

# UC Santa Barbara

## NCGIA Technical Reports

### Title

Multiple Representations: Scientific Report for the Specialist Meeting on Initiative 3 (89-3)

### Permalink

<https://escholarship.org/uc/item/3rn7b748>

### Authors

Buttenfield, Barbara P.  
DeLotto, Joseph S., editors

### Publication Date

1989-02-01

**MULTIPLE REPRESENTATIONS**  
**SCIENTIFIC REPORT FOR THE SPECIALIST MEETING**

**18-21 February, 1989**

**Buffalo, New York**

**Barbara P. Battenfield, Initiative Leader**  
**Joseph S. DeLotto, Senior Rapporteur**

**NCGIA, Department of Geography, SUNY at Buffalo, Buffalo, NY 14260**

**National Center for Geographic Information and Analysis**

**Report 89-3**

## Table of Contents

Preface and Acknowledgements.....	
1. Framework for the Initiative Topic .....	
1.1 Scope and Purpose of the Initiative .....	
1.2 Objectives of the Specialist Meeting.....	
2. Participants .....	
2.1 External Participants.....	
2.2 NCGIA Participants .....	
2.3 Rapporteurs.....	
3. Meeting format.....	
3.1 Preparation .....	
3.2 'Small Group' Format.....	
4. Prioritizing a Research Agenda.....	
4.1 Major Themes.....	
4.2 First Day -- Impediments to Multiple Representations.....	
4.2.1 Impediments of Consistency.....	
4.2.2 Physical Impediments.....	
4.2.3 Cartographic Impediments.....	
4.3 Second Day -- Research Goals.....	
4.4 Third Day -- Research Priorities.....	
4.4.1 Database Issues.....	
4.4.1 Generalization Issues.....	
5. Products of the Initiative.....	
5.1 Research Projects and Publications .....	
5.1.1 Scale-Dependent Geometry .....	
5.1.2 Digital Terrain Issues.....	
5.1.3 Digital Definition Schemes.....	
5.1.4 Map Generalization.....	
5.1.5 Hierarchical Data Structures.....	
5.1.5 Formalizing Database Links.....	
5.2 Multi-Agency, Multi-Scale Database.....	
5.3 Other Deliverables .....	
5.3.1. Visiting Scholars to NCGIA- Buffalo.....	
5.3.2. Symposium "Towards a Rule Base for Map Generalization".....	
5.3.3. National Conference Sessions at Toronto AAG Meetings.....	
6. Summary .....	

## APPENDICES

- A. Bibliography of Recent Papers by Participants Related to Multiple Representations.....
- B. Agenda for the Specialist Meeting.....
- C. Mailing List for the Specialist Meeting.....
- D. Position Papers Submitted by Participants.....

## **Preface and Acknowledgements**

This is a report on a meeting entitled "Multiple Representations", held in Buffalo, New York, 18-21 February, 1989. The organizers wish to thank the management and staff of the Darwin D. Martin House and of the Center for Tomorrow (both affiliated with the State University of New York at Buffalo) for providing excellent conference facilities and arrangements during the meeting. Particular thanks are due to the staff of the University Office of Conferences and Special Events for last minute arrangements. The assistance of Amy Eglowstein at the NCGIA office in Buffalo is also acknowledged.

Notes for the draft of this report were drawn from the notes of seven rapporteurs, including Shou-Wen Chen, Michael Gould, Marcin Jasinski, and Mark MacLennan of NCGIA-Buffalo, and Bud Bruegger and Doug Hudson of NCGIA-Maine. Joe DeLotto consolidated rapporteur notes to begin preparation of this report. Barbara Bittenfield organized and compiled the draft, adding items from her own notes taken at the meeting, and appended appropriate material. Initial and final drafts were sent to NCGIA and external participants for comments and revision. We are grateful for the editing contributions of all participants who submitted revisions and corrections.

The assistance of Fred Broome of US Census is gratefully acknowledged for providing large portions of the text in Section 5.2, and for providing Figure 3 on very short notice.

The meeting and this report represent parts of Research Initiative #3, "Multiple Representations", 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.

## 1. Framework for the Initiative Topic

### 1.1 Scope and Purpose of the Initiative

The term "Multiple Representations" might be applied in many ways to the study of geographic information and analysis. One could consider the multiple versions of a single place or event that are studied as remembered geography in cognitive mapping and spatial preference analysis. Alternatively, the term could be applied to temporal issues, as in the multiple stages evident in monitoring environmental change, or more abstractly, in studying successive stages of an iterative model or simulation.

For this report, the scope of the term "Multiple Representations" will encompass changes in geometric and topological structure of a digital object that may occur with the changing resolution at which that object is encoded for computer storage, analysis and depiction. The conceptual basis for this scope of the term is perhaps most succinctly summarized in arguing that cartographic and digital data are at best only a sample of the geographic features they are intended to represent. As discrete approximations of a continuous reality, each encoded object may capture at most a subset of the feature for subsequent representation and analysis. The resolution at which features are captured by digital encoding methods will often bias the amount and types of details about the features that are encoded. The implications of this for geographic information and analysis are quite complex.

The challenge of storing digital versions of geographic features for representation at many spatial resolutions continues to obstruct the efficiency of GIS operations and the quality of cartographic products. We lack methods to maintain multiple representations of a single geographic object and this constrains the discipline in three ways. First, data acquisition is constrained, for example in reconstructing feature geometry from scanned data. Second, constraints to data representation are evident as obstacles to generating maps at many scales from a single detailed database. Finally, constraints to data analysis occur during map overlay, sheet matching tasks, and in search and query operations. We need capabilities to develop feature descriptions at each desired level of resolution, and to describe formally the connections between levels.

The purpose of Research Initiative 3 is to enumerate impediments associated with digital representation of geographic features, and particularly those impediments associated with problems of producing multiple graphical depictions from a single digital database. The topic relates to two of the five bullets originally solicited by NSF in calling for NCGIA research. The relation to visualization involves the discrete encoding of continuous features, as described above. Problems of completeness, of positional accuracy and of recognizability are of great concern. Multiple representations research also

relates to the development of theory for spatial relations, and involves impediments of consistency and redundancy. If the solution is determined to be creating a separate digital file for each desired level of analysis, one must accommodate the cost of access and sheer volume of storage required for multiple versions and depictions. Error propagation issues become important, especially during database updates. If a single digital file is deemed the appropriate solution, one is confronted with the question of what is the optimal level of precision at which to encode, and with preserving a balance between computational efficiency and digital data quality. These and related issues form the conceptual framework for multiple representations research.

## 1.2 Objectives of the Specialist Meeting

The goal of the Specialist Meeting was to bring together researchers from the disciplines of cartography, geographical analysis, computational vision, and spatial database management with representatives from federal mapping agencies and private mapping houses. For a period of four days, the group discussed technological and conceptual impediments associated with multiple representations, and defined research goals that might be addressed to alleviate existing problems. The intention to prioritize a research agenda from these discussions was largely successful, due in large part to the willing cooperation of participants.

Discussions were intended to focus largely upon the relations between geometry and scale. One objective was to emphasize the need for improved digital feature descriptions, including object-oriented and spatially addressed methods. Conversion between digital models was a related concern. Database considerations formed a second objective, especially in relation to error propagation and to methods of inference between levels of resolution. A third objective focused upon automatic scale-changing and problems of cartographic generalization. Intended topics for discussion included questions of pattern recognition for features that change with scale, feature identification and extraction, as well as the conventional cartographic issues of simplification and map depiction. The Specialist Meeting participants were chosen to bring together individuals with research proficiency in topics directly related to these objectives.

## 2. Participants

### 2.1 External Participants

The intention was to bring together a wide variety of academicians and representatives from public agencies and private sector firms, all having

common interest in multiple representation issues. Participants were sought from the ranks of GIS researchers with interests in spatial data structures, researchers in computational vision and pattern recognition, geographers with interests in spatial statistics, digital cartographers, geodesists, and photogrammetrists, private vendors with interests in computer graphics and database management, and representatives from federal agencies with central mandates in mapping.

# **TABLE 1** **SPECIALIST MEETING PARTICIPANTS**

(Participants are listed below with the topics for which they led discussions)

## GEOGRAPHY AND SPATIAL ANALYSIS

Michael Batty, U. Wales-Cardiff, UK	urban structure, visualization
Stewart Fotheringham, NCGIA-Buffalo	spatial analysis
David Mark, NCGIA-Buffalo	GIS algorithms / data structures
Robert Marx, US Census Bureau	census attributes / data structures

## COMPUTER SCIENCE

Renato Barrera, NCGIA -Maine	computational geometry
Richard Berg, Defense Mapping Agency	database management
Geoffrey Dutton, PRIME Computer, Inc.	hierarchical data structures
Andrew Frank, NCGIA - Maine	databases and data structures
Oliver Gunther, NCGIA - Santa Barbara	database management
James Little, U. British Columbia, Canada	computational vision
Hanan Samet, U. Maryland, US	hierarchical data structures
Terry Smith, NCGIA - Santa Barbara	very large spatial databases
Deborah Walters, NCGIA-Buffalo	digital vision, psychophysics
Marvin White, ETAK Corp.	electronic navigation

Leonard Uhr, U. Wisconsin-Madison	computational vision
-----------------------------------	----------------------

(note: Dr. Uhr submitted a position paper, but was unable to attend the Specialist Meeting due to illness. However, his ideas contributed to discussions, and his name is included as a participant for this reason)

## CARTOGRAPHY, GEODESY, PHOTOGRAMMETRY, REMOTE SENSING

Kate Beard, NCGIA - Maine	categorical map analysis
Fred Broome, US Census Bureau	census mapping
Barbara Battenfield, NCGIA-Buffalo	analytical cartography



Michael Dobson, Rand McNally Corp.	cartographic symbolization
Michael Domaratz, USGS	digital terrain models
Manfred Ehlers, NCGIA - Maine	photogrammetry
John McAvoy, Rand McNally Corp.	digital cartography
Robert McMaster, Syracuse University	generalization algorithms
Charles Schwarz, NOS/NOAA	geodesy
David Simonett, NCGIA - Santa Barbara	remote sensing
Waldo Tobler, NCGIA - Santa Barbara	analytical cartography
Robert Weibel, U. Zurich, Switzerland	digital terrain analysis

Academic participants came from nine universities in four countries. Federal agency participation was very strong, including representatives from the US Bureau of the Census, US Geological Survey, National Ocean Survey (NOAA), and the Defense Mapping Agency. Private sector participation included large and small mapping house representatives and a computer vendor with strong interest in geometry and graphical depiction. External participants are listed in Table 1.

## 2.2 NCGIA Participants

The National Center includes a large number of research faculty with research interests related to multiple representations, and Center participation was very good, including researchers from all three NCGIA sites in departments of Geography, Computer Science, and Surveying Engineering. NCGIA participants are also listed in Table 1.

## 2.3 Rapporteurs

Seven rapporteurs annotated and taped discussions during the Specialist Meeting, and their notes have been included in preparing this report. All rapporteurs are graduate students at NCGIA institutions, with research interests in GIS and related topics. Several of the rapporteurs have direct interest in multiple representations issues. The senior rapporteur helped to condense and consolidate other rapporteurs' notes, as well as administering many rapporteur duties during the Specialist Meeting. Rapporteurs are included in the mailing list of participants (Appendix C).

## 3. Meeting format

### 3.1 Preparation

Participants were asked to prepare for the Specialist Meeting by submitting a position statement of 1-3 pages, focusing on their view of the scope of the multiple representations problem, on important research topics that should be pursued, and on which of these topics could be addressed immediately given current technology and existing knowledge. Participants were also asked to prepare by reading the position statements that were mailed to them roughly two weeks before the Specialist Meeting, and arriving at the Meeting prepared to discuss impediments that were most commonly mentioned in the position statements. Position papers are included in Appendix C of this report.

### 3.2 'Small Group' Format

The venue for the Specialist Meeting iterated between morning meetings of the entire group and small group discussions each afternoon, and followed the formal agenda closely. During the first two days' morning sessions, individuals were given the opportunity to take the floor, with the first day devoted to federal agency and private vendor perspectives on impediments, and the second day devoted to the academic perspectives on research goals. The third and fourth morning sessions were open discussions, guided by the outcome of the small group discussions occurring every afternoon, and were considered very productive.

Afternoon discussions in small groups proved to be one of the most effective parts of the meeting. The topics were pre-selected, and the small group format encouraged intensive interactions. Small groups changed from day to day, to maximize interaction and avoid formation of discipline-specific cliques. Each group chose a single spokesperson to report the discussion back to the larger group, and this helped to direct the larger group discussions on following days. Spokespeople also changed from day to day. Setting up the meeting in this way was a gamble, and it paid off. The initial hesitance to pursue discussion given a general topic without much structure passed quickly. By the last day, the discussion and group bonding had progressed to the point that small groups formed without much direction from the Initiative Leader, and the two groups that formed on the last day selected their own topics.

For the first afternoon session, 4 groups were formed by Bittenfield and Mark to insure maximum variability of interests, i.e., each group had some computer scientists, some geographers, some federal agency representatives, etc. All 4 groups were directed to discuss impediments. An evening session brought the entire group together to hear reports of the group leaders. The next day's small groups were formed (again insuring variability of interests) by Bittenfield, Smith, Mark and Frank, to discuss research goals, and group leaders reported back to the full group the next

morning. The third day, two small groups formed to discuss researchable topics on database issues and on generalization.

## **4. Prioritizing a Research Agenda**

### **4.1 Major Themes**

The purpose of the Specialist Meeting was to first discuss impediments and second to prioritize a research agenda. Throughout the Meetings, discussions focused upon the relations between geometry and scale in spatial data management, analysis, and representation. Three topics were of particular interest, including digital description techniques, database considerations, and automatic scale-changing problems. These themes are introduced briefly below, and will be seen to occur and recur throughout the 3 days of Specialist Meeting discussions.

Digital descriptions of map features are of interest as a tool for automating decisions currently made manually. Most often these methods are parametric, and as they discretize a continuous feature for a particular purpose, one might consider them as a class of mathematical transformations. Walters' rho-space resolves ambiguities of spurious line intersections; Psi-s transformations simplify computations for some affine transformations (eg. rotation). Hough transforms, Witkin's fingerprints and Buttenfield's structure signatures extract resolution-based differences in line shape, and Richard Pyke's geometric signatures extend this to two dimensions. Fourier transforms extract both frequency and amplitude of detail. Other methods could be enumerated. Research goals call for formalized understanding of the relations between data structures (both object-oriented and spatially addressed) and models for digital description. Formalizing a typology of digital models was also proposed, to better understand their role in data structuring and representation. Also of interest are the error propagation properties associated with digital description schemes, and specifically the errors that may be introduced during conversion between digital description models.

Database considerations relate to many aspects of the multiple representations problem, and most important are considerations of error propagation in converting between representations, and efficiency of storage and access. Lack of error propagation models is a severe impediment, and provides a direct link between Initiative 1 ("Accuracy of Spatial Data") and Initiative 3. Efficiency is manifest in problems of attribute coding and database update; we lack ready methods for attaching an attribute to an object at one level of representation and providing for automatic inheritance of that attribute to components of the object. For example, one may use a semantic dat model, and construct a class of objects called 'buildings' that may be

subdivided into 'residential', 'commercial', 'industrial', etc. Next, lesser objects are incorporated into the database description, for example roofs and foundations. The provision that walls must meet at right angles, or that buildings must meet certain requirements to withstand seismic tremors of a specified minimum, or other attributions of the database should be inherited automatically from one level of representation (types of buildings, eg.) to lower levels (foundations and roofs, eg.) to preserve consistency and maintain a database efficiently.

Automatic scale-changing problems continue to be a focus of interest for database management and for cartographic depiction. Electronic vehicle navigation is a prime example of a situation in which speed of data access and display are both paramount. In fact, many spatial decision-making situations utilizing GIS methods mandate efficiency as a first priority, for example, implementing evacuation routes, dispatching rescue vehicles, or alerting individuals about an environmental crisis. The tasks involved in automatic scale-changing operations relating to multiple representations include pattern recognition, feature extraction, simplification and representation. All of these are fundamental components of map generalization, and point out the important fact that generalization is a problem involving data organization, analysis, and interpretation. Conventional perspectives on cartography often limit the scope to issues of display alone, creating a subtle impediment by exclusion.

## 4.2 First Day -- Impediments to Multiple Representations

On Saturday morning, the group listened to statements from private vendors and federal mapping agencies. Afternoon small group discussions focused upon the impediments that currently exist in professional mapping houses to limit the use or avoidance of multiple representations for data acquisition, manipulation, and mapping. Four small groups were formed, and in the evening the entire group reconvened to hear reports from the spokesperson selected by each small group. The uniformity of impediments discussed by each small group was remarkable, given the breadth of topics covered in the morning, and given the diverse backgrounds of participants. The major impediments were seen to be consistency of data, physical impediments in the database, and the obstacles imposed by cartographic tradition.

**4.2.1 Impediments of Consistency.** To understand the complexity which problems of consistency impose, Dick Berg presented a schematic of operations involved in the production of geographic information at DMA (Figure 1). A database is compiled from multiple sources of information, including field survey and observation, existing maps, text or other documentation, and photography or digital remote sensing. Agencies

develop extraction specifications to guide data collection and compilation, and to document the lineage of source data (eg. date, level of geodetic control, map scale, resolution of scanner or digitizer, etc.). Extraction specifications may vary from one extraction format to the next, and this will affect consistency of the database information. The products generated from a database may include maps, digital data in a variety of formats, or other databases, and each one will be generated by a specific product specification. Often, these are user-specified, further complicating the preservation of consistency. The GIS community lacks a formal definition of consistency, and research on formalizing basic concepts crept into the discussion many times during the Meeting.

Two points were emphasized in the discussion on consistency. First, it is important to remember that multiple representations issues arise during extraction as well as during production (display) of data, although the extraction context has been largely ignored. The assumption that a single database can be generated at some very fine level of precision to solve the multiple representations problem may be a fallacy, because a database is frequently compiled from multiple representations. In some circumstances, the process of populating a database involves resolution of inconsistencies in the source information. A second point of importance relates to the transfer of data between agencies, and the need to remember that in-house specifications for extraction and data production may be based upon sound criteria, but still be incompatible with specifications developed in other mapping houses for other purposes. A call for complete standardization of data formats and specifications is probably unreasonable, given the variety of purposes for which spatial data is produced. However, problems of consistency can be minimized when specifications are carefully designed and clearly documented, and this effort can only improve the transfer of spatial data from one database to another.

A different impediment related to consistency is the lack of comparability of research results evaluating algorithms for database access and data generalization. For example, little of the published research on line simplification algorithms has been benchmarked using a single data set. Variations in scale and

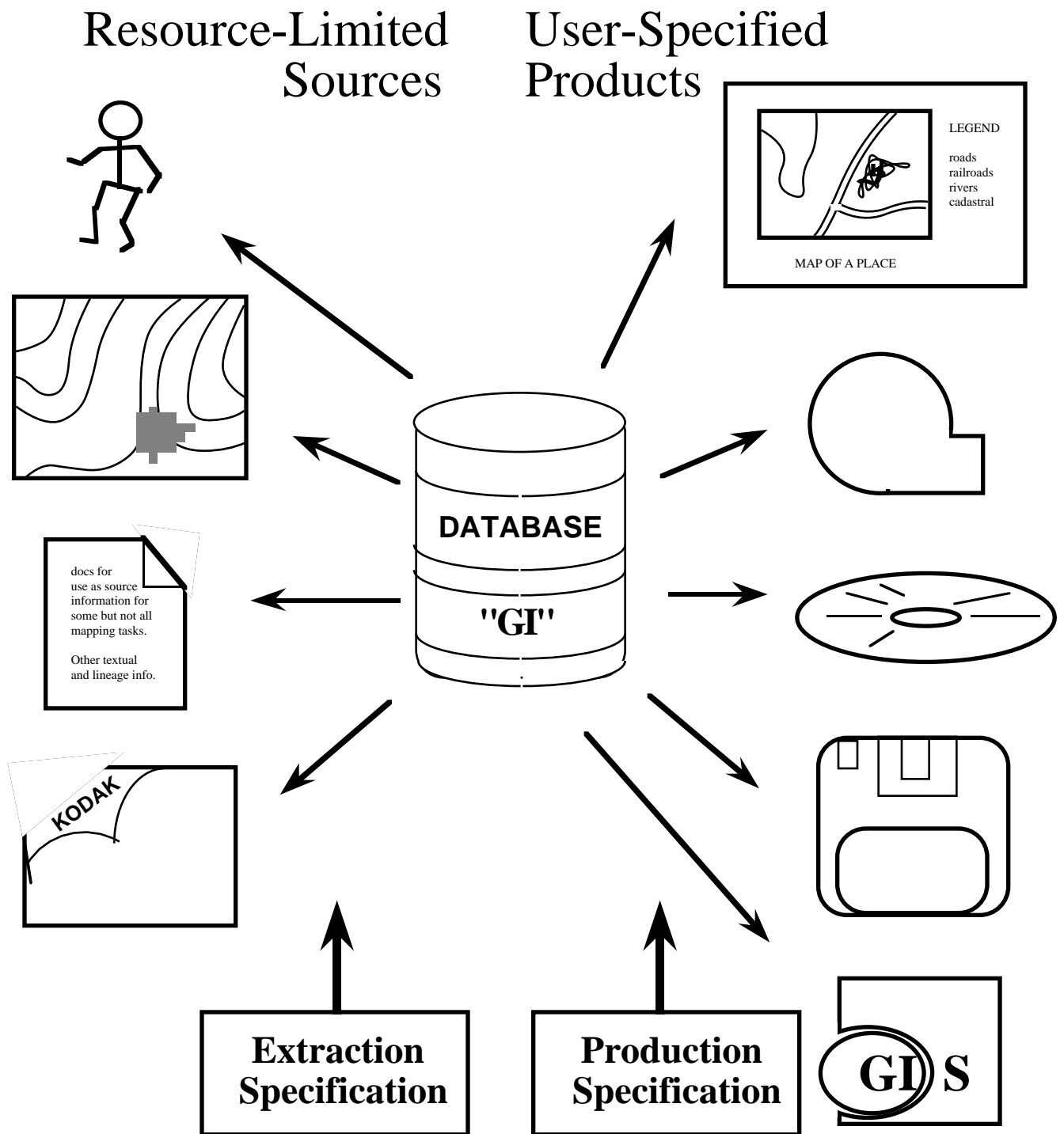


Figure 1  
(submitted by Dick Berg at the Specialist Meeting)

data resolution are compounded by the fact that algorithm performance probably varies according to the complexity of the data used to test it. Because so little is known about the geometry of features contained in data sets, it is

difficult to predict which data will optimize algorithm performance. In fact, definitions of feature complexity have not been formalized. As a result, it is difficult to compare the results of research performed on unique data sets. The lack of consistent and comparable research results continues to obstruct our progress on generating data at multiple scales, and in implementing updates consistently across related representations.

**4.2.2 Physical Impediments.** Physical impediments were discussed, including speed of database access, and the computational overhead associated with maintaining multiple versions of spatial information. The cost of database updates is multiplied when updates must be repeated for every version that is archived, and this emphasizes the importance of developing methods for automatic propagation of updated information throughout an institutions' data archives. Storage volume is less of an impediment as mass storage costs drop, and technology improves with developments such as CD ROM and WORM storage.

**4.2.3 Cartographic Impediments.** Perhaps the most surprising impediment to be discussed on the first day related to the limitations imposed by the traditions of cartography. The cartographic perspective that was focused upon is that reality is somehow 'map-like', that it is layered, and continuous within discrete rectangular windows, and that layers can be registered to each other using a finite set of 'well-identified' points. This perspective may lead to some insidious errors in logic, in some circumstances. The logical trap has constrained the development of many GIS packages, although to be fair it has advantageously provided a conceptual framework required for the development of early GIS software. But problems with the realism of layered reality are quite subtle. Object orientation is difficult to implement in a layered reality without the introduction of structured data types; and assumptions of continuity belie the complexity of objects that move in time as well as in space. One is precluded from accommodating the changes in structure that characterize some but not all objects in a compound object, and forced to consider space as a sort of flattened receptacle that must be (by turns) either enumerated, tessellated, or interpolated.

A second impediment discussed by all groups is the cartographic tradition to ignore extraction operations, and to focus most intensively on the display of information. Related to this is the fact that many cartographic operations in design and production are understood only intuitively. "This looks right" is an intellectual trap that impedes the formulation of sound and consistent guidelines for data extraction and display. Without formal definition of criteria by which to design and guide algorithms, automation of the cartographic process will continue to elude the discipline. Conversely, evaluation of map products must be user-driven, whether the products' purpose is accuracy, precision, or recognizability; hence the appropriateness of

"this looks right" cannot be completely ignored. This impediment is both subtle and severe, and may be difficult to overcome by research alone.

#### 4.3 Second Day -- Research Goals

On Sunday morning, the group assembled to hear statements from the academic perspective, and ensuing discussions focused upon research goals associated with the impediments cited on Saturday. The theme threading through this day is best summarized as a consideration of adopting a single or multiple representational solution to given research problems. Adopting the former solution would entail a single database from which multiple versions of information could be generated at selected scales. In adopting multiple representations, other dilemmas would present themselves. Research goals associated with adoption of a single database were quite diverse, and more easily recounted by topical question.

At what level of precision should the information be encoded? Clearly it is unnecessary to record the same level of detail for cultural features and naturally occurring features. Interstate highways, for example, are constructed according to fixed radii of curvature, and below a certain resolution, details become redundant. River channels, on the other hand, become increasingly complex to very fine levels of resolution. To incorporate both types of linear phenomena at a single level of resolution implies one of three approaches. The database could be recorded at an homogeneous level of precision, and volume of recorded detail will be unnecessary for some features. In another approach, variable levels of precision might be used to archive features in the database, and data structures and management strategies developed to accommodate varying levels of resolution. A third approach would record only topological relationships in the database, avoiding geometric details to the extent possible. Each of these is advantageous in particular data situations.

Related to this issue is the development and refinement of digital models of feature description to accommodate scale dependent structures without constraining self-similarity, and to facilitate access and manipulation within the database. As on the previous day, discussion turned to problems of conversion between digital models, and automatic recognition of digitally encoded features. The research goal in this domain is quite subtle. For example, if the same feature is described by two disparate digital models, what techniques may be developed to determine that the two descriptions refer to the same feature? Complicating this problem is the fact that many features will appear different depending on the scale of their encoding.

Attempts to encode digital features are concurrent with research to describe details about particular spatial process. Here, too, the scale at which



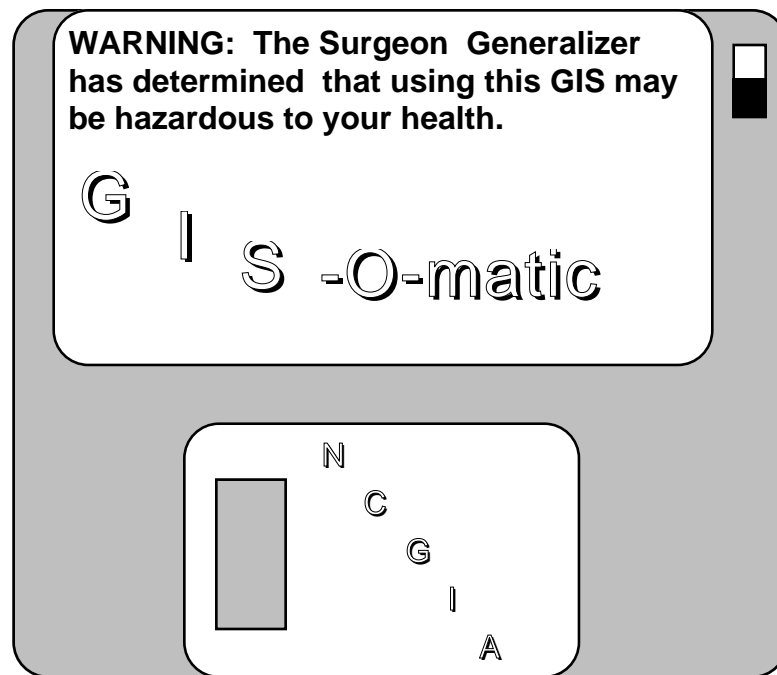
encoding takes place may be problematic. For example, geomorphic process at one scale may incur details of isostatic rebound; at another scale details of erosion become apparent. The same is true for social process, for example where interest might focus at one scale on migration pattern and at another scale on zoning constraints. Can a single database solution store all possible sets of geographic details of form and process comprehensively? The formal understanding of the scale or range of scales at which spatial processes occur must be pursued to address either single or multiple database solutions to the multiple representation problem.

The alternative to a single database involves generating and maintaining multiple versions of information, accepting the ensuing complications as a 'necessary evil'. The biggest issue involves preservation of consistency, especially during update operations. A second issue concerns determining the specific resolutions at which the multiple versions should be stored. In some cases, for example with USGS and DMA data products, the scales are user-specified. Clearly, the cost of generating all possible scales of information is prohibitive, and guidelines to predict a range of usable resolutions for data encoded at a particular scale have not been determined. How far can a database at a given scale be 'pushed' to generate representations for adjacent (both larger and smaller) scales? There is some indication that the range of usable scales is often domain-specific, and that the usable range may be greater for landuse data than for terrain, for example.

Discussion at the end of Sunday afternoon focused on the impact of multiple representations on users and user interfaces. Questions were raised concerning how to determine an appropriate level of resolution for a given analytic or display situation, and for protecting users from selecting inappropriate operations on the database (for example, over- or under-simplification). Several participants felt it would be beneficial to incorporate several levels of user guidance, so as to assist beginning users without placing unnecessary constraints on proficient users. The issue of user guidance versus policing in software interface design was well-summarized by the cartoon presented by Dick Berg and Geoff Dutton (Figure 2). Some attention was paid to explicit inclusion of data quality information in database structures. All of these topics relate to other ongoing or planned NCGIA research initiatives, and point to the interrelatedness of research initiatives on the agenda.

A research topic that arose from this discussion centered on the development of a rule base for both extraction and display operations in GIS. Implicit was the realization that the rule base requires a better understanding of both human and machine vision. NCGIA Research Initiative 8 is currently planned to address 'An Expert System for Cartographic Design', which forms a small part of the rule base. A symposium 'Towards a Rule

Base for Map Generalization' is being planned to take place in April 1990 as an offshoot of the Specialist Meeting (details in Section 5 below).



**Figure 2**  
(submitted by Dick Berg and Geoff Dutton at the Specialist Meeting)

#### 4.4 Third Day -- Research Priorities

On the final full day of the Specialist Meeting, the group turned its attention to research questions that can be addressed now given the current level of technology and software engineering, and given the current state of knowledge. Two sets of issues surfaced for immediate research attention; and in the afternoon, two small groups were formed, to discuss database issues and generalization issues. The two issues are in some ways an extension of initial discussions about the extraction and display operations presented by Dick Berg on the first day of the Specialist Meeting. However, both group reports contain common threads, and it is clear that research in either area will attend to issues mentioned in the counterpart area. It is important to understand that these are not the only two areas, but rather the two felt by the Meeting participants to be of highest priority and most readily pursued at present.

4.4.1 Database Issues. The database group prioritized research needs that must be addressed to accommodate multiple representations in a single or multiple version data management strategy. Of central importance is the need to organize multiple versions for efficient access, and new strategies for ordering spatial variables was cited as an important research topic. Rules to preserve consistency between views generated at different levels of resolution must be refined as well.

A related set of topics focused on database queries, and some of the research questions were also cited as important during Initiative 2 in discussing Spatial Query Language (see Mark et al (1989), "Languages of Spatial Relations", NCGIA Report 89-2). Particular interest was expressed about exploring the stability of query processing in multi-scale database searches. How predictable is the variation in response to a database query made at different levels of resolution? Is the variation more readily predictable for particular data domains, or particular levels of resolution? What range of variation can be considered acceptable for specific types of spatial decision support? To what extent does the architecture of the database impact upon the stability of query processing? These research questions have relevance to the use and value of geographic information, which is the focus of NCGIA Initiative #4. The questions are also related to database architecture, and will be discussed further during the Specialist Meeting for NCGIA Research Initiative 5.

A final topic discussed under the rubric of database issues relates back to initial discussions held the first day of the Meeting, about extracting information from multiple sources during database creation and update. The determination of the scale of a given database is to a certain extent the amalgamation of scales from all data sources. Depending upon the status of updates, the (amalgamated) source information may change from one time to another. The scale of a database is perhaps not meaningfully described by the simple numeric of a Representative Fraction (RF), for example as with USGS "1:24,000 hydrography" (one example from many cited). Asking the scale of a database is in fact making a query about the lineage of the database. The Proposed Standard for Digital Cartographic Data addresses lineage as a data quality issue, but discussions during the Specialist Meeting indicate the issue should be more broadly defined to encompass generation and archival of geographic information.

4.4.1 Generalization Issues. Attention in this group focused upon the nature and display of cartographic data, and touched several times on the constraints imposed by limiting research solely to representation for

illustrative purposes. Visualization for analytical purposes must also drive the research agenda, to refine the use of visual tools in GIS environments.

The need to derive a more comprehensive and flexible definition of resolution drove much of the afternoon's discussions on generalization research. Variations in feature resolution that may occur within a single data set (eg., US county boundaries) must be explored to improve the efficiency and accuracy with which they are utilized for statistical and cartographic applications. Simplification algorithms must be designed that self-adjust according to the local resolution of objects, and this requires clear and formal definition of object resolution. This is a topic of some interest to researchers in Initiative #1, where it is often referred to as the "variable reporting zone" problem.

Waldo Tobler led a detailed discussion on expanding current definitions of resolution, and proposed an object oriented definition that is map-based. He computes resolution as the average size of objects present, that is, the summed area (or length, volume, etc.) divided by the number of objects. Tobler suggests that definitions such as his lend themselves to extensions of variance, standard error, and related types of statistical description. The assessment of error in generalization continues to be an important research priority.

A second item placed on the research agenda involves formalization of digital feature description and categorization models. Digital models must accommodate the complexity of compound and hierarchical objects, and be sensitive to those aspects of geometry that may change with scale. Conversion between digital models must be studied in terms of error propagation, and to associate particular models with specific generalization operations. The properties preserved by a specific transformation often drive the decision to apply it in map projections. Refined understanding of the properties and operations associated with a model or class of digital models will improve decisions about which model best applies to a particular generalization application.

This research issue is not limited to map simplification, but affects iconic as well as graphic information. Problems of consistency in feature codes and the transfer of data between federal agencies attest to the need for consistent attribute assignments in digital databases. Research on the logic of feature code assignments is also a high priority. Categorization of digital data may be improved by attending to cognitive aspects of feature categories, in addition to the current emphasis on accuracy and (perceptual) recognizability so prevalent in database and cartographic research.

A third set of research topics involves rules for map generalization that must be defined explicitly to automate the cartographic process. Rules for setting and modifying tolerance values in feature simplification cannot be developed without improving techniques for automatic feature distinction and identification. Perceptual research will be required to evaluate the automated solutions that may be designed.

Rules for symbolic transformation of map information must also be derived. For example, the decision to collapse an urban area down to a point is determined in part by scale, map purpose, and by dimensionality of the object at its source scale. Other symbolic operations such as coarsening of a categorical coverage, displacement of features during map reduction, etc. are understood only intuitively. Issues of conflict resolution must be defined and prioritized. A typology of generalization operations must be formalized. This aspect of the research agenda will impact upon computer-assisted map production, and also on the types of multiple representations incorporated in digital database generation. Finally, there may be applications to the user interface, particularly concerning rules for appropriate map symbolization and interactive map design.

## **5. Products of the Initiative**

Three types of products have developed since the Specialist Meeting, including research projects and publications, generation of a multi-scale database for research and teaching, and visiting scholars and future symposia. Details will be discussed for each one in turn. During the course of the Initiative, other deliverables will be generated. A comprehensive list of products will be included with the Initiative 3 final report.

### **5.1 Research Projects and Publications**

Categories for the projects and publications that follow are somewhat arbitrary, as is the ordering of sections. This is not a complete listing of the publications emanating from the Specialist Meeting, and represents articles that are in print, in press, and in preparation. Grant proposals in submission are also included. This list will be updated periodically during the course of the Initiative, and a comprehensive listing will be included with the Initiative final report. Within each section below, articles in print and in press are listed first, and followed by articles and proposals submitted or in preparation.

#### **5.1.1 Scale-Dependent Geometry**

Buttenfield, B.P. 1989 "Scale-Dependence and Self-Similarity in Cartographic Lines" In press for **Cartographica**.

ABSTRACT. This paper provides a typology of two classes of geometry for cartographic lines, based on the well-known Richardson plots. The first class contains objects having self-similar geometry, features whose structural characteristics are replicated either precisely or statistically with changes in scale. Self-similar features are currently described by fractal models, which some argue are appropriate for all cartographic objects. The fallacy of this statement will be demonstrated. The second class of features is fully scale-dependent, and contains cartographic objects whose geometry varies distinctly with changing scale. Both models are described and applied to examples of digital line features, to demonstrate their worth in encoding and preserving particular types of cartographic detail during automatic generalization.

Jasinski, M. "Comparison of Complexity Measures for Cartographic Lines"  
MA thesis research in progress - NCGIA funding summer 89  
(to be submitted for publication)

ABSTRACT. Existing geometric measures commonly applied to cartographic line features will be examined for statistical redundancy. The goal is to determine a quantifiable definition of 'line complexity', or specifically to discover the components of a quantifiable definition. It is hypothesized that components will include aspects of distance, angularity, and density of detail, and that redundant measures are being applied by various researchers. A test data set comprised of both cultural and naturally-occurring line features will be measured and compared statistically to evaluate the geometric measures.

Buttenfield, B.P. "Geometry of Cartographic Lines" Research in progress,  
in preparation for submission to **Geographical Analysis** early fall  
(Shou-Wen Chen had NCGIA funding spring 89 to assist, Victor  
Wu has NCGIA funding 1 summer month to assist)

ABSTRACT. The digital abstraction of geographic information for archival, manipulation, and graphic representation precludes storing a unique coordinate file for every desired scale of analysis or map depiction, for reasons of efficiency and consistency. Strategies to provide multiple versions of line features from a single database have conventionally relied upon simplification algorithms, to reduce detail from a high resolution coordinate file, or upon splining algorithms, to reshape the simplified versions. It is commonly assumed that any one of these models will affect all cartographic features in an equivalent manner. The purpose of this paper is to demonstrate first that a single generalization procedure may have differing effects on different types of cartographic features, and second, that geometric guidelines may be used to distinguish between feature types. Examples to demonstrate these points will be limited to cartographic line features, although it will be argued that classes of geometry may be defined for other cartographic objects as well.

## 5.1.2 Digital Terrain Issues

Weibel, W.R. and DeLotto, J.S. 1988 "Automated Terrain Classification for GIS Modeling" **Proceedings** LIS/GIS 88, San Antonio TX, vol.2, p. 618-627. (paper reporting results of early I3 research conducted spring and summer 1988)

ABSTRACT. Automated terrain classification involves partitioning of an area into homogeneous topographic regions through quantitative interpretation of a digital terrain model. By evaluating geometric parameters extracted from raw elevation values of a DTM, it is possible to characterize the topography in general terms ("roughness") or in terms of a specific application (hydrography, or geomorphology). In recent years interest has developed to automate terrain classification for modeling and geoprocessing. This paper reviews the conceptual basis for areal classification of topography and examines specific impediments to the use of DTMS for automated terrain classification.

DeLotto, Joseph S. 1989 "Elevation Matrices as a Data Source for Automated Terrain Classification" **Proceedings**, International Cartographic Association. Conference, Budapest, Hungary (forthcoming, August 1989).

ABSTRACT. Parametric descriptions of terrain geometry have been proposed by a number of authors as a path to automating terrain classification, including measures of slope angle, aspect, hypsometry, and local convexity. Gridded topographic data (DTM's) remain the most widely used format for digital terrain, but are limited as a source for quantitative interpretation of terrain. Because sampling for production of DTM's does not vary with complexity of terrain, DTM's often contain inadequate representations of geomorphically significant features. A number of parameter extraction techniques are compared to demonstrate this limitation.

DeLotto, J.S. 1989 "The Role of Scale in Automated Terrain Classification"  
MA thesis research, August 89 (NCGIA funding spring semester 89)

ABSTRACT. Recent efforts to integrate techniques of automated terrain classification into geographic information systems are hindered by inadequate application of existing knowledge in fields of numerical taxonomy, image processing, and image understanding. The thesis reviews principles of terrain classification and geomorphometry, identifies problems that may occur when classification is performed from gridded elevation data, and proposes techniques to accommodate the scale dependent nature of topography.

### 5.1.3 Digital Definition Schemes

Walters, D.K.W. 1989 "Discrete Contour Descriptors" In **From Pixels to Features**, Edited by J.C. Simon, North Holland, p 287-298.

ABSTRACT. A set of discrete contour descriptors is proposed. The descriptors are able to describe discrete contours, such as those found in digitized maps, using parameters which are independent over the similarity transformations of scaling, rotation and translation. As the descriptors are based on discrete geometry, they do not have the implementation problems of descriptors based on continuous geometric concepts. In addition, the descriptors are perceptually valid: contours which humans perceive to be similar have similar descriptors. The descriptors are much simpler than spline-based or related descriptions, and thus can lead to more compact representations that maintain all of the perceptually relevant information.

Buttenfield, B.P. "How To Tell if Two Cartographic Objects are the Same"  
Accepted for presentation at LIS/GIS 89 in Orlando (**Proceedings**  
forthcoming November 1989)

ABSTRACT. This paper reports validation of a digital definition technique (called a 'structure signature') developed in the author's previous research. The technique accommodates scale-dependent geometry, and has been proposed as a means to automate line feature identification. The stability of the technique can be verified by comparing measures for the same geographic objects collected from two different coordinate databases. If the technique is robust, then measures for the same features archived in each digital file should prove quite similar. The converse issue (whether the digital definitions differ for features that are geomorphically distinct) has been previously determined for small (1500 point) data sets, and must also be verified for a larger data domain.

Buttenfield, B.P. "Recognition and Representation of Cartographic Features" research in progress; solicited for presentation at  
NCGE conference Oct 89 (to be submitted for publication Fall 89)

Buttenfield 1/9 NCGIA funding for summer 89

ABSTRACT. Traditional cartographical depictions of geographical landscapes relied heavily upon an understanding of the underlying geographic (particularly geomorphic) processes that had formed them, and cartographic curricula incorporated generalized instruction in morphometric drawing. The criteria were intuitive, and required a good deal of artistic talent. With the advent of computer mapping and digital storage of geographic features, cartographic depiction has been extended to statistical landscapes, and training in digital representation schemes must be incorporated into the curriculum. A review of



digital definition strategies developed in geography, mathematics, engineering, and computer science will be presented and categorized according to their utility for specific cartographic applications. Conversion between digital models and issues of error propagation will be mentioned.

#### 5.1.4 Map Generalization

Beard, M.K. 1989 "Design Criteria for Automated Generalization"

**Proceedings** ICA, Budapest, Hungary, August 1989 (forthcoming).

ABSTRACT. The history of automated generalization spans approximately 25 years, but solutions developed to date have not yielded satisfactory results. It seems timely to evaluate reasons for this limited success. Although limitations of early computer technology are partially responsible, other factors are hampering progress. These factors include difficulties in formalizing the process, the graphic map and the physical mindset, and introduce persistent hardware and software problems. This paper investigates factors which have hindered both progress and success in automating generalization and to explore directions for advancing efforts in this area.

Beard, M.K. "A Comparison of Two Strategies for Cartographic Generalization" Paper in preparation for submission Fall 1989.

ABSTRACT: This paper examines two different approaches to automated generalization of cartographic objects. A current standard in automated cartographic generalization is the Douglas-Peucker algorithm. This algorithm reflects a particular model of a cartographic object and from this model the algorithm makes assumptions about which points are most crucial to the information content of the line. Perkal provides another model for the generalization of cartographic objects. His model provides a very different approach to what information is eliminated or emphasized during generalization. The effect of these two models are compared on digital coastline data.

Buttenfield, B.P. "Automatic Methods for Cartographic Line Distinction"

Research proposal submitted to USGS

funding 2/9 faculty summer 90, 12 mos. RA

ABSTRACT. The purpose of this research is to evaluate automatic methods to improve cartographic line simplification, and specifically to determine locations in a coordinate file where the geometry of a line feature changes, in order to enable automatic tolerance value modification. Tolerance values must be modified periodically to insure both accuracy and recognizability of graphic details on a generalized map. At present, decisions about where to adjust tolerance values are made manually, and form an expensive bottleneck to map production for government and commercial organizations.

Mark, D.M. (1989) "Conceptual Basis for Geographic Line Generalization"

Proceedings, AUTO-CARTO 9, Baltimore MD, April 1989, p. 68-

77.

ABSTRACT. Line generalization is an important part of any automated map-making effort. Generalization is sometimes performed to reduce data volume while preserving positional accuracy. However, geographic generalization aims to preserve the recognizability of geographic features of the real world, and

their relations. This essay discusses geographic generalization at a conceptual level.

### 5.1.5 Hierarchical Data Structures

Bruegger, Bud P. and Frank, Andrew U. 1989 "Hierarchies over Topological Data Structures" **Proceedings, ASPRS/ACSM Convention, Baltimore, MD., vol.4, p.137-145.** Invited article (revised version of the paper) for **Journal of Photogrammetric Engineering and Remote Sensing**, in preparation.

ABSTRACT. The paper describes a system of multiple, ordered topological representations connected by hierarchical relationships between topological cells. The lowest level representation organizes all spatial objects in detail, while the higher order representations contain only large and important objects represented with less detail. This structure makes the implementation of multipurpose GIS possible, where very small (e.g. parcels) and very large spatial objects (e.g. nations) co-exist. The response time to queries is independent of the size of the objects in contrast to present topological data structure. The spatial objects can be displayed in several resolutions.

Bruegger, Bud P. and Frank, Andrew U. "Hierarchical Cell Complexes" Submitted for presentation at the 12th Annual Applied Geography Conference, Binghamton, N.Y., October 1989.

ABSTRACT. Multipurpose GIS offer considerable cost savings due to the sharing of data acquisition and maintenance efforts among several user groups. The paper demonstrates how a single topological representation impedes the implementation of such a system by causing response times to be unacceptably high when relatively large spatial objects are involved. In order to solve this problem, additional representations with reduced detail have to offer higher level views of the spatial objects. Lattice structures under spatial inclusion and hierarchies over topological cells are shown as examples for such higher level representations.

Bruegger, Bud P. and Frank, Andrew U. "Management of Hierarchies over Topological Cells: Definition of Spatial Objects, Consistency and Query Optimization" Research in progress for submission of a journal article 1989.

ABSTRACT. This paper shows how spatial objects can be organized in a system of multiple topological representations. After describing the basic structure of the representations and the links between representations, methods for defining objects in terms of already defined objects and of topological cells of different representations are proposed. Protocols for inserting these objects in the structure without introducing contradictions between the representations are shown. The system organizes spatial objects multiple times in different level of detail. A query optimizer has to decide from which representations the necessary data shall be accessed in order to minimize processing while still satisfying the level of detail necessary.

### 5.1.6 Formalizing Database Links

Beard, M. K. and Barrera, R. "Linkages Between Multiple Representations"

Research proposal submitted to USGS funding 1/9 faculty  
summer 90, 10% staff release time, 12 mo. RA

ABSTRACT. The proposed research has two components: one focused on linkages between representations, the other on measurements that describe differences in representations of different resolution. The objective of the first component is to develop a logical data structure to manage multiple representations. The proposed database will create linkages between objects representing the same real world entities. The second component of the research is concerned with characterizing data sets without the traditional concept of scale. Different data can be nominally displayed at 1:24,000 scale for example, but exhibit very different geometric and attribute characteristics. Map production specifications for the 1:24,000, 1:100,000 and 1:2,000,000 scale USGS series will be used as a starting point for identifying differences. Empirical measures will capture idiosyncrasies of representations not covered by production specifications.

Bruegger, Bud P., Barrera, Renato and Frank, Andrew U. "A

Framework for Multiple Representations of Spatial Objects"

In preparation.

ABSTRACT. The paper gives an overview of problems that can be treated with a theory for multiple representations. The terms 'representation' and 'multiple representation' are defined. A framework is proposed which makes it possible to describe the characteristics of a representation formally. The system uses this information to relate the different representations, check for consistency, select appropriate mapping operations to propagate changes and updates, decide which representations to access in order to answer a query, etc. Based on this framework, three problems occurring in multiple representation systems are described.

Bruegger, Bud P. "Multiple Representations of Spatial Objects" Ph.D.  
Thesis research in progress, Dept. Surveying Engineering, U. Maine.

ABSTRACT. The thesis proposes a formal framework for the general case of multiple representations of spatial objects. Problems of consistency and mappings between representation are explored. The scope of these studies will concentrate on spatial modelling, quality of shape approximation and abstraction of objects, and excludes inaccuracy, uncertainty, and temporal aspects. The applications of this framework to problems from the field of GIS -- one of which is cartographic generalization--will be described.

## 5.2 Multi-Agency, Multi-Scale Database

Originally suggested by Fred Broome (US Census) at the Specialist Meeting, this product is intended to serve as standardized data domain in which to develop benchmarks for algorithms and in which to establish comparable results for generalization research. Academic researchers do not always have resources to receive and process multiple digital data sets from federal data producers. Additionally, federal agencies do not archive small standard data sets that can be used or compared with other data sets from

other agencies. The intention of collaboration between NCGIA and the federal agencies is to alleviate both impediments, and to provide federal agency experience to NCGIA researchers.

At present, commitments to participate have been established with Census (TIGER files), National Ocean Survey (Shoreline Files), USGS (DLG files), and DMA (DTM data). Some of this information will be available at multiple scales. Census is jointly funding a graduate intern from Buffalo (Joe DeLotto) to spend the summer in DC implementing the database, interacting with all agencies, writing the unpacking software and user documentation. We also have informal commitment from Walt Winn, a member of FICCDC Standards Working Group, to assist in data compression and formatting. DeLotto begins work in Fred Broome's office in Washington DC on 1 June.

The creation of standard data sets need not be a major burden on any of the participating agencies or NCGIA. The level of activity will remain at a manageable level throughout the project, by keeping the scope of the project small, and by preserving the standardized nature of the data product. It is hoped that the final product will also serve an educational purpose. Instructors will have a readily available set of geographic files to incorporate into their curriculum. Files will be in the public domain, by mutual agreement of all agencies and by NCGIA, and distributed by NCGIA on a cost-recovery basis. This should encourage use by students for thesis research as well.

Representatives of the NCGIA and the participating agencies constitute a committee to decide which geographic areas will be included in the data sets. Having too few will limit the number of interested users and the representative nature of the data sets. Having too many will create a production and distribution problem. This will reduce the value of a "standard" data set. Further, this should not become a data clearinghouse type of activity.

Five sets of data are planned at present, each of which will contain data from several three agencies. For example, a data set might include TIGER files from Census, digital elevation data from DMA, DLG files from USGS, chart data from NOS, soils data from SCS, etc. The point is to encourage participation from many agencies. Regions will be selected and prioritized based on current availability of data products. The data sets will be drawn from the following regions; and it is impossible at this time to guarantee that all regions will be incorporated in the current project.

NORTHEAST - Chesapeake Bay  
SOUTHEAST - Southern Florida  
MIDCONTINENT - Eastern Nebraska / Southern Iowa  
MOUNTAIN - Colorado / Wyoming

SOUTHWEST - Arizona / Utah / New Mexico  
NORTHWEST - Washington / Oregon Columbia River Valley

Figure 3 illustrates the Southern Florida region, for which the first data set will be generated. Data to be included in this first set will be provided by Census, USGS, and NOS. As time and funding permit, other data sets will follow. Each file will be compressed to fit onto no more than 2 floppy diskettes, and each data set is intended to comprise 10 - 20 floppy diskettes. Each data set will have a small ASCII instruction file explaining how to unpack and load the data set. The unpacked sets will be in ASCII form, to facilitate preprocessing and use. Data documentation will be a part of the packed data on the disks.

NCGIA will disseminate the data sets in the public domain on a cost recovery basis, in both DOS and Macintosh formats (5 1/4" and 3 1/2" for 800K, 1.2 MB, and 1.44MB diskettes). Dissemination may begin as early as Fall 1989. In addition, NCGIA - organized paper sessions at a national conference are planned (perhaps AUTO-CARTO 10, perhaps LIS/GIS 90) in 18 -24 months to report on actual use of these data sets by agencies and researchers.

### 5.3 Other Deliverables

5.3.1. Visiting Scholars to NCGIA- Buffalo. Professor Jean-Claude Muller, Director of the Institute of Technical Cartography (ITC) in Enschede, Netherlands, has tentatively accepted our invitation to spend February and March 1990 in residence at NCGIA Buffalo, to pursue research on map generalization. Professor Muller will also lead a graduate seminar during his time as visiting scholar, and collaborate with Drs. Battenfield and Mark on a research project on visualization. He has published research on cartographic symbolization, on expert systems, and on map transformations in addition to his work in generalization. His work in Buffalo will relate directly to the research on Multiple Representations, and also contribute to the initial planning of Initiative 7 (Visualizing Data Quality) and Initiative 8 (Expert Systems for Cartographic Design).

5.3.2. Symposium "Towards a Rule Base for Map Generalization". A symposium to address substantive and conceptual issues prerequisite to the development of a knowledge base for cartographic generalization is planned for April 1990, to be jointly sponsored by NCGIA and Syracuse University. The symposium will be hosted by Dr. Robert McMaster (Syracuse) and Dr. Battenfield (NCGIA-Buffalo). Participants will be invited from academia, government and private sector, and each will present a paper focusing on a specific issue (for example, conflict resolution, or preserving the balance between artificial intelligence and amplified intelligence) that must be resolved in order to implement a knowledge base for automated

generalization. The purpose is not to present empirical results but rather to develop conceptual frameworks and research design. A joint budget of \$8,000 has been earmarked for the Symposium. The symposium agenda will be presented this summer to two book publishers (Taylor and Francis, and Oxford University Press) to solicit interest in publication of papers. At the NCGIA Board of Directors meeting in June 1989, a suggestion was made by Robert T. Aangeenbrug that other publishers be contacted as well. This suggestion will be pursued.

5.3.3. National Conference Sessions at Toronto AAG Meetings. Several special paper sessions are planned for the Toronto Meetings, to serve as the National Conference closing formal Initiative activities. At least one of these sessions will include presentations from the rule base symposium discussed above. Other presentations will report on research projects abstracted in previous sections. Participation from external researchers will be encouraged, and invitations have been extended to researchers from Europe, Australia, Canada, and the US. Joint sponsorship of the paper sessions in Toronto has been agreed to by the AAG Cartography Specialty Group and by the AAG GIS Specialty Group.

## **6. Summary**

It is important to realize this provides a starting point for research on multiple representations research, an agenda of research priorities to be discussed, criticized, and expanded by interested researchers in many disciplines. Topics discussed and explored at the Specialist Meeting and in research developing since the Meeting are of course colored according to special interests and knowledge of the particular researchers. Many of these topics form a continuing research interest, and other items indicate directions in thinking developed as a result of discussions held at the Specialist Meeting. The formal endpoint of the Initiative (currently scheduled for spring, 1990) is not intended as the terminus of the research, but rather as point at which the establishment and refinement of a research agenda provides impetus for the rest of the research community.

## Appendix A

### Recent Papers by Participants Related to Multiple Representations

Participants were asked to submit one or more recently published articles of relevance to the topic of Multiple Representations. It was intended that these articles would be xeroxed and distributed to all participants. However, the volume of submissions and the delay imposed by copyright permission requests precluded distributing articles at the meeting. The following lists the citations for articles submitted by participants; copies of reprints may be requested directly from the participants at the addresses listed in Appendix C.

- Barrera, R. and Vazquez, A.M. (1984) A hierarchical method for representing terrain relief. In: **Spatial Information Techniques for Remote Sensing Today and Tomorrow**, ed. R.M. Haralick, IEEE Service Center, p. 87.
- Batty, M. Fotheringham, S. and Longley, P. (1989) Urban growth and form: Scaling, fractal geometry and diffusion-limited aggregation. Submitted to **Environment and Planning A**.
- Batty, M. and Longley, P.A. (1988) The morphology of urban land use. **Environment and Planning B**. Planning and Design 15: 461-488.
- Batty, M. and Sikdar, P.K. (1984) Proximate aggregation-estimation of spatial interaction models. **Environment and Planning A**. 16: 467-486.
- Brassel, K.E. and Weibel, R. (1988) A review and conceptual framework of automated map generalization. **International Journal of Geographical Information Systems** 2: 229-244.
- Buttenfield, B.P. (1989) Scale-dependence and self-similarity of cartographic lines. **Cartographica**, forthcoming.
- Buttenfield, B.P. (1987) Automating the identification of cartographic lines. **The American Cartographer** Vol. 14: 7-20.
- Buttenfield, B.P. (1986) Digital definitions of scale-dependent structure. In: **Proceedings AUTO-CARTO London 1**: 497-506.
- Chen, Z-T. and Tobler, W.R. (1986) Quadtree representations of digital terrain. In: **Proceedings Auto-Carto London 1**: 517-526.
- Dutton, G. (1988) Modelling locational uncertainty via hierarchical tessellation. Presented at Specialist Meeting 1, National Center for



Geographic Information and Analysis, Santa Barbara, CA December 12-16, 1988.

- Ehlers, M., Beard, M.K. and Jadcowski, M. (1989) The Impact of High Resolution Satellite Data on Geographic Information Systems. **Proceedings**, International Geographic Systems (IGIS) Symposium, Baltimore, Maryland. (Forthcoming)
- Fotheringham, A. S., Batty, M. and Longley, P.A. (1989) Diffusion-limited aggregation and the fractal nature of urban growth. Submitted to **Papers of the Regional Science Association**.
- Gunther, O. (1989) The cell tree: an index for geometric databases. In: **Proceedings**. IEEE 5th International Conference on Data Engineering, February, 1989.
- Gunther, O. and Bilmes, J. (1989) The implementation of the cell tree: design alternatives and performance evaluation. In: **Proceedings**. BTW 89 - Database Systems for Office Automation, Engineering and Scientific Applications, Springer-Verlag.
- Honaver, V and Uhr, L. (1988) A network of neuron-like units that learns to perceive by generation as well as reweighting of its links. In: **Proceedings** of the 1988 Connecticut Models Summer School, D. Touretsky, G. Hinton, T. Sejnowski, eds., Morgan Kaufmann, San Mateo, CA.
- Li, Z.N., and Uhr, L. (1987) Pyramid vision using key features to integrate image-driven bottom-up and model-driven top-down processes. **IEEE Transactions on Systems, Man and Cybernetics**, Vol. SMC-17 Z
- Little, James J., Bledsoe, G.E., and Cass, T.A. (1989) Algorithmic techniques for computer vision on a fine-grained parallel machine. Accepted for publication by **P.A.M.I.**, March, 1989.
- Mark, D.M. (1986) The use of quad-trees in geographic information systems and spatial data handling. In: **Proceedings** AUTO-CARTO London 1:517-526.
- Poggio, T., Gamble, E.B. and Little, J.J. (1988) Parallel integration of vision modules. **Science** 242: 436-440.
- Samet, H. (1985) Reconstruction of quadtrees from quadtree medial axis transforms. In: **Computer Vision, Graphics and Image Processing** 29: 311-328.

- Samet, H. (1985) Using quadtrees to represent spatial data. In: **Computer Architectures for Spatially Distributed Data**. H. Freeman, G. Pieroni, eds. Berlin: Springer-Verlag, 391 pp.
- Samet, H. (1985) Data structures for quad-tree approximation and compression. In: **Communications** of the ACM 28: 973-993.
- Samet, H. and Rosenfeld, A. (1985) **Application of hierarchical data structures to Geographical Information Systems** (Contract report). College Park: Maryland Univ. 71 pp.
- Samet, H. and Rosenfeld, A. (1983) **Application of Hierarchical Data Structures to Geographical Information Systems** (contract report). College Park: Maryland Univ. 84 pp.
- Samet, H. and Shaffer, C.H. The segment quadtree: a linear quad-tree based representation for linear features. In: **Proceedings IEEE Conference on Computer Vision & Pattern Recognition**. San Francisco: pp. 385-389.
- Tobler, W.R. (1988) Resolution, resampling and all that. Paper presented International Geographical Union Global Database Planning Project, England, May 9 - 13, 1988.
- Tobler, W.R. Frame independent spatial analysis. (Abstract)
- Uhr, L. (1983) Pyramid multi-computer structures, and augmented pyramids. In: **Computing Structures for Image Processing**, Academic Press, London, pp. 95 - 112.
- Uhr, L. (1985) Pyramid multi-computers, and extensions and augmentations. In: **Algorithmically Specialized Parallel Computers**, Academic Press, Inc., London, pp. 117 - 186.
- Uhr, L. (1986) Parallel, hierarchical software hardware pyramid architectures. In: NATO Advanced Study Institute Series (F), **Computer and System Sciences**, Vol. 25, eds. V. Cantoni, S. Levialdi, Springer-Verlag, Berlin, p. 1.
- Uhr, L. (1987) Highly parallel, hierarchical, recognition cone perceptual structures. **Technical Report #688**, Computer Sciences Department, University of Wisconsin-Madison.
- Walters, D. (1987) Selection of image primitives for general-purpose visual processing. **Computer Vision, Graphics and Image Processing** 37: 261-298.

Walters, D. (1988) Discrete contour descriptors. In: **From the Pixels to the Features**, J.C. Simon, Ed., North Holland Elsevier.





AGENDA  
18 - 21 February, 1989

*NCGIA Specialist Meeting on Multiple Representations*

## **APPENDIX B**

### **Agenda for the Specialist Meeting**

The following pages contain the agenda followed during the I3 Specialist Meeting in February. This agenda differs somewhat from the spiral bound booklet distributed 2 weeks prior to the Meeting, due to last minute changes in arrangements and participants. Dr. Leonard Uhr of the University of Wisconsin was unable to attend the meeting due to illness; and Dr. Joel Morrison of USGS also had to cancel his participation. Mr. Michael Domaratz attended in Dr. Morrison's place.



# **NCGIA**

## *Initiative 3 - "Multiple Representations"*

*18-21 February 1989*  
*Darwin D. Martin House*

*Barbara P. Battenfield, Initiative Leader*

*National Center for Geographic Information and Analysis*  
*State University of New York at Buffalo*

**AGENDA**  
18 - 21 February, 1989

*NCGIA Specialist Meeting on Multiple Representations*

**Friday 17 Feb**

Arrival in Buffalo

Buffalo Hilton  
at the Waterfront  
120 Church Street  
Buffalo NY 12020  
(716) 845-5100

**Saturday 18 Feb**

8:00 am	Vans at Buffalo Hilton for pickup Depart for Darwin D. Martin House at 8:15	Darwin D. Martin House 125 Jewett Parkway Buffalo NY (716) 831 - 2406
8:30 am	Continental Breakfast	
9:00 am	Introduction and Format <b>Barbara Battenfield, NCGIA - Buffalo</b>	
9:30 am	Private Vendor Perspectives <b>Michael Dobson, Rand-McNally</b> <b>Geoffrey Dutton, PRIME</b> <b>Marvin White, ETAK</b>	
10:30 am	Coffee Break	
10:45 am	Federal Agency Perspectives <b>Richard Berg, DMA</b> <b>Michael Domaratz, USGS</b> <b>Robert Marx and Frederick Broome, US Census</b> <b>Charles Schwarz, NOS</b>	
12:30 pm	Lunch	
1:30 pm	General Discussion of Impediments Formulate Small Groups	
2:45 pm	Coffee Break	
3:00 pm	Break into Small Groups	
5:00 pm	Wine and Cheese Reception	

**AGENDA**  
18 - 21 February, 1989

*NCGIA Specialist Meeting on Multiple Representations*

**Saturday, 18 Feb (con't.)**

- 6:00 pm      Buffet Dinner
- 7:00 pm      **David Simonett, Co-Director NCGIA**  
General Discussion - Summarize Impediments  
Reports of Small Group Discussants
- 10:00 pm     Vans Depart for the Buffalo Hilton

**Sunday 19 Feb**

- 8:00 am      Vans at Buffalo Hilton for pickup  
Depart for Darwin D. Martin House at 8:15
- 8:30 am      Continental Breakfast
- 9:00 am      Academic Perspectives
- Cartography -
- Waldo Tobler, NCGIA Senior Scientist**  
Kate Beard, NCGIA-Maine  
Barbara Battenfield, NCGIA-Buffalo  
Robert McMaster, Syracuse University  
Robert Weibel, U. Zurich
- Geographical Analysis -
- Michael Batty, U. Wales**  
**Stewart Fotheringham, NCGIA - Buffalo**  
David Simonett, NCGIA-Santa Barbara
- 10:30 am     Coffee Break
- 10:45 am     Academic Perspectives (con't.)
- Computer Vision -
- James Little, U. British Columbia**  
**Deborah Walters, NCGIA - Buffalo**  
Renato Barrera, NCGIA-Maine

**AGENDA**  
18 - 21 February, 1989

*NCGIA Specialist Meeting on Multiple Representations*

**Sunday 19 Feb (con't.)**

Data Models -

**Hanan Samet, U. Maryland**  
Manfred Ehlers, NCGIA-Maine  
Andrew Frank, NCGIA-Maine  
David Mark, NCGIA- Buffalo

Database Queries -

**Oliver Gunther, NCGIA- Santa Barbara**  
Terence Smith, NCGIA- Santa Barbara

12:30 noon Lunch

1:30 pm General Discussion of Research Goals  
Formulate Small Groups

2:45 pm Coffee Break

3:00 pm Break into Small Groups

5:00 pm Vans depart for Buffalo Hilton

6:30 pm Vans depart for Dinner at Buffalo Brew Pub  
Buffalo Brew Pub  
6861 Main Street  
Williamsville, NY  
(716) 632-0552

**Monday, 20 Feb**

8:00 am Vans at Buffalo Hilton for pickup  
Depart for Darwin D. Martin House at 8:15

8:30 am Continental Breakfast

9:00 am General Discussion -- Summarizing Research Goals  
Reports of Small Group Discussants

10:30 am Coffee Break  
(Vans available for those who have to leave the Specialist Meeting)

10:45 am General Discussion of Research Priorities  
Formulate Small Groups

**AGENDA**  
18 - 21 February, 1989

*NCGIA Specialist Meeting on Multiple Representations*

**Monday, 20 Feb (con't.)**

12:00 noon Lunch

1:00 pm Break into Small Groups

3:15 pm Coffee Break

3:30 pm Small Group Discussants Prepare Final Reports

5:00 pm Vans depart for Buffalo Hilton

6:30 pm Dinner at Crawdaddy's

Crawdaddy's Restaurant  
2 Templeton Terrace  
Buffalo, NY  
(716) 856-9191

**Tuesday 21 Feb**

8:00 am Vans at Buffalo Hilton for pickup  
Depart for Darwin D. Martin House at 8:15

8:30 am Continental Breakfast

9:00 am General Discussion Summarizing Research Priorities  
Final Reports of Small Group Discussants

10:30 am Coffee Break

10:45 am Summarize Specialist Meeting  
Evaluate Specialist Meeting Format

11:45 am Close of Specialist Meeting

12:00 noon Vans depart for Buffalo Hilton



## **APPENDIX C**

### **Mailing List for the Specialist Meeting**

#### **PARTICIPANTS**

Dr. Renato Barrera  
NCGIA  
Dept. Surveying Engineering  
University of Maine  
Orono, ME 04469  
(207) 581-2149  
RENATO@MECAN1.BITNET

Dr. Michael Batty  
Head of Department  
Department of Town Planning  
University of Wales, Cardiff  
P0 Box 906, Cardiff CF1 3YN  
Wales, UK  
BATTY@  
TOWP.ABCY.CARDIFF.AC.UK

Dr. M. Kate Beard  
NCGIA  
Dept. Surveying Engineering  
University of Maine  
Orono, ME 04469  
(207) 581-2149  
BEARD@MECAN1.BITNET

Dr. Richard A. Berg  
HQ DMA Bldg. 56  
US Naval Observatory  
Washington, DC 20305-3000  
RBERG@WSMR-SIMTEL20.ARMY.MIL

Mr. Frederick Broome  
Geography Division  
US Bureau of the Census  
Washington, DC 20233  
(301) 763-3973

Dr. Barbara P. Battenfield  
NCGIA  
Department of Geography  
SUNY/Buffalo  
Buffalo, NY 14260  
(716) 636-2283  
GEOBABS@UBVMS.BITNET

Dr. Michael Dobson  
Digital Cartography-Rand McNally  
8255 N. Central Park  
Skokie, IL 60075  
(312) 673-9100 Ext 3267

Mr. Michael Domaratz  
US Geological Survey  
National Mapping Division  
516 National Center  
Reston, VA 22092  
(703) 648-4639

Mr. Geoff Dutton  
150 Irving Street  
Watertown, MA 02172  
(617) 923-2307

Dr. Manfred Ehlers  
NCGIA  
Dept. Surveying Engineering  
University of Maine  
Orono, ME 04469  
(207) 581-2149  
EHLERS@MECAN1.BITNET

Dr. A. Stewart Fotheringham  
NCGIA  
Department of Geography  
SUNY/Buffalo  
Buffalo, NY 14260  
(716) 636-2289  
GEOASF@UBVMS.BITNET

Dr. Andrew U. Frank  
NCGIA  
Dept. Surveying Engineering  
University of Maine  
Orono, ME 04469  
(207) 581-2149  
FRANK@MECAN1.BITNET

Dr. Oliver Gunther  
FAW - AI Laboratory  
University of Ulm  
Postfach 2060  
7900 Ulm, West Germany  
(011-49-731) 501-500  
GUENTHER@DULFAWLA.BITNET

Dr. James Little  
Dept. Computer Science  
University of British Columbia  
Vancouver, BC CANADA V6T 1W5  
(604) 228-3061  
LITTLE@CS.UBC.CA

Dr. Ross D. MacKinnon  
NCGIA  
Department of Geography  
SUNY/Buffalo  
Buffalo, NY 14260  
(716) 636-3101  
PROROSS@UBVMS.BITNET

Dr. David M. Mark  
NCGIA  
Department of Geography  
SUNY/Buffalo  
Buffalo, NY 14260  
(716) 636-2283  
GEODMM@UBVMS.BITNET

Mr. Robert Marx  
Geography Division  
US Bureau of the Census  
Washington, DC 20233  
(301) 763-5636

Mr. John McAvoy  
Digital Cartography  
Rand McNally  
8255 W. Central Park  
Skokie, IL 60075  
(312) 673-9100

Dr. Robert B. McMaster  
Dept. Geography  
343 HB Crouse  
Syracuse University  
Syracuse, NY 13244-1160  
(315) 443-1870  
MCMASTER@SUNRISE

Dr. Hanen Samet  
Dept. Computer Science  
College Park, MD 20742  
University of Maryland  
(301) 454-7987  
HJS@ALV.UMD.EDU

Dr. Charles R. Schwarz  
Chief ADP Staff  
National Ocean Service  
Office of Charting and  
Geodetic Services--NOAA  
6001 Executive Blvd.  
Rockville, MD 20852  
(301) 443-8381  
`[C.SCHWARZ/OMNET]MAIL/USA%TEL  
EMAIL'@INTERMAIL.ISI.EDU

Dr. David S. Simonett  
NCGIA  
Department of Geography  
U. California/Santa Barbara  
Santa Barbara, CA 93106  
(805) 961-8224  
SIMONETT@SBITP.BITNET

Dr. Terence R. Smith  
NCGIA  
Department of Computer Science  
U. California/Santa Barbara  
Santa Barbara, CA 93106  
(805) 961-8224  
SMITHTR%NIMBUS.UCSB.EDU

Dr. Waldo R. Tobler  
NCGIA  
Department of Geography  
U. California/Santa Barbara  
Santa Barbara, CA 93106  
(805) 961-8224  
TOBLER@SBITP.BITNET

Dr. Deborah K. Walters  
NCGIA  
Dept. Computer Science  
SUNY/Buffalo  
Buffalo, NY 14260  
(716) 836-3187  
WALTERS@CS.BUFFALO.EDU

Mr. W. Robert Weibel  
Dept. Geography  
University of Zurich  
Winterthurerstrasse 190  
CH-8057 Zurich  
Switzerland  
K491170@CZHRZU1A.BITNET

Mr. Marvin White  
ETAK, Inc.  
1455 Adams Drive  
Menlo Park, CA 94025  
(415) 328-3825

## **RAPPORTEURS**

Joseph DeLotto, Senior Rapporteur  
NCGIA  
Dept. Geography  
SUNY/Buffalo  
Buffalo, NY 14260  
(716) 636-2545  
ESRIJD@UBVMS.BITNET

NCGIA  
Dept. Geography  
SUNY/Buffalo  
Buffalo, NY 14260  
(716) 636-2545

NCGIA  
Dept. Surveying Engineering  
University of Maine  
Orono, ME 04469  
(207) 581-2149

Michael Gould  
V010KKYE@UBVMS.BITNET

Bud Bruegger  
BUD@MECAN1.BITNET

Marcin Jasinski  
V103RBXV@UBVMS.BITNET

Douglas Hudson  
HUDSON@MECAN1.BITNET

Mark MacLennan  
V008HK8J@UBVMS.BITNET

Shou-Wen Chen  
ESRISTO3@UBVMS.BITNET

## **APPENDIX D**

### **Position Papers Submitted by Participants**

The following pages contain the complete set of position papers submitted by participants about three weeks prior to the Specialist Meeting. This set was spiral bound and distributed to participants with the request that their preparation for the Meeting include reading the papers and arrive prepared to discuss them. One participant brought a position paper to the Meeting, and this has been added to the set.

Batty, Michael	Position Statement
Berg, Richard A.	Research Initiatives Needed by the Defense Mapping Agency (DMA)
Buttenfield, Barbara P.	Thoughts on the Nature of Things That Change with Scale
Dobson, Michael W., and McAvoy, John M.	Position Statement
Dutton, Geoffrey	The Fallacy of Coordinates
Fotheringham, A. Stewart	Position Statement
Gunther, Oliver	Database Support for Multiple Representations
Maine Surveying Engineering Department	Research Topics on Multiple Representations
Mark, David M.	Multiple Views of Multiple Representations
Marx, Robert W., and Broome, Frederick	Multiple Representations
McMaster, Robert B.	Position Statement
Schwarz, Charles R.	Position Statement on the Problem of Multiple Representations
Uhr, Leonard	Some Preliminary Thoughts on Multiple Representations
Weibel, Robert	The Fuzzy Nature of Spatial Data

# **POSITION STATEMENT FOR THE NCGIA SPECIALIST MEETING ON 'MULTIPLE REPRESENTATIONS', BUFFALO, 18-21 FEBRUARY 1989**

**Michael Batty**

**University of Wales at Cardiff  
PO Box 906  
CARDIFF CF1 3YN, UK**

Most urban land use models are concerned with allocating quantities of demographic and economic activity to zones of an urban region which are seldom smaller than typical census tracts or enumeration districts. Consequently, insofar as the morphology of the urban area is represented and simulated, its physical shape can only be inferred from locational data given by treating census tracts as points. In one sense, such models are predicated at a level of abstraction which precludes any representation of urban form, but at another level, the predictions from such models can only be satisfactorily communicated if these can be visualised in map form. Such visualisation is clearly useful in communicating somewhat obtuse forecasting methods to planners and their client groups, but it is also a useful way of checking models for consistency with respect to what their predictions imply about urban density and structure. It would thus appear that one strategy for progressing these ideas would be to link traditional model predictions to models which are able to convert these traditional 'point data' representations into aerial representations with a morphology embedding land uses into one another, characteristic of real urban structures. A preliminary method for effecting this has been reported in Batty and Longley (1986) using a fractal morphology as a framework within which traditional predictive location models are embedded.

Notwithstanding the motivation for researching such morphologies, the first step is to find some digital representation of the morphology which is consistent with the processes that generate it. The simplest starting point is to examine the shapes of various land use parcels and to simulate their form across a range of scales. Accordingly this work has begun using conventional techniques which enable the degree to which the form in question depends on scale, to be measured. The use of the fractal dimension in classifying land parcel boundaries has been developed. However, the idea of a single process generating a land use structure which is both invariant to scale and time is somewhat fanciful. Our work has shown that such boundaries are invariably multi-fractal with transient but regularly varying self-similarity, thus suggesting the action of many processes at different scales. Moreover, our techniques also suggest that many related examples and applications of self-similarity in various fields dealing with physical phenomena should be reworked in the light of subsequent research emphasising multi-fractality (Batty and Longley, 1988).

There is an immediate pragmatic payoff from our work in suggesting methods for data compression where such data involves map boundaries. Such generalisations raise the problem of to whom such maps are for and to what purpose they may be used. There have been several similar ideas applied to map generalisation, but so far these have only been suggestive of appropriate technique. Much work still needs to be done, however, in how morphologies are created through diverse processes operating in space and through time, how land use patterns interlock with each other, and how 'composite' boundaries are formed. Not only need different submorphologies for different land uses be defined, but rules for how the overall urban morphology might be created need to be identified. Fractal representations of interlocking objects (land use) need to be explored, and, in particular, this analysis needs to be extended to two-dimensional representations which define shape in a more formal way.

Many issues emerge from this research which I hope may strike a chord with the work of others at this Specialist Meeting. These involve the following :

to what extent can simple models leading to the digital representation of artificial in contrast to natural/physical phenomena be defined

to what extent do such representations vary with scale

to what extent can scale-related representations be developed for the plane as well as the line. For example, can we find convincing methods such as Iterated Function System analysis to generate maps and enable data compression

to what extent need we be guided by formal geographic theory, for example, hierarchical central place theory etc. in representing space, and visualising spatial form

These are some of the broader issues worthy of discussion but there are a host of issues of a more technical nature. I hope the meeting will concentrate on commonalities among multiple representations of geographic space, as well as upon their differences.

## References

- M. Batty and P. Longley (1986) The Fractal Simulation of Urban Structure, *Environment and Planning A*, 18, 1142-1179
- M. Batty and P. Longley (1988) The Morphology of Urban Land Use, *Environment and Planning B*, 15, 461-488

## **Research Initiatives Needed by the Defense Mapping Agency (DMA)**

**Richard Berg**

The Defense Mapping Agency has currently one of the community's largest "R&D" activities underway in a production line conversion program called MARK 90. MARK 90 provides an operational production environment which will create over time a digital mapping, charting and geodesy (MC&G) data base of unprecedented size - some thousands of terabytes (400 million college textbooks).

DMA's business is to provide MC&G data to the user (military) community. For the most part, we don't provide software to manipulate the data, although today we do maintain data transformation programs supporting specific systems. We also don't develop systems which use our data. Our only responsibility is to make sure that MC&G data is robust enough for successful use, and is provided in an acceptable, standardized format for those systems.

How each system manipulates DMA-supplied MC&G data is the responsibility of system developers, not DMA. There are many, new "exploitation" systems continually appearing in the market place. Advancing technologies demand that DMA produce more accurate, higher-resolution and greater quantities of digital MC&G data for these systems to process and interpret. These systems also employ a variety of algorithms for datum and projection transformations, and other geographic information data manipulations, for which there is already clearly a lack of consistency and accuracy in application.

We know that

Commanders in the field need the ability to visualize and interact with the battlefield environment. (You can also read this as, "State and local governments interact with their jurisdictional environment.") Markets for these capabilities are unbounded, both in the civil and the defense communities!

- More powerful workstations that assist environmental analysis are on the way, and they require digital geographic information-

- Large data bases are not being fully exploited. It is difficult for the user to comprehend the wealth of information available.

- DoD systems using new technologies require the support of TAILORED digital MC&G products developed from integrated digital geographic information data bases.

- Many users are digitizing hardcopy maps and charts to support their needs, resulting in a proliferation of data that has no standard and does not "interoperate".

To meet increasing demands for MC&G data, DMA production processes will continue to evolve beyond MARK 90. Our role as provider of data will also expand to cope with the proliferation of systems requiring MC&G software tools to manipulate the data we provide. DMA is developing a capability to access, aggregate and distribute all digital geographic information holdings, and to tailor data to standard formats and structures for distribution as functional products. A functional product will be a tailored package of distributable data layers, serving a range of users, that meet requirements for common geographic coverage and feature selection, including utility software permitting basic manipulation and update of these layers by field users, and provided on a specific media.

To do this effectively, significant research activities need to be undertaken in:

- data standards - to facilitate conversion of user-oriented "objects" from internal types, structures and attributes to an exchangeable, interoperable data set (this is where "representations" are important);

- multilevel security - to accommodate accessibility and classification differences among linked data files and elements;

- knowledge base/rule base systems - to significantly shorten the time for exploitation of sources which feed an MC&G data base, the time for generation of products from an MC&G data base; and to improve the usability of GIS systems in the field;

- management of VERY LARGE data bases - to enhance performance of storage, retrieval and concurrency operations; and

storage media - to provide for improved data densities and data transfer rates in support of provisioning field users with MC&G data and GIS software.

DMA has great interests in GIS initiatives. This includes strong support of efforts to develop formal standards which will, in turn, support cost-effective, interoperable data, data bases and GISs. Standards are needed for data structures, media, spatial data layers, computer graphics, MC&G utility software, communications, procedures and training. Additional considerations of quality assessments and maintenance of data are also important.



# Thoughts On the Nature of Things that Change with Scale

Barbara P. Battenfield

## Position Paper for NCGIA Special Meeting on Multiple Representations

The challenge of storing a single digital version of geographic features for representation at many levels of resolution continues to obstruct the efficiency of cartographic operations and the recognizability of cartographic products. Priorities for the research initiative include development of digital models to accommodate the possibility of scale-dependent structures, calibration of these models for feature identification, feature abstraction, and spatial inference in large data bases, and development of methods to preserve computational efficiency and accuracy in spatial data structures.

At first glance, this description of the multiple representations problem (paraphrasing the original NSF proposal) makes our task at the specialist meeting seem straightforward. Basic issues must be addressed, however, that offer subtle intellectual traps. No single position paper can cover all the issues and impediments, of course, and my own view is biased somewhat by my avocation to cartography. The traps I see in the cartographic discipline include a blind faith in the relationship between precision and accuracy, an ongoing conviction that map error is primarily positional, and adoption of digital models of scale change that can accommodate only self-affine geometric properties.

What the cartographer means by 'accuracy' has until recently been measured in terms of horizontal and vertical position. In the context of map making and map analysis, accuracy refers as much to visual appearance as to locational precision. For computer mapping or analysis of GIS map information the distinction becomes more clear, as analytic computations are performed on the digital version stored on disk, and not necessarily on the version displayed on the CRT screen. The graphic depiction seen by a GIS user may appear at a different scale than the digital version, and may contain only a subset of the coordinates stored on disk. Brassel and Weibel (1988) apply the term "statistical generalization" to operations of data reduction and filtering the stored digital information. "Cartographic generalization" focuses on the displayed version, and emphasizes what Brassel and Weibel refer to as visual effectiveness, or recognizability.

The heart of the cartographic challenge is that the visual appearance of features often varies with the resolution of their map depiction. This is true for geographical and statistical landscapes alike. For example, the features along a coastline will change with the resolution of the satellite image on which they appear—bays and inlets evident in the SPOT image may not be resolved in the Thematic Mapper or LANDSAT image. This is not a case of pixel replication or magnification, but rather introduction of newly resolved details. The same phenomena will be evident for vector data, for example in looking at various scale USGS maps of a single region. To cite a statistical example, Tobler has compared migration flows collected by state, county, census tract, and block face. At each level of aggregation, the pattern and rate of migration vary.

Quite a lot of attention has been paid in recent geographic literature to the fact that things change with scale change, and to questions about how to model the rate of change. Most often, the models adopted by geographers assume that the change is self-affine, within a given range of scales or resolution. The computational efficiency of this assumption is obvious. For one thing, it allows formulation of linear models to describe the progression of values (migration flows, coastline length, etc.). For example, in map simplification, if one assumes that details are replicated (either statistically or precisely) at finer levels of resolution, then tolerance thresholds may be chosen in proportion to the scale change.

For example, in applying the Radical Law (Topfer and Pillewizer, 1966), assumptions of homogeneous detail are coupled with the equiprobable importance of each coordinate. The Radical Law assumes self-similarity in simplification. The rate of coordinate reduction is constant for scale reduction from 1:62,500 to 1:250,000 or from 1:10,000,000 to 1:40,000,000, regardless of the differing geomorphic structures that are visually evident across one range of map scale but not the other. And herein lies the trap. Geomorphic process tends to vary with scale, and its variation is often reflected in the visual appearance of the mapped feature. For example, evidence of isostatic rebound may be apparent in a small-scale (1:40,000,000) representation of Hudson Bay; but at much larger scales, nearer 1:100,000, erosional features may become evident, and isostatic structure obscured.

It is difficult to predict what types of cartographic features display self-affinity, and what other types will be characterized by scale-dependent change. Digital methods for describing structure across progressions of scale have been reported (e.g., Witkin, 1986; Battenfield, 1986), and need to be refined and expanded. Automatic identification of scale-dependent cusps in such digital description schemes may provide an intelligent means to guide selection and modification of tolerance thresholds during map simplification. The geomorphological changes are commensurate with changes in geometry, and this is easy to demonstrate, although we have only rudimentary understanding of the associations between geometry of map features and recognizability.

Rules to determine which subset of detail to incorporate into a given scale of feature depiction must be explicitly defined, and incorporated into our analytical and display software automatically, to preserve consistency and accuracy in GIS displays. Additionally, cartographers must develop models for cartographic and statistical generalization that can accommodate the geometry of features that change with scale change, while preserving the geometry of things that do not change. The point to be made is that until we better understand the nature of relationships between geographic structure, graphical appearance and measurable geometry, it will be difficult to automate the production of multiple scale representations from a single detailed data base.

Here is the scope of one problem in multiple representations, and an all too brief and general enumeration of research directions. I see some pieces of the puzzle clearly, struggle mightily with others, grapple with computational issues, discover (often after the fact) that other disciplines consider as trivial those aspects I had thought insurmountable. I need to talk, ask questions, get the advice of researchers in other disciplines.

BRASSEL, K.E. AND WEIBEL, R. (1988) A Review and Framework of Automated Map Generalization. International Journal of GIS, vol.2(3), p.229-244.

BUTTENFIELD, B.P. (1986) Digital Definitions of Scale-Dependent Line Structure. Proceedings AUTO-CARTO London, vol. 1, p-497-506.

TOPFER, R. AND PILLEWIZER, W. (1966) The Principle of Selection. Cartographic journal, vol.3(1), p.10-16.

WITKIN, A.P. (1986) - Scale Space Filtering. Chapter I in Pentland, A.P. (1986) From Pixels to Predicates. Ablex Publishing Corp., Norwood, N.J. p.5-19.

## **Position Statement**

**Rand McNally and Co.**

**Dr. Michael Dobson and John M. McAvoy**

The multiple representation problem as it applies to Rand McNally involves the commercial production of maps from a single master database. The master database will be updated and maintained to provide the information necessary to produce a variety of maps with different scales, symbologies and content. Though there are many benefits in using a single database, products generated from it are limited by the scale and generalization of the data as it was captured. Depending on the scale and purpose of the map, the generalization of linework, selection of text and symbology is going to vary.

Although various rule based procedures have been developed to automatically vary the generalization depending on scale, the solutions do not always apply themselves to the model of a single master database. Once an algorithm is used to produce some form of generalization, editing is usually required to handle the special cases before it can be used for output. Thus, if the same product is to be generated at different times, the same edits will have to be repeated. If the newly generated product is saved in a separate database, a cost in storage is incurred and the benefits of making updates to a single master database are lost.

Another area of concern is the range of map scales that can be produced from a database before the level of generalization becomes unacceptable to the map user. If algorithms are used to automate generalizations, what are the differences in computational and editing costs when using the algorithms on drastic scale changes compared to slight scale changes? When does it become more efficient to start a new database for a different range of scales?

Algorithms that perform generalizations when going to smaller scales tend to leave numerous special cases unresolved. Though algorithms may smooth lines and simplify data, improvements are needed to determine what data is to be simplified and how it will affect the relations of different lines and symbols. Can databases be designed that will take into account the importance of map features at different scales, enabling intelligent algorithms to more readily make cartographically correct generalizations? Relations between the lines themselves and symbology must also be preserved when the scale is reduced. Rule based algorithms have been used to approach this problem and automate name placement: yet, how efficiently can these algorithms be applied to a database and be used in a practical manner for the production of maps?

The various methods of storing cartographic linework in a database is another issue of concern. The scale at which linework is displayed, can influence the way it is stored in the database. Curved lines may be stored as line strings if enough points are placed on the curves and the lines are not displayed at too large a scale. If the lines are to be displayed at large scales, a greater number of points will be needed to represent the curve in an appropriate fashion. This will incur greater storage costs. An alternative method is to store curved lines as "curved lines" in the database using splines, or Akima or Bezier curves. By using these methods, a savings in computer storage can be achieved and the linework can be displayed at greater scales. Yet, if curve functions are going to be used, the lines must be entered into the database by digitizing them manually: making data collection more expensive. Thus, is it more efficient to use an algorithm to smooth line strings that represent curves at large scales or to actually store the curve?

The issues concerning the multiple representation problem are numerous. Relevant research topics should include the further development of generalization algorithms, with an emphasis on making generalizations based on neighbor relationships. It would also be beneficial to see how these algorithms can be practically applied to the production of complete maps. Investigations should also look into methods of constructing databases that allow for automated generalization procedures to run more easily on them. Using present generalization algorithms, what data structures will provide the greatest flexibility in producing various scaled maps in a cost efficient manner?

# **The Fallacy of Coordinates**

**Geoffrey Dutton**  
**Prime Computer, Inc.**

This is a personal response to NCGIA's call for position papers on the subject of its third research initiative, Multiple Representation. This seems to be an appropriate occasion to step back in order to get a larger perspective on what spatial data is and what we want to do with it. I feel strongly that we need to take time to examine some basic assumptions we rely upon, if we hope to advance the state of our art. During the years since I last advocated doing this, the issues have been sharpened, the technology has almost matured, the problems we must handle are more acute and the misconceptions we share remain the same.

The title given to NCGIA's third research initiative, "Multiple Representation" (can you say "multirep"?), is too cryptic and vague to describe a concrete research effort. It admits to multiple interpretations of intent, such as:

- A. Assembly of cartographic features from sets of primitive elements,
- B. Properties of alternative model surfaces and coordinate systems,
- C. Scale-specific feature generalization and enhancement
- D. Selection of appropriate cartographic symbolism,
- E. Pattern recognition and feature identification,
- F. Effects of spatial decomposition,
- G. Geometric versus topological methods,
- H. Alternative spatial data structures,
- I. Temporal aspects of shape,
- J. (Your favorite jargon).

This is a proverbial Ball of Wax. While NCGIA's provisional description offers a slant on multirep, it fails to satisfactorily specify assumptions about the state of the art, the major open issues or the contexts in which they are relevant. As NCGIA is proposing to spearhead what amounts to a crash program to address multirep issues, a better definition of the problem domain is needed before we can begin to describe aspects of a solution space, much less populate it with paradigm

What we have is a description of some dilemmas that current spatial data models, structures and techniques must address (effects of scale, structuring cartographic data, relating dissimilar representations), couched in terms of various potential solutions ("computer vision", "pattern recognition", "hierarchical. structures"). This does not define a proper research agenda-, rather, it tends to encourage further development of technique that may or may not solve cartographic problems, without any useful criteria for success or failure- We can't quite say what we want to do; even if we could and tried, we couldn't tell how well it worked.

I take the position that many of the multirep problems alluded to are inherent to fundamental paradigms of computer cartography that have too long been taken for granted. These prejudices include:

- cartographic features = coordinates + topology + attributes
- maps as phenomenological models
- the very idea of coordinates

It is entirely possible that the "challenges" of appropriately representing geographic phenomena (both computationally and visually) largely derive from a confusion of conceptual models and technical approaches. That is, the spatial data elements we work with and the ways in which we structure them have infected our comprehension of the real world; we really believe that objects and regions of interest in the world are features, and that features have coordinates, attributes, topology and symbology. Having built a maze, we have challenged ourselves to find a way out.

So much of what GIS is about derives from automated cartography, which in turn derives from mapmaking-cum-computer technology. While spatial data handling has become much more sophisticated than, say, SYMAP, it is still in the grip of the same old mindset. Consider how frequently GIS, having overcome representing regions as spaceship polygons, still try to model space as "acetate" layers of digital map sheets; that terrain is still handled as "contour lines"- that points are dimensionless; that lines are arrays of coordinate points, each independent of the others (some of this may stem from an accident of engineering called Fortran).

Working at the bleeding edge, we are increasingly lacerated by continuing to rely upon such paradigms. Arguably, the most unconscious of these is what I term the fallacy of coordinates. Our spatial models behave as if features "owned" coordinates just as they

possess other attributes, such as length, area, land use, population, etc. This is a natural consequence of organizing phenomena into the paradigm Berry called the "geographic matrix" (but with variable-length rows). The principal problem with this model is that there is no explicit indication of how items in rows or columns might be related or interdependent; each fact is a tight little island. By coding topological relationships, we overcome some of this isolationism at the primitive feature level. But over-reliance on the arc-node-polygon mode of representation has led to excessive reductionism, and strange answers to Marvin White's ontological question: "What exists?" (what he meant was "What are the elementary objects on a map?"). Clearly, most answers to this question are highly subjective and contextual. In my view, whatever exists in the real world, there is no such thing as coordinates there, only locations. Furthermore, at some scale and specificity, every location is shared by different objects; rather than features possessing coordinates, locations have features that somehow characterize and distinguish them.

So what the hell are locations anyway? To me, they are places in the real world, spots on the planet. While they may be identified as points, locations are areas of specified influence (more like the triangles engraved in brass benchmarks rather than the dots placed inside them). This sounds suspiciously like a description of raster data, of a grid dividing up space. As a rule, rasters and quadtrees embody the notion of location better than vector coordinate lists do. But raster data structures are by nature surficial constructs, and are awkward envelopes for sparse data such as points and lines. They also exhibit indifference to their coordinate space, other than to assume it is planar, cartesian and consists of uniform discrete chunks.

Coordinates are convenient notations for locations, but being dimensionless, do not convey scale, only precision (which tends to be fixed and unrelated to accuracy or importance). Once digitized, they somehow become reified, even sanctified. If this seems overstated, consider how we manipulate line vectors: when digital lines are generalized, it is nearly always by selecting points to retain or discard, and methods which move coordinates or fabricate points are regarded with great suspicion. This prejudice persists despite wide agreement that it is cartographic lines that should be modelled, not their individual coordinates. Work by Poiker, Ballard, Battenfield, Chrisman, Douglas, Little, Mark and others has refined our view of lines and their loci, but the fallacy that coordinates actually exist continues to hold sway.

Surveyors and photogrameters please take note: I am not contending that measurements do not matter, I am on your side. In measuring spatial phenomena, as much care should be taken as warranted, and as much precision as possible should be retained. But by the time most source data measurements take up residence in a GIS, the context of their capture has been lost, and often their precision has been ironed flat. Even in the field, with all the primary data capture technology now available, positional accuracy in tracing boundaries may be rather moot: a parcel boundary might start at a bridge abutment, run along the East bank of a stream to a willow tree, then follow a fenceline which skirts a foundation corner, turning right at a survey marker and again at a section corner, to proceed East 350 feet, where it attempts to follow an old hedgerow through a new woodlot to a crumbling boulder, where it beelines to an old elm tree that died and was taken down when the road was widened. Each of these monuments has a certain character, degree of permanence, and may play multiple roles in, the local geography. All have areal extents save the survey marker and the section corner, and these are fictional objects (we probably can't locate the section corner within 10 meters, and the monument's location may be locked up in the files of a deceased surveyor). Surveyors, photogrameters and courts of law understand how to treat evidence of this nature, but no GIS does, even were it to be provided, which would be unusual (whatever exists, it isn't that elm tree). Such is the stuff of coordinates.

The foregoing example is a microcosm of multirep: the "features" that define this putative parcel also participate in other "themes", such as geology, hydrology, transportation, administration, land cover and land use. The themes may be regarded as features of locations, just as the boundary monuments are features of sets of locations. Some GIS can describe such complexity, using constructs such as "shared primitives" which allow individual points, lines and polygons to participate in multiple features. While this is advantageous, it burdens users with identifying key primitives and deciding which ones to include in a given set of coincident features. Shared primitives also tend to complicate operations like spatial overlay (especially if nodes are allowed to wander or coalesce) and feature generalization (purely geometric methods can eliminate otherwise important elements).

Now that we are blessed with fruits of aerospace geodesy such as WGS 84 and GPS, anyone who really wants to know where something is can rent a GPS mobile unit, truck it into the cutback and come back with high quality coordinate data, ready to be read into their GIS. This is real progress; for the first time in history, field surveys can relate directly to a planetary grid, rather than to local plane coordinate systems (and in 3D to boot). This technology will help governments correct and densify control networks, and begin to unify horizontal and vertical control, so that the twain shall finally meet. It will also create a resurgence of private control networks (surveyors have always set out secret monuments for their own use). So, when a GIS needs to establish whether similar primitives from separate sources are actually the same, it will be able to guess more accurately, but it still won't know if they are based on the same evidence or not. This is not progress.

Some words from, our sponsor: I issued most of the above complaints once before, at the 1984 ACSM/ASP Spring meeting in Washington. Five years later, it still has to be said: Better coordinates do not make better models; they may reduce certain types of

uncertainty, but do not dispell confusion about what exists The approach I then advocated, geodesic planetary tessellation, has yet to be implemented (I've had to do other things until recently, and nobody else seems to have championed the cause in the meanwhile). So far as I know, no better approach has come along, and the power of this paradigm remains largely misunderstood and the usual reaction to geodesic planetary modelling remains something like "It sounds like a neat idea for handling global datasets, but it doesn't seem to relate to garden variety GIS applications." Yes and no: it is well-suited for sorting out facts at global and continental scales, but its real leverage comes from its capacity to identify locations (places where things exist) by name (i.e., geocode), at any appropriate scale, uniquely and with certifiable accuracy. Isn't that special.

Rather than floating unanchored, like a grid of Landsat pixels (which are never the same twice), the cells of a geodesic tessellation are specific, simple polygons which tile together perfectly and subdivide regularly. Such structures can be described with a few parameters to fully specify the ordering of vertices, edges and faces, and fix their locations on the surface of a spheroid (or even a geoid). Geodesic hierarchies can grow at various rates; if each facet blossoms into four, the data can be managed as a quadtree. If one knows the geocode of a tile where something is located, latitude and longitude can be derived, from which map coordinates may be projected. Larger tiles (shorter geocodes) have fuzzier locations than smaller ones (longer geocodes), and thus have less positional certainty. This property can help one distinguish boulders and fenceposts as monuments. More progress.

The number of useful ways in which spheres may be geodesically tessellated is not great, but is large enough to engender confusion about how best to do it. In all cues, however, the process is rooted in an initial regular polyhedron, each facet of which is divided into tiles in some uniform way. Virtually all procedures for doing this generate tiles which are triangular (although square ones may be had if one is sufficiently perverse). After a few tessellations, the polyhedra grow quite spherical, and things start to fall into place. Identifying the location a thing falls into is entirely a function of the initial polyhedron selected, the method of subdividing it, and the way in which elements of the resultant tessellation are numbered.

We have progressed to the point of identifying a geodesic location model that seems to possess a number of desirable properties and few undesirable ones. It starts out as an octahedron, which is superposed on a planet with its six vertices and three axes aligned to cardinal points. Each of its eight facets is then recursively subdivided into four triangular tiles by connecting the midpoints of its three edges. This set of vertices, edges and facets constitutes a tessellation we refer to as a Quaternary Triangular Mesh (QThO). This deceptively simple structure provides a rich matrix of locations that can represent the details of objects far better than coordinates now do.

Each point fixed by a GFS or terrestrial survey can be mapped to a QTM grid node or facet in a few milliseconds (actual times may vary, but should be 'proportional the log of linear resolution achieved). Likewise, digitized maps (if coordinates are referenced to the globe and there is evidence of their locational certainty) can be converted to QTM representation (using as yet unspecified hierarchical data structures). Once this is done, each will have a well-defined extent that is part and parcel of its location. Should more precise points have to be fixed within the area of influence of any such location, each can be similarly refined, appending its digits to the location's geocode. While we have determined a number of interesting properties for QTM, we are sure that many more remain to be discovered. We welcome your reactions to this analysis, and hope we can work together to shape better ways to represent, manage and utilize spatial information.

**POSITION STATEMENT**  
**NCGIA INITIATIVE 3'MULTIPLE REPRESENTATIONS'**  
**BUFFALO, 18-21 FEBRUARY.1989**

**A. STEWART FOTHERINGHAM      SUNY BUFFALO**

My interest in the area of multiple representations- concerns the degree to which various distributions exhibit properties such as self-similarity, self- affinity and scale dependence. In conjunction with Mike Batty and Paul Longley from the University of Wales, Cardiff I have recently investigated various aspects of the growth and representation of urban structures. On a large-scale map urban areas are usually represented as having a uniform density whereas on smaller scale maps they are seen increasingly as the complex mixtures of developed and undeveloped land they are. We have been concerned with how well a process known as DIFFUSION LIMITED AGGREGATION produces structures with this and other properties of urban areas.

The type of scale-dependence described above Will be present in any spatial distribution which becomes less concentrated towards the margins. Indeed, the fractal dimension of a spatial distribution, a measure of spatial decay, can be used to convey information on the degree to which a distribution exhibits scale dependence. Following this line of reasoning, it may be possible to characterise the boundary between two distributions. A very sharp boundary, say between water and land, will occur whenever the two distributions have fractal dimensions close to 2.0 and where there is very little spatial decay. Increasingly fuzzy boundaries will occur, such as between various soil types, as the two distributions exhibit increasingly lower fractal, dimensions and greater spatial decay.

To date in this research, our concentration has been more on the properties of simulated fractal distributions than on empirical applications, although a rather brief examination of the digitised urban form of Taunton in South West England proved promising. We see the following avenues of research in this area as worth pursuing:

To what extent can the properties of self-similarity, scale dependence and spatial decay be replicated by diffusion limited aggregation in urban areas ?

To what extent can the fractal dimension of a spatial distribution be used as a useful description of the degree to which that distribution exhibits spatial decay ?

Can these ideas be transferred to generate an indicator of the sharpness or fuzziness of boundaries ?

# Database Support for Multiple Representations

Oliver Gunther  
Department of Computer Science  
University of California, Santa Barbara, CA 93106  
January 1989

Modern database systems are no longer limited to business applications. Non-standard applications such as CAD/CAM or geographic data processing are becoming increasingly important, and geometric data (such as points, lines, polygons, polyhedra, or splines) play a crucial role in these areas. In many of these applications, custom-made file systems are still preferred over off-the-shelf database systems for reasons of both functionality and performance. These file systems often fail, however, if the amount of data to be managed becomes very large, if multiuser access is desired, or if a system crash occurs. Database systems, on the other hand, are known for their ability to scale up well to manage very large amounts of data. They provide concurrency control techniques to allow multiuser access and recovery mechanisms to deal with system crashes. Other useful features include a set-oriented relational query interface, normalization techniques for redundancy avoidance, query optimization and indexing techniques to speed up query processing, data independence, and a well-defined data model. We therefore believe that database systems with increased functionality and performance to handle non-standard applications will remain a research topic of crucial importance.

Over the past few years, it has become clear that a vast array of special-purpose database systems is not the best way to meet these challenges. Instead, it has been proposed to develop extensible core systems which can be adapted by the user to support a given application. Examples include the extensible relational database systems EXODUS [DeWi84, Can86], POSTGRES (Ston86), PROBE [Daya85], GENESIS [Bato86], STARBURST [Lind87], DASDBS (Paul87), or XRS (Meie87).

To provide optimal support for non-standard applications it is important to choose efficient schemes to represent the data. This is especially true for applications that involve geometric data. A representation scheme is the mapping of the original data objects into a set of objects that are convenient to store and that facilitate the computation of a particular class of operators. Research on representation schemes for geometric data has produced a wide array of data structures to support various classes of operators.

Some of these representations are indices which organize large sets of geometric data objects (Gun84, Niev84, Sell87, Gunt89). They can often be used as database indices, in analogy to Bayer's B-tree [Baye72]. Many extensible relational database systems allow the user to define and embed indices which specifically support certain non-standard applications. These indices are usually used in connection with abstract data types and are therefore termed abstract indices [Ston83]. The construction of an index may require a considerable amount of computation. Once constructed, however, an index may represent the underlying data in such a way that search operators can be computed very efficiently. Any update to the database may cause the index representation to change as well, which brings up the need for efficient index update algorithms. One may also maintain more than one index structure for the same database, each supporting different kinds of search queries. In this case, query optimization becomes somewhat more complicated because the optimizer must take into account various index structures, estimate their performance with respect to the given query, and select the most promising candidate. We are currently working on the design of such an automated selection mechanism, which involves a major comparative study of several geometric index structures.

Other geometric data structures are intended to represent one particular data object, such that certain operators can be computed on these objects in an efficient manner [Marr80, Broo81, Gunt88]. We are interested in how one may embed these structures into a database environment and how databases can handle the increased internal complexity of the stored objects. Extensible database systems such as POSTGRES (Ston86) offer various possibilities to embed complex representations, such as abstract data types (ADT), non-first normal form (NF2) relations, or procedural data types, and a comparison of these mechanisms seems a promising area of further research.

In a numerical computing environment, it is often sufficient to maintain only one representation of the data. In geometric computing, on the other hand, it is often necessary to store multiple representations of the same data in order to facilitate the efficient computation of a great variety of geometric operators. Multiple representations cause a significant overhead to ensure availability and consistency of the data, and it is yet another subject of future research to see how extensible database management systems can be used efficiently in this context.

One way of supporting multiple representations in databases are database views. Here, some representation is declared the main representation and stored explicitly in the database. Other representations are views of this main representation: they are derived from the main representation whenever needed. In order to improve the availability of these representations, the views may be precomputed and stored as well. The consistency of these representations is then monitored by demons that invalidate or update a



precomputed view if necessary (i.e. if the corresponding main representation changes) [Blak86]. Conversion routines and rules for view updates can be stored directly in the database using procedural data types-

## References

- [Bato86] Batory, D., GENESIS: A reconfigurable database management system, University of Texas at Austin Technical Report TR-86-07, 1986.
- [Baye72] Bayer, R. and McCreight, E. M., Organization and maintenance of large ordered indices, *Acta Informatica* 1, 3 (1972). pages 1-21.
- [Blak86] Blakeley, J. A, Larson, P.A., and Tompa, F. Wm., Efficiently updating materialized views, in *Proc. Of ACM SIGMOD Conference on Management of Data*, Washington, DC, June 1986.
- [Broo81] Brooks, R. A.. Symbolic reasoning among 3-D models and 2-D images, *Artificial Intelligence* 17 (1981), pages 285-348.
- [Care86] Carey, M., DeWitt, D. J., Richardson, J., and Shekita, E., Object and file management in the EXODUS extensible database system, in *Proc. 12th International Conference on Very Large Data Bases*, Kyoto, Japan, Aug. 1986.
- [Daya85] Dayal. U., Buchmann, A., Goldhirsch, D., Heiler, S., Manola, F., Orenstein, J., and Rosenthal, A., PROBE - A research project in knowledge-oriented database systems: preliminary analysis, Computer Corporation of America Technical Report CCA-85-03, July 1985.
- [DeWi84] DeWitt, D. J. and Carty, M. J., Extensible database systems, in *Proc. 1st International Workshop on Expert Database Systems*, Kiawah, S.C., Oct. 1984.
- [Gunt88] Gunther, O., Efficient structures for geometric data management, *Lecture Notes in Computer Science* No. 337, Springer-Verlag, Berlin, 1988.
- [Gunt89] Gunther, O., The cell tree: An object-oriented index structure for geometric databases. in *Proc. IEEE Fifth International Conference on Data Engineering*. Los Angeles, Ca, to appear, Feb. 1989.
- [Gutt84] Guttman, A, R-trees: A dynamic index structure: for spatial searching, in *Proc. of ACM SIGMOD Conference on Management of Data*, Boston, Ma, June 1984.
- [Lind87] Lindsay, B, McPherson, J., and Pirahesh, H, A data management extension architecture, in *Proc. of ACM SIGMOD Conference on Management of Data*, San Francisco, Ca., June 1987.
- [Marr80] Marr, D., *Vision*, W. H. Freeman & Co., San Francisco, Ca., 1980.
- [Meie87] Meier, A.; *Erweiterung relationaler Datenbanksysteme für technische Anwendungen*, Informatik-Fachberichte No. 135, Springer, Berlin, 1987.
- [Nie84] Nitevergelt, J., Hinterberger, H. and Sevcik, K. C., The grid file: An adaptable, symmetric multikey file structure, *ACM Trans. on Database Systems* 9, 1 (March 1984), pages 38-71.
- [Paul87] Paul, H.-B., Schek, H.J., Scholl, M. H., Weikum, G, and Deppisch, U., Architecture and implementation of the Darmstadt database kernel system (DASDES), in *Proc. of ACM SIGMOD Conference on Management of Data*, San Francisco, Ca., May 1987.
- [Sell87] Sellis, T, Roussopoulos. N, and Faloutsos, C, The R+-tree: A dynamic index for multidimensional objects, in *Proc. 13th International Conference on Very Large Data Bases*, Brighton. England. Sept. 1987.
- [Ston83] Stonebraker, M, Rubenstein, B, and Guttman A, Application of abstract data types and abstract indices to CAD data. in *Proc. Engineering Applications Stream of ACM SIGMOD Conference*, San Jose, Ca., May 1983.

[Ston86] Stonebraker, M. and Rowe, L., The design of POSTGRES, in Proc. of ACM SIGMOD Conference on Management of Data, Washington, DC, June 1986.

# **Research Topics on Multiple Representations**

## **Position Paper for NCGIA Initiative 3**

Bud P. Bruegger,                      bud@mecan1.bitnet  
Renato Barrera                      renato@mecan1.bitnet  
Andrew U. FTank,                      frank@mecan1.bitnet  
Kate Beard,                      beard@mecan1.bitnet  
Manfred Eblers                      ehlers@mecan1.bitnet  
National Center for Geographic Information and Analysis  
and  
Department of Surveying Engineering  
University of Maine  
120 Boardman Hall  
Orono, ME 04469

### **Abstract**

Initiative 3 of the National Center for Geographic Information and Analysis (NCGIA) deals with multiple representations. Research in the NCGIA is designed to investigate impediments to the more widespread implementation of geographic information system (GIS). Multiple representations are one of the key problems in current and even more so in future GIS. Specifically, the lack of understanding on how to deal with multiple representations in the same system is a major impediment to the development of multipurpose GIS. This paper serves as a basis for discussion in the initial specialist meeting of initiative 3. Goals of the initiative and research methods to be applied are briefly discussed. The main chapter of the paper gives an overview of multiple representations and research problems in this field. Based on this understanding of multiple representations and a study of related initiatives, a partial research agenda for initiative 3 and constraints on the scope are proposed.

### **1. Introduction**

#### **1.1 Purpose**

Initiative 3 of the National Center for Geographic Information and Analysis (NCGIA) deals with multiple representations. This paper gives an overview of problems related to multiple representations and proposes part of the research agenda for initiative 3. It serves as a basis for discussion in the initial specialist meeting and proposes research to be done in the working groups of Initiative 3.

#### **1.2 Scope**

Research in the NCGIA is designed to investigate impediments to the more widespread implementation of geographic information systems (GIS) [Abler 1987]. It is organized in twelve research initiatives planned to start during the first three years of the Center's operation [NCGIA 1988]. The concerns of several initiatives are closely related. It is therefore necessary to limit the scope of the initiatives to prevent duplicated research. This paper addresses the related initiatives and suggests a demarcation between them and initiative 3. In-progress seminars, and national and international conferences will guarantee the eventual integration of the findings of the different

#### **1.3 Definition**

The term multiple representation expresses the idea that several models of the same real world objects co-exist within a single system. For example, Penobscot County is represented several times, in different scales in order to meet the needs of different agencies using the System. The use of several representations always introduces redundancy. Typically, the representations capture some of the same aspects of the objects. However, the concepts of representation used in the different representations can vary in several aspects as will be described later. Future GIS will typically have to deal with several representations of the same objects in order to be flexible and satisfy all the user needs.

## 1.4 Importance

Multiple representations are one of the key problems in current and even more so in future GIS. The lack of understanding on how to deal with multiple representations in the same system is a major impediment for the development of multipurpose GIS. They will always have to deal with multiple representations, as different disciplines and applications model reality differently. Even if the system includes only a single physical representation of the data, it deals with several conceptual representation while interacting with different users or different parts of the software system. The following example demonstrates how different user groups model the same object differently. The greatly differing concept of road width among the user groups for example, demonstrates the difference in the conceptual models: A GIS designed to organize information about roads offers services to surveyors, civil engineers, transportation companies, the fire department, and other user groups. The surveyor sees the road as a two dimensional object determined by its legal boundary. The civil engineer perceives the same road in three dimensions including materials used, the drainage system, dynamics of vehicles using the road, etc. For transportation the only matter of interest is which points the road connects and how much time it takes to travel between these two points. The fire department is particularly interested in whether the road can be used by its fire trucks.

The following examples show common problems encountered in GIS, which involve several representations of the same objects in the same system:

- integration of data from different sources
- exchange of data between systems and parts of systems
- cartographic generalization<sup>1</sup>
- multi-resolution data bases

## 1.5 Recommendations

Research should result in a general formal theory dealing with multiple representations. This will provide solutions to several problems of GIS design. Of particular interest is research on consistency between representations and how to do modifications without losing it. For example, propagation of changes to other representations fall in this arm. Another central issue is the access of information. Several representations can contain the data from which the information requested can be derived. The system has to decide which representations can most efficiently provide the information requested in a query. Further, the logical design of multi representation systems is of concern. Which kind of representations should be included in a system tailored to a certain purpose?

We recommend that research concentrates on the following aspects of representation:

- variations in spatial resolution
- variations in abstraction level of spatial objects
- variation in concepts of space

Based on a study of overlaps with other initiatives, we recommend not dealing with display aspects in initiative 3. The initiative should concentrate on conceptual aspects rather than dealing with cognitive and psychological questions. Further, aspects of representation such as temporal abstraction, inaccuracy and uncertainty can most effectively be studied in other initiatives.

## 1.6 Organization

Chapter 2 discusses the goals of initiative 3 and the research methods that should be applied. Chapter 3 gives an overview of multiple representations and research problems in this field. Based on this understanding of multiple representations and a study of related initiatives, chapter 4 proposes a partial research agenda for initiative 3 and limitations of the scope.

## 2. Goals and Research Method

### 2.1 Theoretical Understanding as Primary Goal

The primary goal of initiative 3 shall be the theoretical understanding of multiple representations. Research should deal with multiple representations on a conceptual level before addressing implementation aspects. This allows the treatment of representations with the same theory irrelevant of their physical implementation.

---

<sup>1</sup> As will be shown later, the part of cartographic generalization dealing with the display of objects is excluded here.

## 2.2 Need for Formal Methods

Results of research in the field of multiple representations should be expressed in a formal way, to guarantee an easy integration of findings within the initiative and with other initiatives. Further, formal methods from related fields such as geometry or temporal logic can easily be applied to multiple representations. The formal approach further allows the proof of theorems in the theory and avoids ambiguity in communications.

Research in this initiative is expected to lead to computer-based solutions to problems of multiple representations in GIS. As computers are formal systems for symbol manipulation, a formal approach guarantees an easy translation of the results to computer implementations.

## 3. Multiple Representations

This chapter presents some terminology and context necessary for the understanding of the proposed partial research agenda.

### 3.1 Different Concepts of Representation

In the following, a classification of representations is proposed. It is based on different representation concepts used. This classification is the first step towards a formal description of representations, as well as the differences and relations between representations.

Representations will be classified according to the following criteria:

- spatial resolution
- inaccuracy<sup>2</sup>
- uncertainty
- level of abstraction of spatial objects
- concept of space
- temporal abstraction

The list is not necessarily exhaustive. The discussion of these concepts is based on the idea of time variant spatial objects that are described by their geometry and their non-spatial properties.

• **Spatial Resolution**<sup>3</sup>: The geometry of spatial objects in reality is infinitely detailed. During data acquisition, the real geometry has to be mapped to a discrete representation. This limits the parameters describing geometry to a finite number and thus, the representation is only of finite resolution. Cartographic generalization can further reduce the resolution in a set of data. The formal definition of certain geographic features is resolution dependent [Frank 1986].

• **Inaccuracy and Uncertainty of Spatial Data**: Both the geometry and the nonspatial attributes are normally not absolutely accurate and certain. 'Inaccuracy' corresponds to the term 'type I uncertainty' proposed in [Robinson 1985]. It describes the uncertainty of data describing a conceptually exact characteristic or event. Examples of inaccurate data are measurements of geometry, which are always affected by the noise of the measuring system.

'Uncertainty' corresponds to the term 'type II uncertainty' proposed in the same paper. In this case, the concept to be represented by data is intrinsically uncertain. An example is the boundary between soil types, which is not an exactly defined line.

• **Level of Abstraction of the Spatial Objects**: Abstraction mechanisms known from semantic modelling [Peckham1988] [Brodie 1984] can be applied to spatial objects. Using classification, generalization, aggregation and association allows views of spatial objects on different levels of abstraction. For example, the association mechanism permits the view of a certain spatial object as a set of parcels or a community. Multiple representations can differ in the level of abstraction of objects. The abstraction mechanism that relates objects on adjacent levels determines the relations between both the geometry and the non-spatial attributes of the higher- and lower-level objects.

---

<sup>2</sup> Italic writing style indicates that the exclusion of these concepts of representation from initiative 3 is recommended. The topic should be addressed in related initiatives.

<sup>3</sup> Spatial resolution is only defined in metric spaces and thus not applicable to every concept of space.

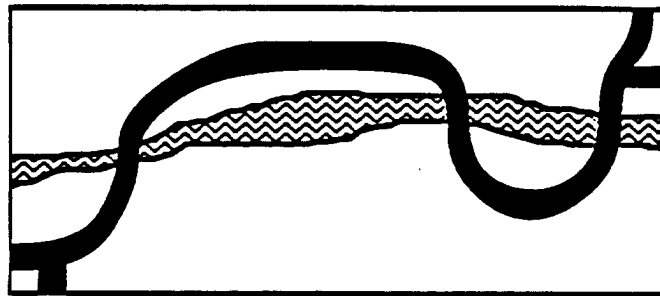
- **Concept of Space Used:** Several concepts of space can be used to express the spatial properties of real world objects. Two common concepts are raster and vector spaces. Other concepts are topological spaces, connectivity graphs, etc. A multiple representation system can organize the same objects using several concepts of space, either because of the characteristics of the data sources, or because several concepts have been included to satisfy different requirements.

- **Temporal Abstraction:** In reality, geometry and non-spatial properties of objects vary with time. Several concepts of time are known today, in which temporal properties of objects can be expressed [Zeigler 1984][Semadas 1980]. Time can be continuous or discrete, and changes can be modelled as events or continuous changes. Representations can either be snapshots of the real world situation, or they can express an average state over a certain interval of time. Information systems dealing with dynamic properties of the world will always have to use multiple representations of objects.

### 3.2 Example: Cartographic Generalization

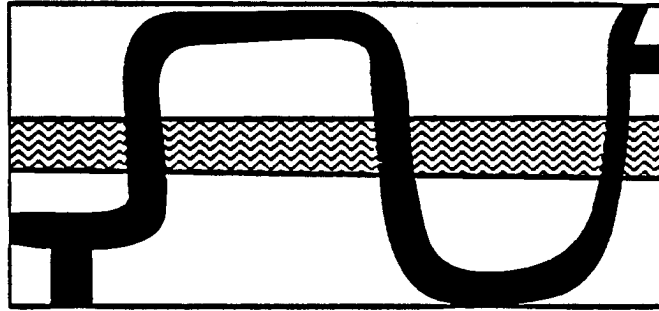
Three representations of the same real world objects will be described in the following example. They can be interpreted as representations related by a process showing some aspects of Cartographic generalization. Thus, representations 2 and 3 can be derived from representation 1 by a reduction of information. All three representations describe a road between two junctions, crossing a river several times. The three representations differ only in the concept of space used.

Representation 1 models both the river and the road as polygons in a vector space. The metrics of their boundaries correspond to the metrics of the road and the river in the real world—apart from the discretization of the boundaries and thus a limitation of the spatial resolution. Figure 1 visualizes this representation.



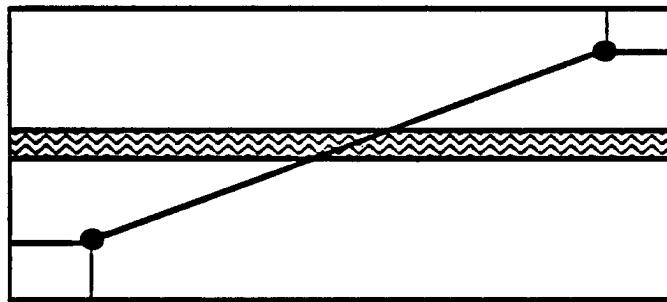
**Figure 1: Visualization of a representation in a metric concept of space.**

Representation 2 again models road and river as polygons, however this time represented using a topological concept of space. No metrics are contained in the representation. The topology corresponds to the real world situation. It can be derived from the information of representation 1. Obviously this representation contains less information than the first. Figure 2 shows a visualization of representation 2. The metrics used for this display do not correspond to the real world situation, but show major displacements compared to the visualization of the first representation.



**Figure 2: Visualization of a representation with a topological concept of space.**

Representation 3 models the same situation based on a graph theoretic concept of space. Figure 3 shows its visualization. Neither the metrics nor the topology of the visualization correspond to the real world situation. Compared to figure 1, displacements and simplifications of the topology occurred.



**Figure 3: Visualization of a representation with a graph theoretic concept of space.**

This example demonstrates that cartographic generalization can involve different concepts of space. The generalization problem can be expressed in terms of multiple representations, where different concepts of representations are used for different degrees of generalization.

### **3.3 Problems in Multiple Representation Systems**

Three groups of problems can be identified in multiple representation systems:

- Logical Design
- Consistency
- Access of Information

#### **Logical Design**

Some of the representations in an CIS are given by the characteristics of the data sources available. Most internal representations, however, can freely be chosen during the design of a system. This design decision addresses the number of representations and their concept of representation. In this case, logical design answers the question of which representations to include to tailor the system to a certain purpose. Multi-resolution data bases are a typical example of an application where logical design is crucial.

## **Consistency**

The multiple representation of objects often introduces redundancy in the system. Thus, a central concern must be to keep the system in a consistent state.

From a static point of view, a system must be able to detect inconsistencies between representations. An application where such tests are required is the integration of data from different sources.

The dynamic aspect of consistency addresses the problem of how to do modifications and updates without introducing inconsistencies. Protocols for modifications and the automatic propagation of a change from one representations to all others have to be considered. A typical application dealing with these dynamic aspects is a GIS organizing representations of different degree of cartographic generalization which have to be kept up to date.

A thorough treatment of consistency problems relies on formal descriptions of the concepts used in a representation, and of relations between different representations. The translation of data from one representation to another one seems to request more features than just a test for inconsistencies. It is not yet clear, however, whether general solutions can be found in all cases.

## **Access of Information**

In multiple representation system, often the decision has to be taken which representation to access in order to answer a certain question. Different representations provide different information as answers. In a multi-resolution system, for example, the answers can differ in their spatial resolution. The method and cost of access depend on the representation chosen.

The information extracted must satisfy certain minimal requirements (e.g. spatial resolution, accuracy), which can be expressed in terms of a concept of representation. The system tries to minimize response time under the conditions of meeting the requirements. Such an optimization process is crucial in query languages which involve spatial reasoning. Without optimization the response times would be unacceptable.

A possible approach to the problem of information access is to impose an order on the representations of the system. A possible structures to express such an order is the partially ordered sets, with hierarchies and lattices as special cases. It is not clear, whether an order relation for the general case can be found. Maybe it is only applicable to specific sets of representations.

## **3.4 Exclusion of Display Aspects**

The scope of this paper is limited to conceptual aspects of representations, thus excluding display aspects. This limits the complexity of the problems considerably. We believe that solutions, including the display of representations, must be based on sound conceptual solutions. Solutions to displaying problems can be added later without affecting the conceptual level. Problem of displaying representations are not specific to multiple representations. They occur all over in GIS as well as in other fields, such as CAD applications.

If only conceptual aspects of representations are considered, all the data reflect properties of the corresponding real world objects. On the other hand, visualizations and displays of such representations often show differences to the real world properties. This is caused by the limitations of the display media and the space available. Typical examples are displacements and exaggerations known from cartographic generalization. Displays often show non-geometric features as geometric objects. Name tags or bar graphics known from thematic cartography are good examples. These objects compete for display space with the 'real' geometric objects and thus increase problems caused by the limitations of the display.

The example of cartographic generalization demonstrates how the complexity of problems decreases significantly if the display aspects are disregarded in the first (conceptual) step of the solution. The example in paragraph 3.2 already showed how cartographic generalization can be modelled by a series of representations using different concepts of representation. On this conceptual level, cartographic generalization is nothing but a reduction of information. A second step deals with the display: dependent on the limitations of the display media, a selection of objects and the concept of their representation has to be chosen. Different decisions can be taken for different locations. After adding the non-geometric objects, displacements and exaggerations have to take place.



### 3.5 Implicit Information

In a multiple-representation system, the same property can be only implicitly contained in some representations, while others explicitly model it. A fjord, for example, could be implicitly represented in the geometry of the coastline or explicitly by some data expressing its existence. Human interpreters can easily cope with both kinds of representations. Formal systems, however, do not 'recognize' an implicit object unless they are taught to do so.

The problem of implicit information occurs when dealing with consistency in multiple representations. Comparisons of representations and transformations between them can only deal with explicit data. For example, two representations are related by a transformation that simplifies the geometry. If a fjord is implicitly contained in the geometry, it is not clear whether the fjord will still be there after applying the generalizing transformation, whether it will disappear, or whether an estuary-like coastline will appear instead.

Besides implicit representations of objects, there are implicitly-contained relationships between objects. Typical relationships are 'parallel', 'intersecting in a right or pointed angle', or 'homogeneously oriented'. In many cases, transformations between different representations should preserve these relationships.

## 4. Proposed Research

### 4.1 Related Initiatives

The research topics of initiative 3 partially overlaps with those of several other initiatives. This paragraph briefly mentions initiatives related to initiative 3 and comments on their kind of relation. Some aspects of multiple representations will be dealt with in other initiatives. This leaves the possibility of concentrating research efforts on other central problems of multiple representations. However, the topics dealt with in other initiatives should not be left out of consideration.

One possible way of treating overlapping research topics is to consider some initiatives clients of other initiatives. The service initiative receives formulations of the problem from client initiatives, reflecting their specific view of the problem and context. Further coordination with client initiatives will, be necessary. The client initiatives will have to integrate the findings of the service initiatives in a late stage.

A list of the initiatives related to initiative 3 follows:

- i1: This initiative deals with inaccuracy and uncertainty. These are also aspects of concepts of representation. Initiative 3 could be a client of initiative I in this respect.
- i2: This initiative deals with among other topics mathematical concepts of space and with formal languages to express representations. Results in these fields can be used while formally expressing results of initiative 3.
- i7: This initiative deals with visualization of quality of spatial information. A part of the display aspects (recommended to be excluded from initiative 3) could be treated here.
- i8: The goal of this initiative is to develop an expert system to design car graphic displays. All the display aspects occluded from initiative 3 are recommended to be treated in this initiative. Together with the results of initiative 3, this initiative should solve problems of cartographic generalization.
- il0: This initiative deals with models of time. The temporal aspects of concepts of representation are recommended to be dealt with in this initiative.
- i11: This initiative deals with space-time statistical models and one of its objectives is to deal with multiple representations in the time domain. As this initiative is planned to start only in October 1990, it would be an ideal forum for the integration of findings of earlier initiatives in a consistent whole.

### 4.2 Research Topics recommend addressing the following research topics in initiative 3:

- **Framework for multiple representations:**  
A framework is needed to facilitate the coordination of efforts within the initiative. In an early stage, this might just be an agreement on terminology. In the end stage the framework should consist of a formal theory of concepts of
- **Logical design of multiple representation systems:**

Which representations should be included in a system tailored to a specific purpose? This research topic includes order structures to organize representations such as lattices, hierarchies and partially ordered sets.

- **Consistency in multiple representation systems:**  
This research topic should include static and dynamic aspects of consistency. Propagation of updates and translations of data from one concept of representation to another are of particular interest.
- **Access of information in multiple representation systems:**  
Which of the representations should be accessed in order to answer certain questions?  
This research topic is closely related to the logical design and order structures for representations.
- **Methods to deal with implicit information:**  
Methods to make implicit information explicit need to be found in order to deal with implicit information in multiple representation systems. Feature extraction is well known in raster representations but far less common in vector space.
- **Special applications of multiple representations:**  
The findings of the initiative should be applied to some of the common applications of multiple representations, e.g. cartographic generalization.

#### **4.3 Limitation of the Scope of Research**

Based on the study of related initiative, the following limitations of the scope of research are proposed.

- Cognitive and psychological aspects should be considered secondary.
- Research on concepts of representation should not go into details of inaccuracy, uncertainty or temporal abstractions.
- Display aspects should be disregarded.

#### **References**

- [Abler 1987] Ronald Abler. The National Science Foundation National Center for Geographic Information and Analysis. International Journal of Geographic Information Systems, Vol. 1, No. 4, 1987.
- [Brodie 1984] Michael L Brodie and Dzenan Ridjanovic. On the Design and Specification of Database Transactions. In: Michael L Brodie, John Mylopoulos and Joachim W. Schmidt On Conceptual Modelling. Springer-Verlag. New York, 1984.
- [Frank 1986] Andrew U. Frank, Bruce Palmer and Vincent B. Robinson. Formal Methods for Accurate Definitions of Some Fundamental Term in Physical Geography. Invited paper at the Second International Symposium on Spatial Data Handling, Seattle (Washington), July 5-10, 1986.
- [NCGIA 1988] The National Center for Geographic Information and Analysis. Summary of Research Plan. University of California - Santa Barbara, State University of New York - Buffalo, University of Maine, September 1988.
- [Peckham 1988] Joan Peckham and Fred Maryanski. Semantic Data Models. ACM Computing Surveys, Vol. 20, No. 3, September 1988.
- [Robinson 1985] Vincent B. Robinson and Andrew U. Frank. About Different Kinds of uncertainty in Collections of Spatial Data. In: Auto Carto 7 Proceedings, Washington (D.C.), 1985.
- [Sernadas 1980] Amílcar Sernadas. Temporal Aspects of Logical Procedure Definition. Inform System, Vol 5, pp. 167-187, 1980.
- [Zeigler 1984] Bernard P. Zeigler. Multifaceted Modelling and Discrete Event Simulation Academic Press, New York, 1984.

# **MULTIPLE VIEWS OF MULTIPLE REPRESENTATIONS**

## **Position paper for the Specialist Meeting for NCGIA Research Initiative #3 "Multiple Representations"**

**David M. Mark  
NCGIA, SUNY at Buffalo**

This research initiative on "Multiple Representations" itself has many views or components. In this brief position paper, I will introduce and develop three themes which I believe to be central to the research agenda on this topic.

### **1. Scale-dependent Geometry**

It is well-established in biology that the forms of organisms are in part dependent upon their sizes or characteristic scales. The principle of allometric growth, as discussed by D'Arcy Thompson, Steven Jay Gould, and many, many others, predicts that in many situations, form (or shape) must change with changing size in order to preserve function. Such scale-dependent effects are termed allometry, in contradistinction to isometry, in which shape would be size-independent. As an example, the bearing strength of mammalian limbs is approximately proportional to the square of the cross-sectional area of the limb, whereas the weight of an organism (of constant shape) increases in proportion to the cube of the length or height. Preservation of function (in this case, bearing strength of the limb) requires the legs of large animals to be relatively thicker than those of small animals of the same group. Effects such as this are common, and isometry is rare. Similar scale-dependent aspects of form have been identified for drainage basins (cf. Woldenberg), cities (cf. Nordbeck, Dutton), and many other systems.

Although perhaps less often recognized, it is also usually the case that the form of a particular object varies with the scale of observation, or the resolution of measurement. Mandelbrot's fractals model of statistically self-similar or self-affine forms is controversial, and may not exist in nature. Nevertheless, the fractals approach has drawn attention to the fact that measured shape and form do depend on measurement scale. The -scale-dependent nature of the geometry of a "true fractal" could be represented by a single parameter, termed the fractal dimension; however, geographic phenomena depart from such fractals in varying degrees. Classification of geographic features according to whether their geometry is approximately fractal, scale-dependent, scale-free, or variable is critical to the development of effective methods to handle geographic data.

### **2. Cartographic Generalization**

Generalization is one of the fundamental processes in the design and production of maps. If a map is reduced in size (and scale) using photographic techniques, detail often becomes crowded or confused. This is especially true when the map is redrafted using a pen- or linewidth similar to that on the original map. Cartographic generalization requires the selection of those features that are to be shown on the reduced scale map, and the simplification of the forms of the features to be retained.

Manual cartographic generalization has traditionally relied upon the skill, judgement, and knowledge of a trained cartographer; it has been motivated by a need to preserve graphic clarity in a static, printed medium. In computer handling of geographic data, however, 'generalization' often is motivated by a need for data compression. This sort of "statistical generalization" attempts to preserve positional accuracy while reducing data volumes. Generalization to reduce data volume can be avoided if sufficient computer storage and processing power is available. But, for printed maps, the generalization problem must always be addressed in order to achieve or approach clarity.

It seems that the GIS community has implicitly assumed that similar generalization is needed for real-time, on-line map display in a GIS environment. It is perhaps heretical to suggest the alternative: with an on-line GIS, for many applications, perhaps we should just plot everything, allowing features to overlap or merge, producing blobs of clutter in high density areas, and then allow the user to zoom in to resolve the ambiguity. I'm confident that a lively discussion could be generated around this suggestion!

The concepts of scale-dependent or scale-independent geometry were noted above as properties of real-world geographic features. However, few if any of the processes which produce geographic features are scale-independent. Thus, features such as spits, or fjords, or drumlins, occur only in particular sizes or size ranges. Conceptually, most of these features can be considered to be involved in part-whole relations: each is composed of parts that may themselves be landforms of particular named kinds, and each may form a part of a landscape assemblage of a particular type. In many cases, successful cartographic generalization will require that compound features be 'recognized', be 'parsed' into their component features, and be generalized in context-dependent and phenomenon-dependent ways.

Realistically, a complete solution to cartographic generalization, for all types of cartographic features, for all mapping situations, for all users, and under all graphic limits imposed by hardware or reproduction techniques, is a very long way off. And without such a solution, a single, "scale-free" cartographic data base to support mapping and analysis over a wide range of scales seems impossible. Thus, for some time to come, we may be forced to deal with explicit multiple representations of cartographic features over a common region. These can best be thought of as map coverages of the same region at different plotting scales, which in fact they probably will be. It is a non-trivial problem to manage such databases. Since the "same" feature may appear in more than one database, there are severe risks that editing or updates will be out of phase among the coverages.

For example, a major through street may appear as a named feature on several maps. If the town changes the street's name, how can we ensure that the name change is made at the same time to the several representations of that street? Similar questions come up if the town itself changes its name. Or if a new highway bypass is constructed. Or if a creek is dammed to produce a reservoir. What we need are links between the various representations of the same feature. A central characteristic of object-oriented programming is that objects communicate, sending 'messages', and possibly 'inheriting' some characteristics from other objects. If we link the several objects, representing the same real-world feature at different scales of observation, in an appropriate way, then perhaps just one would 'have' the feature's name, and the others would 'request' the name when needed. If the name really is stored only once, then any name change would appear to happen concurrently across all map-scale layers in the database. Since at this point I have exhausted my knowledge of object-oriented programming, I await discussion of the practicality of such an approach by more informed participants.

### **3. Multiple Representations in Spatial Inference**

This initiative is not concerned with the fact that different human minds may represent the same geographic region in very different ways. However, the ways in which an individual mind maintains multiple representations of the same region, to be used in different spatial inference tasks, should be of some relevance. (A central theme of NCGIA research initiative #2, 'Languages of Spatial Relations' is that a careful study of how human minds represent spatial phenomena may prove valuable in designing formal computer representations of the same phenomena.) People planning driving routes will usually rely on habitual paths when possible. When using maps, they may ignore most streets, and concentrate only on red (major) roads, or even on freeways. The multiple representations of a city in an Etak Navigator data set have exactly this effect. When the user zooms out to an overview of the city, only major routes are shown; as the user nears their destination and zooms in, minor streets appear. Hierarchical planning, with different levels of detail for different parts of the task, seems to be common in human way-finding. Also, people often reason spatially using a sort of inheritance / part-whole logic which usually works, but which sometimes makes errors, such as those which make "Reno is West of San Diego" counter-intuitive. Natural language studies should also provide clues. For example, in expressions of the form "A is located at B", the preposition "at" acts as an operator to convert a possibly extended geographic feature to a dimensionless point. Coordination of some research topics within 12 and 13 should prove productive.

### **Summary**

The size-dependent nature of the geometry of geographic features is almost universal. This obviously means that, among other things, measured form depends of the resolution of the measurements. In measurements made from maps, the graphic limits and cartographic generalization will also influence results at some scales, but not at others. Cartographic generalization must integrate this with featurespecific, phenomenon-based generalization methods. Multiple representations for spatial inference are also of interest.

# **MULTIPLE REPRESENTATIONS**

**U.S. Bureau of the Census Position**

**January 30, 1989**

The Census Bureau views the problem of multiple representations from three perspectives: 1) updating or adding to an existing GIS data base, 2) GIS data base structure policy and 3) use of GIS data base for different scales of output. The key to the problems associated with multiple representations lies in a good data base policy. Significant research topics related to this problem are associated with creating or adding to an existing data base, and displaying data base contents or responding to non-graphic queries at different scales.

The problems of multiple representations in a GIS data base are avoidable with a proper data base structure policy. The Census Bureau chose to avoid the problems of multiple representation altogether by using a fully integrated topological data base structure, the TIGER File structure. Rather than store various geographic entities in the more traditional manner on separate levels, the Census Bureau vertically integrated all entities into a single level. The result is a structure of primitive units of areas, lines, and points representing all possible combinations of geographic levels. Therefore, if, for example, a particular line is the boundary of a city and also a street, the line exists in the data base only once, and the type of boundary or feature is an attribute attached to the line. Similarly, primitive areas have attributes attached to them to identify the census tract, city, and other geographic units to which the area belongs. The theory underlying the Census Bureau's approach is outlined in recent papers [Saalfeld and Meixler, "Storing, Retrieving and Maintaining Information on Geographic Structures: A Geographic Tabulation Unit Base (GTUB) Approach"]. Thus, we have avoided the problems of multiple representation while at the same time providing a flexible and easily changeable means of creating or adding new geographic entities.

Updating or adding to an existing data base does pose significant problems in multiple representation. The likelihood is low that a source will represent geographic entities the same way that they are represented in the data base. To reduce human intervention in the updating operation, it is important that the same entities in both the source and the GIS be identified automatically even though the entities are represented by different scales, symbology, or nomenclature. The Census Bureau is interested in ways to test for similarity between different representations of geographic entities as evidenced by staff papers. [Saalfeld, "Shape Representation for Linear Features in Automated Cartography"]. A theory of geographic entity identification will lead to vastly improved GIS data base quality because of the ease with which the data base can be updated or new relations added.

Graphic and non-graphic output from a GIS must take into account the scale of the output. There are certain obvious efficiencies to maintaining multiple representations of entities at various output scales: It is more efficient to extract all the county boundaries from a GIS file with a maximum best positional accuracy (MBPA) of 1:24,000 source, and to link and generalize them once for display at 1:5,000,000, than it is to extract, link, and generalize each time a map is needed at the smaller scale. Policy is used at the Census Bureau to avoid problems of maintaining multiple representations. The Census Bureau treats extracts as "snapshots" of the master data structure. Replacement extracts are generated only when changes to the master data base have occurred that might affect the extract. Therefore, problem of multiple representation are taken out of the area of GIS and are more properly dealt with under cartographic generalization and symbolization. Both are critical problems, but they are tied more closely to the operating environment (hardware and software) than to GIS structure.

A similar, and perhaps more GIS-related problem, is one of geographic aggregation. For example, when creating higher-level aggregations of soil types, what percentage of a soil type must be present before it is reported as the type for the aggregated area? Geographic aggregation-"generalized truth"-is important because it helps the human to understand results derived from "detailed truth." A well established theory of geographic generalization is necessary to allow models to run with input from many different-scale data sources.

## **Position Statement for NCGIA Initiative Three:**

### **MULTIPLE REPRESENTATIONS**

**Robert B. McMaster**  
**Department of Geography**  
**Syracuse University**  
**Syracuse, New York 13244-1160**  
**(315) 443-2605**  
**(315) 443-5599**  
**BITNET: MCMaster AT SUNRISE**

A personal perspective on the problem of multiple representations of digital data emphasizes the importance-of generalization. Since it is not possible to maintain digital data at all possible scales for display, the process of cartographic generalization is necessary. within the digital environment of a Geographic Information System a significant, if not the dominant, control on the graphic output is the role and effect of cartographic generalization. Based on this, an obvious concern is how to design, implement, and control generalization operations within a GIS. If the generalization process is viewed as it truly is--subjective, interactive, certainly idiosyncratic in its perception and execution--then we are attempting to implement, in an objective manner, a very complex and ambiguous process. In order to replace at least some portion of the human generalization process with computer algorithms, we must first understand the needs of the generalization process in a digital sense.

#### **The objectives of generalization**

To begin, it is first necessary to understand the objectives of generalization. Precisely, what are we trying to accomplish in numerical cartographic generalization? In a recent paper, McMaster and Shea (1988) have identified three sets of general objectives for the generalization of digital data: [1] philosophical objectives, [2] application objectives, and [3] computational objectives. For instance, from a philosophical perspective, generalization is used to counteract the undesirable consequences of scale reduction since it is not possible, under any circumstances, to reduce a map scale and yet maintain the original level of detail. To guide the generalization process in the digital domain, six philosophical objectives may be identified: [1] reducing complexity, [2] maintaining spatial accuracy, [3] maintaining attribute accuracy, [4] maintaining aesthetic quality, [5] maintaining a logical hierarchy, and [6] consistently applying generalization rules.

For instance, complexity, for the purposes of these objectives, may be defined as a measure of the interaction of various graphic elements within a map. Specifically, the number and/or diversity of these elements within a given area impacts the efficacy with which the mapped information is communicated to the reader. Complexity results, of course, as the scale is reduced and features become cluttered in appearance. Identifying, analyzing, and defining appropriate levels of complexity is perhaps the most difficult problem in generalizing maps in digital mode, for it requires that many of the spatial and attribute transformations be applied either iteratively or simultaneously. in order for successful generalization in a digital environment, we must have a clear understanding of the objectives and how to fulfill the objectives with generalization operators. In a similar manner, we must focus upon the application and computational objectives.

#### **Spatial and Attribute Transformations in Generalization**

The generalization of cartographic features in order to support scale reduction must obviously change the appearance of these objects. Additionally, data sources for map production and GIS applications are typically of variable scales, resolution, and accuracy. Each of these factors contribute to the method in which cartographic information is presented at a given map scale.- The information has two components--location and meaning--and generalization affects both. Specifically, this process involves the application of what are called generalization operators. These are the actual mathematical transformations as applied to both the attribute and geographic data. For instance, a simplification operator will reduce the number of coordinate pairs along a digital line while the amalgamation operator will join two areal features. We must have a clear structure or conceptual framework for these generalization operators and, more importantly, we must have a clear understanding as to how they interact at a variety of scales. These, then, are important questions that must be answered.

- [1] Does a clear conceptual framework for generalization operators exist?
- [2] How do the operators interact? is there a logical consistent sequencing that must be used?
- [3] What is the effect of scale change on the sequencing of the operators.

For instance, a preliminary "conceptual framework" based on the actual algorithms would classify routines as to raster or vector. Raster types of operators would include those used to smooth and enhance the digital types of imagery used in remote sensing. A moving 3 x 3 kernel that removes the high frequency noise is such an example. Another type of raster operator is designed to manipulate non-numerical (categorical) types of information such as a classified land use / land cover image. Monmonier (1984) has focused on such approaches.

Those operators designed for vector data types focus either on the geographical component of the data or the attribute component. For instance, researchers have developed algorithms for point data (omit, aggregate), linear data (simplify, smooth, enhance, displace, merge), and areal data (amalgamate, displace). Other operators classify and symbolize the statistical component of the data. Assuming that such a detailed framework can be established, a next significant problem in numerical generalization is understanding how the generalization operators interact with each other. An even more enigmatic problem is the effect of scale change on the operator sequencing. In summary, I view these as the pressing research issues associated with digital generalization and, ultimately, multiple representations.

### **Sources**

- Bertin, Jacques. (1983). *Semiology of Graphics*. translated by William J. Berg. (Madison, WI: University of Wisconsin Press).
- Coulson, Michael R. C. (1987), "In the Matter of Class Intervals for Choropleth Maps: with Particular Reference to the work of George F. Jenks," *Cartographica*, 24(2): pp. 16-39.
- Eckert, Max (1908), "On the Nature of Maps and Map Logic," Translated by W. Joerg. *Bulletin of the American Geographical Society*, 40(6): pp. 344-351.
- Jensen, John R. (1986). *Introductory Digital Image Processing*. (Englewood Cliffs, N.J.: Prentice-Hall, Inc.).
- McMaster, Robert B. (1987), "Automated Line Generalization," *Cartographica*, 24(2): pp. 74-111.
- McMaster, Robert B. (1988), "Cartographic Generalization in a Digital Environment: A Framework for Implementation in a Geographic Information System," *Proceedings, GIS/LIS Conference*, pp. 240-255.
- Monmonier, Mark S. (1983), "Raster-Mode-Area Generalization for Land Use and Land Cover Maps," *Cartographica*, 20(4): pp. 65-91.
- Shea, K. Stuart and R. B. McMaster. (1989), "Cartographic Generalization in a Digital Environment: When and How to Generalize," forthcoming in *Proceedings of the Ninth International Symposium on Computer-Assisted Cartography (AUTOCARTO-9)*
- Tobler, Waldo. (1966). "Numerical Map Generalization," In John Nystuen. ed. *Michigan Inter-University Community of Mathematical Geographers. Discussion Paper 8*, Ann Arbor: University of Michigan.

## **Position Statement on the Problem of Multiple Representation.**

**Charlie Schwarz  
Office of Charting and Geodetic Services  
National Ocean Service NOAA**

Multiple Representations exist because we know that we will want to represent a data set at a variety of scales. This exists even in the traditional, non-digital world. The National Ocean Service suite of nautical charts is schemed in -an irregular pattern. There are no set series at specific scales. However, the range of scales is large and there are many instances of overlapping charts. The extreme case is Key West, Florida, which is shown on nine different nautical charts, ranging from 1:10,000 to 1:2,160,000 scale.

The existence of map products at a variety of scales is relevant because the production of these products continues to be the most important application of our digital systems. We build our digital systems and data bases to be capable of supporting ad hoc queries and unspecified products, but the truth is that most of the products in automated cartography systems are the products we made before. We see some potential demand for new products, such as digital chart data to support electronic navigation systems, but the present demand for these data sets is not strong and the future demand is uncertain.

In addition to the chart products we need to generate a set of graphical chart indices. Such an index might show the scheme of charts in an area against a very highly generalized geographic background. The most intense use of chart indices is during document evaluation, where we ask the question: "Does this document, describing changes in a certain geographic area, contain anything relevant to any of our charts?" If so, the data base will be updated to reflect changes to the individual features that it describes. Chart indices are also needed at a variety of scales. The geographic background is only used to allow the analyst to place the chart boundary in the real world. Accuracy of representation is not important. Speed of presentation is important.

The problem of multiple representations is that it is difficult to maintain consistency. One of the most important concepts in data base management is that it is preferable to keep only a single copy of a data item, so that consistency is automatically insured. The alternative is controlled redundancy, where one attempts to exercise control procedurally. This is difficult to enforce.

The problem of maintaining consistency across the representations at various scales is also an issue in non-digital nautical cartography. It is important that any feature that shows on a small scale chart also be portrayed on all corresponding larger scale charts. The National Ocean Service attempts to ensure consistency procedurally by a process of "application through the scales." The chart maintenance schedule is developed so that large scale charts are revised before the small scale charts.

Another problem is that we have no guidance as to how many representations should be maintained. Some agencies maintain a separate representation for each product. Others maintain representations at large, medium, and small scales. When (and if) it is necessary to interpolate among the representations this is done with generalization algorithms.

A related issue is that we are unwilling to give up the idea of constructing a truly scale independent data base, which would use only a single representation of all features and thereby ensure consistency. There have certainly been some successes in computerized line and feature generalization procedures, but they do not seem to point the way toward more successes. We look for generalization procedures that are fast enough to be useful, and especially are robust, in the sense that they consistently produce good results, even in the hands of non-professional users. NOS, like many other agencies, has a suite of computer programs for line and feature generalization. These programs are not widely used or trusted, because the results are not reliably good.



## Len Uhr: Some Preliminary Thoughts on Multiple Representations

How can images, icons, symbols, and words best be combined, to inform viewers about spatial entities, as in maps of Earth, Venus, oceans, or sides? How can viewers move around and about the space, moving in and out, changing and comparing different types of representations for different types of information?

Potentially, computers can solve the problems of projecting near spheres or other N-dimensional objects onto 2-dimensional surfaces, like wall-maps or video displays. Rather than rotate a physical globe on one axis, and use magnifying glasses to tease out fine details, the image can be rotated and magnified in any way desired. In addition to displaying the appropriate sequences of images, computers can display many possible representations, and mixtures.

Ideally, all this should be placed under the viewer's Control, So that with minimal effort, in what quickly becomes a natural, unconscious, interaction, the viewer can get whatever mixtures are desired. There should be carefully developed packages for standard displays of different types. And users should be given facilities to develop and store their own packages, and to improvise.

To achieve all this, the following appear to be needed:

A. A data-base that contains:

1. The basic set of units: images, icons, symbols and words, each pointing to all its equivalences.
2. All the rules for combining them, and adjusting them for interactions.

B. Convenient procedures for:

1. Interactively adding to, modifying, and deleting from this stored information.
2. Building a total map, and embedding structures of units within this map.
3. Rotating, magnifying, and translating the whole map or specified parts.
4. Choosing, displaying, and overlaying appropriate representations for appropriate features -sometimes simultaneously (integrated, overlayed, or side-by-side), sometimes sequentially (from split-second switches, to interactively on-demand).

Color, shading and texture, along with different types of symbols and fonts, should be used to make information apparent. When a user chooses a particular combination of icons and symbols, the system should compute how best to handle potential interactions, and display the results along with tools and suggestions for user

If the result must be printed on paper, choices must be made to maximize information and viewability and eliminate interfering clutter, and different maps must be printed for different purposes. If computers in every office and home can be used to display, as well as to construct, maps, then viewers should be helped to zoom blow up, and switch among whatever representations they choose.

All this entails new mixtures of systems for image processing, semantic memory searches, computer-aided design, graphics, and human-computer interactions.

# **MULTIPLE REPRESENTATIONS OF WHAT? or THE FUZZY NATURE OF SPATIAL DATA<sup>1</sup>**

**Robert Weibel**  
**Department of Geography**  
**University of Zurich (Irchel)**  
**Winterthurerstr. 190**  
**CH-8057 Zurich (Switzerland)**  
**phone: 011-41-1-2.57 52 56**  
**e-mail: K491170@CZHRZUIA.bitnet**

## **Scope of Multiple Representations:**

I start explaining my position with a brief description of what my understanding of the scope of multiple representations is (figure 1). I restrict my discussion to multiple representations of spatial data. I will use the diagram shown in figure 1 throughout the text to explain some of my thoughts.

Spatial data may be multiply represented in terms of scale (or resolution) and/or thematic meaning. The scale axis represents an ratio sequence (ranging from "atom" to "galaxy"). The measurement scale of the thematic dimension may be of nominal, ordinal, or ratio/interval character (remark. the size of the sections to show the different measurement scales does not relate to their importance; it just. turned out this way as the diagram evolved). Abstraction in the thematic dimension usually leads from ratio/interval data to ordinal or nominal classifications. Thus, the general direction of abstraction (or generalization) follows a path from top-left to bottom-right on the diagram (i.e. in a small-scale database/map we can only represent ordinal or nominal - i.e. classified - data).

Spatial phenomena may be defined as spatial objects (or classes of objects). The definitions of these objects and their properties change continuously depending on the scale level and on thematic meaning. Per se, objects do not hold discrete states (Le. their is always a certain amount of fuzziness involved). We only assume the existence of discrete states for model simplification.

Objects are transformed in resolution and thematic meaning by processes. Some examples (i.e. aggregation/ disaggregation, symbolization, thematic generalization, modeling natural phenomena, labelling and classification, and change of measurement scale) are shown schematically in figure 1. They will be used for some further explanation later on in this discussion.

## **General Research Questions:**

A specific database (or its graphical representation, Le. a map) has so far been restricted to a particular scale and/or thematic meaning. The vertical crosssection in figure 1 relates to a multivariate database/map of equal scale. The horizontal cross-section corresponds to a multi-scale database (map: perhaps a map with insets of different scale, or an atlas) for a given theme. The intersection of the two cross-sections relates to a univariate database/map at a given scale.

However, on the long run scale-independent and multi-purpose databases will have to be developed. Such abilities are of prime importance to the enhancement of current GIS's for a number of reasons, including the following: data acquisition at different levels of scale is both costly and error-prone; some models require data at different scales, or with multiple attribute definitions; data display has to be appropriate and meaningful relative to scale of the display (as compared to the scale of the data capture and modeling), to name just a few.

In order to develop databases that can hold multiple geometric and thematic representations of spatial data, we will have to solve a multitude of research questions. Clearly, the task of developing a database that spans a significant portion of the domain of representations for spatial data (as shown in Fig. 1) is a monumental 'one. Note that this diagram does not yet include time, which would add another dimension of complexity. It seems to me that solving this problem essentially means solving the fundamental problem of the organization of space ( ... and thus of geography - this is a geographer speaking) and a good deal of philosophical questions. In order to cope with this problem, we will have to identify a few sub-tasks of fundamental importance and first focus on their solution.

---

<sup>1</sup> Position Paper for Workshop to NCCIA Initiative 3 Multiple Representations, held in Buffalo, NY, February 18-21, 1989

To, exemplify just a few of the problems that pertain to the multiple representation of spatial data, I want to briefly discuss the examples of my diagram. Of course I am aware that this selection may be fairly atypical and skewed by my interests.

**Aggregation / Disaggregation (A):** Aggregation is straight-forward for data that form a strict functional hierarchy (e.g. political units). However, in the case of natural resources data, where such hierarchies do rarely exist, aggregation also involves reclassification and is related to thematic generalization. Disaggregation forms the reverse process. Although it is possible to disaggregate data into smaller units, there is substantial error involved in the derived variables. Modeling or display therefore has to take place at a scale level that is appropriate for the particular application.

**Symbolization (B):** Depending on measurement (or display) scale, objects may be represented as point or line data (or symbols). Processes have to be developed to derive adequate representations for a particular scale.

**Thematic Generalization (C):** This process applies to the merging of large-scale objects into classes at smaller scale. Definitions of objects (or classes of objects) are highly scale-dependent, but also context-dependent (not every tree eventually turns into a forest; a meadow becomes part of a forest if located in a generally wooded area). Therefore, it is difficult (if not impossible) to define "typical" objects. Studies such as the ones relating to self-similarity of objects may thus also be polluted by problems of object definition.

**Modeling Natural Phenomena (D):** At very large scales, conventional representations of natural phenomena (e.g. trees, clouds, hydrological features, etc.) are inadequate because they relate to graphical display at medium scales (e.g. a tree is represented as a point) rather than to the natural properties. Representation of natural phenomena has to be based on simulation processes for deriving scale-dependent geometry.

**Labelling / Classification (E):** Labelling applies to tagging nominal data. Classification involves classifying data that are originally of ordinal or ratio/interval character into nominal or ordinal classes. In both cases, misclassification may occur; these errors depend on scale and on the definition of objects and classes. For instance a coniferous tree may be grouped into a deciduous forest on a smaller-scale (e.g. on a remote sensing image), or that forest may be labelled as not affected by a parasite although it contains trees that indeed have been affected.

**Change of Measurement Scale (F):** This is a whole complex of processes that transform data from higher scales of measurement to lower ones. These processes are used in quantitative modeling (eg. typical GIS models). They are also important to spatial scale-changing, since at smaller scales nominal or ordinal representations are preferable. Data may therefore be represented on higher measurement scales at high spatial resolution, and on lower measurement scales at low resolution. Some error always exists even on higher scales of measurement (e.g. due to resolution at sampling time, due to the sampling process, inadequate object definition, etc.). Further errors are introduced in the process of changing the scale of measurement. Keeping these errors as small as possible is a problem of suitable strategies for sampling as well as transformation (which relates to spatial statistics and analysis).

In principle, there are two ways to build a scale-independent multi-purpose database. The data may either be organized as a multi-level (relating to scale) and multi-layer (in the thematic dimension) database. Or they may be derived ad hoc from a single detailed database. The first approach seems inflexible to me for a number of reasons (cf. above discussion): problems of object definition, problems of hierarchical grouping, inability to adjust to a wide variety of modeling or mapping purposes at different scales, etc. I believe pursuing the second approach is more promising. Such a strategy would involve identifying processes that transform spatial data from large (i.e. detailed) scale to small scale, and from one meaning to another. Models and procedures for these processes would have to be developed. Thus, it would indeed be possible to flexibly derive a particular configuration from the detailed fundamental database. This would also allow for fuzzier definitions of objects and their properties.

### **Specific Research Questions:**

Within the broad scope of the subject of multiple representations of spatial data we (i.e. myself, and other researchers in Zurich: Kurt Brassel, Martin Heller, Adrian Herzog) are mainly interested in problems that directly or indirectly relate to the graphical aspects of multiple representations.

database  $\Leftrightarrow$  its graphical representation (ie. map)

Only rarely may the contents of a database be directly translated into graphics (i.e. only involving the process of symbolization). In order to display spatial data in an appropriate and meaningful way, some processes of graphical optimization have to be applied. That is, most often some generalization processes have to take place to transform the internal representation of the

original data into a readable graphical representation. Generalization processes may also be applied to the database (for reduction of accuracy/ resolution., and of content); this process was termed statistical generalization by Brassel and Weibel (1988). For a discussion of statistical vs. cartographic generalization see the attached copy of (Brassel and Weibel, § 3).

In particular, we are interested in:

- Procedures to extract the basic structure and meaning from spatial data in order to select and control generalization processes (cf. "structure recognition", § 3 of attached paper). This task relates to computer vision.
- More flexible and powerful data models that allow to hold structural and semantic information relating to scale-changing processes (or enable easy extraction of such information); data models that support complex generalization processes that require synoptic processing and result in topological distortions. This point relates to data structuring.
- Procedures for controlled statistical generalization of databases.
- identification of generalization processes pertaining to the generalization of different classes of spatial objects (eg. transportation features, terrain, landuse, etc.) as well as to different scale levels (cf. "process recognition", §3); identification of communalities among those processes.
- development of algorithmic or heuristic models for these processes (cf. "process modeling", §3)
- evaluation of the performance of different generalization approaches, perceptual studies.
- integration of the results of the above points (i.e. tools and knowledge) into knowledge-based systems following the amplified intelligence approach (i.e. stepwise knowledge extraction and refinement based on systems under human control)

Apart from this stream, two other topics are of interest to us:

- models to control error due to scale-changing and its propagation (e.g. models to estimate the amount of error introduced by aggregation or disaggregation). This task relates to Initiative I (Accuracy of Spatial Databases) of the NCGIA.
- modeling natural phenomena in database as well as graphical representations (especially growing and changing objects such as trees, forests, clouds, rivers, etc.).

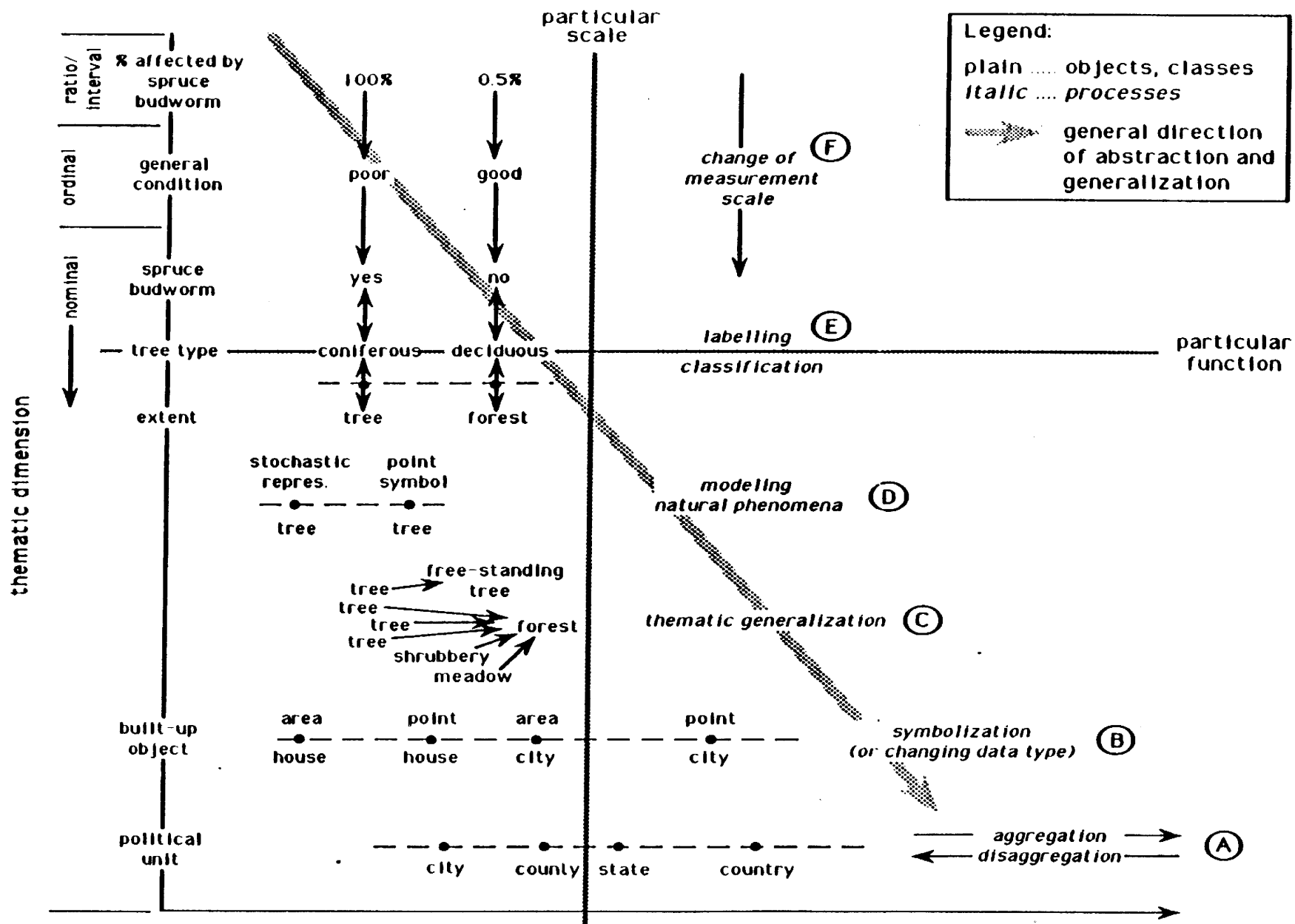
#### **Some Remarks about the Attached Article:**

Attached to this position statement, there is a copy of an article (Brassel and Weibel 1988: A Review and Framework of Automated Map Generalization) that relates to the topic of multiple representations. It presents a conceptual framework of map generalization (or scale-changing in general), and a review of existing work in this particular area. Relevant parts of the article address the following points:

- generalization as a basic human activity, and its relevance to human life (§ 2),
- a conceptual framework for the process of map generalization (§ 3),
- the role of pattern recognition processes to scale-changing (called "structure recognition" in the text; § 3),
- the necessity of mathematical-statistical models (termed "statistical generalization" in the paper) as well as heuristic approaches (called "cartographic generalization") for the solution of problems within multiple representations (§ 3),
- the importance that we (i.e. the authors) assign to processes of graphical optimization (such as cartographic generalization) within the context of multiple representations (§ 3),
- and lastly a review of existing research in the area of automated map generalization and related fields (§ 4), and suggestions for further research (§ 5).

#### **Literature (Attached):**

Brassel, KE., and R. Weibel (1988): A Review and Framework of Automated Map Generalization, Int. J. of GIS, Vol. 2, No. 3, pp. 229-244.



## ADDENDUM TO ORIGINAL POSITION PAPER

-----  
**Robert Weibel**

I wrote my original position paper (PP) in December 1988, answering the call for PP's of the invitation letter to the 13 Specialist meeting. This original version does not explicitly address the points that have been asked for in the second letter (date: 11/1/89) by Babs Buttenfield. In order to approximate that common layout, I would like to rephrase and emphasize some of the points of the original statements.

### **1) Scope of MR:**

See the section "Scope of Multiple Representations (1)" in the original PP.

### **2) Obstacles and Impediments:**

This aspect is implicitly covered in the section "General Research Questions". I do not think it is important to go any deeper into this. I would just like to add the general remark that technological obstacles are irrelevant relative to theoretical and conceptual impediments. I.e. to date the major impediment is the lack of understanding of processes of MR, not missing computational power etc.

### **3) Important Research Topics**

Some aspects of this point are implicitly addressed in sections "General Research Questions (2)" and "Specific Research Questions (3)". In (3), I talk about which specific topics I myself find particularly INTERESTING research-wise. These need not necessarily be the most IMPORTANT research questions to the GIS research community. However, I still think there is a good deal of overlap. The main goal for improved GIA capabilities is the development of scale-independent multi-purpose databases. To reach that objective, we will have to develop

- procedures for information-oriented data reduction to reduce the contents of databases to produce databases of reduced size and accuracy. This is a necessary feature for database transition in analysis and modeling (and called "statistical generalization" in the attached article)
- procedures for graphics-oriented data reduction to optimize graphics and map output from GIS for a specific scale and purpose, because graphics will still form the major man-machine interface for GIS. This form of data reduction is called cartographic generalization.

These two main problems involve a lot of research questions pertaining to: understanding scale and its relevance to GIS models and graphics display, pattern recognition for generalization procedures, appropriate data models to allow consistent MR's, and facilitate the application of generalization algorithms, etc.etc. Yet, an imperative prerequisite for the evolution of MR's of spatial objects and procedures to switch states is the development of more appropriate data models for individual representations. Today's GIS do not have data models suited to represent fuzzyness, growth and change, or complex spatial relations. Improving these data models will be not only a prerequisite for MR's, but will also increase our insight of the way spatial objects and their relations may be digitally represented. more details can be found in the section "Specific Research Questions".