Combining State Route and Local Road Linear Referencing System Information

Andrew J. Graettinger, Xiao Qin, Gabriel Spear,
Steven T. Parker, and Susie Forde

The Wisconsin Department of Transportation maintains two separate geographic information system databases: one for state roads and one for local roads. Both databases employ linear referencing system (LRS) theory to manage and locate information. Combining data from state and local roads into one system is desirable for the purpose of data management and analysis. This paper's approach combines information from the state route LRS (link ID and offset distance) with information from the local road LRS (node ID) to produce a table that can be employed to transfer data between the systems. A computer program was developed to use this table, along with existing tables in each LRS, to move information from the state route system to the local road system. Quality assurance and quality control techniques are presented along with a long-term implementation approach; both help to update the table bridging these two systems. The approach described does not interfere with the current operation of either system; therefore, no interruption of business practice will occur as this data transfer approach is deployed. Although this approach can transfer any LRS data, a case study of crash data for Dane County, Wisconsin, is used to demonstrate and test the approach. In the case study, crash points on the state routes are combined with local road crashes to produce a complete data set of crashes within the county.

The Wisconsin Department of Transportation (WisDOT) developed and currently maintains two geographic information systems (GISs) based on two linear referencing systems (LRSs). The State Trunk Network (STN) LRS is used for state-designated roads, while the Wisconsin Information System for Local Roads (WISLR) is applied to local roads. These LRSs evolved from different data, to meet different business needs, and both perform well for their intended purposes. STN focuses on state routes and contains no local roads, while WISLR contains both local roads and state routes, but maintains no data over state routes. There are approximately 12,000 and 100,000 mi of state routes and local roads represented in STN and WISLR, respectively. It is the intersection of these data sets (i.e., the 12,000 mi of state routes in both systems) which are of interest.

Transportation data are the reason for having a GIS, many of which employ LRS theory. Issues relevant to the use of data include convenience, cost, and reliability. Having state route data stored, analyzed, and displayed in STN while local road data is stored, analyzed, and displayed in WISLR can undermine the benefit of a GIS. With two LRSs focused on two different road networks, it is easy to envision how a statewide data set, such as crash data, could be less convenient, more costly, and less reliable than if that same data were in a single system. Dueker and Butler (*1*) provide additional information on the benefits and challenges with GIS data integration.

The objective of this work is to explore the extent to which linear locations in two overlapping LRSs can be reconciled to improve data convenience, reduce data cost, and improve data reliability. Although any point or line data could be used to develop and test an approach to unify transportation data from two LRSs, a case study using crash location data in Dane County, Wisconsin, is presented. Crash data integration is not an issue unique to Wisconsin and has been researched by others, for example, O'Neill and Harper (*2*). Currently, Wisconsin state route crashes are maintained in STN while local road crashes are managed by WISLR. For data convenience, it is desirable to combine these two data sets into one system for analysis and display. Because crash location data are entered on crash reports using an on–at linear referencing method, which is then hand translated to a reference point system for state routes, the potential to reduce data cost associated with hand coding crashes on state routes exists. Finally, because local road data, such as road names, are continually updated by local officials in WISLR, potentially more reliable information about state-to-local road intersections in STN can be envisioned with a unified approach.

BACKGROUND

Generalized Linear Referencing System

The field of transportation often uses an LRS to identify locations. The LRS specifies locations as a start point and an as-driven distance along a segment of a network. Ries (*3*) identified the advantages of an LRS when compared with map-based coordinate referencing for transportation data management. Because an LRS exists independent of a map, it is able to represent small features, such as an intersection with separated turning lanes, in more detail than a map can show. Different LRSs may vary in design, but all can be compared with a general conceptual form.

A report by the National Cooperative Highway Research Program, NCHRP 218 (*4*), describes a general conceptual LRS and how the LRS relates to both transportation data and cartography. Figure 1 shows the conceptual LRS (middle) along with transportation data

---

A. J. Graettinger, Department of Civil, Construction, and Environmental Engineering, University of Alabama, 260 H. M. Comer, Box 870205, Tuscaloosa, AL 35487-0205. X. Qin, G. Spear, and S. T. Parker, Wisconsin Traffic Operations and Safety Laboratory, B239 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706. S. Forde, Wisconsin Department of Transportation, 901 State Farms Transportation Building, 4802 Sheboygan Avenue, Madison, WI 53702. Corresponding author: A. J. Graettinger, andrewg@eng.ua.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2121, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 152–159.
DOI: 10.3141/2121-17

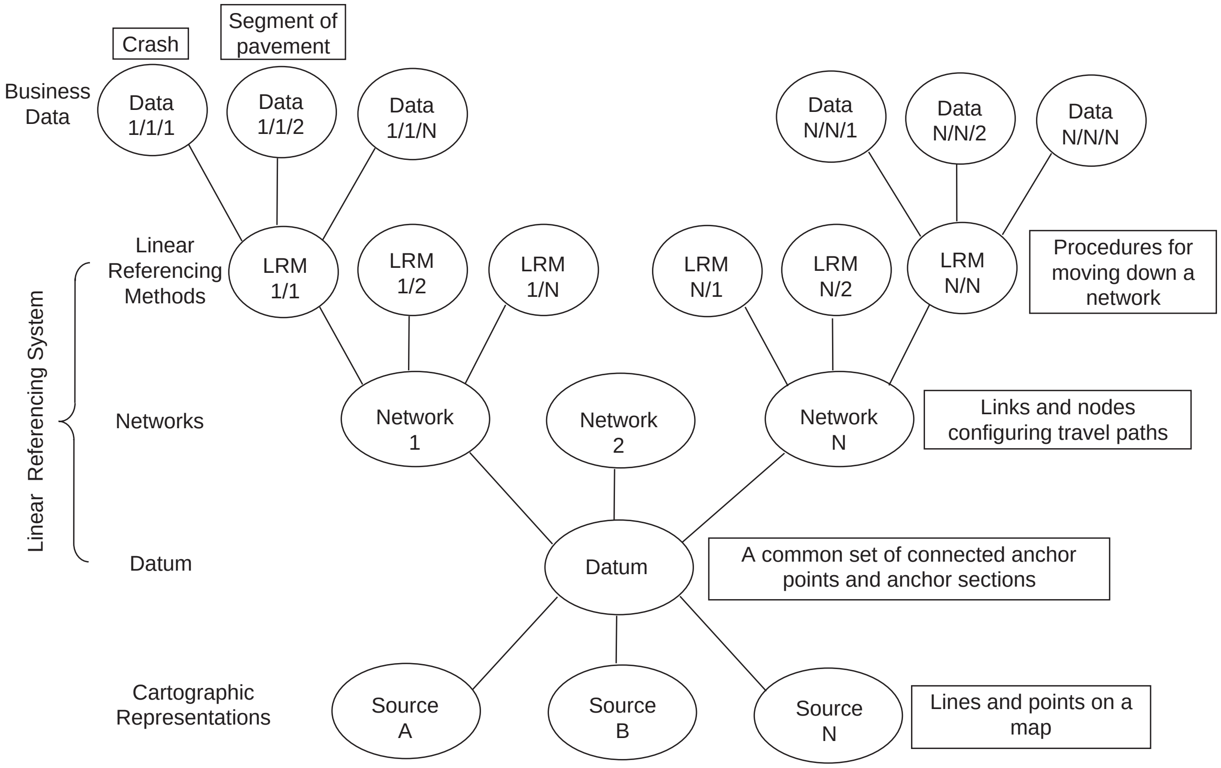


FIGURE 1 Transportation GIS with generalized LRS that allows for multiple methods (*4*).

(top) and cartography (bottom). The conceptual LRS is comprised of three functional levels, labeled along the middle left of Figure 1: (*a*) a datum establishing linear location, (*b*) a network or system of travel paths, and (*c*) a linear referencing method (LRM) employed for choosing start point and direction. A datum can be the reference for multiple networks and each network can have multiple LRMs, as indicated by "N" in Figure 1. Each branch of Figure 1 is a perspective on a virtual model of the road system, and multiple branches can exist and share data within a transportation organization.

The datum and associated networks provide the ability to perform data management activities. The datum can be thought of as a set of anchor sections and points that are connected. Typically these sections and points are widely spaced and would not represent all transportation facilities, but rather provide a spatial framework or structure to locate other more detailed information. The appropriate maximum length of the datum links can be determined through geospatial techniques. The only information that should be stored in a datum is the length of the anchor section (*5*).

Above the datum in Figure 1 is the network. The network stores topology or connectivity information about the roadway network. Part of a datum may represent a distance of roadway through many intersections, while the network at that same location would represent the intersections and also the segments of road between the intersections. Multiple networks can be included in a single LRS. One network may store information about local roads and another may store information about state routes. Any network in this conceptual model is referenced back to the common datum and therefore is automatically connected to all other networks.

On top of any network are linear referencing methods used to reference data to the network. An example of an LRM is the route–milepost method in which a location can be found by first identifying the route and then moving down the route to the correct milepost. Another LRM might be on–at–distance–direction where a location is described by an on road (on), a crossroad (at), a distance from that intersection (distance), and finally a direction (direction). Multiple LRMs can be used with a network and multiple networks can use the same LRM.

As shown at the top of Figure 1, business data sit on top of the linear referencing system and include entities such as pavement, bridges, crashes, and so forth. All of these can be located through an LRM to a network that is referenced to the datum; therefore, all information can be shared throughout this model.

Finally, visualization of the spatial relationships among these data can be accomplished through a cartographic representation shown at the bottom of Figure 1. There are no constraints on where map lines start or stop and overshoots and undershoots are allowed. The starting and ending points of the map line work is linearly located against the anchor sections through appropriate offset distances (*5*). Multiple maps having different line work can exist in one LRS.

WisDOT Referencing System

WisDOT first developed a GIS called STN as a result of efforts dating back to the early 1990s. STN maintains data for state-designated roads, including state and federal highways and several off-mainline facilities such as frontages and rest areas. The creation of STN involved hand compiling a network of links and nodes, representing travel paths and state-to-state road intersections, respectively. As-driven distances along links were then measured in the field, a data collection procedure that WisDOT still employs.

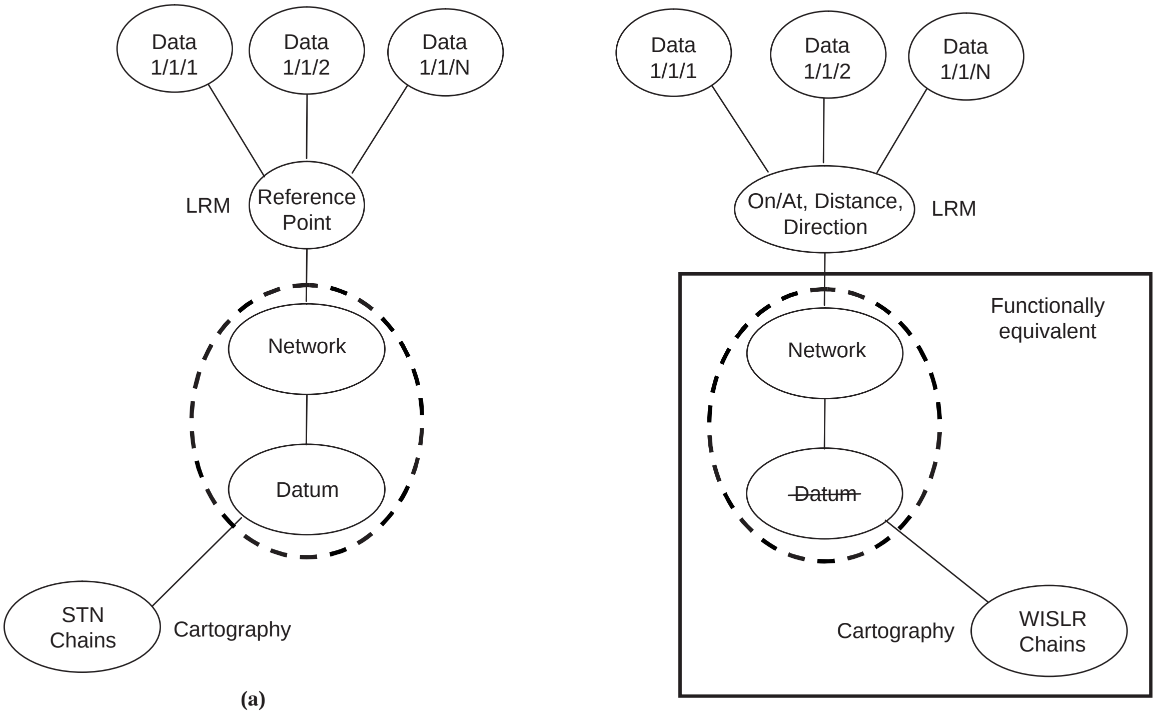
Approximately 10 years after STN, WisDOT created WISLR. The WISLR LRS was generated from digital cartography of both state and local roads in Wisconsin. The digital cartography was assembled from existing local, state, and federal sources. The cartography was then broken at intersecting lines and nodes were added. Each line was converted into a pair of links pointing in opposite directions representing the two directions of travel. This method worked well for local roads but led to some erroneous situations along divided highways and at overpasses. Linear distances along local roads in WISLR are collected by WisDOT personnel, whereas linear distances along state routes are not maintained in WISLR, because WisDOT already maintains state route distance data in STN.

The NCHRP conceptual form may be applied to both STN and WISLR as shown in Figure 2. STN can easily be described within the NCHRP linear referencing model with the exception that the network and datum are the same in STN. Notice that there is only one network associated with each system, in contrast to many potential networks in the NCHRP conceptual model in Figure 1. Because there is only one network per system, information required for both the network and datum could be combined into a single “link” table with rows identifying every link, link length, and link from node and to node. Because the WISLR network (links with nodes at intersections) was derived from cartography, the network and cartography are essentially interchangeable and functionally equivalent. The lack of as-driven distances over the state roads in WISLR means that there is technically no datum in WISLR.

STN and WISLR also differ in the LRM employed, as shown in Figure 2. STN uses a reference point LRM based off of a route and a set of ground-identifiable locations familiar to WisDOT data collectors. In contrast, WISLR uses an on–at–distance–direction LRM. WISLR start points for data locations are intersections identified by road names. The on–at LRM is intuitive to data collectors and also to GIS users performing queries; however, there is difficulty applying the on–at LRM to state roads because route names and crossroads are not always obvious for complex areas such as an interchange. This shortcoming of WISLR is acceptable because an accurate system for state routes, STN, exists.

To better understand the differences between STN and WISLR, it is convenient to look at the same real-world objects represented differently in each system. Figure 3 compares STN and WISLR to each other in Figure 3*a* and to an aerial photo of the same interchange in Figure 3*b*. In Figure 3*a* STN is shown as straight links terminating at square STN nodes, while WISLR is shown as curving links terminating at circular WISLR nodes. Because this interchange is on a state route, it is represented in both systems. Each WISLR link shown in Figure 3*a* is actually two overlapping links representing two directions of travel, even if only one direction of travel truly exists. An analysis of the interchange in detail shows that at location A in Figure 3*a* a WISLR node exists at a bridge where no turns are possible. At location B the WISLR node connects to both directions of travel along the highway, even though turning onto the ramp is only possible from one direction. In contrast, no WISLR node is found for the convergence of ramps at location C because the cartographic lines do not intersect. These examples show how networks can be configured differently even when describing the same area.

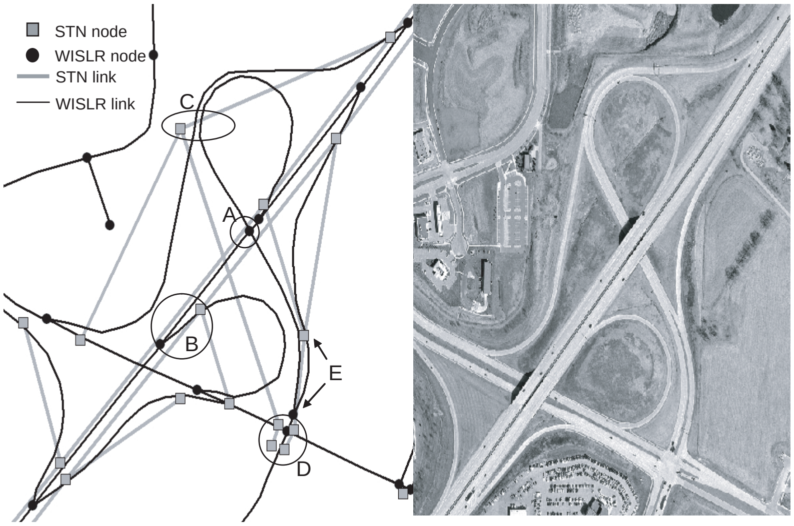
Figure 3*a* also shows the influence of map scale and resolution. The intersection at location D is represented as a single node in WISLR, whereas it is represented by four nodes and two links in STN. Resolution also has a significant influence over the location of tapered intersections that often occur on state roads. As a result there is uncertainty in the location of ramp intersections, as shown at location E in Figure 3*a.* These examples show how specific locations are defined differently in STN and WISLR. These “definitional” differences can potentially affect linear locations along all links.



(a)

(b)

FIGURE 2 NCHRP conceptual form applied to (*a*) STN and (*b*) WISLR.



(a) (b)

FIGURE 3 Interchange with corresponding STN and WISLR networks.

APPROACH

Reconciliation of the differences between STN and WISLR along state routes requires two steps. First, the starting point from which data are located must represent the same real-world location in either system. Second, an appropriate offset value in the LRS receiving the data, WISLR, must be determined. To identify starting points in each system, the location of each WISLR link with respect to the STN system must be determined. In addition, the definitional differences caused by scale and resolution wherever starting point nodes in STN and WISLR represent the same location also must be reconciled. After identifying an appropriate starting point, the data offset value must be prorated into a WISLR map distance value. This involves knowing the lengths of WISLR links in the STN measurement system.

Knowing the location of every WISLR node along state routes based on an STN link and an STN as-driven distance creates a "bridge" between STN and WISLR. Fortunately, some STN to WISLR data already existed in a STN table called Site_STN. The Site_STN table consists of three columns: STN link ID, WISLR node ID, and STN offset distance. While not all necessary STN link–WISLR node combinations were present in the Site_STN table, the table provided initial information to build on and develop a usable relationship between STN and WISLR.

Site_STN Approach

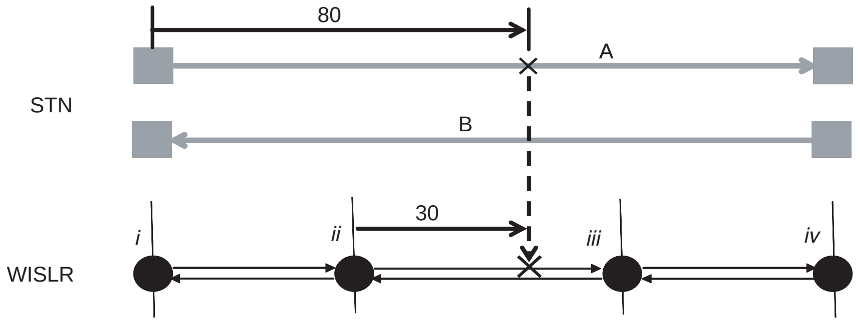
An approach was developed that utilizes the Site_STN table to move STN point data from STN to WISLR. Figure 4 illustrates the approach in the most fundamental case where multiple WISLR links overlap one STN link. Figure 4 shows a fictitious stretch of state route that is represented in STN (top of Figure 4) as two links indicating two directions of travel, while the same stretch of road is represented by six links terminated by four nodes in WISLR. The additional WISLR nodes represent state-to-local road intersections (black dots in Figure 4) that are absent from STN (gray lines in Figure 4).

In this example, the data object in STN represents a crash on a state route, although any point or even line data could be transferred in a similar manner. The crash is originally located in the STN crash inventory table by STN link A and an offset of 80. Using the STN link from the crash inventory table, the approach first finds the rows in the Site_STN table that correspond to the STN link, which are shown in bold in the link column in Figure 4. These selected rows are then sorted by offset value, and the two rows that the crash falls between are selected, shown in bold in the offset column in Figure 4. The two WISLR nodes (*ii* and *iii*) associated with the selected offset values (50 and 100) are the from-node and to-node of the WISLR link on which the crash. Subtracting the from-node offset distance of the WISLR node from the STN crash offset distance ($80 - 50 = 30$) produces the new crash offset in STN as-driven distance. The offset still must be prorated into a WISLR map-based distance, which is done by accessing the length of the WISLR link in WISLR.

Figure 5 illustrates an alternative situation in which there is no WISLR node at an intersection that is represented in STN. To avoid negative offset values in the Site_STN table, a convention was used to include only the upstream STN link in the Site_STN table and add the length of the preceding links together for the from-node offset. Therefore, link B does not appear in the Site_STN table and link A has an offset distance that is a combination of the lengths of links A and B.

For the example crash data in Figure 5 that are located 20 units down link B, the approach initially searches for link B in the Site_STN table. When link B is not found in the Site_STN table, the preceding link that has a to-node that is the same as the from-node of the link where the crash occurred is identified. The preceding link ID, in this case A, is then used to select data in the Site_STN table, shown in bold in the STN link column in Figure 5. The crash offset is recomputed by adding the length of link A to the crash offset distance, $50 + 20 = 70$, and the new crash location can then be transferred to WISLR. The map-based distance must then be calculated to accurately display the crash location.

Differences caused by map scale are also resolved with this approach. These situations often occur at exit ramps, such as the portion of a diamond interchange shown in Figure 6. In the center of the figure is STN link C, which represents a small road segment crossing a divided highway. WISLR represents that same location with



Crash Inventory

STN Link	Offset
A	80

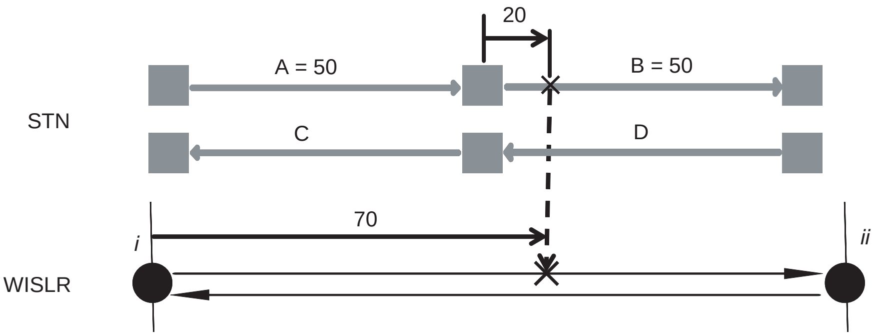
Site_STN Table

STN Link	WISLR Node	Offset
A	i	0
A	ii	**50**
A	iii	**100**
A	iv	150
B	i	150
B	ii	100
B	iii	50
B	iv	0

FIGURE 4 Site_STN program calculates WISLR location based off of Site_STN table.

a single node, node *ii*. Link C is not in the Site_STN table so the approach identifies the preceding link, which is link I or link B in this example. Link B could be considered a preceding link, but link B is not in the Site_STN table; therefore link I is the preceding link for link C. The offset calculated for data occurring along link C will be greater than the offset of the last row for link I in the Site_STN table; the length of link I plus link C is obviously greater than the length of link I alone. The approach will therefore replace the calculated offset with the offset of the last row for link I. This produces the best location for STN data in the new WISLR LRS. This example does not result in a loss of linear precision because scale is being compressed across the road and not along the road.

Referring again to Figure 6, STN link E represents a turn lane that is not included in the WISLR system. As a result, the as-driven lengths of WISLR links *i* to *ii* and *ii* to *iv* are ambiguous because the node at the start of links B and E as well as the node at the ends of links E and F are not present in WISLR. The relatively short STN links in Figure 6 (B, C, F, and E) may all be treated in the same way



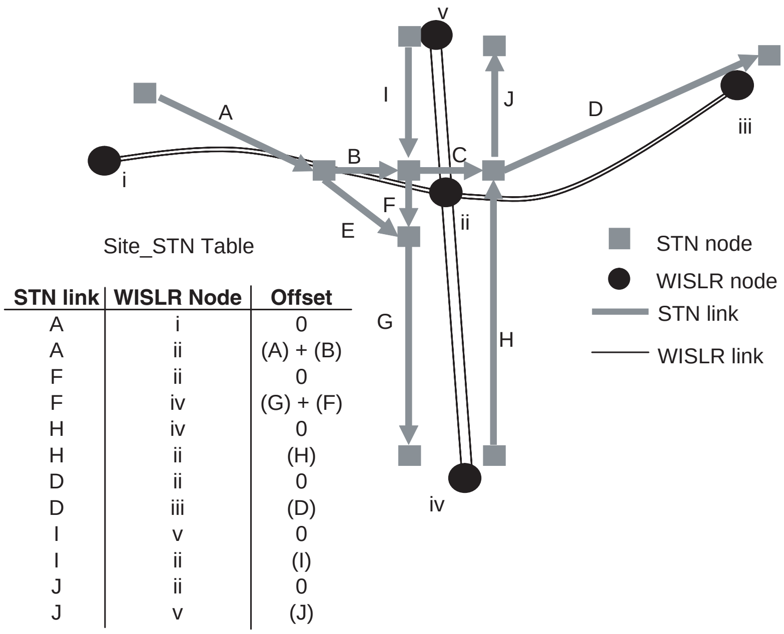
Crash Inventory

STN Link	Offset
B	20

Site_STN Table

STN Link	WISLR Node	Offset
A	i	0
A	ii	100
D	iii	0
D	iv	100

FIGURE 5 Site_STN program calculates crash location by adding preceding link length to crash offset.



Site_STN Table

STN link	WISLR Node	Offset
A	i	0
A	ii	(A) + (B)
F	ii	0
F	iv	(G) + (F)
H	iv	0
H	ii	(H)
D	ii	0
D	iii	(D)
I	v	0
I	ii	(I)
J	ii	0
J	v	(J)

FIGURE 6 Differences between STN and WISLR caused by map scale are handled in Site_STN table.

as link C and compressed to node *ii* in WISLR; however, this would result in a loss of precision. The best match to the WISLR map here excludes the turning lane in favor of STN links B and F as shown in the Site_STN table in Figure 6, which does not include any information for link E. By using only one network trajectory to compute offsets, both missing STN nodes have stable locations in WISLR. Data on link E will occupy the same WISLR location as that of link B, and data from links F and G will be distributed along link *ii* to *iv* in WISLR.

WISLR nodes that represent the ending or beginning of an STN link or combination of links are in the Site_STN table multiple times. This is shown in the Site_STN table in Figure 6, where WISLR node *ii* is in the table six times: once for each of the eastbound STN links (links A and B combined and link D), and once for each of the four north–south STN links (links I, J, and H and the combination of links F and G).

The data integration approach based on the Site_STN table was theoretically evaluated at many STN–WISLR inconsistent situations. The approach performed robust enough to warrant coding and testing on actual data. The Site_STN table was populated for example road segments and the data integration program was coded and tested on that data.

Quality Assurance/Quality Control

To help ensure an error-free development of the Site_STN table a second program was coded for detecting errors. The error-checking program first checks each STN link in the Site_STN table for the appropriate STN starting and ending offsets against the length of each link in STN. Next the program checks for redundancy: the same node at different offsets or different nodes at the same offset. Finally the error-checking program checks the WISLR nodes and offsets in the Site_STN table against the order of links along a road segment in WISLR. This determines if the WISLR nodes will appear in the correct order when sorted. An error report is produced listing STN links and WISLR nodes that have an associated error. These locations can then be selected and manually corrected.

A second technique to check the accuracy of the Site_STN table, as well as the data integration program, was developed. This technique employs visual inspection to evaluate point data coming from STN and moving to WISLR. To ensure that the data integration approach functions everywhere, a data point is placed every 100th of a mile on every STN link. This data point is then moved by the Site_STN data integration program to the appropriate WISLR link and a line is drawn between the two points. Figure 7 shows an example section of a state route where 100th of a mile STN data points (shown as squares) are successfully moved to WISLR locations (shown as circles). It can be seen by looking at the circles in Figure 7 that the divided highway is represented by a single line in WISLR in the right portion of the figure and as a divided line in the left portion of the figure. The lines between STN and WISLR connect matching

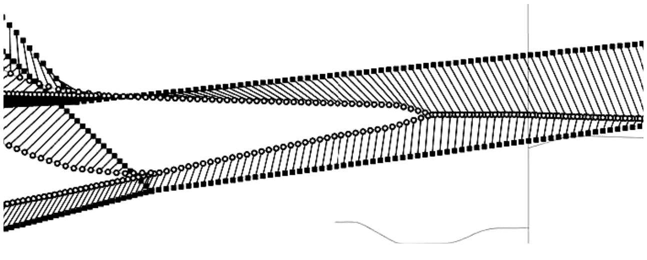


FIGURE 7 Connecting lines between STN data points (squares) moved to WISLR locations (circles) for quality assurance/quality control (QA/QC) along a divided highway.

pairs of points. In this way, the initial location of a data point in STN and the final location of that same data point in WISLR are graphically shown. Errors in the location of points along WISLR can then be identified by simply looking for inconsistencies in the pattern of connecting lines between STN data points and the corresponding WISLR data points.

Implementation

The approach and computer programs developed in this research are capable of achieving the primary goal of data integration, but the Site_STN table is static and therefore will become outdated as new as-driven distances are entered into STN. Additional resources would be required to maintain the Site_STN table as the LRSs evolve over time. This creates the problem of added cost that this data integration approach causes. Fortunately in STN, there exists a state-to-local road intersection table that is already updated with respect to the location of an intersection on an STN link at a specific offset. This table was leveraged to incorporate STN updates into the Site_STN table directly, thereby reducing the need to manually update the Site_STN table.

Starting with the STN intersection table, which contains intersection ID, STN link, and STN as-driven distance, a second table was created that contains WISLR node ID and STN intersection ID. The combination of these two tables provides all the information needed to create the Site_STN table automatically, that being STN link, WISLR node, and STN as-driven distance. To create the STN table automatically from these tables, WISLR nodes that occur at the start or end of a STN link have to be in the Site_STN table multiple times, as shown in Figure 6 for WISLR node *ii*. Therefore, WISLR nodes that occur at the beginning or the end of an STN link or combination of links have to be identified. This is done by accessing the length of each STN link and determining which WISLR nodes occur at the start or end. Then every STN link that shares the same location is found by using the from-node and to-node information that is in the STN link table. A row is then created in the Site_STN table for each STN link that is found.

CASE STUDY

To demonstrate the capabilities of the described approach, a case study was prepared for crashes that occurred in Dane County, Wisconsin, in 2006. Figure 8 shows a map of Dane County with state routes in WISLR as black lines and local roads in WISLR as gray lines. In Dane County there are 1,449 STN state route links, 3,091 WISLR links along state routes, and 1,367 WISLR nodes along state routes. From this information, a Site_STN table was populated for Dane County that contains 3,835 records. The Site_STN table was error checked with the Site_STN error-checking program and also the point transfer capability was visually inspected by moving 80,000 data points, one point for every 100th of a mile along every STN link, from STN to WISLR. After populating the Site_STN table for Dane County and performing the quality assurance/quality control (QA/QC), 3,300 state road crashes were transferred from STN to WISLR. On top of the line work in Figure 8 are WISLR local road crashes (black dots) and STN state route crashes (gray dots) that have been transferred from STN to WISLR.

The combined data set of local road crashes and state route crashes are all in the WISLR LRS. This improves data convenience and analysis because all crash information is in one system. In addition, the combined data reduce cost because the on–at LRM is available

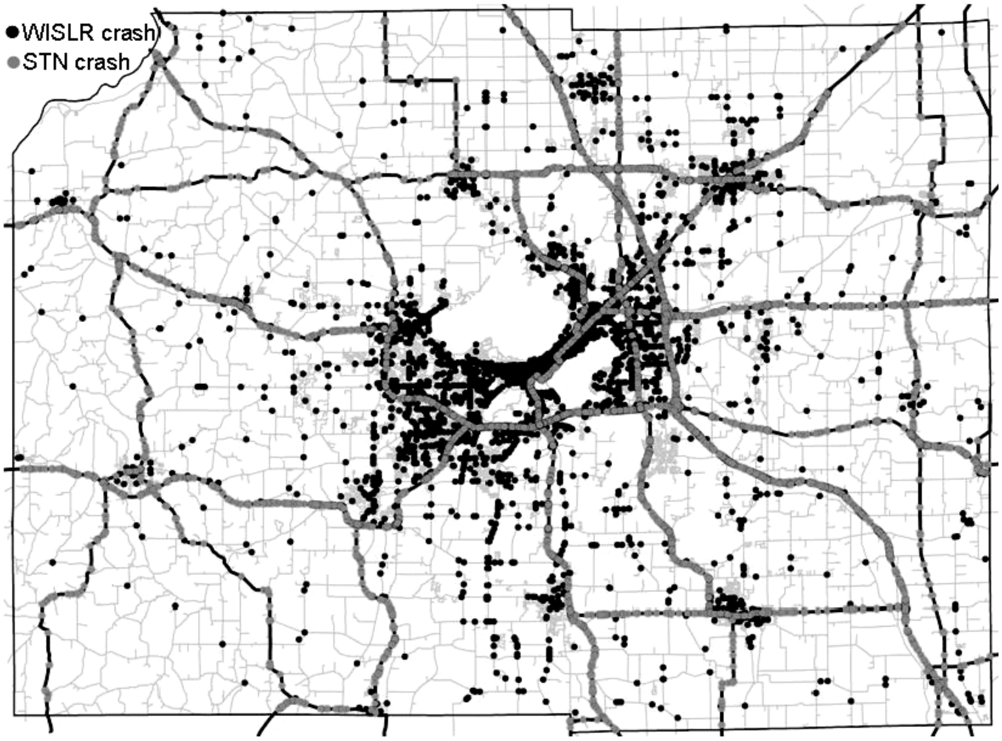


FIGURE 8 Map of combined state and local road crashes in one database for Dane County, Wisconsin.

to identify crash locations. Finally, the combined data improve reliability because road names are continually updated by local officials in WISLR.

CONCLUSION

Two well-designed LRSs were cost-effectively created from different data sources to meet different business needs within WisDOT. Cost-effective data management often involves making use of existing resources. The approach presented here achieves data integration between these two systems without any fundamental redesign of the independent GIS databases. A program was coded to transfer data from state routes in one system to the same state route location in the other system. QA/QC techniques were developed to verify the accuracy of the approach. A method to reduce the burden of updates to the supporting database table was also developed. A case study in Dane County, Wisconsin, was used to demonstrate the data integration approach. The transfer of data has the potential to expand on current applications and analysis, which is expected to increase user demand and in turn provide motivation for improved data management. It is asserted that the most valuable service a database can provide is complete, reliable, and up-to-date information for real-world applications. The approach presented in this paper was demonstrated on the problems associated with crash data mapping, and the approach achieved the stated objectives.

ACKNOWLEDGMENT

The authors acknowledge the Wisconsin Department of Transportation for financial support for this project.

REFERENCES

1. Dueker, K., and J. Butler. A Geographic Information System for Transportation Data Sharing. *Transportation Research Part C: Emerging Technologies,* Vol. 8, No. 1, 2000, pp. 13–36.
2. O'Neill, W. A., and E. Harper. Location Translation Within a Geographic Information System. In *Transportation Research Record 1593*, TRB, National Research Council, Washington, D.C., 1997, pp. 55–63.
3. Ries, T. Design Requirements for Location as a Foundation for Transportation Information Systems. *Proc., Geographic Information Systems for Transportation (GIS-T) Symposium,* Albuquerque, N.Mex., AASHTO, 1993, pp. 48–66.
4. Opiela, K. S. *NCHRP Research Results Digest 218: A Generic Data Model for Linear Referencing Systems.* TRB, National Research Council, Washington, D.C., 1997.
5. Scarponcini, P. Generalized Model for Linear Referencing. *Proc., 7th ACM International Symposium on Advances in Geographic Information Systems,* Kansas City, Mo., 1999, pp. 53–59.

The contents of this paper reflect the authors' opinions and do not necessarily reflect the official views of the Wisconsin Department of Transportation.

The Statewide Transportation Data and Information Systems Committee sponsored publication of this paper.