GEOGRAPHIC INFORMATION:
Aspects of Phenomenology and Cognition

Major R.J.Williams
Department of Geography and Oceanography
Australian Defence Force Academy
CAMPBELL ACT 2600
AUSTRALIA

ABSTRACT

The disciplines of geography and cartography have experienced an on-going debate on the importance of topological structure in geographic data for the past two decades. It is only now that many organizations are realizing the importance of such structures. Therefore, having crossed the 'topology hurdle', the current trend in research is towards 'integration' of various data types and the 'object-orientation' of geographic features.

While these concepts of 'integration' and 'object-orientation' are applaudible, the theoretical approach seems to have some deficiencies It appears that the aim of many researchers is to 'force' all geographic data into one basic structure, or another, such as 'vector-based' or 'raster-based'. Such an approach creates unnatural and illogical constructs of 'real-world' geographic phenomenon.

In response to this dilemma, this paper discusses phenomenological structures of geographic information and aspects of interpretation. Fundamental to this approach is the knowledge representation of phenomenons of the 'real-world' independent of any specific application, and that analysis is based on actual geographic structure and location and not on graphical representations of that data.

INTRODUCTION

The rapidly growing population of the earth and the increasing complexity of modern life, with its attendant pressures and contentions for available resources, has made necessary detailed studies of the physical and social environment, ranging from population to pollution and from food production to energy resources and terrain evaluation. The geographer, preeminently, as well as the planner, historian, economist, agriculturist, geologist, military tactician, and others working in the basic sciences and engineering, long ago found the map to be an indispensable aid. With maps, the geographer may observe or record factual observations, describe the manner in which individual earth's phenomena vary from place to place, develop hypotheses concerning the association of environmental factors, and, in general, study the spatial correlation of the elements of the earth's surface. By their very nature maps are the presentation of geographic spatial relationships (Robinson, Sale and Morrison, 1978; Trewartha, Robinson and Hammond,1967).

Within the last decade, computers have been used with greater regularity to assist in the representation and analysis of geographic phenomena. The management and use of digital data has, however, been influenced by traditional approaches to mapping and resource management. These approaches to the management of data have, in turn, influenced how the geographic data has been structured. Three major classes of data structure have evolved; these being the *unlinked vector model* (used predominately within computer-assisted mapping systems), the *topological structured vector model* (used mainly in Land Information Systems, and with those types of data managed according to societal regulation); and the *grid cell and raster model* (generally used by natural resource managers, and with data obtained from remote sensing scanners).

Until recently these trends were often viewed as non-compatible approaches designed for different markets. But today's mapping market is seeing a major shift toward decision support and operations management. Supporters of the three trends, although continuing to foster their respective approaches, are now agreeing that there is a need to develop interfaces and integration between systems.

TOWARDS INTELLIGENT SYSTEMS

Up until now the *data models* have restricted the use of geographic data, but the major trend in research is to structure data suitable for multiple purposes and using *real-world* relationships (Grady, 1986). Bouillé (1986) asserts that "contrary of what is generally taught, we must emphasize the fact that the structure [of geographic data] is completely independent of the problem we want actually to solve. Moreover, a structure built correctly, and with no *'a priori'* idea, always contains a substructure which immediately answers our problem..." But, in order to achieve these desirable characteristics, far more intelligence has to applied to data than in the past. The theory for these developments will come from the field of artificial intelligence, specifically the area known as *expert systems* (Williams, 1987)

Intelligence can be achieved via two fundamental, but integrated sources. These sources are those of data relationships and structure, and techniques and procedures for manipulating and analysing the data relationships. These sources can be considered as forming *expertise*. *Expertise* consists of *knowledge* about a particular domain (*real-world* geographic structures), understanding of the domain problems, and *skill* at solving some of these problems. *Knowledge* (in any speciality) is usually of two sorts: *public* and *private*. *Public* *knowledge* includes the published definition, facts, and theories of which textbooks and references in the domain of study are typically composed. But *expertise* usually involves more that just this *public knowledge*. Human experts generally possess *private knowledge* that has not found its way into the published literature. This *private knowledge* consists largely of rules of thumb that have come to be called *heuristics*. *Heuristics* enable the human expert to make educated guesses when necessary, to recognize promising approaches to problems, and to check effectively with errorful or incomplete data. Elucidating and reproducing such *knowledge* is the central task in building *expert systems* (Hayes-Roth, Waterman, and Lenat, 1983).

The *phenomena of the real world*, (or data relationships and structure) can be described by an abstraction defined within functional categories, related via topological properties and spatially referenced Data abstracted in such a way could be considered as being in a *geographic knowledge base* The uses of this geographic data are affected by the requirements of a user's purpose. The products for this purpose can be static reports, such as map overlays of categorized and symbolized geographic data, or as a result of analyses or desired view of various aspects of the data. In a sense, these products result from processes by which information is selected and encoded, reduced or elaborated, stored and recovered, and decoded and used. Such products can be considered as having *cognitive* properties. Moore and Golledge (1976) suggest that "environmental cognition refers to awareness, attitudes, impressions, information, images, and beliefs that people have about environments. These environments may be directly experienced, learned about, or imagined. Environmental cognition refers to essentially large scale-environments, from nation and geographic regions down to cities and spaces between buildings, to both built and natural environments, and to the entire range of physical, social, cultural, political, and economic aspects of man's world. Cognition of these environments implies not only what individuals and groups have information and images about the existence of these environments and their constituent elements, but also that they have impressions about their character, function, dynamics and structural interrelatedness, and that they imbue them with meaning, significance and symbolic properties".

In summary, the *phenomena of the real world* are those abstractions which describe the earth's surface, its form and physical features, its natural and political divisions, the climate, the productions and the populations, whereas the *cognition* is the interpretation and analyses of relations between the phenomenological aspects.

THE PHENOMENA OF OUR DATA REALITY

Several researchers have developed and itemized "levels of data abstraction". Notable, within the field of geographic data description, there are contributions by Nyerges (1980) who identifies six levels of data abstraction and, more recently, work by Guptill et al (1987) who identify five levels of abstraction. The important component within both schemes is the stated importance on the *data model* (or Nyerges' information and canonical structures).

Data Reality	The data existing as ideas about geographical entities and their relationships which knowledgeable persons would communicate with each other using any medium for communication
Information Structure	A formal model that specifies the information organization of a particular phenomenon. This structure acts as a skeleton to the canonical structure and includes entity sets plus the types of relationships which exist between those entity sets
Canonical Structure	A model of data which represents the inherent structure of that data and hence is independent of individual applications of the data and also of the hardware and software mechanisms which are employed in representing and using the data
Data Structure	A description elucidating the logical structure of data accessibility in the canonical structure. There are access paths which are dependent on explicit links and others which are independent of links
Storage Structure	An explicit statement of the nature of links expressed interms of diagrams which represent cells, linked and contiguous lists, levels of storage medium, etc
Machine Encoding	A machine representation of data including the specification of addressing, data compression and machine code

Figure 1 Nyerges Levels of Data Abstraction

Spatial Data Model

The phenomena of data reality is considered, in totality, as *entities*. An entity and its digital representation is termed a *feature*. A feature is a set of phenomena with common attributes and relationships. All of the elements of this set of phenomena are homogeneous with respect to the set of selected common attributes and relationships used to define a feature. All geographic features implicitly have location as a defining attribute.

The concept of *feature* encompasses both entities and objects. The common attributes and relationships used to define the feature also apply to the corresponding entities and objects An *entity* is a real world phenomena that is not subdivided into phenomena of the same kind. This 'real world phenomena' is defined by the attributes and relationships used to define the feature. An *object* is the representation of all or part of an entity. The concept *object* encompasses both *feature object* and *spatial object.* A *feature object* is an element used to represent the non-spatial aspects of an entity. A *spatial object* is an element used to represent the position of an element (Moellering, 1987; Guptill et al, 1987; Rossmeissl et al, 1987).

An attribute is a characteristic of a feature, or of an attribute value. The characteristics of a feature include such concepts of shape, size, material composition, form and function of a feature. The attributes assigned to a feature include those inherent in the definition of the feature and additional attributes which further describe the feature. An attribute value is a measurement assigned to an attribute for a feature instance, or for another feature value.

Further, three major groups of rules can be used to formulate the description of feature instance. These groups are rules for defining feature instances, composition rules for representing feature instances, and rules for aggregating feature instances.

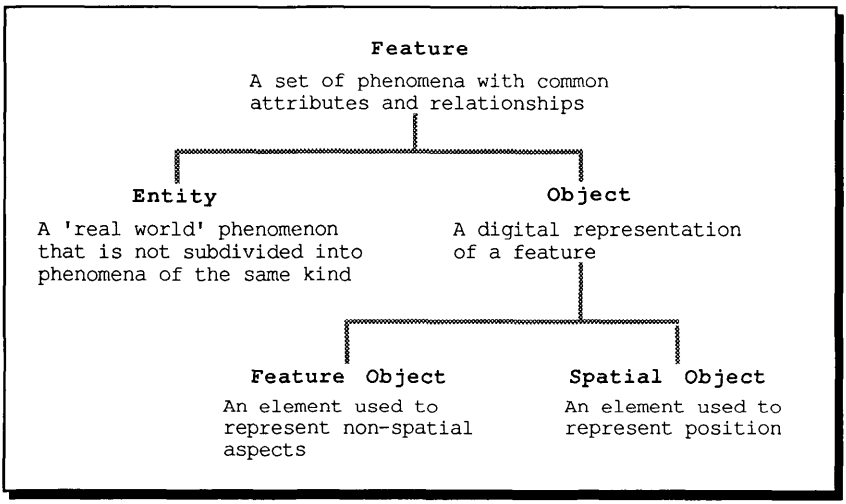


Figure 2 Concepts of feature, entity and object

World Views

Rossmeissl et al (1987) suggest that a methodology to define the domain of features is to use a concept of *World views* of spatial entities. Rossmeissl's proposed views are 'cover', 'division', 'ecosystem', 'geoposition', and 'morphology'. Each view is subdivided into subviews. An alternate approach would be to have *World views* related to functional categories as shown in Figure 3. The alternative approach is more applicable to *real-world* feature categories and corresponds to grouping commonly used in land inventories. Separate and independent research by McKeown (1983) indicates that complex features (spatial entities) can be represented using *schemas.* Each entity is a 'concept map' and can be represented by one *concept* schema and at least one *role* schema. McKeown notes that using such an approach facilitates hierarchical decomposition of features, say, using natural hierarchies such as political boundaries, neighbourhoods, commercial and industrial areas, and so on (Figure 4).

Apart from categorizing the domain of features, aspects of perception (scale relevance) seem to be inadequately addressed by researchers. Rhind (1988) notes that 'existing computer geographic information systems (apart from Domesday) are entirely or very substantially based upon digital storage of coordinate data and their attributes - essentially low level conceptualizations of the objects under consideration". He observes that "human beings evidently store multiple levels of conceptualization of objects, sometimes in a 'soft' or 'fuzzy' fashion". It seems that aspects of conceptualization can be defined using adaptions of Rossmeissl's 'views' and McKeown's 'concepts' to manage data from 'scale related' views.

Under the *expert system* approach, the aspects of *data reality* discussed in this section correspond to a formal definition of *public knowledge* and constitute a 'geographic knowledge base'. Having established this geographic knowledge base, it should be possible to extract information according to different interpretations aligned to the use of the information.

COGNITION OF OUR DATA REALITY

Traditionally, geographic data has been processed using one of three fundamental techniques. These are: (1) through graphic reports, such as standard series maps and thematic maps; (2) by overlay analysis for management and control of phenomena relevant to land management; and (3) using analytical techniques for planning and modeling of terrain related features (Figure 5).

These various interpretations can be produced by using combinations of existing data management and transformation techniques. As an example, standard series and thematic maps would use 'rules' for *world view* overlays and aggregation of features; techniques for generalization, simplification, and symbology; transformations for coordinate change; projection formulae, and processes for scale change, map derivation and update. Management and control products would require operators to process data that is represented with each object having spatial location as an essential property *(vector model)* and locations that have object properties *(tessellation or grid cell models)* ; techniques to overlay and integrate*world views* in both model types; and data base management functions. Planning and modeling products would require an extensive range of operators as mentioned already, as well as mathematical and topological functions; spatial data operators; and heuristic algorithms (Figure 6).

Regional Administration	Includes political, administrative, institutional, statistical facilities and regions; as well as reservations, parks, monuments, etc.
Population	Includes places of human habitation and occupation eg residential, commercial, religious,cultural,entertainment, recreational,educational,etc.
Road Infrastructure	Includes roads, junctions, bridges, overpasses, and related features,etc.
Rail Infrastructure	Includes lines, marshalling yards,bridges, and all related features, etc.
Air Infrastructure	Includes airfields, facilities, navigations aids, etc.
Sea Infrastructure	Includes port facilities, jetties, piers, sea control features, channels,canals,etc.
Telecommunications	Includes communication faclities, structures, networks, etc.
Electricity / Fuel	Includes power plants, facilities, and networks for generation and distribution of power and fuel.
Water Resources	Includes facilities and networks for storage and distribution of water resources.
Industry	Includes manufacturing, mining, agricultural facilities; extraction and disposal complexes; etc.
Health / Medical	Includes hospitals, research institutions, aid posts
Physiography	Description of terrain
Oceanography	Includes environments of the oceans
Vegetation	Includes natural and cultivation plant life
Climatology	Climatic phenomena

Figure 3 World view categories

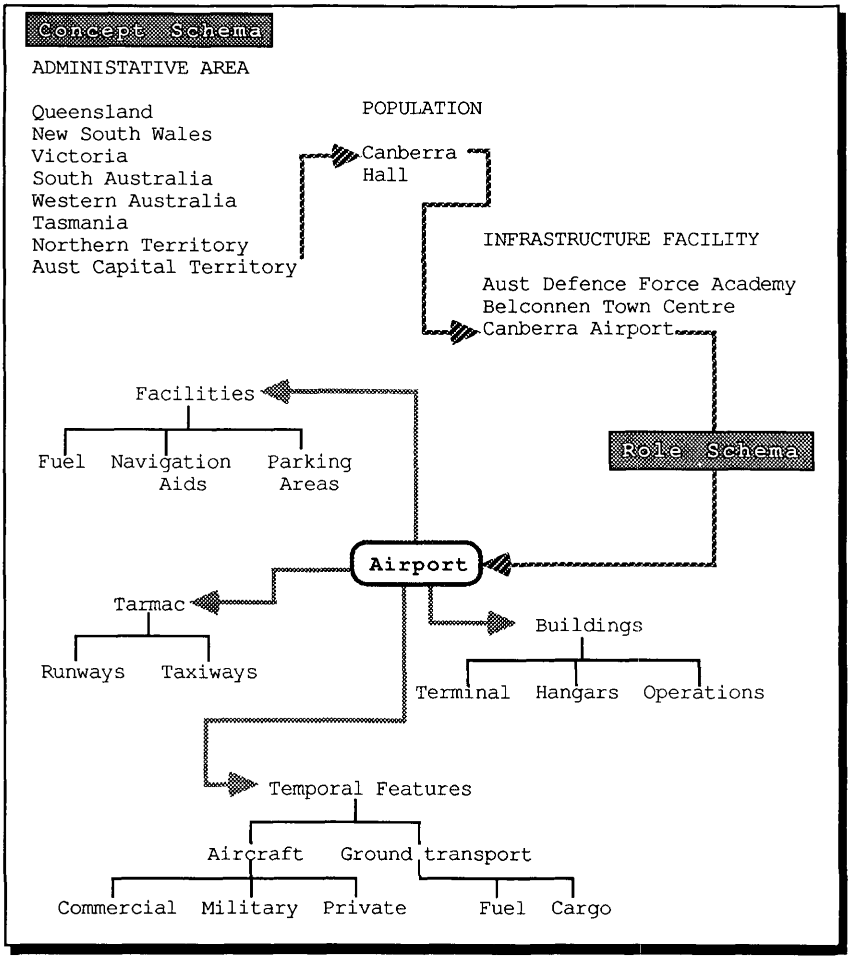


Figure 4 Concept and role schemas

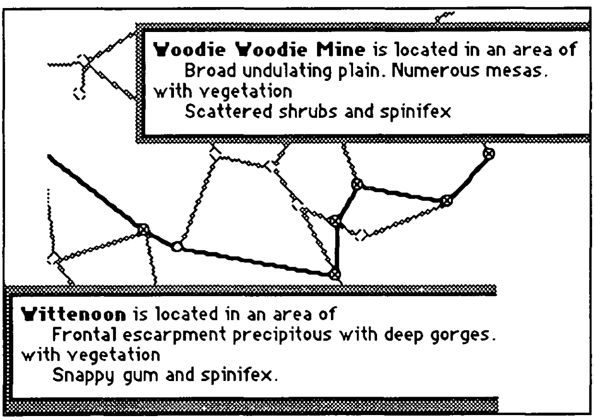
Society is becoming information conscious. Each year there is an ever growing need to know more about the environment. Therefore there is pressure on those involved with land studies to provide accurate information and creditable advice on subjects pertaining to the environment. It seems apparent that the tools required to perform analyses on the *phenomena of the real world*, will come from the *expert systems* area of artificial intelligence but operating on geographic data structured according to *world view* concepts.

Standard Products
- Series Mapping
- Thematic Mapping

Management and Control
- Land Records Maintenance
- Service and Utilities Management
- Resource management
- 'World Views' Integration

Planning and Modeling
- Slope and Viewshed Assessment
- Environment Simulation
- Route and Corridor Determination
- Terrain analysis and evaluation
- Geographic Searching
- Proximity Determination

Figure 5 Some cognitive views of geographic information



An example in the use of heuristic analysis and integration of other data types might to plan routes in the road network system and then to process the route through selected terrain descriptions, such as physiography and vegetation cover

(Williams, 1988)

Figure 6 Analysis of the road transportation network

REFERENCES

Bouillé, F. (1986). "Interfacing cartographic knowledge structures and robotics", **Proceedings - Auto Carto London**, Vol 2, Ed Michael Blakemore, International Cartographic Association.

Grady, R.K. (1986). "New-age cartography", **Computer Graphics World**, 9, 10.

Guptill, Stephen C., Kenneth J.Boyko, Michael A.Domaratz, Robin G.Fegeas and David E.Hair (1987). "An Enhanced Digital Line Graph Design", Internal Paper of the National Mapping Division, United States Geological Survey, Washington, DC

Hayes-Roth, F, D.A.Waterman and D.B.Lenat (1983). **Building Expert Systems**, Addison-Wesley Publishing Company, Reading, Massachusetts.

Moellering, Harold (Ed) (1987). "Issues in Digital Cartographic Data Standards", Report #8, National Committee for Digital Cartographic Data Standards, The Ohio State University, Columbus, Oh.

Moore, G.T and R.G.Golledge (eds) (1976). **Environmental Knowing: Theories, Research, and Methods**, Dowden, Hutchinson and Ross, Stroudsberg,Pa.

McKeown, David M. Jr (1983). "Concept Maps", Technical Report CMU-CS-83-117, Carnegie-Mellon University, Pittsburgh, Pa.

Rhind, David (1988). "A GIS Research Agenda", **International Journal of Geographic Information Systems**, Ed. J.T.Coppock and K.E.Anderson, Vol.2, No.1,Taylor and Francis, London.

Robinson, Arthur, Randall Sale and Joel Morrison (1978). **Elements of Cartography**, Ed. 4, John Wiley and Sons, New York.

Rossmeissl, Hedy, Michael Domaratz, John List, Lynn Usery (1987). "Proposed Definition of Cartographic Features for Digital Line Graph-Enhanced (DLG-E)", Report of Committee Investigating Cartographic Entities, Definitions and Standards, National Mapping Division, USGS, Washington, DC.

Trewartha, Glenn T., Arthur H. Robinson and Edwin H.Hammond (1967). **Elements of Geography**, Ed 5, McGraw-Hill, New York.

Williams, R.J. (1987). "Evolution in cartography: data intelligence", **Cartography**, Vol 16, No 2, The Australian Institute of Cartographers.

Williams,R.J.(1988) "Analysis of the Road Transportation Network", Paper presented to **7th Australian Cartographic Conference**, Sydney