

# Multi-scale Spatial Data Models for Decision Making and Environmental Modeling

Alexandre Sorokine

31st May 2003

Geography Department, University at Buffalo  
105 Wilkeson Quad Buffalo, NY 14261  
Phone: (716) 645-2722x40, Fax: (716) 645-2329  
<mailto:sorokine@buffalo.edu>

## 1 Introduction

From their first college year students are being taught in geography that scale is an intrinsic property of geographic space and geographic objects. However, in the existing Geographic Information Systems (GIS) and desktop mapping scale is merely a ratio between distances measured on the map and on the ground. This interpretation of scale results from technological limitations of the traditional map-making process and of paper as a communication medium [Goodchild, 2000]. Although the use of digital media does not necessarily imply these limitations, existing spatial data models seldom go beyond modeling of a paper map.

This paper represents an ongoing study and focuses on the problem definition and methodological overview. Section 2 introduces the problem of representation of scale in

geographic information science. Methodological approach is outlined in Section 3 that is followed by the Section 4 that depicts authors' views on the conceptualization of geographic objects. The paper is concluded with a section (Section 5) on the preliminary results and planned outcome of this project.

## **2 The Problem of Representation of Scale**

Demands of modern spatial data processing extend far beyond visualization either in on-screen or paper form. On one hand, one of the largest consumers of spatial data is decision making. On the other hand, environmental modeling offers the most sophisticated uses of spatial data. The goal of this study is the development of a class of multi-scale spatial data models that are better suited for decision making and environmental modeling rather than for cartographic visualization. To achieve this goal we will analyze existing and well-established mapping techniques that are used for in several areas such as soil, vegetation and land use mapping.

### **2.1 The term "SCALE"**

Even though in many glossaries and professional geographic dictionaries [Kennedy, 2001, Whittow, 1984] scale is defined as a ratio between distances measured on the map and on the ground, the notion of scale in GIS and geography is much more complex than a simple multiplier for coordinate values. In literature one can find many examples when the word scale is used in a sense that is different from the strict understanding of the term. For a geographer to say that a topographic map of scale 1:24,000 is used means much more than that the lines on the map enlarged by the scale multiplier will match geometric shapes on the surface of the Earth. By saying so a geographer tries to convey other information that is pertinent to coordinate accuracy, composition of features, extent of a single map sheet and many other properties of the map. For other types of cartographic products understanding of scale is also usually wider than its original sense. Numerical scale for a Digital Elevation Model (DEM) automatically translates into the resolution (cell size). For a land use map scale would designate a level of land use classification hierarchy with its own set of types. Scale of a soil map bears approximate size of a

soil map unit and prevailing level of the soil taxonomy. In all these examples scale also reflects a certain time interval that is specific for the process that is depicted on the map.

Such use of the word scale in practice means that scale denominator is typically used to designate several closely interrelated characteristics that include scale *per se* (in the sense of coordinate multiplier), resolution, coordinate accuracy, extent or spatial variability, generalization method or level of taxonomy. In this paper, unless stated otherwise, the term scale will be used in a wider sense of a SCALE LEVEL that implies a variety of scale-related characteristics and their relationships.

## **2.2 The paradigm of geographic scale and its representation in GIS and spatial analysis**

Relation between scale-related variables are conspicuous in the case of paper maps. Due to the finite resolution of paper as a physical media and mechanical measurement instruments scale (in the sense of a ratio) can be directly translated into coordinate accuracy. Physical limits of a single map sheet determine the maximum extent of the area that can be covered. Stringent limitations of the physical space had lead to the development of sophisticated and not totally formalizable generalization methods whose ultimate goal was to maximize the amount of useful information within that space. Advent of computer on-screen visualization had allowed to overcome some of these limitations but only to a limited extent: zoom and pan operations in GIS can be used only on a limited range of scales. Indeed, there are systems that are capable to provide a smooth zoom through the whole practically usable range of scales, however, behind the scene such systems employ sets of maps for several scale ranges.

In the general sense in case of visualization interrelation between scale-related variables can be explained by the limitation of human visual perception. However, this explanation does not hold because scaling issues are important in case of spatial analysis and environmental modeling too. For spatial analysis and modeling scale as a distance ratio does not make much sense because analytical and modeling algorithms are not dependent on it. However, other scale-related variables such as accuracy, resolution, precision, extent neither lose their importance nor interdependence. For example, many modeling algorithms demonstrate a certain resolution threshold after which

reduction of a cell size does not affect its outcome. There is a class of models in hydrology that operate on a whole watershed thus imposing restriction on the extent of a dataset that can be used.

Even such basic notions in vector GIS such as point, line and polygon are in fact scale dependent. In the physical world infinitely thin (1D, lines) or small objects (0D, points) are prohibited. Reduction of the dimensionality of geometric objects is a mean to promote sub-scale significance of a geometric shape up to the scale that is used to operate on. It is very common in GIS and cartography when the same object has to be represented by different geometric shapes on various scales. For example, on large-scale maps rivers are typically represented as polygons but on smaller scales they are generalized to lines.

### **2.3 The problem of boundaries and scale**

Almost all modern GIS operate with a notion of an object with a crisp (infinitely thin) boundary or a field with a rigid association between a value and a location. These representations are limited and their origin can be traced to the paper-based cartography and techniques to convert paper maps into digital form. Indeed, most of environmental objects have indeterminate boundaries or no boundaries at all by smoothly turning one into another in space. Even in case of fiat boundaries, it is not possible to record their positions with an absolute accuracy (or record positions of other objects that interact with objects having fiat boundaries). The meaning of a crisp boundary in GIS should not be understood as an infinitely thin boundary but rather as a boundary whose thickness can be safely neglected for a set of operations that it will be used in. Existing methods simply tend to ignore such cases by reducing them either to measurement accuracy or the accuracy of representing numbers in the computer memory. As a result, the notion of crisp boundaries in GIS implies some sort of the notion of resolution that in turn would depend upon a certain scale level.

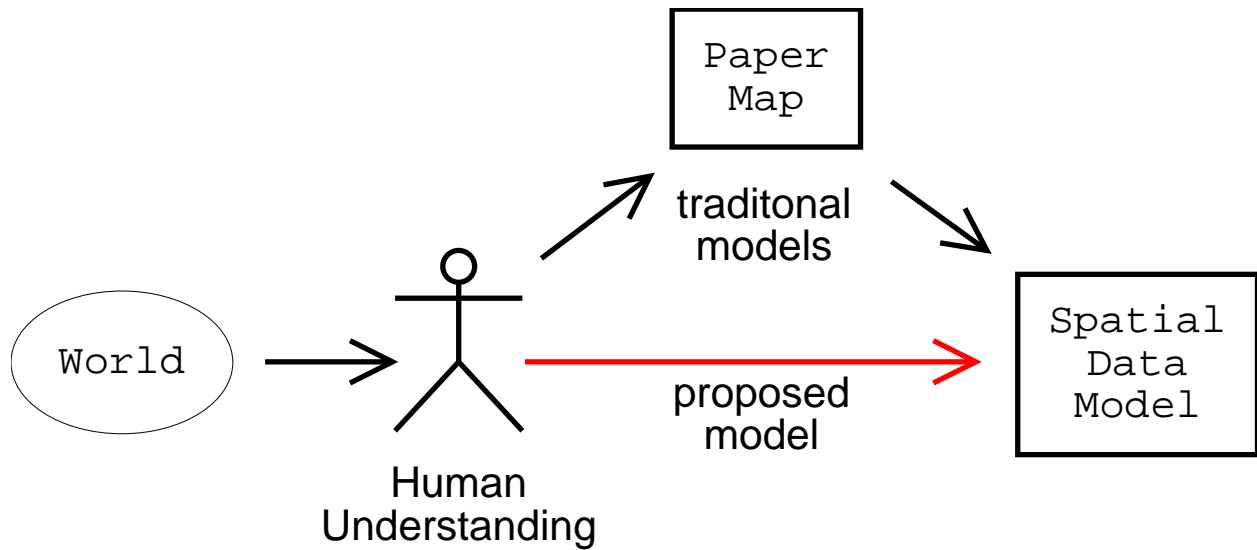


Figure 1: Formation of spatial data models

### 3 Methodology

As it was mentioned earlier, the goal of this study is to create a geospatial data model that would be better suited for environmental modeling and decision making and is not rooted in paper cartography like many of the existing data models (Figure 1). The goal of creating of a new data model will be achieved through an intermediate step of building of a formal ontology for several subject areas.

The term “ontology” in current computer science literature can be used in several various meanings. Most of artificial intelligence, knowledge management and terminology research focuses on information systems ontologies and view construction of ontologies as working applications. This paper represents another line of ontological thought that can be called a “reference ontology” [Smith, 2000]. Our approach is more close to ontology in its philosophical meaning that involves creating of ontological theories for particular domains. In this sense ontology can be paralleled to mathematics that represents the most general laws and rules that underline reality.

Several levels of reference ontologies that are applicable to our study can be distinguished [Papakin1 et al., 2003]:

1. Domain-independent formal ontology that addresses very general notions such as entity, process, identity, part, cause, location, granularity and others;
2. General domain-level geospatial ontology that addresses concepts of a geographic space, geographic object, boundary and topological relations;
3. Area-specific ontology that involves a lot of concepts from sciences that are relevant to a particular mapping technique, such as, for example, soil science on case of soil mapping. Soil taxonomies and notion of pedon, polypedon, epipedon has to be understood and interpreted in the context of the concepts addressed in formal and geospatial ontologies.

The task of building of a formal ontology calls for a deep understanding of the basic properties and conceptualization of (1) a geographic object, (2) boundaries of geographic objects and types of such boundaries and (3) principles behind the systems of geographic objects. Ontological relations will be formalized and used as a basis for developing of general spatial data and query models. Data models in turn will be coded in a functional language in order to achieve a proof of their consistency. The use of the formal ontological methodology as an intermediated step would provide a solid and an unambiguous foundation for the data model. Also it would allow to avoid the limitations of a particular computing platform.

## **4 Conceptualization of Geographic Objects**

Conceptualization of a geographic object has its own peculiarities compared to the conceptualization of, for instance, table-top objects. Geographic objects are intrinsically tied to space and inherit from space many of its structural properties such as topological, mereological (part-of relations) and geometrical ones. Also there are differences in terms of predication. For non-geographic objects space plays a role of accidental predication but it does not affect categorical predication and only the interior of a table-top objects is meaningful for its classification. For geographic objects, on the contrary, its category would depend upon its location and neighborhood [Mark et al., 1999].

## 4.1 Boundary

The first criteria for the individuation of a geographic entity is its boundary. A boundary separates the entity from its environment and is one of the marks of its individuality [Casati et al., 1998]. Notion of a boundary of a geographic object is less obvious than notion of a typical boundary of a table-top object. Often geographic boundaries are indeterminate (vague or fuzzy) and may be dependent upon the scale of observation. The basic typology of boundaries involves two major types of boundaries: bona fide boundaries and fiat boundaries [Smith and Varzi, 1997, Smith and Varzi, 2000]. Bona-fide boundaries involve either some sort of spatial discontinuity or qualitative heterogeneity. These are mostly often observed natural boundaries such as coast lines or watersheds. Fiat boundaries result from acts of human cognition. Examples of fiat boundaries are northern and southern hemispheres or national borders. There are several kinds of fiat boundaries such as social, non-social and mathematical boundaries. The distinction between bona fide and fiat boundaries affects many notions including adjacency, contact, separation and division.

## 4.2 Systems of Geographic Objects

The next question after the concept of a geographic object had been defined is how to organize the systems of objects. Objects can be organized either in taxonomic (is-a, kind-of) or partonomic (part-of) hierarchies. Taxonomic hierarchies can be best illustrated by the Linnean classification of plants and animals or hierarchies of classes in object-oriented programming languages. In taxonomies entities are grouped into classes basing on the commonality of their attributes. Classes at lower taxonomic levels inherit attributes from classes at the higher taxonomic levels. One of the useful functions of taxonomies is the ability to make inferences based on the knowledge of common properties shared by a class of entities.

It has to be noted that genetic taxonomies are different from taxonomies based on the pure commonalities of the entities' attributes. For example, in modern classification of plants and animals the basis of putting species under a single class is their common ancestry in the phylogenetic tree. Typically it means inheritance of some common properties but not necessarily requires it.

As it was mentioned earlier, paronomies reflect part-of relations and also form hierarchies. Paronomies are especially relevant for geographic objects because the latter are often defined in terms of their neighborhood or location within a larger object. In cognition sciences paronomies (unlike taxonomies) do not permit property inferences and are not transitive [Mark et al., 1999]. It can be argued that in the context of soil science paronomies perform both of these functions. Also paronomies that emerge as an outcome of hierarchical natural processes can be viewed as a tool to describe geographic objects at various scale levels.

Geographic object can be characterized using such topological concepts as boundaries, interior, connectedness and separation. Mereotopology is the integration of the theory of part/whole relations with the topological theory [Smith and Varzi, 2000]. An object is considered “closed” if it includes its outer boundary as its part. An object is “open” if its boundary is included in its complement. Table-top objects are always closed but this is not true for geographic objects. Some geographic objects like, for example, lakes are closed, others like bays or mountains may have open boundaries. In the latter case placing of a boundary is not determined by a physical discontinuity and is a matter of fiat [Mark et al., 1999]. In order to describe this kind of relationships mereotopology has to be two-sorted dealing both with bona-fide boundaries in one case and fiat boundaries and non-classical topological principles in the other [Smith and Varzi, 2000].

## 5 Conclusions and Future Work

This is an ongoing project and now we are in the stage of outlining of basic categories and formalizing ontological relations for geographic objects. This stage will be followed by the development on their basis of a new spatial data model in the form of entity-relationship diagrams that later will be coded in functional language “Haskell”.

We hope that the outcome of the study will facilitate more meaningful and computer-friendly spatial databases of soil, land use and vegetation data by providing more advanced data models. Also it will allow more robust description of properties and processes as well as appropriate modeling algorithms in environmental modeling and decision making. The expected results of our analysis can be applied to mapping of other natural phenomenon that demonstrate similar

character of hierarchical spatial organization as soils, vegetation and land use.

## References

- [Casati et al., 1998] Casati, R., Smith, B., and Varzi, A. C. (1998). Ontological tools for geographic representation. In Guarino, N., editor, *Formal Ontology in Information Systems*, pages 77–85. IOS Press, Amsterdam.
- [Goodchild, 2000] Goodchild, M. F. (2000). Communicating geographic information in digital age. *Annals of the Association of American Geographers*, 90(2):244–355.
- [Kennedy, 2001] Kennedy, H., editor (2001). *ESRI Press Dictionary of GIS Terminology*. ESRI Press.
- [Mark et al., 1999] Mark, D., Smith, B., and Tversky, B. (1999). Ontology and geographic objects: An empirical study of cognitive categorization. In Freksa, C. and Mark, D., editors, *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, number 1661 in Lecture Notes in Computer Science, pages 283–298. Springer.
- [Papakin1 et al., 2003] Papakin1, I., Smith, B., Munn, K., and Ceusters, W. (2003). Ontology and bodily systems. <http://ontology.buffalo.edu/medo/OBS.pdf>.
- [Smith, 2000] Smith, B. (2000). Ontology and information systems. Technical report, NCGIA. <http://ontology.buffalo.edu/smith/articles/ontologies.htm>.
- [Smith and Varzi, 1997] Smith, B. and Varzi, A. C. (1997). The formal ontology of boundaries. *The Electronic Journal of Analytic Philosophy*, 5(5). <http://ejap.louisiana.edu/EJAP/1997.spring/smithvarzi976.html>.
- [Smith and Varzi, 2000] Smith, B. and Varzi, A. C. (2000). Fiat and bona fide boundaries. *Philosophy and Phenomenological Research*, 60(2). <http://wings.buffalo.edu/philosophy/faculty/smith/articles/fiatvs.html>.
- [Whittow, 1984] Whittow, J. (1984). *Dictionary of physical geography*. Penguin Books.