Summary of Final Report, BC353-21
December 2002

DEVELOPMENT OF GIS-BASED CONFLATION TOOLS FOR DATA INTEGRATION AND MATCHING

PROBLEM STATEMENT

Every year, the Florida Department of Transportation (FDOT) and many local and regional agencies conduct a large number of transportation planning studies. These studies are critical to the development of the transportation system in the state because they are useful in the identification of needs and prioritization of transportation improvement projects. Transportation planning studies, such as the long range transportation plan updates undertaken periodically by metropolitan planning organizations in each urbanized area, provide information that guide billions of dollars in transportation investment.

Transportation planning studies involve the use of transportation models and procedures that both require a large amount of data as input and generate a large amount of data as output. For example, many transportation planning studies need as input data describing the roadway network and the land use of an area. Planning procedures use these data to generate present and future traffic counts on every link of the roadway network as output. All of this input and output information is contained in a wide variety of databases for each jurisdiction in the state.

Transportation planning studies often require the exchange and integration of information across jurisdictions or databases. For example, FDOT needs current and future traffic counts for all roadway links on the state highway system to develop the annual work program, which is the list of highway construction projects to be undertaken in a given year. This information is available in local and regional transportation models and studies in a wide variety of formats and data structures. At this time, planners and engineers must spend considerable time, money, and resources to obtain, match, and integrate this data into the work program development process. There are no readily available tools to render this data matching and integration procedure (termed “conflation”) seamless and trouble-free.

Likewise, local agencies that develop long range plan updates need a host of data from the state-level databases for refining and validating transportation models and studies. For example, state traffic count data, roadway characteristics inventory data, and pavement condition data are available from the state. As it is extremely time consuming and resource-intensive to obtain, match, and integrate this data into local models and procedures, the need for computer tools that help to automatically match and integrate this data is being felt around the state.

OBJECTIVES

The purpose of this research is to make progress towards filling the need for a comprehensive set of conflation tools that allow the matching, merging, and integration of data and networks across disparate data sources. Specific objectives include the following:

- Identify data items, data sources, and planning/modeling applications in which data and network conflation would be needed

- Develop GIS-based conflation tools that allow the matching and integration of data items and networks from a variety of sources and formats and across varying levels of aggregation

- Provide a mechanism by which conflation tools can be easily updated and adapted as key data sources get modified and updated.

FINDINGS AND CONCLUSIONS

This research provides several key deliverables that would be of immense use to state and local agencies interested in conflating networks and databases. The developed conflation tools are GIS-based and operate on standard database formats so that they are applicable across a wide variety of contexts and applications. Some of the key findings/results are as follows:

- A comprehensive algorithm for network matching has been developed. This algorithm involves a system of geometric matching of nodes and edges across networks so that the correspondence between two networks can be established. The node matching and edge matching algorithms employ state-of-the-art distance and angle based measures and criteria to establish the correspondence between networks. The algorithm has been tested using a portion of the Tampa Bay Regional Planning Model network and the corresponding portion of the Tiger file. The algorithm is able to automatically match nodes and edges while providing the user the ability to check the matches at each step, make manual corrections, and then re-run the automated procedures to obtain complete match coverage.

- The algorithm developed in this research project is capable of providing a correspondence between networks so that attributes can be matched as well. The correspondence mechanisms accommodated in this research project include one-to-one, many-to-one, and one-to-many.

- A prototype conflation software program that implements the algorithm was developed in this study. The conflation software, called NETMATCH, is user-friendly software developed as an extension of ArcGIS 8.x. The software has a graphical user interface and maximizes the use of in-built functionality of ArcGIS 8.x. Users can utilize the software to develop correspondence maps between two networks. A simple user's manual was produced.

BENEFITS

This research project has resulted in the development of a set of computerized tools that provide for the seamless, trouble-free, efficient, and accurate transfer, matching, and integration of data across jurisdictions and among multiple databases and formats. These data matching and integration tools, often referred to as "conflation" tools, will greatly aid planners and engineers conducting planning studies, long range transportation plan updates, and work program development. The ability to match and integrate data and information from a wide variety of sources will greatly improve the accuracy and efficiency of planning studies in the state. In turn, this ability will lead to improved identification of

transportation system needs, project prioritization, and planning decisions. The traveling public will greatly benefit from improved planning decisions as scarce highway dollars are allocated judiciously to the most critical projects. They should see reduced travel times, decreased congestion, enhanced safety and mobility, and improved accessibility. The use of these tools will also result in dollar savings to agencies that are already short on planning funds. It is anticipated that agencies that use these data matching tools could save up to 20 percent on their planning studies. For example, a long range transportation plan update that currently costs about $500,000 may be accomplished for $400,000 by incorporating these tools into the planning process.

This research was conducted by Ram Pendyala, Ph.D., at University of South Florida. For more information, contact Vidya Mysore at (850) 414-4294, vidya.mysore@dot.state.fl.us.