GIS PROCESS OVERVIEW
for the
RAILROAD LINE-HAUL CAPACITY PROJECT

Geographic Information Systems Group
Engineering Services
Tennessee Valley Authority
Norris, Tennessee

May 1997

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AML - Arc Macro Language

BTS - Bureau of Transportation Statistics

DEM - Digital Elevation Model

FRA - Federal Railroad Administration

FTP - File Transfer Protocol

GIS - Geographic Information System

NTAD - National Transportation Atlas Databases

ORNL - Oak Ridge National Laboratory

USGS - United States Geological Survey

INTRODUCTION

Members of the TVA Norris GIS Group accepted a project from the TVA Navigation Team in February 1997 to assist them in determining the line haul capacity of selected railroad lines in the United States. The objective of the GIS phase of the project was to merge *attribute* information from multiple transportation and topographic data sources. This was a pilot project to be accomplished in the least amount of time and finances possible — not to provide a topologically correct routing network.

The primary attributes requested by the customer were:

- specialized route identification numbers
- railroad ownership names / abbreviations
- USGS Digital Line Graph major and minor attribute codes
- density categories
- signaling system types
- slope information
- railroad grade crossing identification numbers and street names

A specialized route identification number was manually added by an undergraduate student interning with the TVA Navigation Team. The railroad ownership names and major and minor attribute codes were taken from a 1:100,000 scale railroad network provided by the Bureau of Transportation Statistics. The density and signaling information was taken from a 1:2,000,000 scale railroad network also provided by BTS. Slope information was calculated from USGS Digital Elevation Model data. Railroad grade crossing data were acquired from the Oak Ridge National Laboratory.

Because of the lack of common attribute information (no key fields), it was necessary to use a Geographic Information System to *spatially* join each database together. For instance, the 1:2,000,000 scale network arc attributes were joined to the 1:100,000 scale network arc attributes based on their proximity. Figure 1 and Tables 1, 2, and 3 illustrate the process of joining an arc from the 1:2,000,000 scale network to an arc from the 1:100,000 scale network. An example arc (Arc #1) from the 1:100,000 scale network is shown in Figure 1 and its attributes in Table 1. An example arc (Arc #99) from the 1:2,000,000 scale network is also shown in Figure 1 and its attributes are depicted in Table 2. In this example, Arc #99 is the arc nearest to Arc #1, therefore its attributes are appended to the Arc #1 attributes. The resulting attribute table is shown in Table 3.

Members of the GIS Group used this type of process to merge all of the initial databases together to produce the final output for the project (Figure 2). The GIS Group used Arc/Info® 7.0.4 and ArcView® 2.1 running on a network of Sun Ultra Workstations. The final digital data files were transferred to the customer on a network Pentium PC.

Unlike most GIS tasks, the final products of this pilot project were listings of attribute information only. In most GIS transportation applications, the primary objective is to produce a topologically correct network at a maintained scale. In this case, the emphasis was not on the connectivity of the geographic data, but on the amount of time taken to merge the attribute

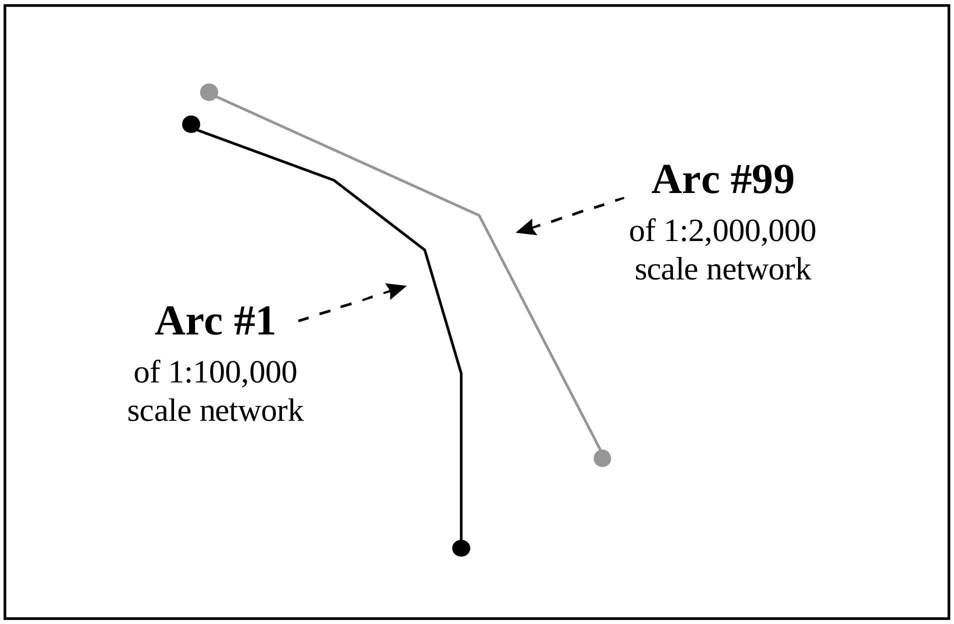

Figure 1. Spatail Join Example

Arc100k #	Route #	Owner	Major Code	Minor Code
1	2462	WC	180	208

Table 1. Example 1:100,000 Scale Railroad Network Attributes

Arc2m #	Density	Signaling
99	1.0	Manual

Table 2. Example 1:2,000,000 Scale Railroad Network Attributes

Arc100k #	Route #	Owner	Major Code	Minor Code	Arc2m #	Density	Signaling
1	2462	WC	180	208	99	1.0	Manual

Table 3. Resulting Railroad Join Attributes

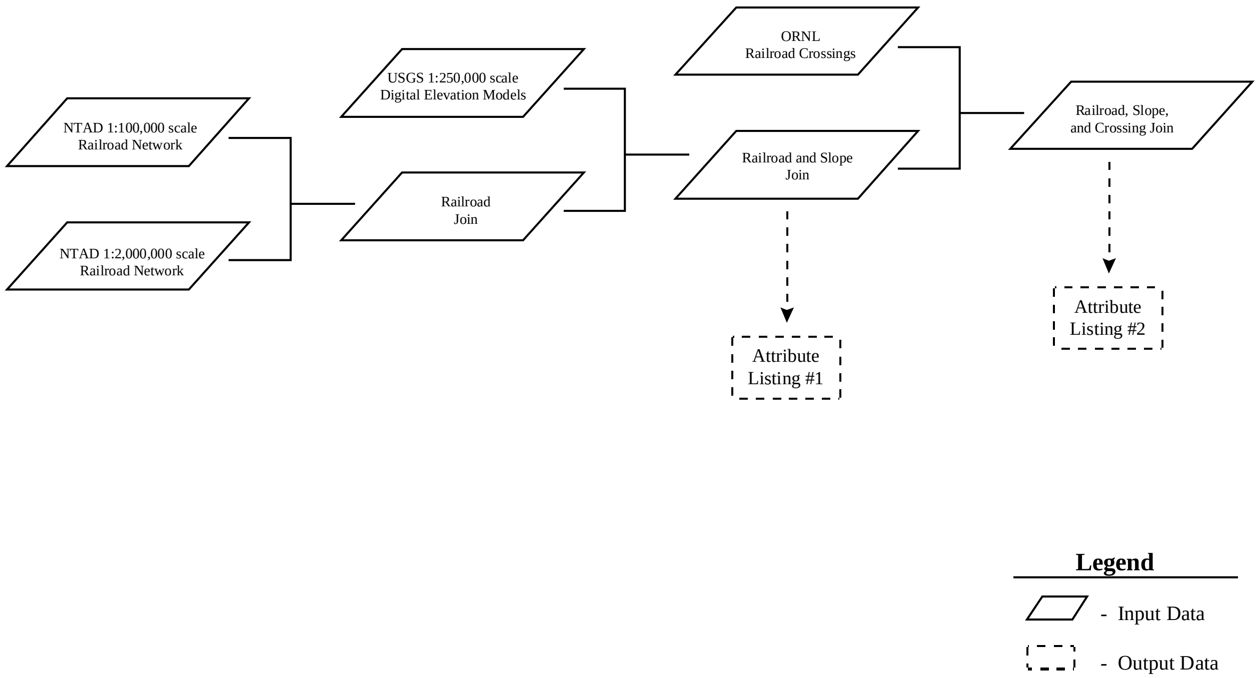

Figure 2. Multiple Joining of Input Data to Produce the Final Output Data

information together. Therefore, although the 1:100,000 scale railroad network did not maintain connectivity, it was chosen as the base network for the project since the 1:2,000,000 scale network did not contain secondary routes. For the next phase of the project, however, a topologically correct 1:100,000 scale railroad network should be available.

INPUT DATA

There were four main input data sets used for the project:

1). 1:100,000 scale railroad network taken from the 1995 National Transportation Atlas Databases (NTAD) compact disc. The CD was ordered via the Department of Transportation's Bureau of Transportation Statistics website: **http://www.bts.gov**
C code and ARC Macro Language (AML) routines were written to import the data into Arc/Info®.

2). 1:2,000,000 scale railroad network taken from the 1996 National Transportation Atlas Databases (NTAD) compact disc. The CD was ordered via the Department of Transportation's Bureau of Transportation Statistics website: **http://www.bts.gov**
The data was imported into Arc/Info® using an AML macro downloaded from the internet (btsarc.aml) and a user written AML macro routine.

3). 1:250,000 Digital Elevation Models downloaded from the United States Geological Survey website: **http://edcwww.cr.usgs.gov/glis/hyper/guide/1_dgr_demfig/index1m.html**
The DEMs were downloaded from the internet and copied to recordable compact disks. Another set of CDs was also made which contained only those DEMs thought to be necessary for the project. AML macros were written to copy each of these DEMs from CD to a disk drive, uncompress them, and use the Arc/Info® DEMLATTICE command to convert them to an Arc/Info® LATTICE.

4). Railroad grade crossing data received from Bruce Peterson of the Oak Ridge National Laboratory via FTP. The railroad crossing data were imported into Arc/Info® manually using Info™ commands.

In addition, the Navigation Team student used an Arc/Info® coverage of the 1995 NTAD Place Names provided by the Norris GIS Group, and a list of railroad routes along with a PC Rail© network provided by the customer.

PROCESS OVERVIEW

A simplified graphical description of the GIS process is shown in Figure 3. Crucial network routes were first extracted from the NTAD 1:100,000 scale network to create a new, reduced network. Attributes from the NTAD 1:2,000,000 scale network were then joined to the new 'crucial route network'. Slope attributes were calculated for each arc in the new network and an output listing was created which contained all attribute information for every arc. Afterwards, the network arc attributes were joined to the railroad crossing point attributes and another output listing was created. Both output listings were then delivered to the customer.

Selecting Crucial Routes

A list of crucial railroad routes was defined and provided by the customer along with a PC Rail© railroad network to the Navigation Team undergraduate intern. For each route on the list provided, the intern used the origin, destination, and ownership names to visually locate the route on the PC Rail© network on a desktop PC. A Sun workstation running Arc/Info® was used to visually locate the identical route on the 1:100,000 scale network. The arcs for the route were <u>select</u>ed[1] in ArcEdit and <u>put</u> into (appended to) a new data layer. A unique identification number was manually assigned to each route via the listing received from the customer. The final Arc/Info® output coverage containing the crucial routes was then given to the Norris GIS Group. Because of some of the following problems, not all routes were matched.

1. The two networks were not displayed in the same projection. The 1:100,000 scale network was in a Geographic coordinate system (latitude and longitude in decimal degrees) and the PC Rail© network projection was unknown.
2. There were discrepancies amongst railroad ownership names.
3. The topology differed between the two networks.

[1] Arc/Info® commands are underlined

GIS PROCESS OVERVIEW

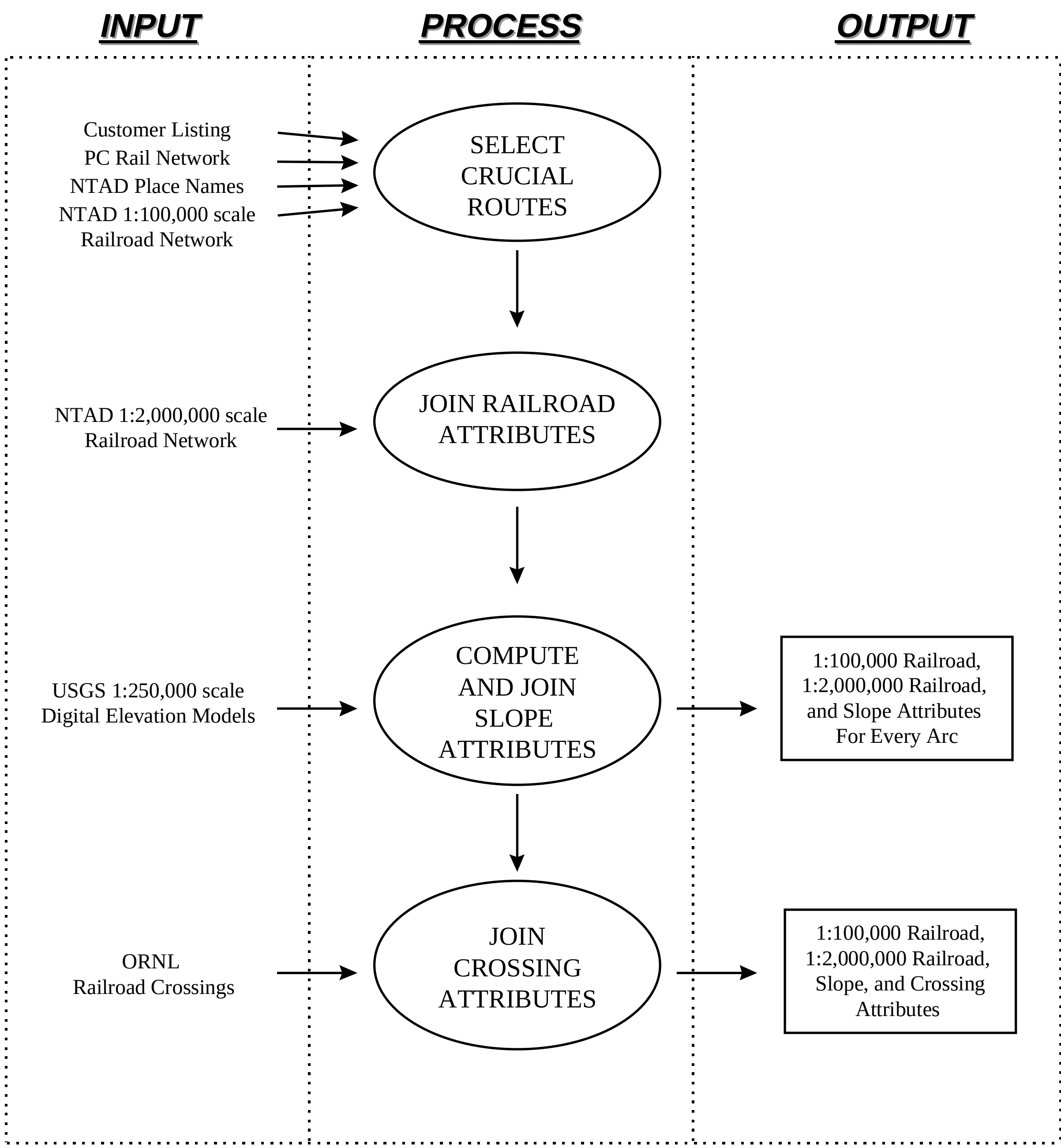

Figure 3. GIS Process Overview

Joining Railroad Network Attributes

After receiving the crucial route network, the Norris GIS Group joined the 1:2,000,000 scale NTAD railroad network attributes to it through a two step procedure. First, the following Arc/Info® commands were used to automate matching the attributes:

- ARCLABEL - to create a coverage containing the midpoint of each arc in the crucial route network.
- BUILD - to build the point topology for the coverage.
- NEAR - to place a pointer in the midpoint coverage to the nearest 1:2,000,000 scale arc (within a specified tolerance).
- JOINITEM - to join the 1:2,000,000 attributes to the midpoint coverage.
- JOINITEM - to join the midpoint coverage attributes (now containing the 1:2,000,000 attributes) back to the 1:100,000 scale crucial route network.

The primary challenge encountered in the first step was to determine the tolerance level setting so that as many attributes as possible from the 1:2,000,000 scale NTAD network could be joined without creating incorrect matches at intersections or near parallel lines. The poor topology (lack of connectivity and duplicate arcs) of the original NTAD 1:100,000 scale network was also a factor. The Arc/Info® CLEAN command was used in an attempt to lessen the problem.

After finishing the automated procedure, the second step was to make a visual pass of the network and manually correct any problems, i.e. verify that the correct 1:2,000,000 scale attributes had been joined. The crucial route network was divided into two separate coverages so that two GIS technicians could correct it simultaneously. AML macros and menus were written to aid the technicians in transferring attributes. Attributes for arcs in which a match could not be determined were set to zero. Problems encountered were mainly due to the differing topology and scale between the 1:100,000 scale crucial route network and the 1:2,000,000 scale NTAD network. The crucial route network was re-appended upon completion of the manual corrections.

Computing and Joining Slope Attributes

The next phase of the project was to compute the slope for each arc in the crucial route network and join the slope attributes to the network. As mentioned before, AML macros were used to copy each DEM from CD to a disk drive, uncompress them, and convert them to an Arc/Info® LATTICE. The slope (in percent) was then computed for each LATTICE using the Arc/Info® GRID function SLOPE. These slope LATTICEs were written to a set of 8 recordable compact disks.

ArcView® was used to review each file on the CDs for anomalies. Many of the files had ‘streaks’ which originated from the USGS data collection procedures, but they were not corrected (filtered) as part of the pilot project because of time constraints. There were also two anomalous rectangular areas originating from the downloaded DEMs. One was near Texas in DEM files: Brownfield-E, Clovis-E, Lubbock-W, and Plainview-W. The other was in the Norfolk-W file. Therefore, slope was not computed for network data overlaying these areas.

Arc/Info® was used to extract slope data from multiple points along the crucial route network. The following Arc/Info® commands were used:

- PROJECT - to place the network in the same coordinate space as the slope LATTICE files. (Also, the USGS quad map boundaries and names were projected so they could be used as background data).
- DENSIFYARC - to place a vertex at least every 90 meters along the crucial route network (since the slope data was based on 90 meter DEM data).
- ARCPOINT - to create a point coverage from all of the nodes and vertices contained in the crucial route network.
- BUILD - to build point topology for the new point coverage.
- SELECT & PUT - to manually divided the point coverage into smaller coverages to correspond with each DEM slope LATTICE. (Because of the large amount of data, it was necessary to process the slope data one file at a time.)
- LATTICESPOT - to extract slope values for each point along the crucial route network, therefore providing a slope value at least every 90 meters. This command was used in a series of AMLs which cycled through each point coverage alphabetically and extracted the slope data values from the set of 8 CDs.

The Tables module of Arc/Info® was used to reselect all data that did not have undefined slope values and unload them into ASCII text format files. UNIX commands were used to concatenate all of these files into one large file. The file was imported into ArcView® and the Summary Table Definition function was used to compute the minimum, maximum, variance, and average slope for each arc identification number. This tabular data was exported as an Info™ file and the Arc/Info® JOINITEM command was used to permanently join the slope information to the crucial route network. ArcView® was used to sort the network by arc ID number, add a flag for determining railroad crossing availability, and export all the arc information in ASCII comma-delimited and also dBASE format.

During this phase of the project a few files had to be reprocessed (mostly because of incorrect file names), but the main challenge was managing disk space. The GIS Group used one 4 gigabyte hard drive and four 2 gigabyte hard drives, as well as a CD writer, two CD readers, and an 8 mm tape drive.

Joining Railroad Crossing Attributes

Many of the railroad crossing data points received from ORNL did not have latitude and longitude information and, consequently, were deleted. The crucial route network attributes were then joined to the existing railroad crossing data points using the following Arc/Info® commands:

- NEAR - to place a pointer in the railroad crossing coverage to the nearest crucial route network arc (within a very small tolerance).
- JOINITEM - to join the crucial route network attributes to the railroad crossing coverage.

ArcView® was used to sort the railroad crossings by their associated arc ID number and export all the point information in ASCII comma-delimited and also dBASE format.

OUTPUT DATA

The following data were produced from the pilot project:

- 2 CDs containing USGS DEMs in GNU Zip compression format
- 2 CDs with pilot project DEMs in GNU Zip compression format
- 8 CDs with slope data for the project in Arc/Info® LATTICE format
- 9 sets of 8mm archival tapes containing pilot project data
- 2 final output files:

    1). A file in dBASE format containing attribute information from all the possible arcs considered important for calculating the line haul capacity of selected railways. See Appendix A for attribute descriptions.

    2). A file in dBASE format containing attribute information from all the railroad grade crossing points located near crucial route arcs and the attribute information from those arcs. See Appendix B for attribute descriptions.

The customer imported the two final output files into SAS, deleted any unnecessary fields, and merged the data together with other FRA data to perform the final analyses. The customer was made aware that the final output contained 78 arcs without slope data attributes. Slope attributes had not been computed for these arcs because they overlayed the anomalous DEM areas mentioned earlier. (These arcs comprised seven partial routes and one whole route.) There were also 687 duplicate arc ID numbers. Unfortunately, these had been created from the Arc/Info® CLEAN command which was used to clean up the poor topology from the base network. This problem, however, was not a serious detriment to the customer's needs since his main analysis was route-based, not arc-based.

FINAL REMARKS

There were three major difficulties in accomplishing this pilot project:

1) the lack of a topologically correct railroad network which included secondary routes,
2) the challenge of utilizing given GIS tools to accomplish an unconventional task, and
3) the lack of contiguous disk space.

As technology improves, the integrity of input data, the capability of software packages, and the speed and capacity of computer hardware will increase, thus, making a project such as this a much simpler task. Even so, we will continue to push our resources to their fullest capacity to try to solve more complicated problems.

APPENDIX A

Output File #1 Attribute Descriptions

OUTPUT FILE #1 ATTRIBUTE DESCRIPTIONS

1: 100,000 Railroad Network Attributes

MARKFINAL#: Record number generated by Arc/Info®
MARKFINAL-ID: Arc ID number taken from the original 1995 NTAD 1:100,000 scale railroad network
[2]FROMNODE: Node ID in rail_100.pnt
[2]TONODE: Node ID in rail_100.pnt
[2]LINKID: Unique identification number
[2]LINKLEN: Link length
[2]DIRECTION: Always 0
[2]MAJORATT: Major attribute code from USGS digital line graphs

180 Transportation systems - railroads
181 Railroads: minor attribute indicates number of tracks
188 Best estimate of position or classification
189 Coincident feature

[2]MINORATT: Minor attribute code from USGS digital line graphs

0001 Bridge abutment
0002 Tunnel portal
0007 Drawbridge
0100 Void area
0201 Railroad
0202 Railroad in street or road
0204 Carline
0205 Cog railroad, incline railway, logging tram
0207 Ferry crossing
0208 Railroad siding
0209 Perimeter or limit of yard
0210 Arbitrary line extension
0211 Closure line
0400 Railroad station, perimeter of station
0401 Turntable
0402 Roundhouse
0600 Historical
0601 In tunnel
0602 Overpassing, on bridge
0603 Abandoned
0604 Dismantled
0605 Underpassing
0606 Narrow gauge
0607 In snowshed or under structure
0608 Under construction
0609 Elevated
0610 Rapid transit
0611 On drawbridge

---

[2] Taken from the 1995 NTAD rail_100.lin file. Refer to the CDs rail_100.txt file for further description.

0612 Private
0613 U.S. Government
0614 Juxtaposition
0000 Photorevised feature

Note: If major attribute is 181 then minor attribute is number of tracks.

[2]OWNER: Alphanumeric identifier of the owning railroad
ROUTEID: Mark Burton's route ID number added by Cathy Adams

1: 2 million Railroad Network Attributes

NRAIL2M#: Record number generated by Arc/Info®
NRAIL2M-ID: Arc ID number taken from the original 1996 NTAD 1:2,000,000 scale railroad network
[3]LRECTYPE: Link record type: always 'L'
[3]LVERSION: Link file version number
[3]LREVISION: Link record revision number
[3]LMODDATE: Link record modification date
[3]LINKID2M: Unique sequential line identification
[3]FEATUREID: Unique line identification
[3]ANODE: Node identification for the beginning node of the line
[3]BNODE: Node identification for the ending node of the line
[3]DESCRIPT: Name or identification for the line feature
[3]STFIPS1: Primary State FIPS Code
[3]STFIPS2: Secondary State FIPS Code
[4]RECTYPE: Text record type: Always 'T'
[4]VERSION: Text file version number
[4]REVISION: Text record revision number
[4]MODDATE: Text record modification date
[4]OVERLAY: Country marker
[4]RROWN1: First railroad owner name
[4]RROWN2: Second railroad owner name
[4]RROWN3: Third railroad owner name
[4]TR1: First railroad having trackage rights
[4]TR2: Second railroad having trackage rights
[4]TR3: Third railroad having trackage rights
[4]TR4: Fourth railroad having trackage rights
[4]TR5: Fifth railroad having trackage rights
[4]TR6: Sixth railroad having trackage rights
[4]TR7: Seventh railroad having trackage rights
[4]TR8: Eighth railroad having trackage rights
[4]TR9: Ninth railroad having trackage rights
[4]SSRR: Subsidiary railroad

---

[3] Taken from the 1996 NTAD rail2m.lnk file. Refer to the CDs rail_2m.met file for further description.
[4] Taken from the 1996 NTAD rail2m.tl1 file. Refer to the CDs rail_2m.met file for further description.

[4]PRR1:	First previous Railroad owner
[4]PRR2:	Second previous railroad owner
[4]ABDN:	Abandoned flag
[4]PASS:	Type of passenger rail flag
[4]MIL:	Military importance flag
[4]STATE:	Postal Code
[4]USGS_REG:	USGS Region Code
[4]FRA_REG:	FRA Region Code
[4]DENSITY:	Density Category
[4]RR_CLS:	Railroad classification
[4]SIGNALS:	Type of signaling system
[4]ABDYR:	Abandonment Year
[4]STFIPS:	State FIPS Code

Slope Attributes Generated from USGS DEM Data

COUNT:	Number of slope sample points for this arc
MIN_SLOPE:	Slope minimum for this arc (percent rise)
MAX_SLOPE:	Slope maximum for this arc (percent rise)
VAR_SLOPE:	Slope variance for this arc
AVE_SLOPE:	Slope average for this arc

Railroad Crossing Attributes

XING: Flag for determining if this arc has associated railroad crossing data:

> "1" - means associated railroad crossing data exists, otherwise the field is blank.

APPENDIX B

Output File #2 Attribute Descriptions

OUTPUT FILE #2 ATTRIBUTE DESCRIPTIONS

Railroad Crossing Attributes

MARKXINGS#:	Record number generated by Arc/Info®
MARKXINGS-ID:	Arc/Info® Point ID number
[5]GCIS_ID:	Railroad Crossing ID number (same as FRA ID)
[5]X_DD:	Longitude of the railroad crossing
[5]Y_DD:	Latitude of the railroad crossing
[5]SOURCE:	"V" - means located by Paul Cheng in TIGER with a street name or railroad match "M" (by milepoint) - interpolated between V's
[5]RR:	Ownership name abbreviation for the railroad crossing
[5]DIVISION:	Division
[5]SUB_BRANCH:	Sub/branch
[5]MP:	Milepoint
[5]STREET:	Street name of the railroad grade crossing

1: 100,000 Railroad Network Attributes

MARKFINAL#:	Record number generated by Arc/Info®
MARKFINAL-ID:	Arc ID number taken from the original 1995 NTAD 1:100,000 scale rail network
[6]FROMNODE:	Node ID in rail_100.pnt
[6]TONODE:	Node ID in rail_100.pnt
[6]LINKID:	Unique identification number
[6]LINKLEN:	Link ID number
[6]DIRECTION:	Always 0
[6]MAJORATT:	Major attribute code from USGS digital line graphs 180 Transportation systems - railroads 181 Railroads: minor attribute indicates number of tracks 188 Best estimate of position or classification 189 Coincident feature
[6]MINORATT:	Minor attribute code from USGS digital line graphs 0001 Bridge abutment 0002 Tunnel portal 0007 Drawbridge 0100 Void area 0201 Railroad 0202 Railroad in street or road 0204 Carline 0205 Cog railroad, incline railway, logging tram 0207 Ferry crossing

---

[5] Taken from data provided by Bruce Peterson of the Oak Ridge National Laboratory.

[6] Taken from the 1995 NTAD rail_100.lin file. Refer to the CDs rail_100.txt file for further description.

0208	Railroad siding
0209	Perimeter or limit of yard
0210	Arbitrary line extension
0211	Closure line
0400	Railroad station, perimeter of station
0401	Turntable
0402	Roundhouse
0600	Historical
0601	In tunnel
0602	Overpassing, on bridge
0603	Abandoned
0604	Dismantled
0605	Underpassing
0606	Narrow gauge
0607	In snowshed or under structure
0608	Under construction
0609	Elevated
0610	Rapid transit
0611	On drawbridge
0612	Private
0613	U.S. Government
0614	Juxtaposition
0000	Photorevised feature

Note: If major attribute is 181 then minor attribute is number of tracks.

[6]OWNER: Alphanumeric identifier of the owning railroad

ROUTEID: Mark Burton's route ID number added by Cathy Adams

1: 2 million Railroad Network Attributes

NRAIL2M#: Record number generated by Arc/Info®

NRAIL2M-ID: Arc ID number taken from the original 1996 NTAD 1:2,000,000 scale railroad network

[7]LRECTYPE: Link record type: always 'L'

[7]LVERSION: Link file version number

[7]LREVISION: Link record revision number

[7]LMODDATE: Link record modification date

[7]LINKID2M: Unique sequential line identification

[7]FEATUREID: Unique line identification

[7]ANODE: Node identification for the beginning node of the line

[7]BNODE: Node identification for the ending node of the line

[7]DESCRIPT: Name or identification for the line feature

[7]STFIPS1: Primary State FIPS Code

[7]STFIPS2: Secondary State FIPS Code

[7] Taken from the 1996 NTAD rail2m.lnk file. Refer to the CDs rail_2m.met file for further description.

[8]RECTYPE:	Text record type: Always 'T'
[8]VERSION:	Text file version number
[8]REVISION:	Text record revision number
[8]MODDATE:	Text record modification date
[8]OVERLAY:	Country marker
[8]RROWN1:	First railroad owner name
[8]RROWN2:	Second railroad owner name
[8]RROWN3:	Third railroad owner name
[8]TR1:	First railroad having trackage rights
[8]TR2:	Second railroad having trackage rights
[8]TR3:	Third railroad having trackage rights
[8]TR4:	Fourth railroad having trackage rights
[8]TR5:	Fifth railroad having trackage rights
[8]TR6:	Sixth railroad having trackage rights
[8]TR7:	Seventh railroad having trackage rights
[8]TR8:	Eighth railroad having trackage rights
[8]TR9:	Ninth railroad having trackage rights
[8]SSRR:	Subsidiary railroad
[8]PRR1:	First previous Railroad owner
[8]PRR2:	Second previous railroad owner
[8]ABDN:	Abandoned flag
[8]PASS:	Type of passenger rail flag
[8]MIL:	Military importance flag
[8]STATE:	Postal Code
[8]USGS_REG:	USGS Region Code
[8]FRA_REG:	FRA Region Code
[8]DENSITY:	Density Category
[8]RR_CLS:	Railroad classification
[8]SIGNALS:	Type of signaling system
[8]ABDYR:	Abandonment Year
[8]STFIPS:	State FIPS Code

Slope Attributes Generated from USGS DEM Data

COUNT:	Number of slope sample points for this arc
MIN_SLOPE:	Slope minimum for this arc (percent rise)
MAX_SLOPE:	Slope maximum for this arc (percent rise)
VAR_SLOPE:	Slope variance for this arc
AVE_SLOPE:	Slope average for this arc

[8] Taken from the 1996 NTAD rail2m.tl1 file. Refer to the CDs rail_2m.met file for further description.