This article is published in GeoPlace.com

Weekly features February, 2002

How *Good* Are Your Maps?

*By **Bruce Joffe**, principal, GIS Consultants; e-mail: 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 last map you glanced at, and suppose you were to write down every fact depicted on it. 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, plus information about things not shown on the map, plus all the relative relationships, etc.

Maps are such concise conveyors of information, because they're scaled representations of reality. As such, their use and value depends on their accuracy or, more generally, their quality. Although maps vary in accuracy and quality, few can be classified as inherently having "high" or "low" accuracy or "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's good enough. If the same map, or mapping database, can correctly inform a lot of queries, it's 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's designed.

Map Quality

Although positional accuracy often is 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's updated.
- Representation--whether the mapped objects are connected, and if the symbology and annotation portray them understandably.
- Lineage--includes identification and quality assessment of the source documents as well as the methods by which the data were collected and multiple data sources were compiled into an integrated mapping database.
- Accuracy--includes factors such as: 1. Positional accuracy--how closely the mapped coordinates correspond to the "actual" coordinates of a

mapped object. 2. Referential accuracy--how closely the portrayed distances between mapped objects compare with field measurements. 3. 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. 4. Attribute accuracy--how reliably the attributes associated with a mapped object represent the actual object.

Assessing the quality level of a given mapping database 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 isn't 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 understanding felt by a statistical sample of the map-using population.

Map lineage assessment requires a standardized classification system of map compilation techniques, such as the one proposed by Rudy Stricklan (see "Cadastral Reference Databases: Categorizing and Certifying the Conversion Should Be Standardized," *PoB Magazine*, January 2000), which includes the following categories: direct observation, constructed from legal source documents, constructed from derived maps, trace digitized, measurements and adjustment factors recorded to enable "auto-refineable" update, measurements recorded as text and stored as attribution, measurements shown as text only, and 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 map scale. The NMAS standard says that maps with scales larger than 1:20,000 must locate at least 90 percent of their objects correctly, within one-thirtieth of an inch (at scale). Therefore, 90 percent of the objects mapped at a scale of 1:1,200 (one inch = 100 feet) must be correctly located within 3.33 feet (100-foot scale divided by one-thirtieth of an inch) of their "actual" location. Objects mapped at scales smaller than 1:20,000 must be correctly located within one-fiftieth of an inch.

The development of increasingly accurate methods and techniques for locational measurement, 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 in 1998, offers a methodology for comparing sample, mapped points with independent measurements of their location to derive a statistical assessment that's valid for 95 percent of the points. Generally, the standard defines the following methodology:

- Decide whether to test for horizontal or vertical accuracy, or both.
- 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.
- Select corresponding points from an independent, more accurate dataset. In many cases, the dataset will be created by onsight measurements. The independent dataset should be three times more accurate than the expected accuracy of the sample dataset.
- Record the measurement values.
- Calculate the sum of all "error radius" measurements, then calculate the average, and then take the square root to yield the root mean square (RMS) error.
- The NSSDA standard calls for an RMS error that includes 95 percent 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 percent confidence level is xx feet." This metadata statement enables the map database user to expect locational errors no worse than "xx" feet 95 percent of the time, regardless of the scale at which the map is displayed.

Accuracy Statements

When one considers city-, county- or state-wide GIS-based map data, one can expect there to be domains where the locational accuracy is better than the NSSDA 95 percent level as well as areas where the accuracy is worse. Older urbanized areas and less-developed areas tend to have less accurate mapping, and 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's 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.

The spatial relationship among floodplain and parcel boundaries is no better than the accuracy of the least-accurate data layer.

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.

Nevertheless, methods of adjustment often are applied to clearly identifiable points on the less-accurate layer that registers 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 isn't the same as accuracy. Responsible metadata should describe the adjustment transformations that were applied.

Complete Maps

Although data currency and accuracy are important factors of map quality, perhaps the most important aspect of a map's usefulness is the most obvious: its area of coverage. Just as the regime of paper maps seemed to follow a natural law that caused one's area of interest always to span multiple map sheets, so it seems that digital maps follow the "donut law"--when compiling multiple map files, there always seems to be a hole with no coverage.

Just as the USGS paper-based quad maps cover the entire United States, the agency now plans to create nationwide digital map coverage. The project, appropriately named the "National Map," will "provide the nation with current, accurate and nationally consistent basic spatial data." Moreover, although the average currency of USGS quad maps is 23 years, the National Map's intent is "to deliver spatial information that is not more than seven days old."

Such lofty ambition should grab the attention of our geographic information community, but the reaction may be characterized as no more than passively supportive. We know how much difficulty USGS has had in funding its current mapping programs, so how is the National Map going to be built so effectively? Rather than build the entire digital product itself, USGS is positioning itself to stimulate and organize collaboration with state and regional governments, universities and private industries to coordinate and compile all relevant spatial datasets. The 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 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 public and private data producers that have map data to contribute to the public-domain National Map. I proposed the "Open Data Consortium" project to organize a series of workshops and discussions in which producers of geographic information can meet with data distributors to formulate Model Data Distribution Agreements. Currently, the Urban and Regional Information Systems Association and the GeoData Alliance have endorsed the initiative. (People interested in participating in the Open Data Consortium project may download a brief description from *ftp://joffes.com,* under "Open_Data_Consortium_concpt.pdf.")

Geographic information can be used effectively only when its quality characteristics are appropriate for the intended use of the data. Whether data are provided from a single source or from a compilation and integration of sources, intended users must know what the quality characteristics are for each and all constituent data objects, layers or themes. Such 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.