GeoInformatica: An International Journal on Advances of Computer Science for Geographic Information Systems

GeoInformatica is abstracted and/or indexed in *Zentralblatt für Mathematik/Mathematics Abstracts* and *Geographical Abstracts : Physical Geography GEOBASE.*

Subscription Rates

The subscription price for 1998 Volume 2 (4 quarterly issues) is:

Institutions:	US $354.00	NLG 690.00
Individuals:	US $75.00	NLG 130.00

The above rates are inclusive of postage and handling. The individual rate is not available to institutions, libraries, or companies.

For airmail delivery, please add US $8.50/NLG 22.00. The journal is shipped to the U.S.A. and Canada in bulk airfreight at no extra cost.

U.S. mailing agent: Mercury Airfreight International, Ltd., Inc.
2323 Randolph Ave.
Avenel, NJ 07001, U.S.A.

Postmaster: Please send all address corrections to: *GeoInformatica* c/o Mercury Airfreight International, Ltd., Inc., 2323 Randolph Ave., Avenel, NJ 07001, U.S.A.

Ordering Information/Sample Copies:

Subscription orders and requests for sample copies should be sent to:

Kluwer Academic Publishers
Post Office Box 358
Accord Station
Hingham, MA 02018-0358, U.S.A.
Phone: 781-871-6600
Fax: 781-871-6528
email: Kluwer@wkap.com

or

Kluwer Academic Publishers
Post Office Box 322
3300 AH Dordrecht
THE NETHERLANDS
Fax: 011-317-833-4254

Subscriptions may also be sent to any subscription agent. Private subscriptions should be sent directly to the publisher at the above addresses.

ISSN 1384-6175

GeoInformatica 2:3, 211–214 (1998)

Editorial

MICHAEL F. GOODCHILD, GUEST EDITOR
University of California, Santa Barbara
good@ncgia.ucsb.edu

ROBERT JEANSOULIN, GUEST EDITOR
Centre National de la Recherche Scientifique, Marseille
robert.jeansoulin@lim.univ-mrs.fr

Problems of data quality have long threatened to be the technology's Achilles heel. We know that the data on which GIS analyses and reasoning are based are not perfect, and yet we frequently lack the means to describe their level of imperfection, and the means to determine the effects of imperfection on results. It is clear that most numerical results from GIS analyses and reasoning are not sufficiently accurate to match the precision with which the results can be reported (roughly 1 part in 10^7 for single precision arithmetic, 1 part in 10^{14} for double precision). But too often it has been almost impossible to turn that into a positive statement: what exactly is the accuracy of GIS analysis, and how reliable is GIS-based reasoning?

Much has been written on the subject of quality in GIS in the past 10 years since the appearance of *Accuracy in Spatial Databases* [1], and much progress has been made. At least one thing is sure: that the problem is much more complex than we had earlier thought. Assessment of accuracy requires a comparison between the data and the truth, but there are abundant problems in defining truth in connection with geographic data. And if *truth* is replaced by the weaker *source of higher accuracy* then the definition takes on an awkward aspect of circularity. Moreover, accuracy so defined is only one of numerous ways of thinking about the broad issue of quality.

The four papers in this issue on Quality in GIS cover much of the spectrum of current research, and create an excellent picture of this research field. They have been developed from the best of the papers given at an international workshop on quality in GIS held in Paris in April 1997, under the auspices of the Cassini research network, and through the initiative of Robert Jeansoulin. A more complete record of the workshop can be found in the published proceedings [2].

One of the most exciting aspects of recent research on data quality in GIS is its sheer diversity of approaches. Unlike traditional error analysis in science, which deals with the differences between measurement and truth, the analysis of data quality in GIS must deal with the enormous complexity in the methods devised by humans for describing the world around them. Thus, it is not surprising that current research is dominated by several apparently incompatible paradigms. Perhaps the simplest of these is what one might term the *scientific measurement* paradigm, which attempts to recruit traditional error analysis to the geographic context. It has led Heuvelink, for example, to a very successful analysis of error propagation in GIS, using a variety of analytic and simulation methods [3]. The scientific measurement paradigm also makes heavy use of *probabilistic* methods, in order to characterize distributions and as the basis for testing hypotheses.

But while error analysis and probability may satisfy the needs of much of science, they fall far short of the needs of the complicated world of GIS. Specifically, it has proven impossible to develop a theory of data quality in GIS that does not reflect the intimate involvement of the human user and observer in the GIS process.

- The human user needs to know about the quality of data and results, so these must be presented in ways that are intelligible to the user, who may lack formal scientific training.
- Much observation in GIS is inherently subjective and uncertain, and thus not replicable in a scientifically rigorous sense. Human uncertainty may be better analyzed using methods of subjective probability and fuzzy sets.
- While scientific facts can be submitted to empirical test for verification, many facts in GIS represent human belief, and yet may be just as significant and useful in GIS applications.

Today, research in GIS data quality may appear to the outsider to be hopelessly split by apparently incompatible paradigms. There appears, for example, to be little hope of reconciling fuzzy and probabilistic approaches, since they disagree at the level of their axioms as well as in their interpretations. If one takes a *field* view that the geographic world can be characterized by a series of functions over its spatio-temporal domain, then quality is a matter of measuring the differences between functions; but if one takes a *discrete object* view, that the geographic world is an empty space littered with discrete point, line, area, or volume objects, differentiated by their attributes, then positional quality and attribute quality can be separated.

An empiricist might demand impatiently that one resolve these differences by empirical test: can't we decide whether the world is continuous or discrete, or whether the axioms of fuzzy sets are more correct than the axioms of probability? Can't one side *win*? But empirical test is only one part of the GIS story. An empiricist might argue that boundaries between soil types never exist, in the sense that they can never be defined exactly on the ground, and that any theory or system that allows a point to be *on* such a boundary is unnecessarily complex. But a technology for use by humans must reflect to some degree the priorities of humans, and if humans believe in soil boundaries then so be it. The legal system perhaps comes closest to a compromise between field and discrete object views, in its insistence that for some purposes ownership is a field, having a nominal value z at every point (x, y); while for other purposes the parcel belonging to an owner is a discrete object, which may or may not match perfectly with other objects at its edges.

Thus we find ourselves after a decade of intensive research into GIS data quality with many approaches, each useful for some purposes, and with deep incompatibilities between them. The four papers in this issue reflect this situation in microcosm. They include two that are firmly embedded in the probabilistic, scientific measurement paradigm; and two that are focused on the logic of discrete objects. There may well be no prospect of merging these two approaches; but on the other hand the presence of such divergent approaches within one field is very exciting, and will lead to very fruitful cross-stimulation. Moreover, *GeoInformatica* as a journal that overlaps both geography and computer science, seems an ideal forum for this process.

Fisher's paper addresses the question of DEM accuracy, and examines how a set of independently determined spot heights might provide the basis for a comprehensive geostatistical model of error that successfully incorporates both spatial dependence and spatial heterogeneity; the paper can also be seen as an interesting way of *integrating* two data sets that provide information about the same geographic phenomenon, but at different scales and accuracies. Jones presents an extremely useful analysis of the complete set of algorithms for estimating slope, based on a simple statistical model of error, and provides methods that could be readily incorporated into GIS software, especially if they could be combined with Fisher's model of spatial dependence and heterogeneity in the error field.

Papers by Worboys and by Hölbling, Kuhn, and Frank show us the other side of the data quality research coin. The user's expectation that a GIS operation computes a geometrically exact result of course contradicts the GIS reality, that locations can be stored and intersections computed only to a finite level of precision. While such problems may be largely hidden from the user, they become surprisingly visible in certain artifacts, such as the spurious slivers of a polygon overlay. The authors show that a logically consistent, finite-resolution system can be built based on the principle of simplicial complexes, by reducing every task to a set of operations on planar-enforced triangles. A system based on these principles would never produce spurious artifacts, and would be consequently much easier to build and maintain. Worboys presents a comprehensive theory of imprecision in objects, with interesting and useful treatment of hierarchical relationships across scales. Such formal treatments may seem esoteric, but they have the potential to do several very useful things for the field of GIS: to standardize its terminology; to provide it with sound theoretical underpinnings; to integrate the treatment of objects with imprecise (numerical imprecision) or uncertain boundaries (conceptual imprecision, like in the soils example mentioned earlier); and to provide the basis for much simpler, more comprehensive, and more internally consistent programming environments.

In writing this introduction to four papers, which divide so neatly into two pairs, embedded in two distinct and apparently incompatible paradigms, we are reminded of how Noah saved the biosphere by taking two of each species into the Ark. Members of different species cannot produce offspring with merged characteristics; but species are interdependent, and all contribute to the success of the ecosystem. We hope the juxtaposition of two such distinct approaches to data quality, and the excellent quality of all four papers, will similarly reinforce the notion that an understanding of data quality in GIS must draw on many different perspectives, and need not attempt to resolve between them.

References

1. M.F. Goodchild and S. Gopal. (Eds.). Accuracy of Spatial Databases. Taylor and Francis: London. 1989.
2. M.F. Goodchild and R. Jeansoulin (Eds.). Data Quality in Geographic Information: From Error to Uncertainty, Editions Hermes: Paris, 1997.
3. G.B.M. Heuvelink. Error Propagation in Environmental Modelling. Taylor and Francis: London. 1998.

Michael F. Goodchild is Professor of Geography at the University of California, Santa Barbara; Chair of the Executive Committee, National Center for Geographic Information and Analysis (NCGIA); Associate Director of the Alexandria Digital Library Project; and Director of NCGIA's Varenius project. He received his BA degree from Cambridge University in Physics in 1965 and his Ph.D. in Geography from McMaster University in 1969. After 19 years at the University of Western Ontario, including three years as Chair, he moved to Santa Barbara in 1988. He was Director of NCGIA from 1991 to 1997. In 1990 he was given the Canadian Association of Geographers Award for Scholarly Distinction, and in 1996 the Association of American Geographers award for Outstanding Scholarship; he has won the American Society of Photogrammetry and Remote Sensing Intergraph Award and twice won the Horwood Critique Prize of the Urban and Regional Information Systems Association. He was Editor of *Geographical Analysis* between 1987 and 1990, and serves on the editorial boards of ten other journals and book series. His major publications include Geographical Information Systems: Principles and Applications (1991); Environmental Modeling with GIS (1993); Accuracy of Spatial Databases (1989); GIS and Environmental Modeling: Progress and Research Issues (1996); and Scale in Remote Sensing and GIS (1997); in addition he is author of some 300 scientific papers. He is currently Chair of the National Research Council's Mapping Science Committee. His current research interests center on geographic information science, spatial analysis, the future of the library, and uncertainty in geographic data.

Robert Jeansoulin is CNRS research fellow in the "Logic" team of the LIM, Computer Science Lab in Marseille, and Director of "Cassini", the French public research consortium on GIS (CNRS-IGN).

He holds a B.Sc. in mathematics (Marseille), a B.Sc. in computer science (Toulouse), a Dr. Ing. degree and a Ph.D. in computer image processing (Toulouse, 1982). He was Visiting Scientist at the USC (Los Angeles, 1983). He worked several years with the CNES (Image Processing Division), the INRIA (Data Base Project) and was Associate Professor at Univ Paris-Sud (Orsay).

He is member of several French Scientific programs: Remote sensing, Environment, and CNRS Social Sciences sector.

His experience covers remote sensing and geographic data: from image processing (image registration, fuzzy edge detection, region growing, neural networks classification), to data base (raster-vector object model). He was IGN 1992 laureate with its software 'Goodies'. His current interest focuses on AI approaches to geographical information.