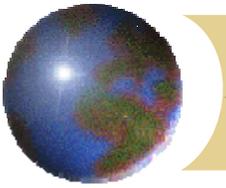
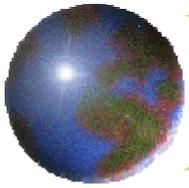


GIS Data: Classification, Potential Misuse, and Practical Limitations



Goals & Objectives

- Develop an easy to use geospatial data classification system
- Educate users on the potential misuse of geospatial data
- Explore practical limitations in using geospatial data



GIS Data: Classification

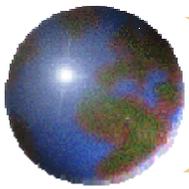
United States National Map Accuracy Standards (NMAS)

<http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/nmas/NMAS647.PDF>

Defines accuracy standards for **published** maps, including horizontal and vertical accuracy, accuracy testing method, accuracy labeling on published maps, labeling when a map is an enlargement of another map, and basic information for map construction as to latitude and longitude boundaries.

Horizontal accuracy. For maps on publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch.

Vertical accuracy, as applied to contour maps on all publication scales, shall be such that not more than 10 percent of the elevations tested shall be in error more than one-half the contour interval.



GIS Data: Classification

ASPRS Accuracy Standards for Large-scale Maps

https://eserv.asprs.org/PERS/1990journal/jul/1990_jul_1068-1070.pdf

A major feature of these ASPRS standards is that they indicate accuracy at ground scale. Thus, digital spatial data of known ground-scale accuracy can be related to the appropriate map scale for graphic presentation at a recognized standard.

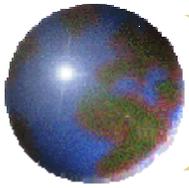
Table 1E. — Planimetric Coordinate Accuracy Requirement (Ground X or Y in feet) for Well-defined Points - Class 1. Maps

PLANIMETRIC (X or Y) ACCURACY ³ (limiting rms error, feet)	TYPICAL MAP SCALE
0.05	1:60
0.1	1:120
0.2	1:240

0.3	1:360
0.4	1:480
0.5	1:600
1.0	1:1,200
2.0	1:2,400
4.0	1:4,800
5.0	1:6,000
8.0	1:9,600
10.0	1:12,000
16.7	1:20,000

*indicates the practical limit for aerial methods - for scales above this line, ground methods are normally used.

Map accuracies can also be defined at lower spatial accuracy standards. Maps compiled within limiting rms errors of twice or three times the those allowed for a Class 1 map shall be designated Class 2 or Class 3 maps respectively. A map may be compiled that complies with one class of accuracy in elevation and an other in plan. Multiple accuracies on the same map are allowed provided a diagram is included which clearly relates segments of the map with the appropriate map accuracy class.



GIS Data: Classification

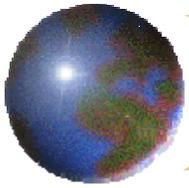
National Standard for Spatial Data Accuracy (NSSDA)

<http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3>

Implements a statistical and testing methodology for estimating the positional accuracy of points on maps and in **digital** geospatial data, with respect to georeferenced ground positions of higher accuracy.

The NSSDA uses root-mean-square error (RMSE) to estimate positional accuracy. RMSE is the square root of the average of the set of squared differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for identical points¹.

Accuracy is reported in ground distances at the 95% confidence level. Accuracy reported at the 95% confidence level means that 95% of the positions in the dataset will have an error with respect to true ground position that is equal to or smaller than the reported accuracy value. The reported accuracy value reflects all uncertainties, including those introduced by geodetic control coordinates, compilation, and final computation of ground coordinate values in the product.

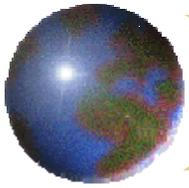


GIS Data: Classification

More on NSSDA

A **minimum** of 20 check points shall be tested, distributed to reflect the geographic area of interest and the distribution of error in the dataset.⁴ When 20 points are tested, the 95% confidence level allows one point to fail the threshold given in product specifications.

If accuracy reporting cannot be provided using NSSDA or other recognized standards, provide information to enable users to evaluate how the data fit their applications requirements. This information may include descriptions of the source material from which the data were compiled, accuracy of ground surveys associated with compilation, digitizing procedures, equipment, and quality control procedures used in production.



GIS Data: Classification

A need to categorize and certify digital landbases

R.Stricklan, 1998

Basic premise: There is a need for standardizing the uncertainty in digital landbases, enabling legislated definition of creation/permitted uses for those cadastral references.

Although it would be impossible to verify positional accuracy of all digital landbase points, there should be some type of grading system that would quantify how closely any such landbase reference matched its original source documents and/or professionally-interpreted intent of those documents.

A proposed system of categorizing digital landbases based on their creation methodology and database structure (Stricklan, 1998):

Class 1A: Temporally-complete, auto-refinable, COGO-created digital landbase using legal source documents, with inclusion of record/adjusted measurements as attribution linked to every cadastral line.

Class 1B: Auto-refinable, COGO-created digital landbase using legal source documents, with inclusion of record/adjusted measurements as attribution linked to every cadastral line. (no temporal transactions included)

Class 1C: COGO-created digital landbase using legal source documents, with inclusion of record/adjusted measurements as attribution linked to every cadastral line. (no auto-refinement nor temporal transactions included)

Class 2A: Trace digitized from record plat maps; record measurements shown as annotation and stored as attribution.

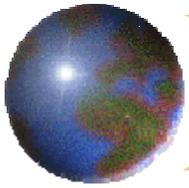
Class 2B: Trace digitized from record plat maps; record measurements shown as annotation only.

Class 2C: Trace digitized from record plat maps; but no record measurements shown.

Class 3A: Trace digitized from composite derived documents (e.g. assessor maps) other than record plats and descriptions; record measurements shown as annotation and stored as attribution.

Class 3B: Trace digitized from composite derived documents; record measurements shown as annotation only.

Class 3C: Trace digitized from composite derived documents; but no record measurements shown.



GIS Data: Classification

APLS Geospatial Data Standards

http://www.azpls.org/associations/1444/files/AZ%20Spatial%20Data%20Standards_v3.1%20final.pdf

Standards for the acquisition and evaluation of geospatial data

The main purpose of these standards is to provide guidance for developing the scope or statement of work for projects involving collection of spatial data.

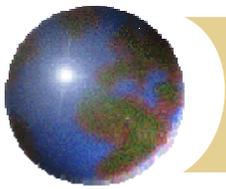
In all cases, accuracy shall be determined at the 95% confidence level both horizontally and vertically, including all accuracy values cited within these Standards.

For datasets larger than or equal to 100 points but less than 1000 points, a minimum of 20 check points shall be used.

For datasets of 1000 or more points, the number of check points shall be at least 2% of the total number of points in the dataset.

For datasets of less than 100 points, the number of check points shall be specified in the project scope.

The positional accuracy of polyline and polygon features shall be evaluated in the same manner as point features, by observation of discrete points on the perimeter of the feature. If more than one point is evaluated for a single polyline or polygon feature, the mean accuracy of the set of validation measurements shall be used to estimate the overall accuracy of the feature. The minimum number of check observations for each polyline or polygon feature shall be based on the number of vertices using the same criteria as for individual point features.



GIS Data: Classification

Proposed Data Classification System (DRAFT)

The purpose of the rating system is to guide users in assessing the accuracy of geospatial data, thereby indicating to them the appropriate use of the data.

Three categories of criteria:

Who created the data.

How they created the data.

Measured Spatial Accuracy.

Equipment and Process will determine the Class, also:

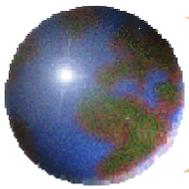
(examples)

A: High-order GPS/post-processed to Class accuracy range.

B: Medium to low-order GPS/optional post-processing; high to medium-grade orthophotography/field verified to stated accuracies.

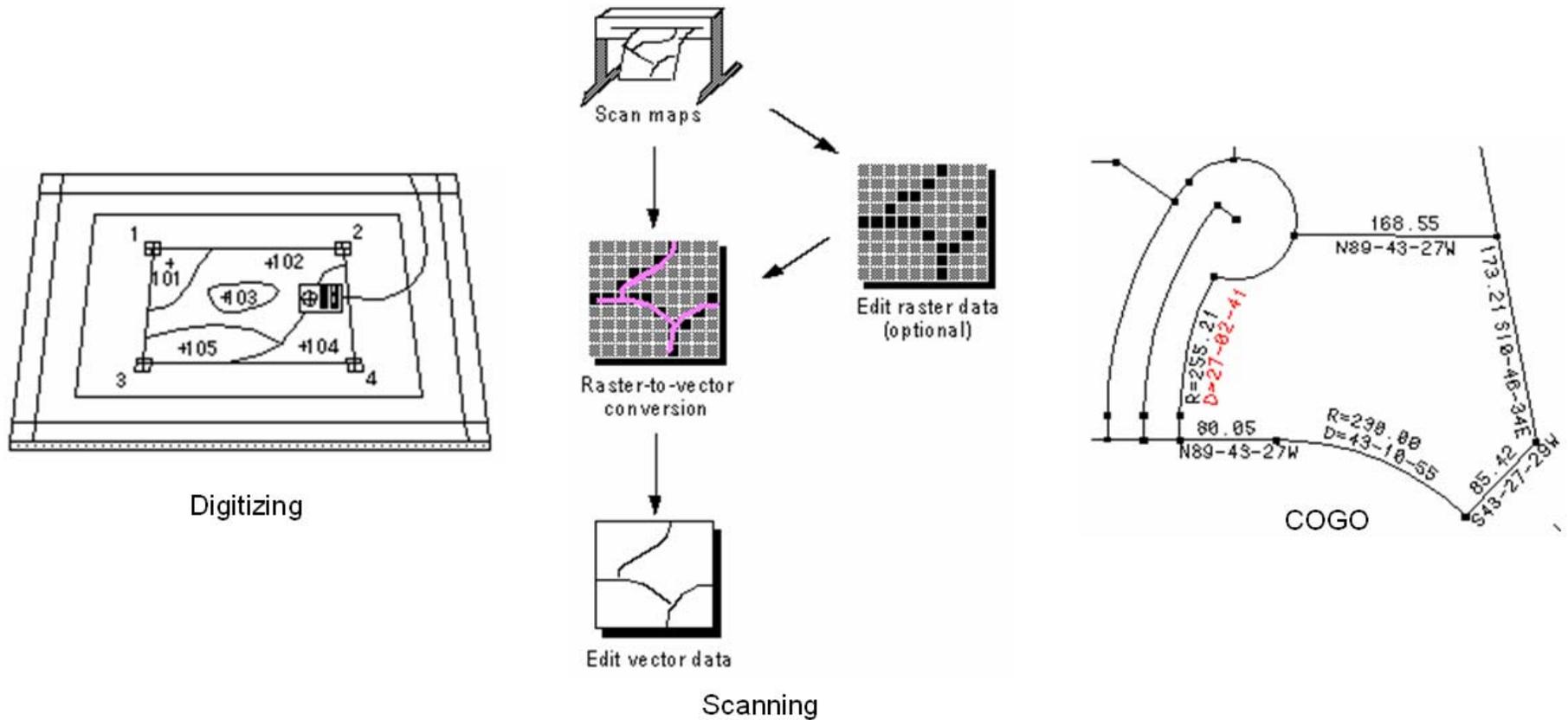
C: low-grade orthophotography/non-field verified; small-scale maps/heads-up or tablet digitized.

CLASS	GRADE	WHO	HOW	ACCURACY
A	1	RLS	Field derived source - PP ¹	+/- 1cm
	2	RLS	Field derived source	+/- 1dm
	3	RLS GeoSpatial	Field derived source	+/- 1m
B	1	RLS GeoSpatial	Field derived source Ortho derived source Digitizing on A1	+/- 3m
	2	RLS GeoSpatial	Field derived source Ortho derived source Other geo-referenced GIS Sources	+/- 5m
	3	RLS GeoSpatial	Field derived source Ortho derived source Other geo-referenced GIS Sources	+/- 10m
C	1	GeoSpatial Field Techs GIS Techs	Ortho derived source Satellite imagery (geo-referenced) USGS 7.5'	
	2	GeoSpatial Field Techs GIS Techs	USGS 15', 1:100K Other geo-referenced GIS Sources	
	3	GeoSpatial Field Techs GIS Techs	USGS 1:250K	
D	1	Non-GeoSpatial	3 rd party GIS data sets of stated accuracy	unknown
	2	Non-GeoSpatial	Unknown 3 rd party GIS data sets of unknown accuracy	Unknown



GIS Data: Classification

Data Input Methodologies



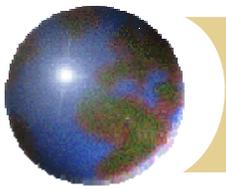
Heads-up digitizing

- Screen digitizing over imagery (orthogonal, oblique, etc.)
- Semi-automatic feature extraction (ARAN van imagery, street-level views, etc.)

- Measured vs. visual

GPS

- Resource-grade
- Survey-grade
- Techniques for high accuracy
- Post-processing
- Smart phones



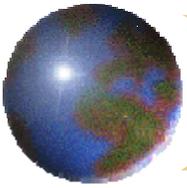
GIS Data: Classification

Metadata

Minnesota Geographic Metadata Guidelines	
Attribute Accuracy	An explanation of how accurately the entities have been identified or how accurately values have been assigned in the data set. This can be the results of quantitative analysis, steps taken to ensure accuracy during development, or known deficiencies.
Logical Consistency	An indication of topological problems such as overshoots, undershoots, unwanted intersections, unclosed polygons, missing or duplicate labels, etc.
Completeness	Information about omissions, selection criteria, generalization, definitions used, and other rules used to derive the data set that may affect the completeness of content of the data.
Horizontal Positional Accuracy	An explanation of the accuracy of the horizontal positions (coordinates) of spatial objects and a description of the tests or line of reasoning used to arrive at the estimate.
Vertical Positional Accuracy	An explanation of the accuracy of the vertical positions (coordinates) of spatial objects and a description of the tests or line of reasoning used to arrive at the estimate.
Lineage	Information about the sources of data used to construct the data set and steps used to process the data. For each processing step provide, where possible, the following: source data name, source data scale, source data date, description of processing steps performed, scanning or digitizing specifications, hardware & software used, processing tolerances, etc.
Source Scale Denominator	The denominator of the representative fraction of source map(s). Please leave out commas (e.g. 24000 not 24,000). Example: on a 1:24000 scale map, the Source Scale Denominator is 24000; on a "100 scale" map (1 inch = 100 feet) the Source Scale Denominator is 1200. If no source maps were used, enter zero (e.g. GPS data). If multiple source map scales were used, enter the Source Scale Denominator of the smallest scale map (largest denominator).

Other important information:

- Descriptive name/ description/abstract
- Known errors and qualifications
- Feature type
- Projection
- File name
- Dates – creation, update
- Contacts – general, maintenance
- Availability
- Field metadata
- Feature metadata



GIS Data: Classification

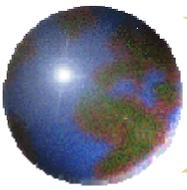
Proposed Data Classification System (DRAFT)

To do:

- Complete the accuracy ranges
- Provide examples
- Emphasize the use of ground-based features and/or control for evaluating existing data, e.g. Ground Control Points, Photo Control Points, etc.
- Provide references to the Geospatial Data Standards for guidance
- Figure in imagery as a data creation/enhancement resource
- Propose as an addendum to the Spatial Data Standards

Suggestions:

- How accurate do you “think” your data is?
- Add a metadata column to the matrix



GIS Data: Potential Misuse and Practical Limitations

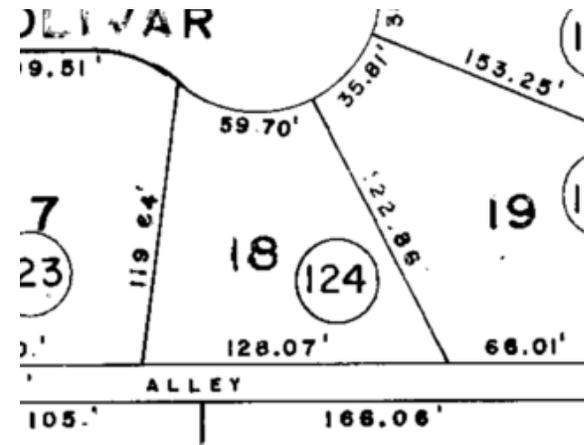
Boundary misinterpretation

- Especially when displayed on top of imagery

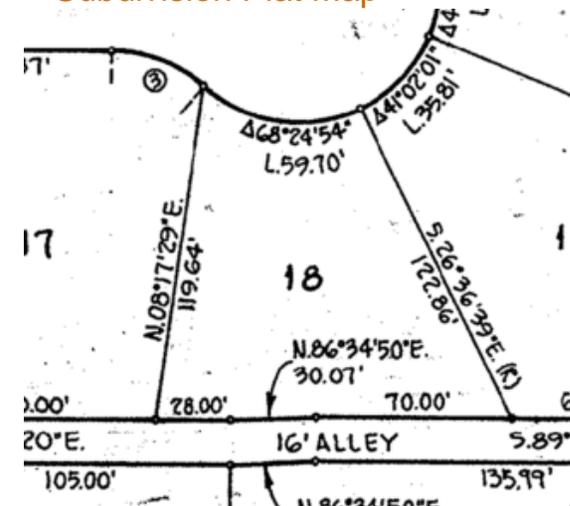


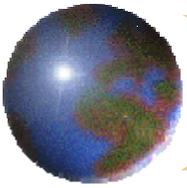
- How good is the data?
- How good is the imagery?
- Mixing scales?
- Different projections/coordinate systems?

- Assessor Record Map



- Subdivision Plat Map

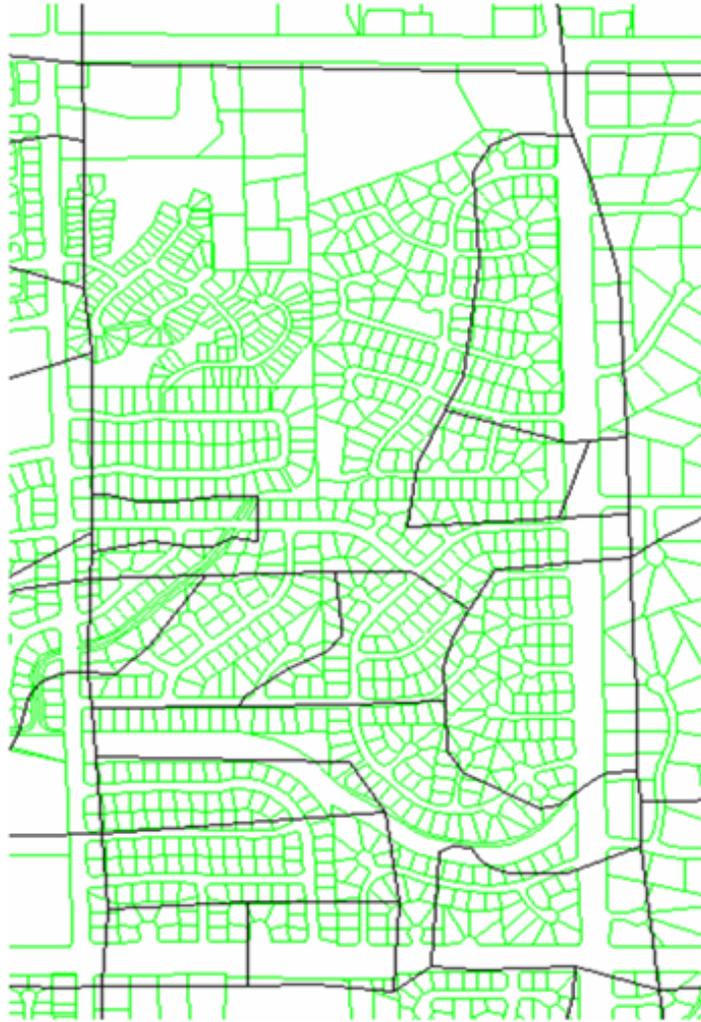




GIS Data: Potential Misuse and Practical Limitations

Mixing Data Scales/Accuracy/Quality

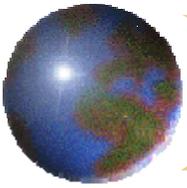
- Census Blocks on Parcels



- Parcels on NAIP imagery (+/- 15-feet)



- Miss-registered layers



GIS Data: Potential Misuse and Practical Limitations

Mixing Data Scales/Accuracy/Quality

- Building roof lines, as captured from aerial imagery

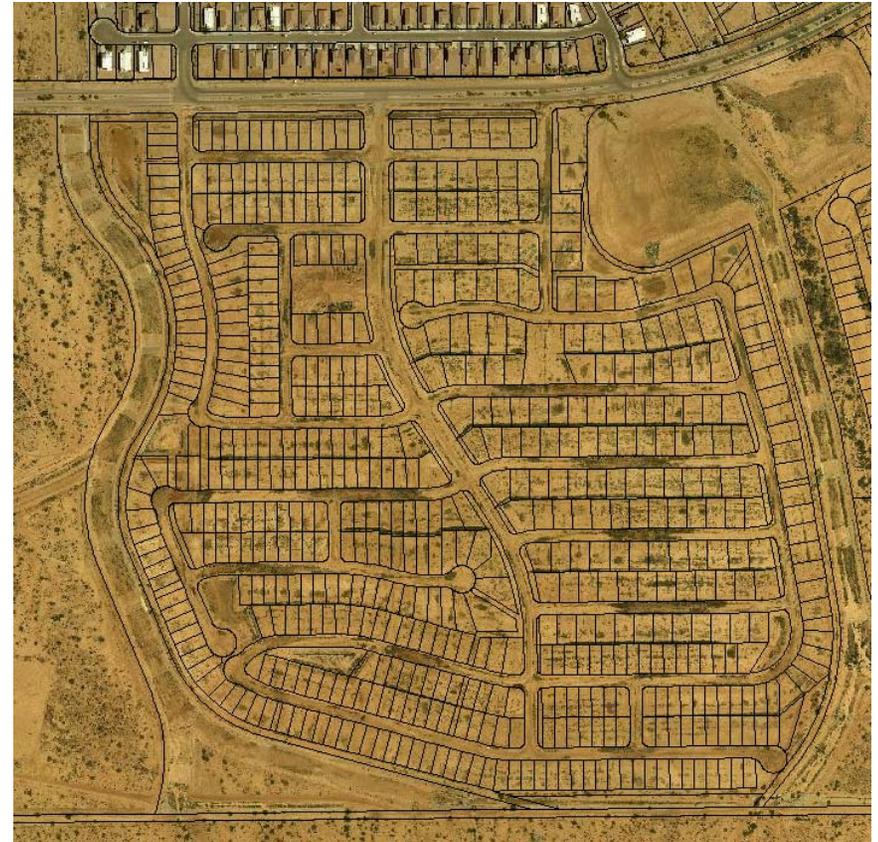


2010 imagery, +/- 10-foot accuracy

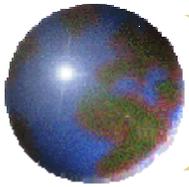


2008 imagery, +/- 2-foot accuracy

- Planned features



- Spatial location differences between imagery
- Different image specifications
- Different image years
- Why so much imagery?



GIS Data: Potential Misuse and Practical Limitations

Attribute Data Misinterpretation

- Data Normalization

AREA	PARCEL	LANDFCV	IMPFCV	TOTALFCV
245555	20518018C	32580	17582	50162
721363	20518018A	33663	1551	35214
1281323	20518011A	56620	65101	121721
64349	20518011A	56620	65101	121721
33446	20518011A	56620	65101	121721
654160	20518018B	33012	2703	35715
6546707	20516001A	56000	0	56000

- Divide all aggregated values by the frequency

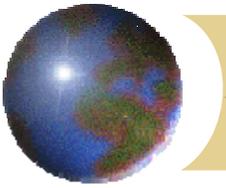
Know your table relationships:

- One-to-one
- One-to-many
- Many-to-many
- Many-to-one

- Data View Settings

TOTAL_POP	POP_DENS	AREA_SQFT
2823	0.00020560103	13730475.77900000000
5700	0.00003385918	168344309.65200000000
4733	0.00003665918	129108194.50100000000
1752	0.00005217327	33580412.85220000000
3361	0.00003769203	89170054.78820000000
5098	0.00008032533	63466902.78090000000
3728	0.00008460112	44065608.42170000000
3353	0.00009063870	36993028.01510000000
3400	0.00000948171	358585121.99100000000
3599	0	27826360419.30000000000
9743	0.00010552884	92325472.62920000000
3519	0.00005985815	58788986.82550000000
6213	0.00008569848	72498368.61260000000
4153	0.00013874760	29932048.43930000000

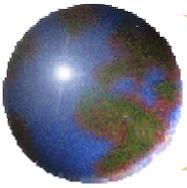
- Data value is stored correctly, but is too small for the display setting



GIS Data: Potential Misuse and Practical Limitations

Color/Symbology Misinterpretation

- Colors that are too alike (and patterns)
- Too many layers displayed together
- Colorblindness
- Grayscale
- Black & white

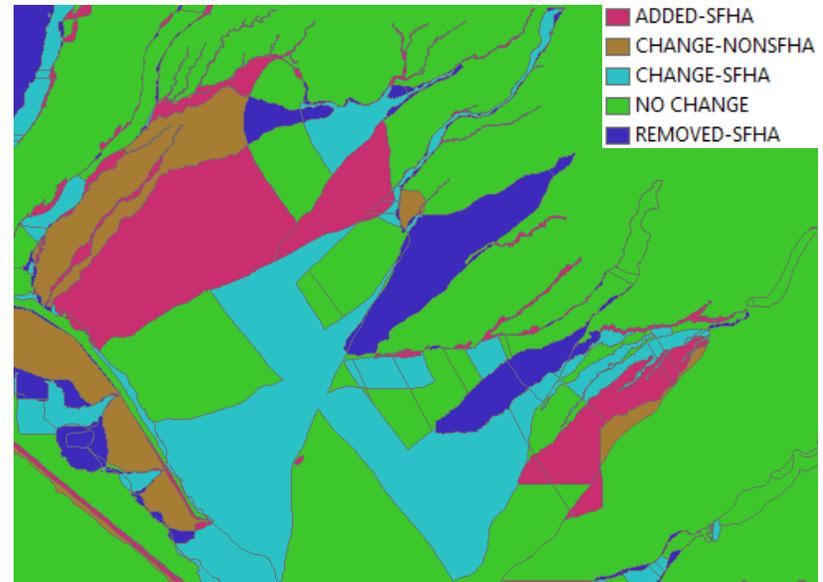


GIS Data: Potential Misuse and Practical Limitations

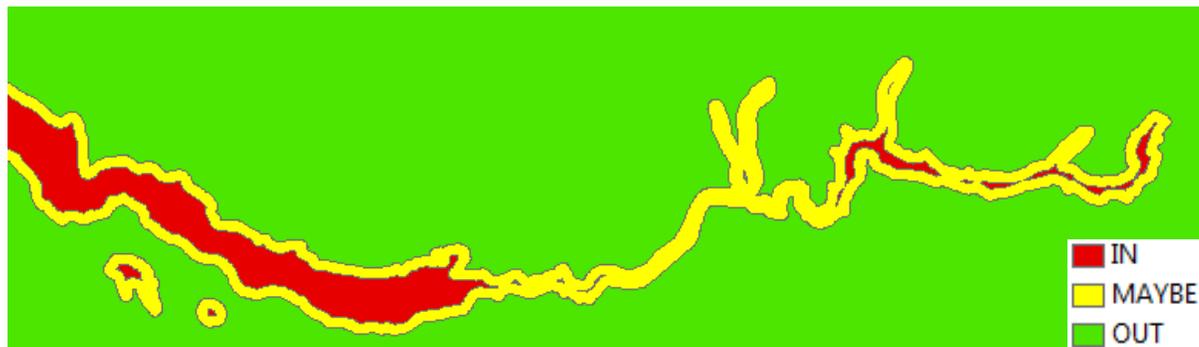
Inappropriate Use for Regulatory Applications

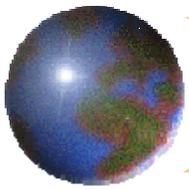
- Floodplains?
- Development/building restrictions
 - Zoning
 - Overlay zones, e.g. hillside, riparian, etc.
 - Building setbacks
 - Density
- In/out/maybe zones of confidence

- FEMA floodplain changes



- Zones of confidence

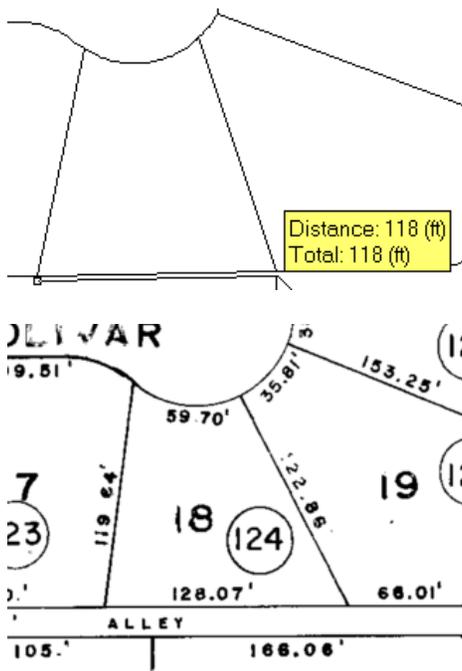




GIS Data: Potential Misuse and Practical Limitations

Inappropriate Measuring

- Inappropriate data accuracy/quality/classification
- Scale is too large
- Interpreted view, e.g. oblique, integrated into applications, street, etc.
- Know the source accuracy/classification to know if you are in the “ball-park” or better

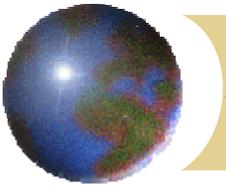


- Oblique image



- ARAN van image

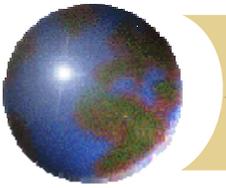




GIS Data: Potential Misuse and Practical Limitations

Bad Analysis Practices

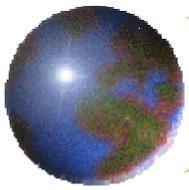
- Overstating the accuracy of the results
 - Data accuracy/quality
 - Vector
 - Raster
 - Elevation
 - Area proportion overlays
 - Mixed scales
 - Attribute accuracy
- Ignoring topological errors



GIS Data: Potential Misuse and Practical Limitations

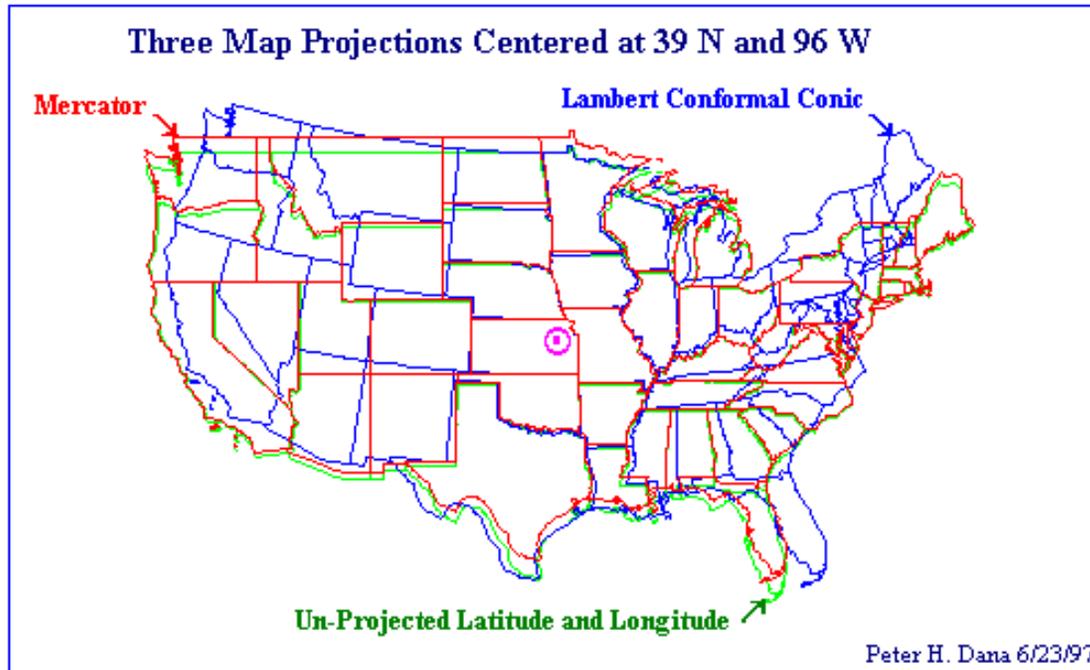
Internal Use Only Data, That Gets Loose

- Critical Infrastructure and other DHS classified data
- Developmental/incomplete data sets
- Non-public data, e.g. health, public safety, etc.



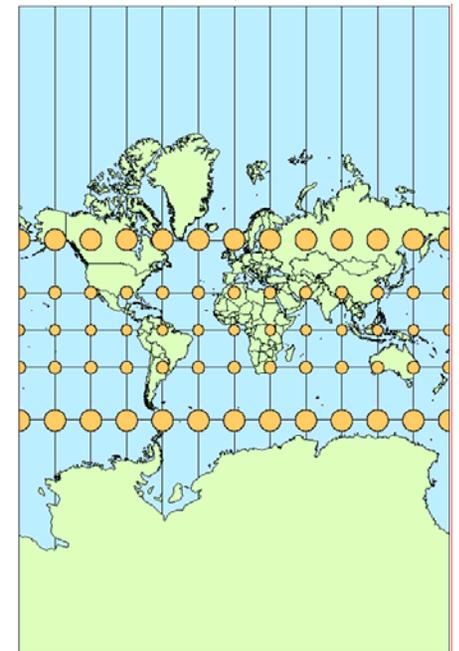
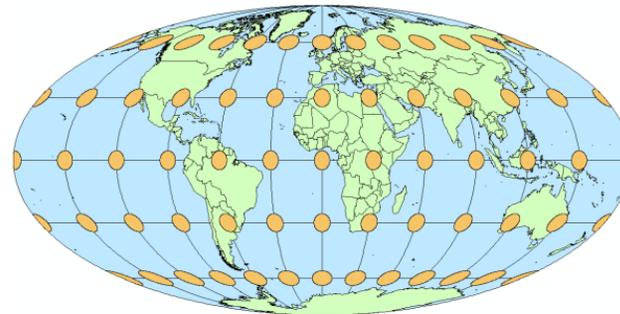
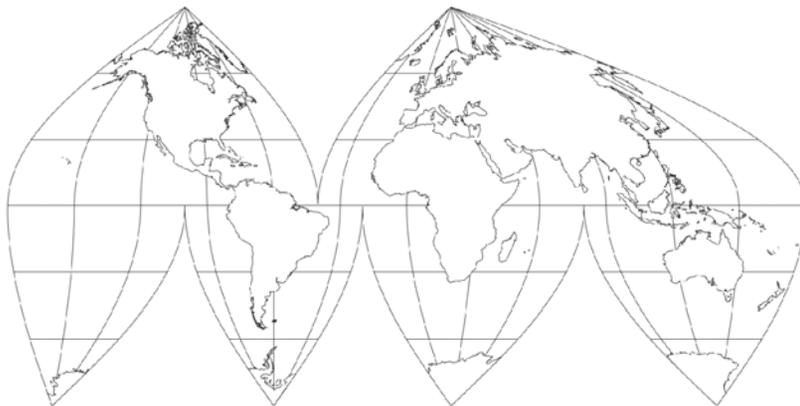
GIS Data: Potential Misuse and Practical Limitations

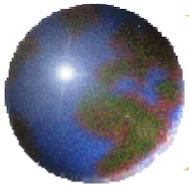
Projections and Coordinate Systems



- Metadata
- Projecting data
 - Use correct Geographic Transformations
 - Datum changes add complexity
 - Projections/coordinate systems have a sweet spot or zone of “best” accuracy

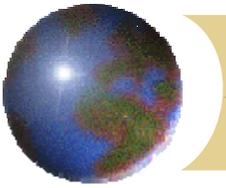
Sinusoidal 6pt Projection Map





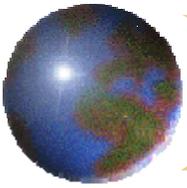
GIS Data: Potential Misuse and Practical Limitations

- Data Creation
 - Data input methodologies
 - Data sources
 - Classification
- Data Enhancement (or Degradation)
 - Conflation
 - Measured adjustment
- Data Dissemination
 - Metadata
 - Classification
 - Disclaimers and data qualifications
 - Scale-dependent viewing, e.g. layers are viewable within defined scale ranges
 - Turn off boundaries at certain scales, especially when displayed against imagery
 - Mixing scales

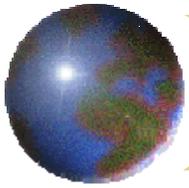


GIS Data: Potential Misuse and Practical Limitations

- Data Analysis
 - Mixing scales
 - Relative accuracy vs. absolute accuracy
 - Projections and coordinate systems
- Data Reporting
 - Attribute data accuracy
 - Analyses results



Questions?



GIS Data: ?

?

- ?