GeoWorld magazine, February, 2002, p. 40-41

HOW GOOD ARE YOUR MAPS?

Bruce Joffe, Principal
GIS Consultants
Oakland, CA
510-238-9771
GIS.Consultants@joffes.com

If a picture is worth a thousand words, how many more words is a map worth? Ten times as many? a hundred times? a thousand times? Consider the very last map you glanced at, and suppose you were to write down every fact depicted on that map. The list would include the names of everything shown on the map, plus the relative distances between everything, plus the spatial relationships between everything and everything else ("Pine Mountain is halfway between Oak Bluff and Grand Junction on Highway 12"), plus information about things not shown on the map ("my favorite camping place is 3/4 mile down the second dirt road off Highway 30"), plus all the relative relationships among everything on the map ("the South Fork of the North River is the longest river shown").

Maps are such concise conveyors of information because they are scaled representations of reality. As such, their use and value depends on their accuracy, or more generally, their quality. While maps vary greatly in accuracy and quality, few can be classified as inherently having "high" or "low" accuracy, "good" or "bad" quality, without reference to the map's purpose. Different levels of accuracy and quality are appropriate for different purposes. If a map correctly informs your query, it is good enough. If the same map, or mapping database, can correctly inform a whole lot of queries, it is better. Because maps model and represent reality, they omit some objects and simplify others. A map's quality depends on how thoroughly and accurately it portrays the objects of interest to the purposes for which it is designed.

Map Quality

Although positional accuracy is often regarded as the primary measure of quality for maps and GIS databases, "quality" includes the following additional important criteria:

- **Currency** – how up-to-date the mapped information is, and how frequently it is updated.
- **Representation** – whether the mapped objects are topologically connected, and if the symbology and annotation portray them understandably.
- **Lineage** – includes identification and quality assessment of the source documents, the methods by which the original data were collected and the methods by which multiple data sources were compiled into an integrated mapping database.
- **Accuracy** – includes such factors as:
  - √ <u>Positional</u> <u>accuracy</u> – how closely the mapped coordinates correspond to the "actual" coordinates of a mapped object . (This criterion always is limited by the means and methods available for measuring "ground truth" coordinates.)

√ Referential accuracy – how closely the portrayed distances between mapped objects compare with field measurements. (This criterion also includes the distance between mapped objects and mapped reference points in the context of a specified projection system and reference datum.)

√ Identification accuracy – how reliably mapped objects are portrayed as the actual objects they represent, and how few relevant objects were omitted from the set of mapped objects.

√ Attribute accuracy – how reliably the descriptive attributes associated with a mapped object represent the actual object.

Assessing the quality level of a given mapping database along these criteria requires different standards and methods for each type of quality factor.

Map Currency assessment requires that the date of data collection be annotated in the descriptive attributes of each mapped object. The date of data compilation is not equivalent, because a map or GIS database may be compiled years after an original observation. Nor is a single date for the creation or modification of the entire map an adequate indication of the currency of each constituent mapped object.

Map Representation assessment requires measuring the degree of clarity or understanding felt by a statistical sample of the map-using population. If a sample of people were given a map and then asked questions about it, would the correctness of their scores be attributable to the clarity of the map or their capability as map readers?

Map Lineage assessment requires a standardized classification system of map compilation techniques, such as the one proposed by Rudy Stricklan, RLS.[1] Stricklan's data lineage classification system includes the following categories:

√ Direct observation
√ Constructed from legal source documents
√ COGO constructed
√ Constructed from derived maps
√ Trace digitized
√ Measurements and adjustment factors recorded to enable "auto-refineable" update (this capability is becoming known as "measurement-based cadastre")
√ Measurements recorded as text and stored as attribution
√ Measurements shown as text only
√ No measurements shown

Map Accuracy assessment requires measuring the statistical variance of a set of sample mapped points from a set of independent reference measurements for those points. This applies to measurements of location for positional and referential accuracy, as well as observations for identification and attribute accuracy.

National Standards

Since 1947, the U.S. National Map Accuracy Standard (NMAS) has governed the way map locational accuracy has been characterized, according to the scale of the map. The

1 "Cadastral Reference Databases: Categorizing and certifying the conversion should be standardized". PoB magazine, January 2000, http://www.pobonline.com/feature2.htm

NMAS standard says that maps with scales larger than 1:20,000 (mapped objects at a larger scale appear larger on the map) must locate at least 90% of their objects correctly, within 1/30 inch (at scale). So, 90% of the objects mapped at a scale of 1:1,200 (1 in. = 100 ft.) must be correctly located within 3.33 ft (100 ft. scale divided by 1/30 inch) of their "actual" location. Objects mapped at scales smaller than 1:20,000 (for example 1:24,000 USGS Quad maps) must be correctly located within 1/50 inch (e.g., 40 ft. at USGS Quad scale).

The development of increasingly accurate methods and techniques for locational measurement (especially horizontal measurement using GPS), and the fact that GIS-based map data can be displayed at any scale, have necessitated a redefinition of map accuracy standards. The National Standard for Spatial Data Accuracy (NSSDA), proposed by the Federal Geographic Data Committee (FGDC) in 1998, offers a methodology for comparing sample, mapped points with independent measurements of their location to derive a statistical assessment that is valid for 95% of the points.[2] Generally, the Standard defines the following methodology:

1) Decide whether to test for horizontal or vertical accuracy, or both
2) Select a minimum of 20 well-defined and identifiable points from the mapped set. The sample points should be representative; more points give better results.
3) Select corresponding points from an independent, more accurate dataset. In many cases, the dataset will be created by on-site measurement observations. The independent dataset should be three times more accurate than the expected accuracy of the sample dataset.
4) Record the measurement values; a matrix is recommended with columns for point number, point description, x-value (independent), x-value (sample), difference in x-value, difference in x-value squared, y-value (independent), y-value (sample), difference in y-value, difference in y-value squared, and "error radius" (the difference in x-value squared plus difference in y-value squared).
5) Calculate the sum of all "error radius" measurements, then calculate the average, then take the square root to yield the root mean square (RMS) error.
6) The NSSDA standard calls for an RMS error that includes 95% (a two-sigma variance) of the sample points. The statistic is calculated by multiplying the RMS error by 1.7308 (for horizontal error) or 1.9600 (for vertical error).

The results of the NSSDA method should be expressed in a statement similar to: "The tested horizontal accuracy at a 95% confidence level is xx feet." This metadata statement enables the map-database user to expect locational errors no worse than "xx" feet, 95% of the time, regardless of the scale at which the map is displayed.

---

2 You can download "Geospatial Positioning Accuracy Standards" from www.fgdc.gov/standards/documents/standards/accuracy/, or from www.fgdc.gov/standards/status/sub1_3.html

Chris Cialek of the Minnesota Planning Land Management Center, has put together an excellent workbook to demonstrate the NSSDA with examples, available from http://www.ot.state.mn.us/ot_files/handbook/standard/std19-1.html and a short description from http://www.mnplan.state.mn.us/press/accurate.html

Accuracy Statements

When one considers citywide, countywide, or statewide GIS-based map data, one can expect there to be domains where the locational accuracy is better than the NSSDA 95% level, and areas where the accuracy is worse. Older urbanized areas, and less-developed areas, tend to have less accurate mapping, while newer, or higher land-value areas, tend to have more accurate mapping. I recommend, therefore, that the NSSDA methodology be applied to "accuracy domain" polygons, each of which is assessed with a separate statistical dataset. There is no reason why a GIS map should have only one accuracy statement. The accuracy of various domains (areas) could be color-coded, or otherwise graphically indicated.

Because a GIS can display various map layers that have different accuracies, what can be said about the locational accuracy of a combination of map-layers? The accuracy of the compilation is no better than the worst layer's accuracy. So, if a floodplain is overlaid on a cadastral theme, the accuracy of the combination probably is no better than the accuracy of the floodplain boundary. Nevertheless, methods of adjustment often are applied to clearly identifiable points on the less-accurate layer that register them to corresponding points on the more-accurate layer. But how are the rest of the lower-accuracy layer's points adjusted? Various mathematical transformations are available (generally referred to as "rubber sheeting") that make the less-accurate layer look more aligned with the more-accurate one. But alignment is not the same as accuracy. Responsible metadata should describe the adjustment transformations that were applied to a map layer.

Complete Maps

Although data currency and accuracy are important factors of map quality, perhaps the most important aspect of a map's usefulness to a particular application is the most obvious: its area of coverage. Just as the regime of paper maps seems to follow a natural law that causes one's area of interest always to span multiple map sheets, so it seems that digital maps follow the "donut law", i.e., when compiling multiple map files, there always seems to be a hole with no coverage. Just as the USGS paper-based quad sheets cover the entire United States, the agency now plan to create a nationwide digital map coverage. The project, appropriately named "The National Map", will "provide the nation with current, accurate, and nationally consistent basic spatial data ..."[3] Moreover, while the average currency of USGS Quad maps is 23 years (!), the intent of The National Map is "... to deliver spatial information that is not more than *seven days* old" (my italics).

Such lofty ambition should grab the attention of our Geographic Information community, yet the reaction may be characterized as no more than passively supportive. We know how much difficulty the USGS has had in funding its current mapping programs (hence the low average data currency), so how is the National Map going to be built and maintained so effectively? Rather than build the entire digital product itself, the USGS is positioning itself to stimulate and organize collaboration with state and regional governments, universities,

3 The National Map project proposal can be downloaded from http://nationalmap.usgs.gov/

and private industries to coordinate and compile all relevant spatial datasets.[4] This data integration project will require a consistent classification system for mapped objects, as well as positional accuracy sufficient to align objects from different data themes and source area coverages. The integration of locally-produced spatial information is envisioned as an ongoing, regular process that will enable changes in local map data to update the National Map databank quickly and reliably.

The Open GIS Consortium, a private, nonprofit association of industry leaders, is relentlessly attacking the technical problems of geographic information exchange and geoprocessing interoperability. Such solutions are fundamental to the seamless integration of thousands of datasets into a consistent National Map.

A related problem also looms: creating standardized data distribution agreements and financial arrangements for the many public and private data producers that have map data to contribute to the public-domain National Map. I have proposed the "Open Data Consortium" project to organize a series of workshops and discussions in which public and private producers of geographic information can meet with public and private data distributors to formulate consensus-based Model Data Distribution Agreements. Currently, URISA and the GeoData Alliance have endorsed this initiative.[5]

Geographic information can be used effectively only when its quality characteristics – currency, representation, lineage, and accuracy – are appropriate for the intended uses of mapped data. Whether data are provided from a single source, or from a compilation and integration of sources, the intended user must know what the quality characteristics are for each and all constituent data objects, layers, or themes. This information is communicated as metadata associated with each map-data file or with mapped object attributes. Inevitably, as creators of geographic information find themselves also needing to access and use data created by others, the importance of rigorously maintaining accurate metadata is becoming accepted as a necessary part of professional practice.

Bruce Joffe is founding Principal of GIS Consultants (in Oakland, CA), providing GIS planning and implementation management services to cities, counties, and utility companies for over 24 years. In recent years, the practice has evolved to include organizational therapy and GIS public policy formulation.

Bruce recently served as President of the BAAMA chapter of URISA, and is a past Chair of California's Geographic Information Coordinating Committee. He currently is a member of the URISA Board of Directors. Mr. Joffe represented California on the FGDC Steering Committee. He chaired the second and seventh annual California GIS Conferences, in 1996 and 2001.

4 National Map data themes include digital orthorectified imagery, surface elevation, hydrography, transportation, structures, boundaries of government units, geographic names, and land cover classifications.

5 People interested in sponsoring or participating in the Open Data Consortium project may download a brief description from ftp://joffes.com, the file name is Open_Data_Consortium_concpt.pdf or, contact Bruce Joffe directly at GIS.Consultants@joffes.com