset" of the state road network, the master network equivalency table was created based on the County Road network for Chester County, with coverage of the Tredyfferin Township. A table was constructed in Oracle that maps each route system to the master network coverage using ArcInfo. IDs for each route were programmatically assigned using a PL/SQLprocedure in Oracle.

V-10.4 Participant Network Preprocessing Steps

Two equivalence tables were constructed: a county road network equivalence table and a state road network equivalence table. Each of these tables was created based on the respective road networks, and have IDs that were generated to map to the master network equivalence table. This was done through a programmatic process.

V-10.5 Base Map Preprocessing Steps

To construct a spatially correct base map in ArcView for the prototype presentation, certain layers had to be reprojected to conform to the NAD 83 Pennsylvania Southern State Plane Projection System. Layers were added to the base map that consisted of county and state boundaries, state highways for the entire state of Pennsylvania, and county roads for Chester County, Pennsylvania.

V-10.6 Point Event Source Data Selection

Two subsets of input data were derived from the provided source data sets. The first was a simulated event file that lists the locations of traffic events on the county road network by address location. This data set is used to demonstrate the mapping of the equivalent locations onto the state highway network. The second subset is a traffic event file that includes locations linearly referenced by state route and milepost on the state highway network. This data set is used to demonstrate the mapping of the equivalent locations on the county road network.

V-10.7 Linear Event Source Data Selection

Two subsets of linear event input data were derived. The county data set contains simulated linear events listed by route, with start and ending offsets. This data set is used to demonstrate the mapping of the equivalent locations onto the state highway network. The second data subset is a traffic event file that has locations linearly referenced by state route and milepost on the state highway network. This data set is used to demonstrate the mapping of the equivalent locations on the county road network.

V-10.8 Prototype Operation Description

The prototype operates as follows:

1. The user starts the client program and connects to the Oracle database.

2. Using the client, an input data set is selected from a list of tables retrieved from the Microsoft Access database.

3. The user selects the desired transformation (address to SRI/milepost, x,y) or (SRI/milepost to address, x,y) and presses the transform button.

4. The transformation algorithms execute, and the results are output to a comma separated value file. The user can specify the name and location of the file prior to performing the transformation.

5. The user then starts ArcView, and opens an ArcView project file that contains the appropriate base map and thematic layers.

6. The CSV file is imported into ArcView as a layer, allowing the results of the transformation to be viewed.

7. Additional CSV files can be imported, which allows comparison of the locations between the source network and the destination network to verify the transformation.

V-10.9 Components in the Prototype

The following deliverables are contained in the prototype:

1. A VisualBasic client program that imports/exports data and manages the transformation process.
2. A Microsoft Access database with tables that contain the input data sets.
3. An Oracle database that contains the road networks and the transformation algorithms.

V-10.10 Benefits of this Approach

Benefits of this approach includes:

- NSDI provides a highly stable transformation model.
- Compatible with many COTS packages, including ArcView and GeoMedia.
- Provides a basic framework for developing temporal references.
- Handles non-spatial networks (TranPlan, EMME/2, etc.).
- Does not require common reference framework (scale, projection, attribution, etc.).
- Provides a collaboration framework for external organizations to participate (FHWA, neighboring MPO's, and states).

V-10.11 Tradeoffs of the Extended NSDI Approach

This approach has some of the following tradeoffs:

- Sponsored centralized management and maintenance of the network is required. This is less of an issue for State DOTs as they would be a likely candidate for stewardship of the centrally managed solution.
- An equivalency table must be developed and maintained for each participating agency.
- To support linear events, the underlying transportation networks must have topological connectivity. This may require additional data investments to be made.

(This page is intentionally blank)

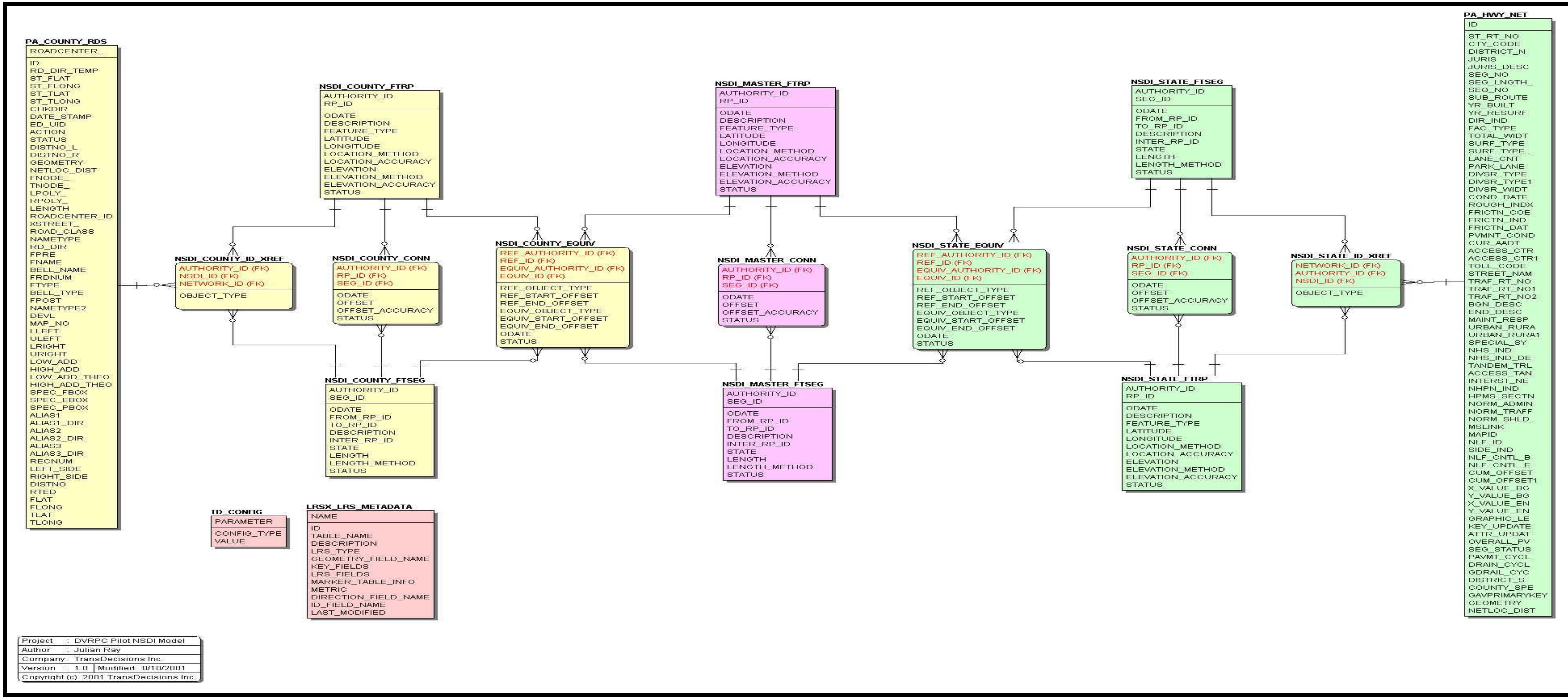

Figure V-8 – Extended NSDI Complete Logical Data Model

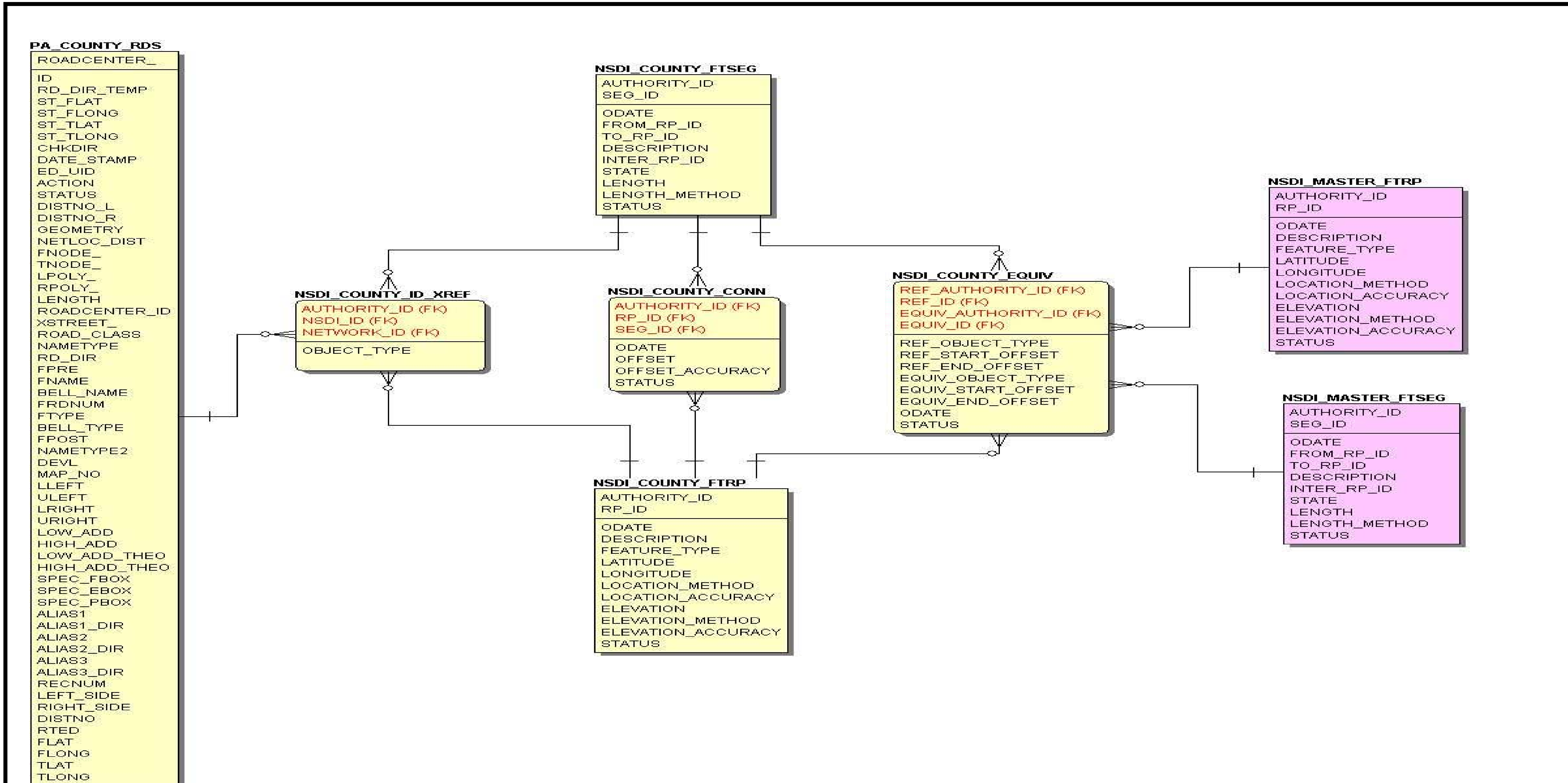

Figure V-9 – Extended NSDI County Data Model Subset

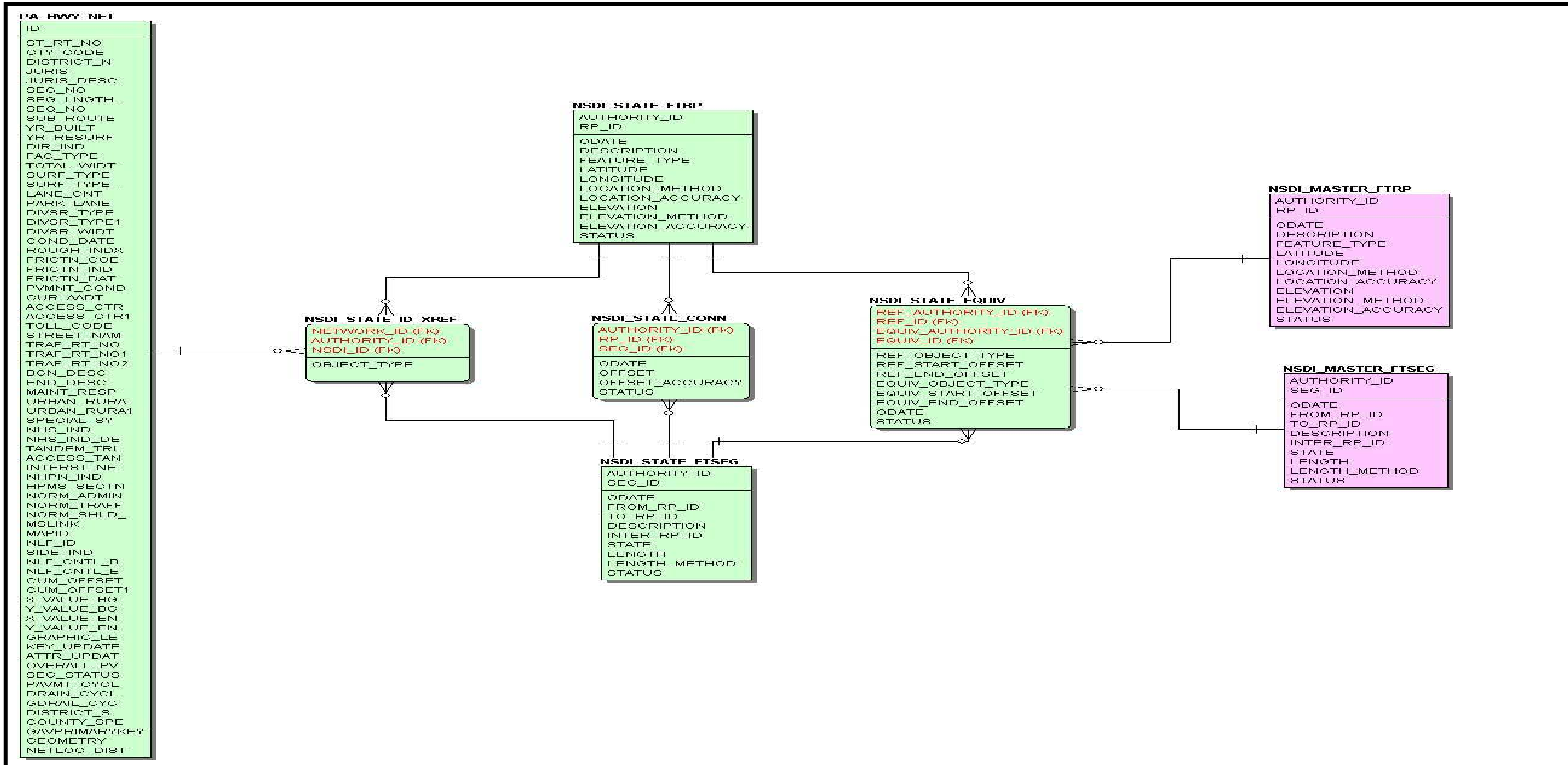

Figure V-10 – Extended NSDI State Data Model Subset

Appendix A - Cost Estimates for Street Centerline Development

Internal Development by the Agency/Contract Development

	Project Manager (Hours)	***GIS Analyst (Hours)***	***Total Hours***
Tasks			
Task 1 – Develop Centerlines			
Rectify Aerial Photographs	*4.0*	*40.0*	***44.0***
Define Feature Classes	*4.0*	*8.0*	***12.0***
Develop Data Collection Rule Base	*8.0*	*40.0*	***48.0***
Collect Centerlines	*80.0*	*800.0*	***880.0***
Collect Attributes	*30.0*	*300.0*	***330.0***
Task 1 Total	***126.0***	***1188.0***	***1314.0***
Task 2 – Prepare Datasets			
Remove Pseudonodes	*0.0*	*40.0*	***40.0***
Match Attribute Names & Data Types	*0.0*	*8.0*	***8.0***
Develop/Edit Manual Conflation Tool	*0.0*	*24.0*	***24.0***
Task 2 – Total	***0.0***	***72.0***	***72.0***
Task 3 – Conflation			
Automatic Conflation	*4.0*	*24.0*	***28.0***
Manual Conflation	*60.0*	*600.0*	***660.0***
Total Task 3	***64.0***	***624.0***	***688.0***
Task 4 – QA/QC			
Find & Correct Null Values	*8.0*	*120.0*	***128.0***
Find & Correct Address Errors	*8.0*	*120.0*	***128.0***
Find & Correct Street Name Inconsistencies	*8.0*	*120.0*	***128.0***
Total Task 4	***24.0***	***360.0***	***384.0***
Task 5 - Maintenance			
Establish Domain Tables for Attributes	*8.0*	*80.0*	***88.0***
Total Task 5	***8.0***	***80.0***	***88.0***
GRAND TOTAL	***222.0***	***2324.0***	***2546.0***

Table A-1 Internal Data Development Level of Effort

Purchase from Data Vendor

GDT – Geographic Data Technology

GDT offers government agencies two ways to obtain their street centerline data. These two options are explained below:

1. Purchase the data: GDT would provide the Dynamap/Transportation product for the following:

Price per County	*Number of Concurrent Users*
$800.00	*1-5*
$1600.00	*6-15*
$2000.00	*16-25*

Note: Add an additional $10 for each user after 25 users.

Table A-2 GDT Data Costs

2. Join "The Community Update Program": The program is a public/private partnership for maintaining the nation's street centerline data layer. Through a data-sharing agreement, local and regional government agencies send new and updated information about their local streets to GDT where it is validated, cross-referenced and integrated with other data, then returned to a central server for downloading by those agencies. The cost of data is free to participating agencies.

The accuracy of GDT data is very high. GDT currently has a level of accuracy of 12 meters or less for 30% of their database. Their efforts will not stop until the whole US is at that mark. It is impossible to say what part of what county is inside 12M, but they can offer "release notes" that will give the number of streets under 12M from last update to the present. GDT uses several sources for the development and maintenance of their data, including:

- Local, state and federal governments
- Original TIGER files
- Community update programs
- Various digital and hard copy maps
- Ortho-enhanced aerial photography

GDT offers annual renewals, semi-annual updates, and quarterly updates of its data, which allows maintenance to be a non-factor. As long as the Agency purchases some sort of update with the initial purchase, we can rate maintenance as a strong positive factor. At the time of initial purchase semi-annual updates are an additional 10% of the base and quarterly updates are an additional 25% of the base. Renewals are completed on the anniversary of the date of the initial license agreement.

Sharing GDT data with other non-licensed agencies is an issue. No one other than the organization listed on the License Agreement can use the data. If other agencies or counties need the data, either a separate agreement needs to be drafted or a DVRPC needs to be the licensee and list each agency that needs access to the data as a user. The pricing will reflect the correct number of users and all will be covered under one agreement.

ADCi – NavTech

Navtech offers a few options on purchasing the NavTech NAVSTREET data. These options are based per county and are listed below:

License Type	***Standard data costs***	***Premium data costs***
Single User	*$1500.00*	*$3,000.00*
Server License (up to 350 users)	*N\A*	*$18,000.00*
Internet	*Price based by processor speed and the number of average users and hits per site.*	

Table A-3 Navtech Pricing

Navtech also offers the option of purchasing the data by State. The prices for the states of Pennsylvania and New Jersey together are listed below:

License Type	***Standard data costs***	***Premium data costs***
Single User	*$7,650.00*	*$15,300.00*
Server License (up to 350 users)	*$45,900.00*	*$91,800.00*
Internet	*Price based by processor speed and the number of average users and hits per site.*	

Table A-4 Navtech State Pricing

NavTech claims to have the best accuracy in the market. Their data is accurate to a 1:24K level (+/- 40 feet) in rural and suburban areas, and a level of accuracy as good as +/- 15 feet in detailed city areas. NavTech uses several sources for the development and maintenance of their data, including:

- Enhanced TIGER files
- Satellite Imagery
- Information provided by county offices
- They also have driven a lot of areas and collected information. They are in the process of driving the whole US.

Navtech offers the option to sign up for updates at the time of purchase. They offer two types of updates, which are explained below:

- Annual Updates at 10% of original price per year.
- Quarterly Updates at 25% of the original price per year.

Data sharing is allowed as long as the correct number of user licenses has been purchased. Each user must have a license regardless of where their office is physically located. This is a similar issue to purchasing GDT data and can be dealt with in the same manner. Again, if other agencies need the data, DVRPC needs to be the licensee and list each agency that needs access to the data as a user. The pricing will reflect the correct number of users and all will be covered under one agreement.

Tele-Atlas (ETAK)

Tele Atlas offers a cost effective method of purchasing multiple counties by providing the data in larger geographic areas, which they call Etak Coverage Areas (ECA). Two ECA's include all the counties in the DVRPC area. The cost of data per Single End User (1-5 concurrent users) Annually per ECA is:

Tele Atlas MultiNet (US)	*Tele Atlas MultiNet (US) w/directions*
$800.00	*$1,800.00*

Note: Tele Atlas also offers a discount for Multi-Year Licenses at 100% of Annual License Fee for year one and 50% of Annual License Fee for years two and three.

Table A-5 Tele Atlas MultiNet Pricing

If you want to purchase the data for more than 5 concurrent users than a multiplier is applied, see the table below for the corresponding multiplier:

Number of Concurrent Users	Multiplier
6-15	2
16-25	2.5
26-50	3
51-100	3.5
101-500	4.5
501+	Call for quote

Table A-6 Tele Atlas Pricing

Tele Atlas claims to have a level of accuracy in urban and suburban areas of 12 meters or better and also a level of accuracy in rural areas of 50 meters or better.

Tele Atlas uses several sources for the development and maintenance of their data, including:

- Municipal maps
- Ortho-enhanced aerial photos
- Demographic information
- Traffic signs
- Tourist information
- Roadmaps
- USGS Topographic 1:24K quad maps
- Department of Transportation maps
- USPS ZIP+4
- Line-of-Travel
- US Census TIGER data
- Differential GPS field data survey and collection
- Department of Transportation engineering diagrams
- Relationships with most local transportation authorities

Tele Atlas data is currently updated twice a year (June & December) and starting quarter one of 2002 updates will be quarterly. This is probably adequate for most agencies' needs, but it may be seen as a negative due to the long intervals between updates. Maintenance plans can be purchased at the time of initial purchase for the following costs:

- For a Single Year Maintenance Fee add 25% of Annual License Fee for semi-annual updates.

- For a Multi-Year Maintenance Fee add 0% for year one and add 10% for years two and three for semi-annual updates.

Region-wide Transportation GIS Project Design and File Architecture
Volume II – Technology Vision and Implementation Options

Publication No.: 02015

Date Published: March 2002

Geographic Area Covered: Delaware Valley Region

Key Words: Linear Referencing System (LRS), Transportation, Geographic Information System (GIS), Technology Ramp, DVRPC Member Governments, Operating Agencies, Transformation Methods, Geometric, NSDI Transportation Framework, NCHP 20-27, Common LRS, Centerline Development Options, Data Model, Prototype

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 II – Technology Vision and Implementation Options presents a matrix of data transformation technologies that can be used to perform data sharing between DVRPC member agencies. Prototypes were developed and tested to evaluate the transformation options including Geometric, Common LRS, NSDI Transportation Framework and NCHRP 20-27. Centerline development options are also presented and estimated costs for development are included in an appendix.

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