GIS Data Verification Seminar

IGUG, Fall 1995

San Jose, CA

Presented by:

John Koltun and Ken Stumpf

Geographic Resource Solutions

1125 16th Street, Suite 213

Arcata, CA 95521

(707) 822-8005

(707) 822-2864 (fax)

Introduction

Workshop Format and Content

- Graphic Problems and Solutions
  - Graphic Problems
  - Graphic Error Resolution
- Database Problems and Solutions
  - Database Problems
  - Database Error Resolution
- Data Development Strategies --
  - Feature Definition
  - Database Design
  - Digitizing Setup and Planning
  - Using MGE Tools
- Summary

The Need for Clean Data

- Topology
- Data Query
- Reporting
- Accuracy
- Portability of Data

The Need for Clean Data (cont.)

- Topology
  - What is Topology?
  - MGE
  - MGA

The Need for Clean Data (cont.)

- Data Query
  - Location
- Reporting
- Accuracy
  - Length
  - Area
- Portability of Data

Graphic Problems and Solutions

Graphic Problems

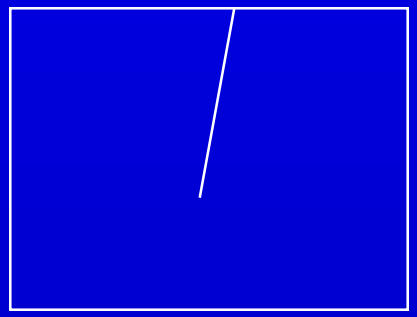
- Dangling End Points
- Unbroken Intersections
- Overshoots
- Duplicate Lines
- Zero Length Lines

Graphic Problems (cont.)

- Missing Centroids
- Multiple Centroids
- Bleeding Polygons
- Edge Matching
- Blunders

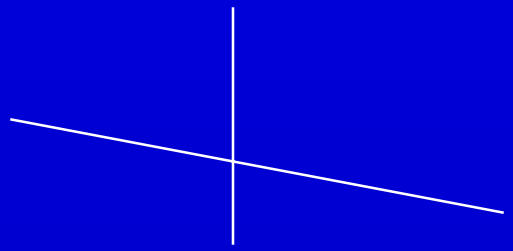
Graphic Problems (cont.)

- Dangling End Points
  - Free or dangling end points occur when a linear element is not connected with another. This may be a valid situation depending on the theme



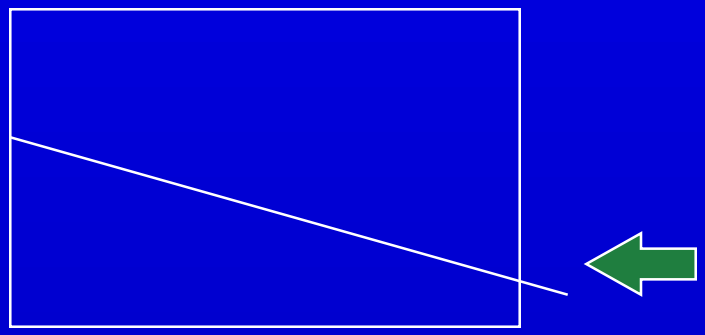
Graphic Problems (cont.)

- Unbroken Intersections
  - Linear elements that cross and are not broken where they intersect.
  - This is never valid for area boundary themes.
  - It may be a valid for some linear themes.



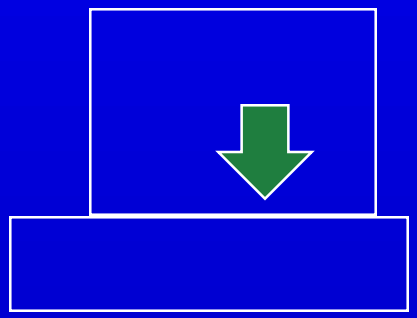
Graphic Problems (cont.)

- Overshoots
  - Overshoots are a combination of a free end point and an invalid intersection where lines are meant to intersect, yet one of the lines extends beyond the point of intersection



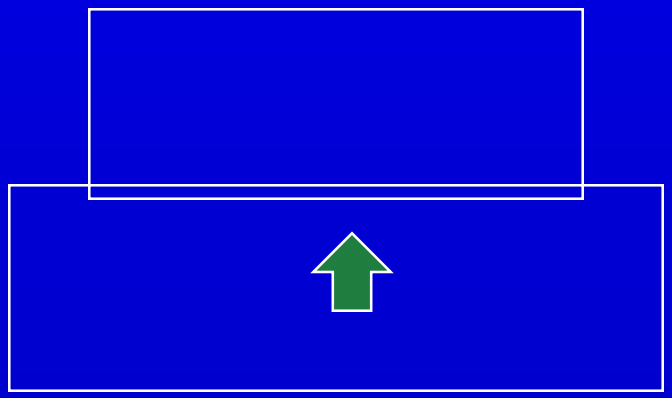
Graphic Problems (cont.)

- Duplicate Lines
  - Coincident
  - Coincident duplicate lines occur when two linear elements that represent the same feature match exactly.



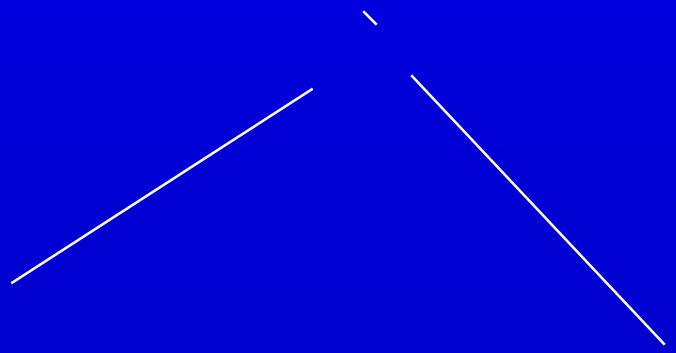
Graphic Problems (cont.)

- Duplicate Lines (cont.)
  - Non-coincident
  - Non-coincident duplicate lines are represented by two or more linear elements which do not match that are supposed to be the same feature.



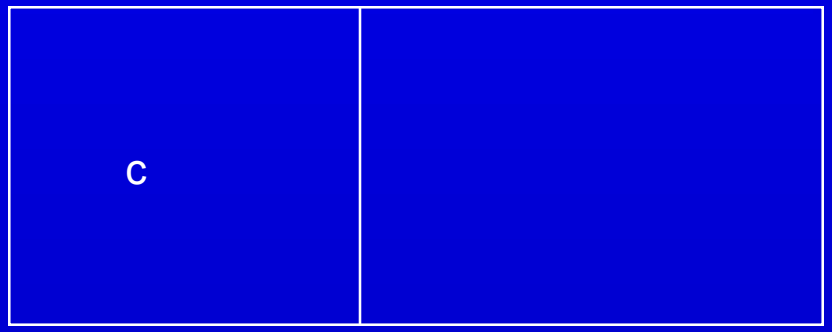
Graphic Problems (cont.)

- Zero Length Lines
  - Zero length lines are linear elements of zero length with two or more identical vertices.
  - How about area Centroids?



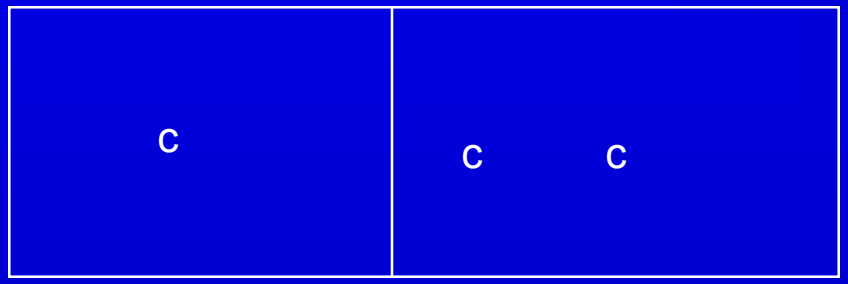
Graphic Problems (cont.)

- Missing Centroids
  - A set of area boundary features lacking a centroid will not be recognized as an area feature.



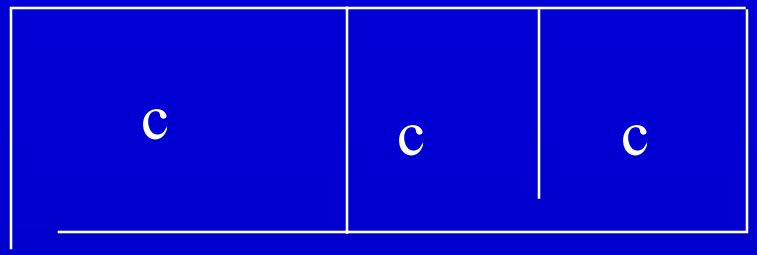
Graphic Problems (cont.)

- Multiple Centroids
  - The presence of multiple centroids within a set of boundary elements (that define one area) indicate an error condition.



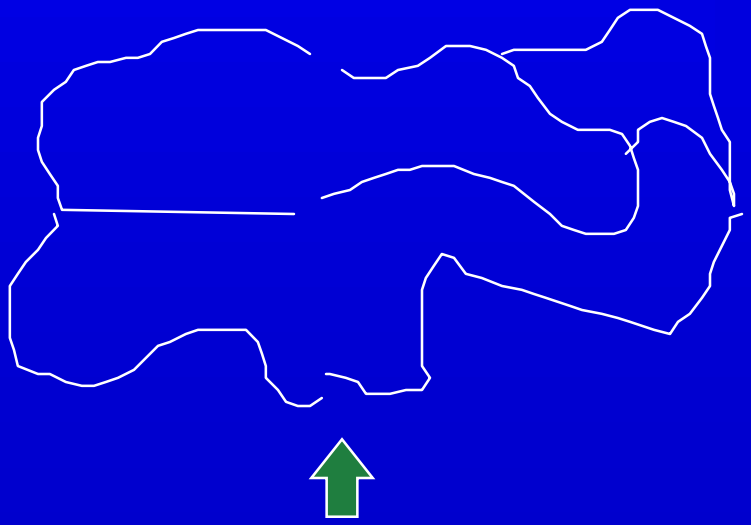
Graphic Problems (cont.)

- Bleeding Polygons
  - A bleeding polygon occurs when the lines that define area features are incomplete.
  - The missing line may define the boundary between areas or bounding an outside edge.
  - This results in multiple centroids or centroids with no defined boundary.



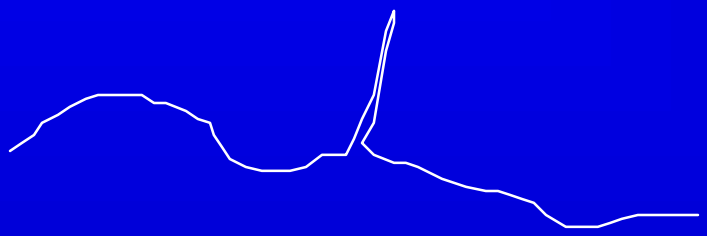
Graphic Problems (cont.)

- Edge Matching
  - When combining data from multiple sources the edges of the data may not tie directly across adjacent edges.



Graphic Problems (cont.)

- Blunders
  - Gross digitizing errors of any type including omissions, poor input techniques, or improper translation.



How's YOUR Data ??

- Dangling End Points _____
- Unbroken Intersections _____
- Overshoots _____
- Duplicate Lines _____
- Zero Length Lines _____
- Missing Centroids _____
- Multiple Centroids _____
- Bleeding Polygons _____
- Edge Matching _____
- Blunders _____

Graphic Error Resolution

- Microstation (Manual Edits)
- Line Cleaning Utilities
- Topology Creation
- Queued Locate

Graphic Error Resolution (cont.)

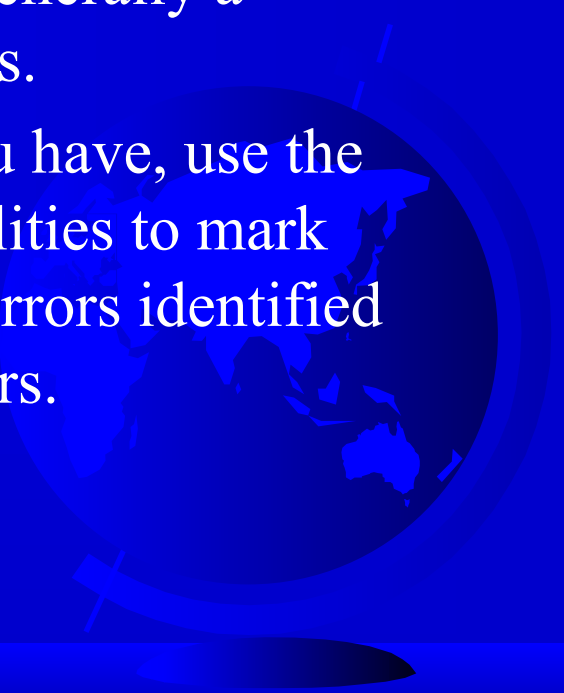
- Microstation (Manual Edits)
  - Modification Tools
  - Construction Tools
  - Use Tentative Point Religiously
- deldup.exe

Graphic Error Resolution (cont.)

- Line Cleaning Utilities
  - linecleaner
  - MRF clean
  - segjoin
  - edgematcher

Graphic Error Resolution (cont.)

- Line Cleaning Utilities
  - Know your Data
  - The proper tolerances are necessary for line cleaning utilities to operate effectively. This is generally a matter of trial and error for new data sets.
  - To get an idea of the types of errors you have, use the flagging capabilities of line cleaning utilities to mark the errors in your files. Then view the errors identified to determine the magnitude of your errors.



Graphic Error Resolution (cont.)

- Line Cleaning Utilities
  - Multiple runs of utilities may be required to make all the corrections necessary.
  - These iterations are facilitated through the use of shell scripts or batch files.

Graphic Error Resolution (cont.)

- Topology Creation
  - arealoader
  - centroidplacer
  - topobuilder
    - invalid.data
    - invalid1.ulf

Graphic Error Resolution (cont.)

- MGE Queued Locate
  - Allows the sequential location of elements in list file, such as error flags.

Graphic Error Resolution (cont.)

Graphic Error Resolution Tools

	Ustn	Cleaning Utilities	deldup	Topol. MGA	Edge Matcher	MGE	MGGA
Endpoints	M	I,A					
Intersections	M	I,A		I			
Overshoots	M	I,A					
Duplicate Coincident Lines	M	A	A	I			
Duplicate Non-coincident	M			I			A,I
Zero Length Lines	M	A		I			
Duplicate Centroids	M			I		I	I
Missing Centroids Bleeding Polygons	M			I			I
Edge Matching	M	A		I	A		A,I

A: Automatic Repair.
I: Identify Error Conditions.
M: Manual Repair Tools.

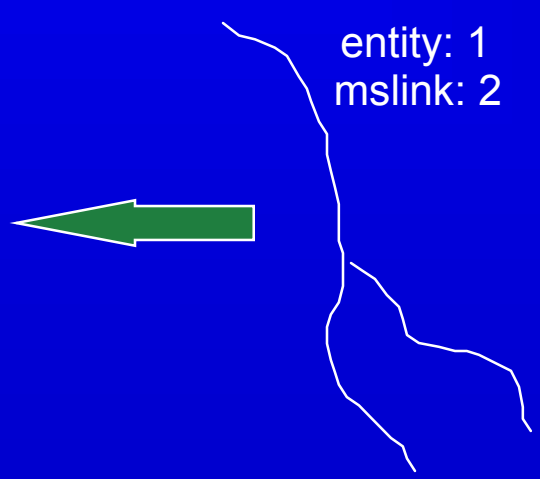
Database Problems and Solutions

MGE Database Structure

- entity
- mslink
- mscatalog

river

mslink	mapid	name...
1	100	fish cr
2	100	cold cr.



Database Problems

- Duplicate mslink values
- Orphan rows in dbs tables
- Orphan graphic elements
- Invalid dbs data
- Invalid entity number
- Invalid mslink number
- Invalid linkage format

Database Problems (cont)

- Duplicate mslink values
  - graphic file(s)
    - Multiple graphic elements pointing to the same record.
    - Caused by user editing or wrong linkage mode.
  - database table
    - Multiple records with same mslink.
    - Caused by errant MGE processes or user editing

Database Problems (cont)

- Orphan rows in dbs tables
  - Database records not connected to graphics.
- Orphan graphic elements
  - Graphics without database records

Database Problems (cont)

- Invalid dbs data
  - Bad data input.

IP
I.P.
D.I.P
DIP

Calif.
California
CA
Ca
Californy

Database Problems (cont)

- Invalid entity number
  - Graphics pointing to wrong table
- Invalid mslink number
  - Graphics pointing to wrong record

Database Problems (cont)

- Invalid linkage format
  - DMRS, RIS, Oracle, Informix...

How's YOUR Data ??

- Duplicate mslink values _____
- Orphan rows in dbs tables _____
- Orphan graphic elements _____
- Invalid dbs data _____
- Invalid entity number _____
- Invalid mslink number _____
- Invalid linkage format _____

Database Error Resolution

- MGE processes
- featurecheck
- recordcheck
- domaincheck
- attribute list domain generator
- mslinkloader
- linkdetacher
- SQL command execution
- 3rd party software

Database Error Resolution (cont.)

- MGE graphic processes
  - GDL
  - Feature Attribute Manager

Database Error Resolution (cont.)

- featurecheck
  - Determines correspondence of graphic elements with database records and feature definitions.
  - Only checks does not repair
- recordcheck
  - Checks for graphic elements for every database record.
  - Must have mapid's loaded correctly

Database Error Resolution (cont.)

- domaincheck
  - Domain check checks values in a database column against a list of valid values for that column.

Database Error Resolution (cont.)

- attrlsdomgen
  - Attribute list domain generator generates a list of unique attribute values from a database column.
  - Can also create the list with an sql statement.
    - select <columnname>, count(*) from <tablename>
    - group by <columnname>;

Database Error Resolution (cont.)

- mslinkloader
  - Loads a mslink value to the database for all records lacking a mslink value.
- linkdetacher
  - Detaches linkages from graphics and optionally deletes associated records.

Database Error Resolution (cont.)

- SQL command execution
  - Useful sql examples
  - To create a column index:
    - create unique index < indexname > on <tablename>(columnname)
  - To search for duplicate mslinks in a table:
    - select mslink, count(*) from <tablename> group by mslink having count(*) > 1
  - To review next mslink information for your tables:
    - select tablename,entitynum,nextocc from mscatalog
  - To update the next mslink number for a specific table:
    - update mscatalog set nextocc = <value> where tablename='<tablename>'

Database Error Resolution (cont.)

- 3rd party software

Database Error Resolution (cont.)

Database Error Resolution Tools

	SQL	feature check	record check	attribute list domain gen.	domain check	mslink loader	MGE Procs	Ustn Cmds	link detach
Duplicate mslinks: in graphics in rdbs table	M	I				A	I,M	M	
Orphan rows in rdbs table	M		IS				I,M	I,M	
Orphan Elements in Graphics		I					I	I,M	A
Invalid Data	M			I	IS		A	I,M	
Invalid Entity Number (Help !!)		I						edg 3rd P	
Multiple Attribute/Feature Attachments							M	M	A,I

A: Automatic Repair.
I: Identify Error Conditions
IS: Identify Error Conditions and Generate SQL File
M: Manual Repair Tools.

Data Development Strategies -- An Ounce of Prevention...

- Feature Definition
- Database Design
- Digitizing Setup and Planning
- MGE Tools for Input

Feature Definition

- Features or Attributes
  - Indicated by Use
- Element types
- Shared graphic elements
  - Graphic Normalization
- Plotting

Database Design

- Database Normalization
  - Reduce Redundancy
  - Increase Flexibility
- Domain Development
  - Constrain Data Entry

Digitizing Setup and Planning

- Document Review
  - Media Quality
  - Appropriate Scale
  - Missing or Conflicting Data
  - Edge Matching

Digitizing Setup and Planning (cont.)

- Customizing the Conversion Environment
  - Special Workflows
  - Digitzer Menus
  - Graphic Menus
  - Cursor Menus

Digitizing Setup and Planning (cont.)

- Input Specifications
  - Scale relative to accuracy
  - Scale of use
  - Precision
  - Coordinate Systems
  - GIS Ready !

Digitizing Setup and Planning (cont.)

- Input Techniques
  - Thematic Hierarchy
  - Snap Locks
  - User Commands
  - Scanning
  - Table Digitizing
  - COGO

Digitizing Setup and Planning (cont.)

- Quality Control
  - Plotting/Overlay
  - MGE
  - MGA
  - Process Analysis

MGE Tools For Input

- MGE Digitizing Environment
- MGE Attribute Processing
- MGA Spatial Analysis
- MGGA Spatial Analysis

MGE Tools For Input (cont.)

- MGE Digitizing Environment
  - Automatic feature coding
  - Semi-automatic attribution

MGE Tools For Input (cont.)

- MGE Attribute Processing
  - featuremaker
  - labelloader
  - pointloader
  - pointplacer

MGE Tools For Input (cont.)

- MGA Spatial Analysis
  - Spatial Joins
- MGGA Spatial Analysis
  - Data correlation
  - Derived themes

Summary

- GIS has different data requirements than other graphic applications
- GIS data MUST be clean for analysis
- It is easier to do it right the first time
- Automate processes to reduce error
- Data is the most valuable asset of a GIS