Title
Two Perspectives on Data Quality (92-12)

Authors
Couclelis, Helen
Beard, Kate
Mackaness, William

Publication Date
1992-12-01

Two Perspectives on Data Quality

Geographic Knowledge Production Through GIS:
Towards a Model for Quality Monitoring

By

Helen Couclelis
University of California, Santa Barbara

and

Data Integrity: The Academic and Research Perspective

By

Kate Board and William Mackaness
University of Maine, Orono

Technical Report 92-12

December 1992

Geographic Knowledge Production Through GIS:

Towards a Model for Quality Monitoring

By

Helen Couclelis
University of California, Santa Barbara

Geographic knowledge production through GIS: Towards a model for quality monitoring.

HELEN COUCLELIS
Department of Geography
University of California
Santa Barbara, CA 93106

Abstract.

GIS is a tool for the production of applied geographic knowledge as well as for its wide dissemination and use. This paper examines the problems of quality control and monitoring resulting from the ease with which non-experts can now produce, using GIS, many kinds of documents thus far obtainable only from experts. Because the quality of a GIS product is largely a function of the background assumptions and specific purpose of each application, it is in general not possible to build appropriate product quality controls directly into the system. A conceptual model is proposed, based on a linguistic metaphor, that would allow users to monitor visually the quality of a GIS product at any stage of deriving an application. The model could be implemented as a "metal'expert system incorporating general rules for the selection of the specific kinds of quality criteria and quality feedback indicators suitable for each application.

1. Introduction

Most current definitions of GIS stress its functions of data storage, manipulation, and analysis, with an increasing emphasis on the latter aspect. Less often discussed is the fact that GIS is also a tool for the *synthesis* of geographic information, and therefore, for the production of applied geographic knowledge. GIS users can combine information from diverse data sources, available in a wide variety of different formats, to generate maps, tables, graphics, and other products that, in many cases, until recently only specialists could deliver. Not only do GIS-generated maps dispense with the cartographer, but any sufficiently complex GIS product may substitute for a document that might have been provided by the regional geographer, the urban geographer, the forestry expert, the surveyor, the agricultural specialist, and so on, often only after extensive field work.

Thus, the much hailed "electronic democracy" brought about by the informational revolution, does not only refer to the wide availability and ease of *use* of information of different kinds; in the case of GIS, it also means the democratization of the *production* of applied geographic knowledge. But democracy has its problems, as we all well know. This paper addresses a critical issue resulting from the ease with which non-experts using GIS can now generate items of geographic knowledge thus far obtainable only from experts: the question of *quality control* of the resulting products.

The paper is structured as follows. Section 2 examines some of the issues resulting from the democratization of geographic knowledge production through GIS. The issue of product quality is contrasted with data quality, which is only one of its dimensions. The other components of product quality are even less tractable, as they have to do with context-dependent factors such as the implicit spatial model and specific purpose of each particular application. In section 3, 1 develop a metaphor likening the interaction of a GIS user with the system to interpersonal discourse, highlighting how human partners in conversation give each other, on a continuous basis, abundant clues as to the quality of what they say. I then outline a framework for a semiology of mostly graphic signs that would play a similar role in the course of the interactive development of GIS applications. The paper concludes with a discussion of the prospects for such a model to be implemented as a "metal'expert system, incorporating general rules for the provision of the appropriate kind of ongoing visual feedback to monitor product quality in different kinds of applications.

2. Problems of quality in "democratic"' knowledge production

Discussions of quality in GIS usually revolve around issues of data quality and error (Goodchild and Gopal 1989; Beard, Buttenfield and Clapham, 1991). While a major aspect of the quality of the final product, data quality is only one of its dimensions. Indeed, it may be argued that from a pragmatic point of view, data quality is not of direct concern to the user. The GIS user intent on getting a good product for his or her purposes is not interested in data quality per se any more than the taster of a chocolate cake is interested in the taste of the flour or the baking soda that went into its making. Just as the proof of the pudding is in the eating, for most practical purposes, the critical question in a GIS application is the quality of the end product. A first task is thus to put the issue of data quality in the proper context. The other, even less tractable dimensions of product quality will be discussed next.

2.1 Data quality vs. product quality

The question of data quality is difficult enough in itself. It is obviously associated with that of data error, although the two are not synonymous: whereas data error can be discussed in absolute terms, data quality is a function of context and purpose. Tell me what you intend to do with them and I will tell you how good your data are. Gross errors in the third dimension are irrelevant if all you want is a two-dimensional map. Bad data may look good if your other data are even worse. Data that are unacceptable at some scale may be just what is needed at a coarser level of resolution, and longitudinal data too noisy for year-by-year analysis may be fine for longer intervals. Slivers and gaps between what should have been adjacent polygons could have homicidal consequences on a cadaster map, but may be welcome on a planner's sketch where hard boundaries are barriers to the imagination. And one of the best known, most widely used, and most useful maps in history, the London subway map, reflects spatial data that would be terrible for most purposes, though obviously right for that particular one. These facts are well known. It is also well known that things can get even worse when you consider what happens to source data that must undergo a complex sequence of manipulations. Years of work on error propagation in models, culminating in the flurry of spectacular graphical outputs from models of chaotic systems, have yielded sobering insights: while in some cases errors may cancel out, in other cases they are so quickly amplified through non-linear dynamics that even the best of data can produce garbage after a while. Putting a spatial data base through a series of GIS manipulations is also like running a complex spatial model of sorts, albeit one that is very largely unexamined and ad hoc in terms of its underlying structure. Tracking the propagation of uncertainty through GIS operations on data is currently an active area of research (Goodchild et al. 1992). This work, along with the vast literature on data error, and on error in spatial data bases in particular (as reviewed by Veregin, 1989), obviously has a direct bearing on the issue of product quality in GIS: still, by itself, it does not even get close to answering the question.

Just as data is not knowledge, data quality is not the same as product quality in GIS. The quality of applied knowledge of any sort is judged not so much by its intrinsic truth or freedom from error as by its capacity to answer practical questions and solve problems without causing other problems along the way (Laudan 1977). This much also holds true, in general, for the GIS product. But an understanding of the GIS product as expressing new applied geographic knowledge must be refined by an appreciation of its hybrid nature as part electronic cartography, and part technological artifact. These two aspects relate to two other major dimensions of quality in the GIS product: these are, respectively, its degree of agreement with the model of space implicit in the application, and its degree of adaptation to the requirements of the purpose at hand.

2.2 Designing applied geographic knowledge

The most characteristic product of the GIS application, the map-like display or hard copy, is on the surface similar to the product of cartographic design. Yet the question of quality for maps does not have anywhere near the same urgency. The cartographic map is the result of a design process which, while still evolving, has been codified and refined over many centuries. Part science and part traditional art, making a good map is like making good wine, produced by a few experts for the benefit of all. The GIS product by contrast is more like what comes out of a kit for do-it-yourself chemistry experiments: it can be tailored to one's desires, it is endlessly varied, often surprising, frequently hard to understand, sometimes insiduously lethal, and the (amateur) maker and (naive) user are often one and the same. Safeguards and warnings are built into the kit to help avoid the most likely mistakes and disasters, but as for every open-ended process, there can be no guarantee as to the quality or safety of the final outcome, and accidents do happen.

Recent work on automated cartography has shown both the promise, and the difficulty, of turning the science and art of the cartographer into sets of unambiguous rules for use in computer-based systems and in particular GIS (Buttenfield and McMaster 1991). There is a distinction to be made between the correct use of the cartographic language on the one hand (such as the appropriate use of hues or symbols), and the correct choice of cartographic representation, including the choice of projection, scale, distance metric, and thematic layout. While there is good hope that the former aspects will eventually be automated to a large degree, this is not the case with the latter, more conceptual and creative aspects of cartographic design. It is too much to expect of the user who has little or no background in cartography or geography to get these notions straight. Even the geographically trained user working with GIS may be excused for forgetting the significance of basic geographic concepts on occasion. It is easy to lose track of the notion of map scale when one can zoom in and out at the touch of a key, of the notion of direction, when it is so simple to rotate the display whichever way one likes, of the notions of distance and shape, when one can "rubbersheet" any map, or of the notion of location, when one works with raster elements. It is just as easy to lose sight of the true texture of terrain when elevations are routinely exaggerated for visual impact, or to accept uncritically the system's determination of a viewshed, forgetting about the thick tree cover. The very same technical capabilities of GIS that promote creative geographic thinking when used properly, are also accountable for many of the subtler quality problems in GIS products.

Qua artifact produced by advanced technology, the GIS product is controlled by a large number of safeguards built into the system to prevent manipulations that are illegal in some well defined sense, sefeguards deriving from requirements of logic, software design, or the underlying geographic data model. This ensures up to a point the technical adequacy of the product, but does little for those aspects of product quality relating to its function and purpose. In his classic monograph on the theory of design, Herbert Simon

(1969) distinguishes between the *internal* and *external* environments of a design product. The internal environment is made up of the physical and logical constraints that a design must meet to be technically viable, while the external environment consists of the set of requirements that it must satisfy in order to be able to fulfill successfully its intended purpose. A good design reflects both what the artifact is made of, and what it is made for. Because the products of GIS are members of an open set of possible designs and purposes, the quality controls relating to the external environment cannot be built in.

The lack of a priori rules and criteria for evaluating their fitness for a specific purpose, is what distinguishes items of knowledge in general, and the products of information technology in particular, from other artifacts and designs. If the cake has too much salt, or if the chemistry experiment blows up in your face, you know it at once; if the model airplane you just built cannot fly, you soon find out. But a bad GIS product does not necessarily carry with it the signs of its own inadequacy any more than a bad theory of navigation does. Bad knowledge is not intrinsically different from good. Information technology differs from ordinary, material technology in that the most critical aspects of the quality of its products are only testable through their indirect consequences.

Because of the difficulty, if not impossibility, of testing the quality of knowledge directly, society has entrusted not only its production, but also its evaluation to experts and specialists. The "electronic democracy" of which GIS is part runs counter to the centuries-old trend for greater and greater specialization of the fields of knowledge. With it, part of the reassurance that an accredited, experienced specialist brings to the task is lost. With GIS, anyone can now be an instant plant ecologist, regional geographer or surveyor. Much is gained, but something - a tentative, at least, guarantee of product quality - may be lost.

3. The conversational metaphor

Society's way of controlling and monitoring the quality of the knowledge produced its reliance on the judgement of specialists and experts suggest how the product quality issue in GIS might be approached in an environment consisting of a GIS system and its user(s). A key aspect of GIS as information technology is its interactive nature. In a GIS application, new knowledge about some aspect of the geographic world is produced cooperatively between a human of varying ability and sophistication, and an electronic partner who is incredibly smart in some respects, and unfathomably stupid in others. Although there is no direct analog of this kind of partnership in the realm of human discourse, it is a case of interactive knowledge production involving two parties with complementary forms of expertise.

A persistent theme in the literature on human-computer interface design concerns the utility of models of human-tohuman communication. I agree with those weary of anthropomorphizing the computer too much, and have little patience for over-friendly systems that address me by my first name. At the same time, it is difficult to ignore the profound analogies between human discourse and personcomputer interaction that several researchers have already pointed out (Laurel 1990; Winograd and Flores 1987).

Viewing the GIS application as an instance of cooperative production of applied spatial knowledge makes questions about the quality of the GIS product analogous to questions about the quality of knowledge produced by a human team through verbal interaction. A prerequisite for meaningful conversation to take place among humans is that the partners share a *schema* of the situation at hand, that is, a cognitive model of the issue discussed along with its background of assumptions, tacit presuppositions, appropriate dispositions for action, and other relevant forms of understanding. Failure to share such a background in human interactions results in the partners talking past each other and using the same words but meaning different things, giving rise to *quid pro quo's* and other predicaments celebrated in farce and drama. The schemas that GIS systems embody, in the guise of digital representations of geographic space, are highly specialized, focussed and controlled (though not necessarily by their users), but can still differ considerably from the spatial schemas of their human partners, or even from those of other systems in the same working environment.

Dutton (1984) argues that the question of data quality in GIS often relates to the "spatial ontology" embodied in the system. This is usually understood to mean a field or object view of geographic space, which practically reduces to the technical raster-vector distinction. In actual fact, there are several more ontologies of space reflected in geographic theory and applications. Though not supported by widely accepted data models, other formal models of space are certainly possible, and indeed several different ones (such as "coverage" or "network" models) are used in different kinds of applications. In an overview of conceptions of space used in geography, Couclelis (1992) distinguishes between geometrical, physical, socioeconomic, behavioral and experiential spaces, few of which can be accommodated comfortably within either a pure field or object perspective as currently understood. The issue of continuous versus discrete space, also discussed in that article, underlies a significant discrepancy in background spatial representations limiting the usefulness of GIS for environmental modeling (Kemp 1992). Old debates about "absolute" *versus* "relative" space in geography have been recently revived, as some see another such discrepancy between the absolute conception of space embodied in GIS on the one hand, and the relative conception underlying most mathematical models in human geography on the other (Couclelis 1991). In an attempt to bridge the gap between absolute and relative space, Getis (1992) recently formalized the notion of "proximal space", each point of which contains information about the points around it.

Despite the proliferation of background conceptions of space, a functional typology of geographic data models is still missing. Frank (1992) sees the definition of a small set of generic spatial data models for reference purposes as a very important research problem, the solution of which will not only benefit the design of task-specific user interfaces, but will also facilitate the adaptation of the system itself to the task at hand. We can expect that the quality of a GIS product (as well as the quality of the data used in it) will be to some extent a function of the degree of agreement or discrepancy between the model of space embodied in the system and that presupposed by the application itself.

Even assuming compatible spatial ontologies, the interactive production of spatial knowledge using a GIS is subject to the same difficulties as the production of any kind of applied knowledge through conversation. Well-meaning conversation partners can arrive at bad conclusions for many different kinds of reasons, having to do with the quality of their beliefs relating to the area under discussion, the quality of the inference processes used, or the quality of the conversational interaction itself. In close analogy, the GIS product may suffer from insufficient (for the particular appplication) quality in the source data, from problematic data manipulations due either to the operation of the system itself or to a poor model of the spatial phenomena analyzed, and from a misleading, inefficient, or inefficiently used interface.

Human knowledge structures are characterized by several kinds of imperfections, of which those discussed in the quantitative literature on error are only a small part. Thrift (1985) discusses five forms of "unknowing": that which is unknown or unsuspected; that which is not understood; that which is undiscussed or undisputed; that which is deliberately hidden; and that which is distorted. All of these can occur in the kinds of spatial knowledge represented and **produced through GIS, although only** the last kind, distorted knowledge, is much discussed in the literature on error.

Processes of inference in humans can go wrong because of faulty deductions, because of hasty inductions, because of unwarranted generalizations, because of improper handling of dubious beliefs, because of the use of inappropriate analogies or heuristics, because of loss of focus on the problem at hand, because of having been led up the garden path, because of the use of bad examples, because of drawing the wrong lessons from experience, and so on. Quantitative work on uncertainty and reasoning, and in particular formal logic, has modeled some of these inference processes in ways that can be directly transferred to computer-based systems (Shafer and Pearl 1990), although many would argue that such models are more appropriate for computer than for human reasoning anyway.

In conversational interaction, beliefs and inferences are expressed in the kinds of speech acts studied by Searle (1969) and others. These are summarized by Winograd and Flores (1987, p.58) as follows:

Assertives commit the speaker to something being the case; directives, including questions, requests and commands, attempt to get the hearer to do something; commissives commit the speaker to a future course of action; expressives express-psychological states, such as relief or apology; and declarations bring about the correspondence between the propositional content of the speech and reality, as when pronouncing a couple married.

A key point in speech act theory is that conversation creates patterns of commitment between the partners, in the sense that each of them must assume that the other is basically sincere in his or her manifest intentions, and uses speech acts appropriately given the context of the interaction. A similar notion is stated in Grice's (1975) "cooperative principle" in human communication, which is expressed in the following maxims of obvious relevance to the issue of controlling product quality in GIS. (Reported by S.E. Brennan: "Conversation as direct manipulation: an iconoclastic view", in Laurel, 1990).

Maxims of quantity
-make your contribution as informative as required
-do not make your contribution more informative than is required

Maxims of quality
-do not say what you believe to be false
-do not say that for which you lack adequate evidence

Maxims of manner
-be perspicuous
-avoid obscurity of expression
-avoid ambiguity
-be brief
-be orderly

Maxim of relation
-be relevant

Failure to meet any of these in conversation may result in poor quality outcomes, because the partners have talked past each other, because they have used terms in different ways, because they have used the wrong body language, have conveyed the wrong signals to each other, because they have witheld critical information, because they have not been as clear and to the point as they might have been, because they have tried to conceal their ignorance, please, flatter, manipulate or mislead their partner, or because they had incompatible intentions from the start.

By and large though human conversations work because the cooperative principle is followed. This involves, among other things, giving your partner information on an ongoing basis about the degree of validity of your speech acts. Humans have a variety of means at their disposal for qualifiying what they say and for marking uncertainty and error in speech or body language. Precision or vagueness in assertions is expressed by means of adverbs such as exactly, precisely, vaguely, approximately, relatively, roughly, basically, and so on. Phrases such as I am sure, I am positive, I think, I believe, it seems to me, on the other hand, let me think, I take this back, are hedges intended to convey to your partner the force of your own confidence in your assertions, while other phrases (is that right? are you sure? how come? come on! ...) invite your partner to give more information about the quality of his or her speech acts. Pauses, hesitations, frowning, head scratching, shrugging, looking away, some kinds of smiling or gesturing, complete this vocabulary, along with a variety of subliminal vocal signs of the kind picked up by lie detectors.

3.2 User-GIS interaction as conversation

The analogy between conversation and human-computer interaction can never be perfect, because human conversation is governed by the intentions of the partners, something that computers and all other non-conscious things cannot have (Dreyfus, 1979). The day is saved by Dennett's (1978) notion of an *intentional-system,* defined as any system that behaves *as if* it had beliefs, desires and intentions. Many systems designed -for specific purposes meet that definition, and a computer programmed to help produce some piece of new knowledge is perhaps the best example of one. The computer behaves *as if* it were a friendly, docile, knowledgeable, sincere, benign, albeit not overly insightful or creative partner, whose intention is solely to help you solve whatever problem you may have. That thorny philosophical problem swept under the rug, most key aspects of conversational interaction have direct analogs in the interaction with a GIS (Figure 1). The intentions of the user can be matched by the intentionality simulated by the system's purposeful, application-directed problem-solving behavior. The requirement for compatible background schemas can in principle be met, on the system side, by a database structure embodying a model of geographic space appropriate for the application at hand. Both intentions and background model may be captured in a judiciously composed "user profile" (see Daniels, 1986), which would help set the basic parameters of geographic representation, required level of detail and accuracy, user interface repertoire, and so on, suitable for the specific application. Beliefs and inferences in humans map rather conveniently into data, data structures and processes in computer systems. Speech acts have their analogs in the graphical displays, icons, menus, windows, messages, commands, warnings, queries and other user inputs supported by the user-system interface. The only thing missing to make the analogy complete is a counterpart to the human compulsion to provide abundant direct and indirect clues as to the quality of what they say in conversation.

It is only appropriate that the shortcomings of applied spatial knowledge of the kind best expressed in map form, and derived through interaction with a user interface that is primarily graphic, should also be expressed primarily through visual means. The analogy with knowledge production through conversation provides a meaningful typology for the kinds of problems that need to be made explicit, if possible in graphic format: problems of imperfectly shared or partially incompatible background schemas (spatial ontologies); problems of inadequate beliefs (including erroneous and missing source data, and inadequately represented spatial relations); problems of faulty inference (inappropriate data transformations due to system operation or implicit model); and problems due to flawed speech acts (unsatisfactory user-system interaction), resulting in misleading, insufficient, or largely irrelevant results.

4. Specifying the conversational metaphor for GIS

The challenge is thus to elaborate a coherent system of signs that can help the user visualize the quality of GIS products at any stage of deriving an application. That visual semiology should be analogous to the rich system of qualifiers of speech acts used in human conversation. It should be a visual language for indicating hesitation, doubt, ignorance, uncertainty and error which, as in human conversation, is a function of background model, purpose and context.

Table 1 summarizes the correspondences between human discourse and user-GIS system that will need to be worked out within a visual semiology. A first difficulty arises, as was briefly discussed earlier, from the lack of a typology of generic data models for different concepts of geographic space to correspond to the background schemas of human discourse. By contrast, considerable progress has already been achieved in formulating and implementing the graphic counterparts of speech acts, through Monmonier's work with graphic narratives, scripts, and phrases. *Graphic scripts* are automated sequences of maps, text blocks, statistical diagrams, and graphs designed to introduce a data set, explore salient trends, or reveal anomalies. They can include *graphic phrases*, which are shorter, more focused, data-driven graphic sequences. Graphic scripts can come together in *graphic narratives*, which "tell the story" of a database and what is to be found in there (Monmonier 1989a).

Thus, the linguistic metaphor has already been successfully introduced into GIS at the level of the interface between user and system. What is proposed in this paper is an expansion of the paradigm from language to discourse, so as to encompass not only the communication of the geographic knowledge stored in a GIS, but also its cooperative production through the interaction of the system and its users.

More subtle will be the problem of finding a vocabulary of visual counterparts to the expressions of utterance quality conveyed by speech and body language. Seeing is believing: the visual medium does not lend itself naturally to the expression of uncertainty and doubt. The work of Bertin (1983) on the rules of correspondence between concept and graphic sign will be relevant there, as will be Tufte's (1990) perceptive examples and commentaries on the visualization of information of different kinds. Work under Intiative 7 of the NCGIA, "Visualization of data quality", is currently exploring similar issues within the narrower scope of data, rather than product, quality, and some interesting results have already been reported both on modeling and displaying uncertainty and error in GIS (Beard, Buttenfield and Clapham, 1991).

What is quite clear from all these efforts is that the graphic symbols for product quality should form a system, that is, there should be a logic to how they are chosen and how they relate to each other and to the task being performed; a grab-bag of piecemeal graphic tricks will not do the job. The conversational metaphor proposed here provides a model for the vocabulary, syntax, and semantics of a system fulfilling these requirements. The bottom half of Table 1 offers suggestions for some possible graphic expressions of quality in a user-GIS discourse, based on a typology of corresponding linguistic expressions.

The different nuances of confidence and uncertainty that need to be conveyed visually (in graphical or simple textual form) may parallel the adverbs, hedges, proddings, and more substantial probes used in conversation for similar purposes. More specifically:

(a) *Adverbs* and adverbial phrases such as exactly, precisely, completely, approximately, quite, somewhat, roughly, more or less, express *precision or vagueness* that could be reflected in the *sharpness or fuzziness* of appropriate lines and symbols on the display.

(b) *Hedges* such as I am sure, I think, I believe, it seems to me, as well as several non-verbal signs of confidence or hesitation indicating *force of belief*, may be expressed in the degree of *color saturation* of the affected parts of the display, with grey tones reserved for the most speculative elements.

(c) *Proddings* such as is that right? come on! are you sure? how come? are used in conversation to express *disbelief* in something a partner is saying, inviting further information or clarification. They could be approximated by a set of *special icons, prompts. or elementary symbols* such as flashing asterisks and question marks, inviting the user to reconsider a recent input or query. Conversely, the user would have access to the same repertoire to probe any element of the current display that appears questionable or unexpected.

(d) More thorough *probes* of any aspect of the current display, corresponding to the digressions that human conversation partners occasionally engage in to explore aspects of their topic that may be *unsuspected. undiscussed, not understood*, or even deliberately hidden (Thrift 1985), may be accomplished through a number of available graphical database exploration and visualization techniques. Of these, *brushing* (Monmonier 1989b), developed specifically for the exploration of electronic map displays, is among the most promising. simple graphic techniques such as partial or total occlusion of problematic elements in the display may also be effective in suggesting ignorance.

Obviously, the effectiveness of such a system of symbols must be tested experimentally with different kinds of users. At this stage, the aim is simply to point out that a fairly complete correspondence can be set up between a quality monitoring system for GIS and human conversation.

5. Conclusion

This paper put forward an outline for a kind of "metal'expert system intended to monitor interactively the quality of GIS products. It is a "metal'system because of the absence of non-trivial generic rules, or recipes, directly for obtaining a quality product through GIS, although such rules may exist for deciding how product quality should be defined and monitored in the context of each kind of application. Figure 2 summarizes the logic of the proposed model in its present tentative form:

A *user profile* helps establish the background model of space implicit in the application (e.g., field, object, network, cell, space-time, relational, proximal, continuous, stratified - or whatever the members of a useful space typology turn out to be), and thus call up the appropriate *data model(s)* from among a limited set of alternative possibilities. At the same time, the *purpose* of the application (such as education, analysis, management, strategic planning, modeling....) is determined. These first two decisions help define a *profile of data quality* as required by the application, including the *thresholds of tolerance* for error and uncertainty in different parts of the data base. An appropriate *repertoire of graphic flags* and other qualifiers is also selected at this stage, out of a larger set of quality feedback indicators. A model of *error propagation* keeps track of error tolerance thresholds throughout the data manipulations, and suitable warnings and other *visual feedback* are issued whenever error or uncertainty in an element of the current display is likely to impact product quality as determined for the specific application. If the user cannot (or does not) correct the problem before moving on to the next operation, the warning is stored and becomes part of the final output.

The realization of such a quality monitoring system is contingent upon the operational specification of the background model and purpose components of product quality in GIS (Figure 3). Thanks to the recent work in the area of data quality, that third component is currently much better understood. Something along the lines of the VPF typology of data quality aspects (source, positional accuracy, attribute accuracy, currency, logical consistency, feature completeness, and attribute completeness: see Department of Defense 1991), may provide a suitable initial structure for the data quality axis. But even if similar high-level classes were to be agreed upon for the model and purpose components, mapping out the 3-dimensional space of product quality would still be a daunting task. Fortunately, not all cells of the space represented in Figure 3 are equally important, or even meaningful. One may begin with the most common kinds of applications in GIS, and figure out where in the model/purpose/data-quality space these may lie.

Irrespective of the prospects of successful implementation of the particular model proposed here, the central message of this paper is the need to face the issue of quality monitoring for GIS products in a systematic way. Current efforts devoted to the critical aspect of data quality should expand at this stage so as to address the problem of *product* quality more explicitly. In addition, we need to understand better both the quality requirements of the diverse *purposes* for which geographic information systems are used, and the different background *conceptions of space* implicit in different problem areas. GIS researchers and developers can take most of the credit for bringing about the geographic electronic democracy; making this an *enlightened* democracy may be more difficult, but we should make that our next task.

Acknowledgements

This research constitutes part of Research Initiative 7 on the Visualization of Data Quality of the National center for Geographic Information and Analysis (NCGIA). An early version of this paper was prepared for the specialist Meeting of Initiative 7 and was partially supported by the NCGIA.

References

BERTIN, J., 1983, Semiology of Graphics. (Madison, WI: University of Wisconsin Press).

BUTTENFIELD, B.P., and McMaster, R.B. (editors) 1991, MAR generalization: making rules for knowledge representation (New York, NY: Wiley).

BEARD, K.M., BUTTENFIELD,B.P, and CLAPHAM, S.(editors) 1991, Initiative 7 specialist meeting: visualization of data quality. NCGIA Technical Paper 91-26, University of California, Santa Barbara.

COUCLELIS, H., 1991, Requirements for planning-relevant GIS: a spatial perspective. Papers in Regional Science, 70, 1, pp. 9-19.

COUCLELIS, H., 1992, Location, place, region and space. Geography's Inner Worlds, edited by R.F.Abler, M.G. Marcus and J.M. Olson (New Brunswick: Rutgers University Press), pp.215-233.

DANIELS, P.J., 1986, Cognitive models in information retrieval - an evaluative review. Journal of Documentation, 42, 272-304.

DENNETT, D.C., 1978, Brainstorms: Philosophical-Essays on Mind and Psychology (Montgomery, VT.:Bradford Books).

DEPARTMENT OF DEFENSE, 1991, Vector Product Format, Interi Militarv Standard (MIL-STD-60006, Defense Mapping Agency).

DREYFUS, H.L., 1979, What Computers Can't Do: the Limits of Artificial Intelligence (New York: Harper&Row).

DUTTON, G. 1984 Truth and its consequences in digital cartography, Technical Papers, 44th Annual Meeting, American Congress on Surveying and Mapping, pp.273-283.

FRANK, A., 1992, Concepts that determine the user interface. Research Initiative 13 Report on the Specialist Meeting: User Interfaces for Geographic Information Systems, edited by D.M. Mark and A.U. Frank. NCGIA Technical Paper 92-3, University of California, Santa Barbara.

GETIS, A., 1992, Identifying spatial heterogeneity in a geographic information system. Presented at the symposium of the commission on Mathematical Models of the International Geographical Union, Princeton, N.J., August 5-7, 1992 (unpublished).

GOODCHILD, M.F. and GOPAL, S., (editors) 1989, Accuracy of Spatial Databases (London: Taylor & Francis).

GOODCHILD, M.F., GUOQIN, S., AND SHIREN, Y., 1992, Development and test of an error model for categorical data. International Journal of Geo4raphic Information Systems, 6, pp. 87-104.

GRICE, H.P., 1975, Logic and conversation. S tax and Semantics 3: Speech Acts, edited by P. Cole and J. Morgan (New York, Academic Press).

KEMP, K., 1992, Environmental modeling with GIS: a strategy for dealing with spatial continuity. Proceedings of GIS/LIS 192, San Jose, CA, Nov. 10-12, 1992 (in press).

LAUDAN, L., 1977, Progress and its Problems. (Berkeley, CA: Univesity of California Press).

LAUREL, B., 1990, The Art of Human-Computer Interface Design (Reading, MA: Addison-Wesley).

MONMONIER, M., 1989a, Graphic scripts for the sequenced visualization of geographic data. Proceedings of GIS/LIS 189, Orlando, Florida, Nov. 26-30, 1989, pp. 381-389.

MONMONIER, M., 1989b, Geographic brushing: enhancing exploratory analysis of the scatterplot matrix. Geographical Analysis 21: pp. 81-84

SEARLE, J., 1969, Speech Acts (Cambridge: Cambridge University Press).

SHAFER, G., and PEARL, J., 1990, Readings in Uncertain Reasoning (San Mateo, CA: Morgan Kaufmann Publishers, Inc.).

SIMON, H,. 1969, The Sciences of the Artificial (Cambridge, MA: The MIT Press).

THRIFT, N., 1985, Of flies and germs: a geography of knowledge. Social Relation and Spatial Structure, edited by D. Gregory and J. Urry (London: Macmillan).

TUFTE, E.R., 1990, Envisioning Information (Cheshire, CT: Graphics Press).

VEREGIN, H., 1989, A taxonomy of error in spatial databases. NCGIA Technical Paper 89-12, University of California, Santa Barbara.

WINOGRAD, T. and FLORES, F., 1987 Understanding Computers and Coanition: a New Foundation for Desicm (Reading, MA: Addison-Wesley).

Table 1. The conversational metaphor in GIS

DISCOURSE	SYSTEM
Intentionality	**Purpose**
Schema	**Background spatial model**
Knowledge/beliefs	**Data, data structures**
Inferences	**Data manipulations**
Speech acts	**Graphic scripts**
assertives	graphic phrases,
directives	commands
commissives	?
expressives	cautions, OK's
declarations	map symbols, icons

EXPRESSIONS OF UNCERTAINTY

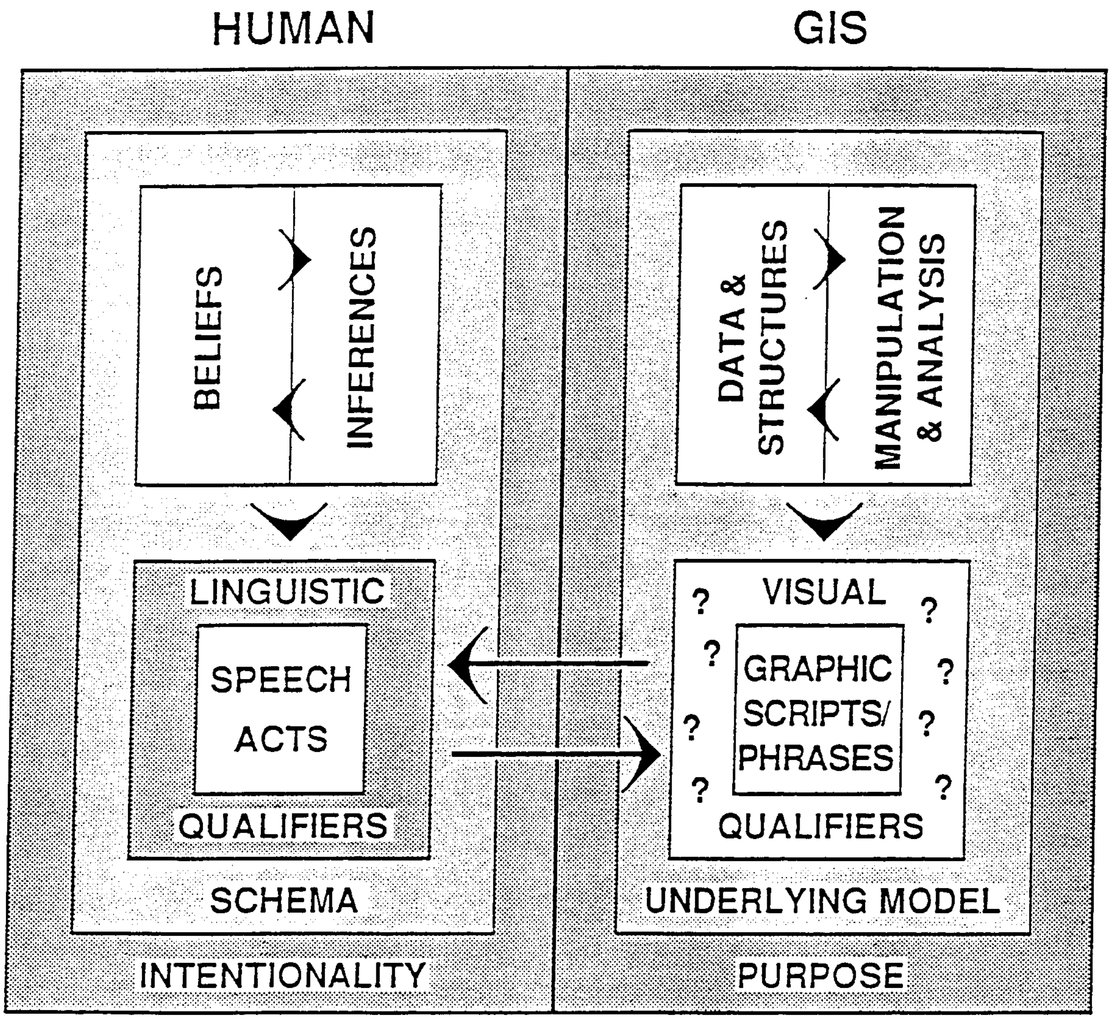
Speech/Body Language		Graphic/Visual
adverbs	**Precision/Vagueness**	sharpness/fuzziness
hedges	**Force of Belief**	color saturation/hue
proddings	**Doubt/Disbelief**	inquiring icons
hesitations	**Un-knowing**	occlusion/partial shape distortion alternating symbols ....

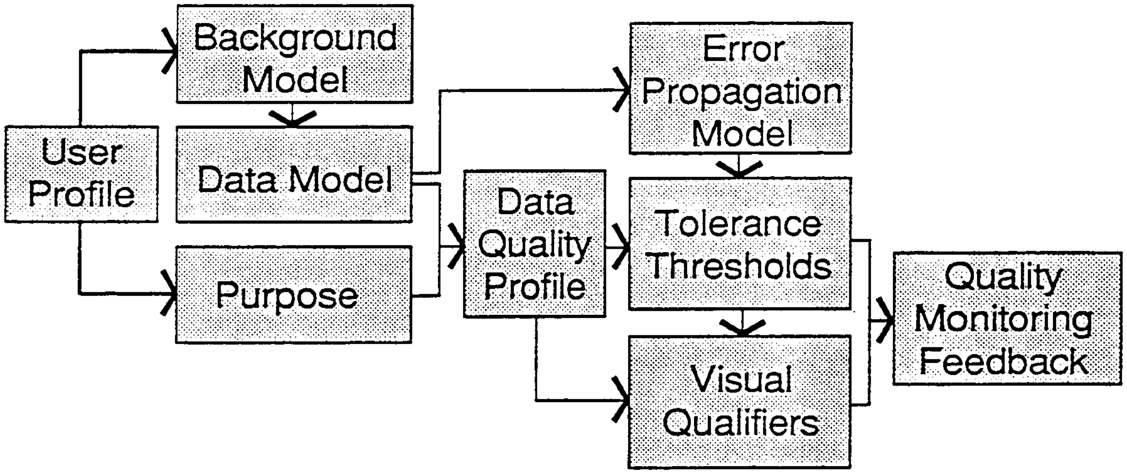
Figure captions

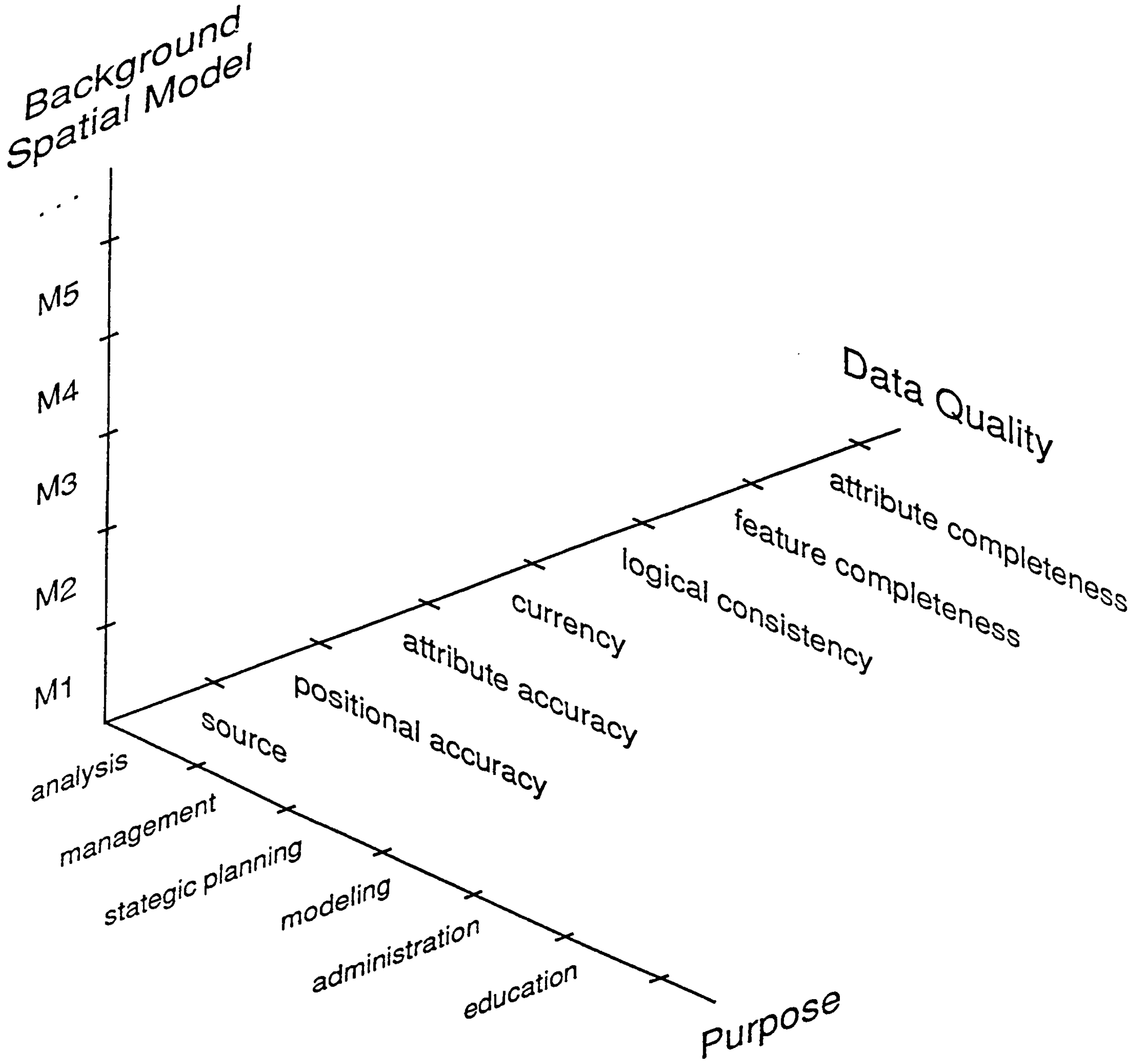
1. The conversational metaphor for GIS.

2. Principle of a "metal' expert system f or monitoring product quality.

3. The product quality space.


HUMAN
GIS
BELIEFS
INFERENCES
DATA & STRUCTURES
MANIPULATION & ANALYSIS
LINGUISTIC
SPEECH ACTS
QUALIFIERS
VISUAL
GRAPHIC SCRIPTS/ PHRASES
? ? ? ? ? ? ? ?
QUALIFIERS
SCHEMA
UNDERLYING MODEL
INTENTIONALITY
PURPOSE


User Profile
Background Model
Data Model
Purpose
Data Quality Profile
Error Propagation Model
Tolerance Thresholds
Visual Qualifiers
Quality Monitoring Feedback


Background Spatial Model
. . .
M5
M4
M3
M2
M1
Data Quality
source
positional accuracy
attribute accuracy
currency
logical consistency
feature completeness
attribute completeness
analysis
management
stategic planning
modeling
administration
education
Purpose

DATA INTEGRITY: THE ACADEMIC AND RESEARCH PERSPECTIVE

Kate Beard
University of Maine, Orono

and

Williamm Mackaness
University of Maine, Orono

Acknowledgement

This paper was presented at a regional surveying engineering conference on "Trends and Issues in Geo-Information Management Systems - Building the National Spatial Data Infrastructure: The Data Integrity Challenge", May 1-2, 1992, Hartford, Connecticut. The conference was sponsored by the American Congress on Surveying and Mapping -New England Section, the Atlantic Institute, and the Institute for Land Information. We gratefully acknowledge the support by a grant from the National Science Foundation (SES-88-10917).

DATA INTEGRITY: THE ACADEMIC AND RESEARCH PERSPECTIVE

Kate Beard
William Mackaness
Department of Surveying Engineering
University of Maine
5711 Boardman Hall
Orono, ME 04469-5711

ABSTRACT

Data integrity is central to the success of spatial information systems particularly where the investments in data are large and where information is being shared. This paper begins by considering the factors that compromise integrity and looks at ways of overcoming them. The nature of geographical phenomenon (its temporal and non discrete behaviour), methods of data capture, data models and data transformations are some of the important factors affecting data integrity. The paper outlines a research agenda based on the identification of key impediments to data integrity and researching those areas where rapid progress is most likely.

INTRODUCTION

Spatial data are expensive to collect and compile and national spatial databases have and will continue to require significant expenditures. Spatial information at any governmental level or within private industry is a valuable resource (certainly as valuable as any other investment) and as such it warrants careful management and maintenance. Spatial data, because of their collection expense and relatively slow rate of change, tend to have a long life span. Managing the integrity of spatial data therefore must be a long and persistent enterprise. Concern for data integrity begins with data collection and the goal should be to assure that data integrity is not compromised over the life span of the data.

The Oxford English Dictionary defines integrity as the state or quality of being unimpaired. In the database community integrity is quite specifically defined as a condition of the database such that it is in compliance with a set of consistency constraints. The implication for data or information is that an unimpaired state must be defined with respect to some reference frame. The reference frame could be a set of consistency constraints imposed on the database, a set of national standards, or a set of specifications identified by a user. We can therefore broadly define data integrity as compliance with some set of rules, standards, thresholds, or consistency constraints. A research agenda which addresses data integrity must consider a few basic questions:

- How is data integrity compromised?
- How can we avoid or minimize any compromise of data integrity?
- If data are impaired how can they be modified to achieve a state of integrity?

This paper begins by first establishing the areas of particular need, that is conditions under which data integrity can be most frequently and seriously compromised. It then outlines research areas which are either currently underway or which are required to address data integrity issues.

FACTORS WHICH COMPROMISE DATA INTEGRITY

Beginning with data collection we can itemize corrupting factors as including:

- data collection methods
- data models
- software
- hardware
- users
- data transfer/sharing
- time

Each of the following sections describe how these factors can lead to corruption of data integrity. Analysis of these factors helps target where research efforts should be concentrated to minimize or overcome these corrupting influences.

Data Collection Methods

During data collection we are typically at our most vigilant in establishing and assuring data integrity. Once we have collected high quality data, attention can lapse and we are less attentive to subsequent actions which can seriously comprise the data. Collection techniques for spatial data have been consistently improving, and consequently data integrity has been improved. The spatial and spectral resolution of satellite sensors continually improves. Research in GPS continues to improve the accuracy of horizontal and vertical measurements. Electronic data recorders eliminate much of the potential for human error in reading and transcribing measurements. Thus in many areas the completeness and accuracy of data collection is becoming less prone to corruption by technology and relatively more prone to corruption by human intervention.

During data collection, data integrity is built on the basis of two components: the accuracy of the definition of the parameter to be measured and the accuracy of the quantification. Unambiguous identification of what was measured is the first requirement and often where we encounter problems with geographic data. Geographic objects are rarely discrete and not easily summarized mathematically. With new technologies we can typically measure the locations with much greater precision than we can identify what exists at that location. Geographic phenomena which vary continuously and are impossible to measure everywhere create the greatest problems. The observation and measurement of these variables can be compromised by poor sample design, inattention to the spatial dependencies among various factors, and classification strategies that mask important information about the phenomena. Another aspect of data collection which can lead to future integrity problems is that not all pertinent information is recorded at the time of collection. If weather conditions, site conditions, and other variables which may bias measurements go unrecorded, the future ability to check for and remove bias is lost. Compromises in data integrity at this stage can be very serious since the only recourse is likely to be recollection of the data.

Data Models

Models can be a corrupting influence at several levels. By definition models are approximations of reality and some obviously describe reality better than others. Model bias will be large when an inappropriate, overly simplistic or incompletely defined model is used. One of the first levels in which models have an effect is in the abstraction from an ellipsoidal model to a planar model. The planar model, most common in geographic information systems, can be a corrupting influence over large geographic areas (Lukatela 1989). We potentially lose the integrity of precise geodetic coordinates by importing them to a GIS and discarding the original measurements.

Data models (raster and vector) also compromise data integrity. The object model (Goodchild 1989) compromises thematic information. This model commonly used to represent soils, geology, vegetation, and land use, aggregates attribute information over areas such that resulting polygons although assumed to be homogeneous may actually contain up to 40 to 50 percent variation (Beckett and Webster 197 1) within any one polygon. Alternatively the raster model potentially compromises the spatial integrity of geographic variables. Regular spatial units of fixed size can masks the true distribution of spatial variables and any variation smaller than the grid cell size is lost.

Database models have been continuously evolving in ways that address data integrity concerns. Complete schema description through data dictionaries document data, valid ranges, and expected relationships. Integrity constraints expressed in the form of rules maintain the consistency of the database and have become a standard database feature (McFadden and Hoffer 1985). Many current geographic information systems, however, do not protect data from abuse and erroneous entries or allow multiple users to access and change data simultaneously.

Statistical and analytical numerical models are also potential corrupting factors. For computational efficiency we often make simplifying assumptions in building models which may not be justified. For example it is common to adopt a linear model without first ascertaining that higher order parameters are really insignificant. By adopting inappropriate models the analysis of the data is corrupted.

Software

The integrity of all digital data is at the mercy of software. If software runs we generally assume it runs correctly and are therefore less vigilant in looking for errors. Aside from outright blunders, software problems can be subtle and difficult to detect especially lengthy code needed for complex geometric computation which have many exceptions. Although generic algorithms appear in the literature and are subject to scrutiny, vendors implementations are proprietary and we generally operate under the assumption that they have been sufficiently checked and verified. Empirical tests, however, have revealed considerable variation in outcomes of common algorithms.

Hardware

Hardware is a potential corrupter of data integrity, but not a significant one. The effects of machine precision and roundoff problems in complex spatial computations have been largely overcome. Hardware failures such as disk crashes have been largely overcome by standard database recovery methods. Long term data integrity is subtly compromised by the rapid obsolescence of hardware. Data may be lost because the media on which it is was stored becomes obsolete. Archived data have also been lost due to media instability. Thousands of magnetic tapes of earlier landsat data are no longer readable. The long term stability of new media such as optical disks is not fully known. Tests of early CDs indicate that they may not be as stable as originally thought.

Users

Users are the least controllable factor and the biggest unknown in the management of data integrity. New systems make manipulation of spatial data very simple. The assumption that users have been appropriately trained and educated to use these systems is not often the reality. Users can quite easily perform invalid operations and rapidly corrupt their databases.

Data transfer

We know that transmission of information by any channel is subject to the addition of noise and corruption. During transfers, data may be corrupted by transmission errors, by incomplete transfers, or by incompatibility between the data models of the sending and receiving systems. The presence of different dam models across a network is referred to as a heterogeneous system. Different data models reflect different conceptual, logical, and physical design specifications. Transfers between these models can result in inconsistent definitions for nominally the same features and lose or corruption of certain relationships. Gallagher and Salazar indicate that data structure translation can be quite costly in a heterogeneous data base environment.

Time

Time is an obvious corrupter of data and most problematic for features which change frequently. Once a certain time interval has elapsed data will be out of date and inconsistent with the situation on the ground. The time period over which data are viable varies both within and across feature types in a database. For example soil information may change relatively slowly with respect to land use but individual soil characteristics can change quite rapidly such as soil moisture content. Unlike checking account or grocery inventory databases, updates or new data collection efforts for spatial data can usually not be carried out as rapidly as data changes occur. For example, people move on average every seven years, so decennial census information can be out of data before new census data is available. Maintaining the temporal integrity of spatial databases, particularly on a national scale, thus becomes a significant challenge.

AREAS OF PARTICULAR NEED

Having considered some of the factors which can corrupt data integrity, research attention should be directed by two principles. Research should:

- focus on the factors which most seriously compromise data integrity
- and address those in which we can achieve the greatest improvement given the current state of technology.

With these two principles in mind the following research areas are identified as currently underway or in need of immediate attention,

- Refined models for spatial data
- Development and expansion of integrity constraints for spatial databases
- Benchmarking for GIS software
- Understanding of error propagation
- Inclusion of metadata and models of our modeling processes
- Informing and controlling user actions
- Expanding standards for data transfer
- Managing the temporal aspects of data

Refined models for spatial data

- Geodetic Models for GIS

With recent advances in technology it has become easier to collect, store, and process geodetic measurements and with the increasing need and interest in building global databases, an important research effort will be to develop geodetic or global data models. Some research is already underway on these models (Lukatela 1987, 1989; Goodchild, Yang and Dutton 199 1). Lukatela advocates a geodetic model based on a spherical vornoi diagram. Goodchild, Yang and Dutton have developed a model based on a triangular hierarchical tessellation of the globe. Additional research is needed to make geodetic models feasible foundations for geographic information systems.

- Measurement-based models

One of the deficiencies of current spatial databases is that they consist of derived or abstracted information. Original observations which were the basis of this data have in most cases been discarded or lost. The original observations are important for consistency checks, for reconstructing the data if they become corrupted, or in contributing required parameters for updates (i.e. adjustment computation for a datum change). Several researchers have been investigating measurement management systems as a way of maintaining and utilizing original measurements in cadastral information systems (Buyong et al. 1990 199 1, Hintz and Onsrud 1990). These ideas could be extended to the development of linked databases whose data contents would differ in level of processing.

In other words at the primary level, raw observations would be stored and managed (all updates would occur at this level). A sequence of databases could be linked to this and contain processed or more abstract information such as plane coordinates, classified imagery, or interpolated data.

- Error Models for GIS

There is a need for development or enhancement of models that explicitly incorporate notions of error and uncertainty. The inclusion of an explicit error component in the data model allows the quality and integrity of the data to be known, not assumed. Dutton's (1989) global model is designed to incorporate the positional uncertainty of information. Classified satellite imagery only acknowledges the classes and any accuracy information on classes is typically discarded. Goodchild and Wang (1989) have looked at raster models that would store a vector of class probabilities with each pixel.

Development and expansion of integrity constraints for spatial databases

Although traditional databases have vastly improved the controls on data integrity, such controls have not extended to spatial databases (Frank et al 1991). Research is needed to expand integrity constraints to include consistency checks for spatial definitions and relationships. Topological consistency checks for graphic primitives have been implemented in most GIS, but higher level checks at a feature or field level do not exist. Transaction management in GIS databases differs from that in commercial database systems since change operations are very complex, take considerable time to specify and affect large sets of data. Researchers are currently investigating object oriented databases systems designed to support long transactions. Under these systems users can check out versions of objects or groups of objects and perform several lengthy edits. Meanwhile other users can continue to use previous versions or even edit their own checked out versions. Versions are later merged back together to reconcile differences resulting from this parallel development (Lamb et al). Parallel research is needed on data integrity management for merging long spatial transactions.

Understanding error propagation

The manipulation of spatial data introduces errors which are not well understood or documented- GIS processes combine data from different sources with different levels of spatial resolution, using rules which are often complex. Veregrin (1989) has looked at error propagation for overlay operations in a raster model. Error propagation in fuzzy vector intersection has also been investigated (Chrisman 1989, Pullar 1990), but is more complex. Work is still needed on the propagation of errors in spatial operations such as generalization and in the combination of several operations.

Benchmarking for GIS software

Currently there are no official benchmark standards for GIS software. We operate under the assumption that vendors have checked their software for defects. Most benchmarks are undertaken to establish performance efficiency but checks must also be made to establish validity. Research needs to be undertaken to develop standardized tests that check the validity and accuracy of common GIS algorithms. These should be designed to test for the robustness of algorithms under complex geometric computations.

Inclusion of metadata and models of our modeling processes

Data are less useful when descriptions of the data including collection methods, database design and processing steps are not included with the data. Data describing data is referred to as metadata. Metadata is usually not included because spatial databases have not been designed to incorporate it. Lanter (1991) has developed a system for recording lineage information on GIS processing, but additional research is needed on database designs that can incorporate metadata on quality an more exhaustive lineage information. Questions include how large volumes of quality information can be efficiently stored, what objects quality information should be linked to, and how quality information is managed on data transfer.

Informing and controlling user actions Users are potentially the greatest compromisers of data integrity. Research therefore needs to be directed towards either informing users of problems or controlling their actions to avoid problems. Research currently underway addresses documentation of data quality to keep users informed of actions which may compromise their data (Beard et al 199 1). More recently visualization techniques have been suggested as a means of representing spatial data quality, and for conveying the repercussions of user actions.

Managing the temporal aspects of data

As databases evolve, age, and become merged with other data the temporal aspects of the data become more complex. We need to adopt temporal constraints which will check that time periods of data are consistent with other time elements in the data base and verify that elements are not inconsistent with the situation on the ground. One potential database solution that has been investigated is the association of a life span with every data element (Elmasri and Wuu 1990). If an element has a projected life span, this life span can be checked against the object's duration in the database since the last update. Research could also investigate the feasibility of automated updates triggered by certain actions or events.

Improving standards for data transfer

We are moving toward the existence of heterogeneous distributed databases. Databases are considered heterogeneous if they do not share a set of common assumptions about the information they contain and distributed if each database is under local control and connected by low-bandwidth links (Silberschatz, et. al. 199 1). 'Me ability to make distributed databases behave as if they were a single database is referred to as interoperability. Currently the integrity of query response from a 'single database' can be compromised by semantic inconsistencies or differences in the local schema. To overcome these problems research will be needed on extended data models that include more semantic information on data in each of the local databases. There are a number of system design issues that must be addressed. Some of these items are common to other types of distributed database systems. Others have special importance with respect to GIS including:

- Classification of data heterogeneity as a prerequisite to a 'common' interface,
- Failure resilient transactions/ processing management,
- Integrating diverse information repositories,
- Automatic indexing (geographic, thematic) of database contents.

CONCLUSION

No agenda is without bias; the headings listed in this paper are intended as 'research leaders'. Data integrity spans a wide number of issues from database design, data capture, and software metrics through to transaction processing, and user interface design. Integrity can be compromised by poor conceptual understanding, inadequate query language, system failure and inappropriate integration of data from different sources.

The research agenda proposed in this paper has focused on those factors that underpin the above issues. These are considered to be the refinement of spatial data models to support geodetic measurements and computations, error and uncertainty, methods for improving heterogenous distributed databases and the study of error propogation and visualisation of quality components.

The study of quality and error propogation will reveal which metadata needs to be gathered at the time of data capture and how processing (either spatial or transactional) affects those quality components. Successful interpretation of mixed data (data sourced from different sites) will necessitate additional geographical information being stored, as well as lineage information.

REFERENCES

Beard M. K, B. Buttenfield and S. Clapham. 1991. Visualization of Spatial Data Quality: Initiative 7 Specialist Meeting Report. Technical Paper 91-26. National center Technical Report

Beckett, P. H. T. and R. Webster. 197 1. Soil Variability: A Review. Soils and Fertilizers 34:1 1-14

Buyong, T. W. Kuhn and A. Frank. 1991. A Conceptual Model of Measurement-Based Multipurpose Cadastral Systems. URISA Journal. 3:2 5-49.

Buyong, T. B . and W. Kuhn. 1990. Local Adjustment for Cadastral Measurement Databases. Proceedings ACSM/ASPRS Annual Convention Denver, CO 3:19-27.

Chrisman, N. 1989. Modeling Error in Overlaid Categorical Maps. Accuracy of Spatial Databases. Ed M. Goodchild and S. Gopal. New York: Taylor and Francis. 21-34.

Elmasri, R and G. Wuu. 1990. A Temporal Model and Query Language for ER Databases. Proceedings IEEE 6th International Conference on Data Engineering. Los Angels CA. 76-83.

Frank, A U. M. Egenhofer, and W. Kuhn. 199 1. A perspective on GIS technology in the Nineties. Photogrammetric Engineering and Remote Sensing 57(11):1431 - 1436.

Gallagher, L and S. Salazar 1984. Report on Approaches to Database Translation. NBS Special Publication 500-115. National Bureau of Standards.

Goodchild, M. 1989. Modeling Error in Objects and Fields. Accuracy of Spatial Databases. Ed M. Goodchild and S. Gopal. New York: Taylor and Francis. 107-114.

Goodchild, M. 199 1. Issues of Quality and Uncertainty. in J. C Muller (ed.) Advances in Cartography. Elsevier 113-139.

Goodchild, M. F. 1988. The Issue of Accuracy in Spatial Databases. Building Databases for Global Science. Ed. H Mounsey and R. Tomlinson. New York: Taylor and Francis 31-48.

Hintz R and H. Onsrud. 1990. Upgrading Real Property Boundary Information -in a GIS. URISA Journal. 2:12-10.

Goodchild, M.F., Y. Shiren and G. Dutton. 1991. Spatial Data Representation and Basic Operations for a Triangular Hierarchical Data Structure. Technical Paper 91-8. National Center for Geographic Information and Analysis.

Lamb, C., G Landis, J. Orenstein, and D. Weinreb. 1991. The Objectstore Database. Communications of ACM. 34:10 51- 63

Lanter, D. P. 199 1. Design of a Lineage-Based Meta-Data Base for GIS Cartography and Geographic Information Systems. 18:4 255-26 1.

Lukatela, H. 1989. GIS Future: Automated Cartography or Geo-Relational Solid Modeling. Proceedings Auto Carto 9. Baltimore, MD. 341 -347.

Lukatela, H. 1989. Hipparchus Data Structures: Points, Lines, Regions in Spherical Vornoi Grid. Proceedings Auto Carto 9. Baltimore, MD. 164-170.

McFadden F. R and J. A Hoffer. 1985. Data Base Management. Reading, MA: The Benjamin/Cummings Publishing Company.

Pullar, D. 1990. A Comparative Study of Algorithms for Reporting Geometrical Intersections. Proceedings, Fourth International Symposium on Spatial Data Handling. 1: 66-76.

Silberschatz, A., M. Stonebraker, and J. Ullman. 1991. Database Systems: Achievements and Opportunities. Communications ofACM. 34:10 111-120.

Smith, J. L. S. P. Prisley, and R. Weih. 1991. Considering the Effects of Spatial Data Variability on the Outcomes of Forest Management Decisions. Proceedings GISILIS '91. 1: p. 286-292.

Veregin, H. 1989. Error Modeling for the Map Overlay Operation. Accuracy of Spatial Databases. Ed M. Goodchild and S. Gopal. New York: Taylor and Francis. 3-18.