VISUALIZING THE QUALITY OF SPATIAL INFORMATION

Barbara P. Buttenfield
NCGIA, Department of Geography, SUNY-Buffalo, Buffalo, NY 14261
email GEOBABS@UBVMS

M. Kate Beard
NCGIA, Department of Surveying Engineering
University of Maine, Orono, Maine 04469
email BEARD@MECAN1

DESCRIPTION AND SCOPE OF THE ISSUE

Technology currently allows us to process and display large volumes of information very quickly. Effective use of this information for analysis and decision making presupposes that the information is correct or reasonably reliable. Information on the quality of data is essential for effective use of GIS data: it affects the fitness of use of data for a particular application, the credibility of data representation and interpretation, and the evaluation of decision alternatives. The credibility of spatial decision support using GIS may indeed depend on the incorporation of quality information within the database and the display. As Goodchild (1990) states the best insurance will be to sensitize the GIS user community to accuracy issues and to develop tools which allow spatial data handling systems to be sensitive to error propagation. Visualization should be explored as a method for capturing, interpreting and communicating quality information to users of GIS. Clearly, the quality of information varies spatially, and visual tools for display of data quality will improve and facilitate use of GIS. At present, those tools are either unavailable (in existing GIS packages) or not-well developed (error models and the process of visualization are only recently beginning to be addressed directly as research topics).

The quality of spatial data and databases is a major concern for developers and users of GIS (Chrisman, 1983). The quality of spatial information products is multidimensional, and relates to accuracy, error, consistency and reliability. Implications for spatial analysis and for spatial decision-making are too complex for a comprehensive inventory, but can be identified in theoretical work (for example in spatial statistics) as well as in GIS applications (for example in resource management). This paper presents an initial framework for discussion of the role of visualization for understanding and analyzing information about the quality of GIS data. The discussion will proceed from and expand upon the ideas presented here in a panel session at the meeting.

This paper represents part of Research Initiative #7, "Visualizing the Quality of Spatial Information", of the National Center for Geographic Information and Analysis, supported by a grant from the National Science Foundation (SES-88-10917); support by NSF is gratefully acknowledged.

Our goal in this research panel is to bring together representatives from academia, federal agencies and the private sector to present their needs for knowledge about the quality of spatial data products. Discussion will focus on effective means to manage and visually communicate components of data quality to researchers, decision-makers and users of spatial information, particularly in the context of GIS. The intention is to consider a variety of perspectives on topics for a research agenda available to the general GIS community, and to hear the various sectors (educational, commercial and applications) express priorities for topics in the agenda.

THEMES FOR RESEARCH

Questions and impediments relating to the visualization of data quality conceivably cover a very broad ground. For example, issues of modeling and sensitivity analysis might be considered to determine what visual tools are appropriate for particular models, the opportunity for visualization to facilitate spatial analysis, and caveats to consider in implementing visual tools in modeling. The role of visualization in geographical analysis and its role in hypothesis testing and data exploration have been recently reviewed (Buttenfield and Mackaness, 1991), but these topics lie beyond a manageable scope for the panel. Instead, impediments and research priorities within four categories will be addressed. These include defining components of data quality, identifying impediments for maintenance of data models and databases, addressing representational issues, and evaluating particular user needs for data quality information.

Data Quality Components. Perhaps the most commonly cited component of data quality relates to measures of error. Commonly recognized errors include those associated with data collection(source error) and the processing of data for map compilation (process error). Information on source error is often discarded with the completion of map compilation. Process errors have proven difficult to analyze in many cases, for example in studies of digitizing error, or in modeling error associated with soil mapping (Fisher, 1991). In statistics, the concept of Least Squares Error has been applied to determine reliability (or what is called 'confidence') in hypothesis testing. A third error component (use error) is associated with the appropriate application of data or data products (Beard, 1989).

By some definitions, error (the discrepancy between measurement and true value) is much more difficult to assess than accuracy (the discrepancy between measurement and a model). The best examples of this may be found in determination of geodetic position, which until the development of GPS systems was limited to (albeit precise) projection of location with reference to a geodetic spheroid and datum. The Proposed Standard for Digital Cartographic Data Quality (Moellering, 1988) incorporates three accuracy measures (positional accuracy, attribute accuracy, and consistency) in addition to lineage and completeness.

A standard definition of data quality and its components may be difficult to agree upon, as the domain of an application will likely impact

the user needs. For soils data, for example, requirements for consistent attribution of soil type are more readily evaluated than requirements for accurate positioning of soil parcel boundaries. For demographic data, where enumeration boundaries are determined by mechanisms unrelated to the particular variable at hand, just the converse may be true. Regardless, there should be consensus about some of the research priorities for this theme:

What visual tools are appropriate for particular error models?

How can visualization facilitate monitoring of error propagation?

Data Models and Database Issues. Management of data quality within a GIS database requires attention during manipulation and update, and will likely impact upon the future architecture of such databases for implementation. Information about the information within a database is referred to as metadata, and has recently become a research issue in its own right (see for example Lanter and Veregin, 1990). The representation of data quality components in a data structure will not only have requirements to facilitate their visual display, but also must be implemented with efficient pointers and links to facilitate update operations. Analysis of error propagation might also be facilitated by visual display, and the design of these graphic tools may not be closely aligned with the design of conventional GIS graphics. This will be covered under the third theme presented below. Other questions arise:

How can the metadata be updated simultaneously with the data?

What database requirements must be implemented to accommodate real-time data quality representations for static GIS products, or for dynamic displays?

Can current data structuring alternatives accommodate changes to data and data quality in effective ways? How can links b/t data and data quality be preserved during database modification or update?

Representational Issues. The ease with which visualization tools may be integrated within GIS packages varies considerably depending on at least three issues, including the domain of the phenomena to be studied, the purpose or intent of the user, and the format of the GIS software (MacEachren, Buttenfield, Campbell, and Monmonier, 1991). This presents a substantial challenge to the system designer. Buttenfield and Ganter (1990) suggest that GIS requirements for visualization include conceptual, technological, and evaluatory solutions, which may be seen to vary over three broad domains: inference, illustration, and decision-making. Each presents a challenge to the integration of appropriate visualization tools.

Maps are a major tool for decision-making with GIS. Current GIS software includes functions to create cartographic output automatically or interactively. However, none of the current turnkey systems include mechanisms to ensure correct use of graphics functions. Poorly designed maps may convey false ideas about the facts represented by the data, and bias

the decision-making process. Weibel and Buttenfield (1988) explore ways to improve the quality of GIS map products, and increase effectiveness of information transfer based on graphics. Their guidelines may provide only a rudimentary implementation for visualizing data quality. Research priorities that come to mind under this theme may involve both system benchmarking and cognitive evaluations, as seen for example by the following questions:

> What design tools are appropriate for graphical depiction of data quality?
>
> Will generation of realtime data displays during database update facilitate monitoring of error and error propagation?
>
> How can the effectiveness of such displays be evaluated? For example, What is the utility of embedding data quality with data in graphic display? Can the two be merged, or is this too much of a cognitive challenge for effective interpretation?

Evaluation of User Needs. Ganter (1990) discusses visualization from a cognitive as opposed to graphical perspective, cautioning readers that discovery and innovation, which have traditionally involved thinking visually and producing images, increasingly benefit from GIS and CAD. He argues for the importance of understanding the human faculties which use pictures as tools in thinking. Science and engineering define problems, explain processes, and design solutions through observation, imagination and logic. Evaluation of user demands for data quality information will require sensitivity to the internal (perceptual and cognitive) mechanisms by which spatial and temporal patterns are interpreted.

Equally important is the need for sensitivity to the domain of the GIS application. For example, reliability associated with a routing of emergency dispatch vehicles will likely vary with each link of the route; this information must be presented with high precision and in a short timeframe. Reliability variations associated with the environmental impact of a timber clear-cut operation cannot be tied to a routed network, and variations may be interpolated as opposed to tabulated raw data. In this context, some research questions may be proposed:

> What are expectations of GIS users regarding data quality displays?
>
> How will visualization of data quality impact upon the reliability and credibility of spatial decision-making using GIS?

SUMMARY

With advances in technology, storage and displays mechanisms are now in place for real-time display not only of spatial pattern but also of the quality of the rendered data. Developments in software provide spatial inference and statistical explanation to the verge of providing models about

the reliability and consistency of spatial interpretation, and this has paved the way for application of GIS to policy-making and decision support. There is a need and timeliness to consider data quality issues in the context of GIS. Our acuity for visual processing indicates that current technology in graphical display may assist our efforts to validate the decisions and results based on GIS analyses. The panel discussion is intended to present multiple viewpoints and to encourage the research and user community to address visualization of data quality as an attainable goal in the development of GIS.

REFERENCES CITED

Beard M K (1989) Use error: The neglected error component. **Proceedings,** AUTO-CARTO 9, Baltimore, Maryland, March 1989: 808-817.

Buttenfield B P and Ganter J H (1990) Visualization and GIS: what should we see? what might we miss? **Proceedings,** 4th International Symposium on Spatial Data Handling, Zurich Switzerland, July 1990, vol.1: 307-316.

Buttenfield B P and Mackaness WA (1991) Visualisation. Chapter II.a.4 in Maguire, D. Goodchild, M.F. and Rhind, D (Eds.) **GIS: Principles and Applications.** London: Longman Publishers Ltd. (in press)

Chrisman N R (1983) The role of quality information in the longterm functioning of a geographic information system. **Cartographica** 21(2): 79-87.

Fisher P F (1991) Modeling Soil Map-Unit Inclusions by Monte Carlo Simulation. **International Journal of GIS** (in press).

Ganter J H (1988) Interactive graphics: linking the human to the model. **Proceedings,** GIS/LIS '88, pp. 230-239.

Goodchild, M F (1990) Spatial Information Science. **Proceedings,** 4th International Symposium on Spatial Data Handling, Zurich, Switzerland, July 1990, vol.1: 3-12.

Lanter D P and Veregin H (1990) A lineage meta-database program for propagating error in geographic information systems. **Proceedings** GIS/LIS '90, Anaheim, California, November 1990, vol.1, 144-153.

MacEachren A E, Buttenfield B P, Campbell J C and Monmonier M S (1991) Visualization. In Abler, R. A., Olson, J. M. and Marcus, N. G. (Eds.) **Geography's Inner World.** Washington, D. C.: AAG (forthcoming).

Moellering, H (1988) The proposed standard for digital cartographic data: report of the digital cartographic data standards task force. **The American Cartographer,** 15(1) (entire issue).

Weibel, W R and Buttenfield B P (1988) Map design for geographic information systems. **Proceedings,** GIS/LIS 88, November 1988, San Antonio, Texas vol.1: 350-359.