LAND RESOURCE INFORMATION SYSTEMS: SPATIAL AND ATTRIBUTE RESOLUTION ISSUES

Kenneth J. Dueker
Center for Urban Studies
School of Urban Affairs
Portland State University
Portland, Oregon

I. Introduction

The tasks of inventory of land uses and the reconaissance of resources for large regions has resulted in the need for and the development of a variety of computer data-base systems. When data bases are organized to handle non-routine analyses stemming from land planning and management questions, the data base and its computerized and procedural environment may be termed a land resource information system.

Land resource information systems stem from diverse origins, disciplinary orientations, and purposes. This diversity makes comparison and analysis of systems difficult, but it is important to compare systems so as to understand better the different approaches to land resource information systems.

The single most important issue in designing a land resource information system is the determination of the appropriate level of spatial and attribute resolution. Other kinds of information systems have discrete basic data such as transactions, persons, or events, whereas geographic space is continuous and choices must be made how to classify activities in that space to manageable categories and how to partition the space into observable spatial units. In addition, the observable spatial units must describe the extent and character of features as well as maintain spatial relationships between units, and thereby features. The choice of spatial and attribute resolution is interrelated. The size of areal unit and the number of categories for a characteristic are interdependent and together determine the volume of data, which in turn is the most important determinant of the hardware and software requirements of the information system.

The spatial and attribute resolution of land data in a land resource information system is crucial to the kinds of questions that can be addressed. It is easy to claim the need for the greatest detail possible, but the volume of data may become overwhelming. Yet, there exists little knowledge as to tradeoffs between detail and volume in terms of cost and use of information for management and planning. Often users of information cannot wait or pay for detailed data and more aggregate data are used. The more aggregate data may be used because the value of more detailed data is not well known in terms of the decision to be made.

Purpose

This presentation serves to illustrate the linkage between alternative ways of capturing and encoding spatial data, levels of spatial and attribute resolution, and uses of the data. An interpretive comparison and assessment of land resource information systems is employed to highlight the linkages. Early rather than current systems are compared because they provide a broader range of approaches that were largely uninfluenced by one another. These early systems are: The Canada Geographic Information System (CGIS), the Polygon Information Overlay System (PIOS), the Minnesota Land Management Information System (MLMIS), the Land Use and Natural Resources Inventory of New York State (LUNR), and the Oak Ridge Regional Modeling Information System (ORRMIS). These systems are described and examined in Tomlinson, Calkins, and Marble (1976). These systems will be reexamined in a framework utilizing a communications system model for geographic data that allows comparison of approaches for data capture and illuminates choices made for spatial and attribute resolution. There has been a considerable amount of learning from early experience and convergence on system design approaches that is discernable and is documented more fully (Dueker, 1978).

A Framework for Comparison

To compare systems, a communications model is employed. This enables explicit recognition of data capture and spatial and attribute resolution choices made by systems (Dueker, 1976).

A generalized communications system is comprised of elements: 1) a data source from which a message is encoded, 2) a transmitter from which a signal is emitted, 3) a channel for communicating the signal, and 4) a receiver for receiving the signal and converting it back to a message at the final destination. This problem of sending and receiving messages through a communications system which is constrained by channel capacity and the presence of perturbances (noise and distortion) is analogous to the capture and encoding of spatial data.

A geographic information system has an attribute message for each spatial unit that derive from a source, are captured, transmitted, communicated, and decoded and received. The problems of channel capacity and transmission cost may be considered as analogous to computer storage size and machine processing cost. In the generalized communication system, information theory is used to measure the amount of information (in units called "bits") that is contained in the data being transmitted and this theory aids in the evaluation of alternative encoding schemes to eliminate redundancy through efficient coding.

Comparison of Early Systems

Although technology has advanced, system designers are faced with making choices as to the extent of preprocessing image data prior to capture in machine records versus the amount of data reduction and editing subsequent to data capture. But foremost, designers must make choices with respect to spatial and attribute resolution in conjunction with the data capture technology.

To illuminate the data capture and resolution interactions, Table 1 provides in a comparative format the five early geographic information systems. Using the communications model analogy, these systems are compared with respect to the process used to prepare, capture, encode, reduce, edit, format, and retrieve geographic data.

In interpreting Table 1, one observes that two systems (CGIS and PIOS) generate vector data and the other three (MLMIS, LUNR, ORRMIS) generate grid data. Two systems (CGIS and ORRMIS) capture data with scanners, PIOS employs a digitizer, and MLMIS and LUNR rely on manual encoding. Although one normally thinks of scanners as generating grid data, the CGIS vectorized grid data is a data reduction process, wherein ORRMIS maintained a grid format but aggregated the fine grid mesh of the scanner to larger geodetic grid units.

With respect to spatial resolution and attribute resolution, the systems again vary considerably. The spatial resolution of CGIS and PIOS depend on the data; that is, the size of the polygons, which is a consequence of the number of attribute levels and the diversity of the study area. Although CGIS and PIOS are theoretically limited by the precision of the hardware, the ability to draw small polygons on source maps and the number of attributes rarely resulted in the hardware resolution as being a constraint as more aggregate data resulted from the process. The spatial resolution of MLMIS, LUNR and ORRMIS were direct consequences of the grid size selected. In MLMIS and LUNR, a single grid size was imposed on all data entered into the system, whereas in ORRMIS one could capture data for several size grid units, depending on the size needed to capture the complexity of data on the source map.

The attribute resolution depends on the nature of data represented. Land use and soils are two difficult data types to classify, map, and capture. Land use is used to illustrate attribute resolution choices made by the five systems and how attribute resolution interacts with spatial resolution. CGIS had a 14-category system for land use and PIOS 20 categories. In diverse areas, this level of attribute resolution resulted in small polygons to capture and process. A finer attribute resolution (or more categories) would have required mapping at a smaller scale, and consequently more data for the same study area.

MLMIS used nine categories of land use encoded to 40-acre units. This coarse level of land use was apparently selected so that predominant (or single users)

Table 1. Comparison of Early Systems

SYSTEM / PROCESS	CGIS The Canada Geographic Information System	LUNR The New York Land Use and Natural Resource Information System	MLMIS The Minnesota Land Management Information System	PIOS The Polygon Information Overlay System	ORRMIS The Oak Ridge Modeling Information System
Date Initiated	1964	1967	1969	1971	1972
Basic Technology					
Data Capture	drum scanner	manual	manual	digitizer	flying spot scanner
Encoding	fine polygon	large grid	medium grid	polygon	small grid
Graphic Data					
Pre-processing	scribe polygon boundaries, number polygons and mark centers	air photo interpret, compile on quad maps	none	polygons and classifications drawn on stable base mylar	polygon, linear, and point data from source map are transferred to a mylar, category by category, and the mylar is photographed each step
Data Capture Techniques	scan for presence of line segment, digitize polygon centers	manual coding to UTM (1 km^2) grid and keypunch	mark sense card	digitize polygon boundaries	scan to detect darkened areas, each darkening representing a category
Encoded Data					
Data Reduction	vectorization of bits indicating presence of lines or boundaries between areas	none	none	sliver and gap removal	reduction of sum cells to 3.75 second or larger grids
Mode	vector; polygon, distance-direction notation	grid, percentages of each grid in particular land uses	grid, predominant use	vector, polygon chain of x, y coordinates	geodetic hierarchal grid, predominant or percentage
Editing	gap and spike removal, edge matching	sample field check	redundant interpretation, machine echo plot of coding	delete, change, rotate, shift polygons	quick look plots

could be assigned to each 40-acre grid.

The LUNR system employed 130 land use categories for larger, 1 km^2 (247 acres), grids. Consequently, predominant assignment of land use to grids was inappropriate and the amount or area of land uses in each grid was encoded. This disparity between attribute resolution and spatial resolution results in a loss of information because the land use detail inherent in a detailed classification is lost when aggregated to a 1 km^2 grid. This loss takes a form where one cannot determine where within a grid unit the different uses occur and how they are related spatially.

ORRMIS utilized both a hierarchy of categories for land use and a hierarchy of geodetic grid cells. A grid cell size of 7½' is used in conjunction with nine categories of land use, 2½' grid cell size with 46 categories, and 30' grid cell size corresponds to 235 categories.

If the technology employed captures a large part of the graphic content of maps or photos, there is an associated and large problem of data reduction, editing, aggregation and formatting. On the other hand preprocessing of map data and/or abstracting (or recording only part of the map content, say by imposing a large grid) simplifies the amount of reducing or processing of encoded data to develop formatted data for use. CGIS is an example of a system which requires preprocessing of map data because of strict input requirements, considerable data reduction software, and severe editing requirements to insure quality data. PIOS requires considerably less pre-processing and data reduction, although the edit requirements are again severe, so severe that strict polygon systems (those that maintain vector data throughout overlay) are rarely used now. MLMIS and LUNR require little pre-processing, reduction, or editing because of the level of abstraction inherent in those systems; they are for coarse level inventories of large regions. ORRMIS, like CGIS, requires considerable pre-processing to prepare for scanning and data reduction to translate an extremely fine mesh of scan line cells to the geodetic grids. Another complication of the ORRMIS is the merging of separate scan data, a separate scan for each land use category, into a single data set containing all land uses. The ORRMIS data capture technology works well when there are few categories (low attribute resolution) per map coverage, but when encoding a high attribute resolution, each category requires a separate scanning process. This is particularly burdensome when inputting a set of unique areas, such as census tracts, counties, traffic zones; each unique areal unit must be transferred to the mylar, photographed, and scanned.

Lessons from Experience

None of the systems is functioning today as described here. CGIS relies more on digitizer data entry; PIOS on single line digitizing and generating polygons rather than double digitizing of lines inherent in direct polygon digitizing, greater attention to preparing higher quality input maps, and conversion to grid prior to overlay analysis; MLMIS on interactive data entry and retrieval; LUNR is not functioning with respect to new data entry; and ORRMIS relies on digitizer input of data. Nevertheless, comparison of the original systems is instructive in gaining an appreciation of approaches to encoding geographic data, especially the interrelatedness of data capture technology, the character of source data, spatial and attribute resolution, and data reduction and editing. To a great extent the designs of current systems have integrated knowledge gained from these earlier experiences. Each of these early efforts contributed knowledge to current approaches. In effect, we have modularized our successes in terms of sub-routines or hardware, and discarded (or postponed subject to further research and development) the less successful features of these early systems.

But most importantly, system designers have learned better through these experiences the appropriate functions to be performed by man or machine. For example, the current version of PIOS requires more carefully controlled source maps and utilizes grid data structures where gridded data are more cost-effective (Dangermond, 1978). Dangermond (1976) posits that manual compositing of slope, soil, vegetation, and geology is necessary due to limitations inherent in the individual source maps; man's judgement is needed to resolve inconsistencies. He calls polygons created in this compositing "integrated terrain units" and argues that they have greater validity than a computer overlay of the separate coverages would provide. PIOS also relies more on chain encoding rather than polygon encoding so as to avoid the "sliver and overlap" problem that occurs when digitizing adjacent polygons, which results in two lines which do not match exactly.

For natural resource applications, polygon encoding is giving way to a streamlined creation of topological data which links "all the segments which constitute a curvy boundary into a list, and to assign to this whole chain of points the topological properties which all the segments have in common...The more irregular the geometry of the network being represented, the more efficient chaining becomes." (Dutton and Nisen, 1978, p. 140).

These early experiences also identified areas needing greater computerization. For example, many of the early systems relied on sequential data stored on cards and tape, which have given way to more complex, but realistic, data structures that are stored in more efficient forms. Originally, CGIS, PIOS, and ORRMIS relied too heavily on the computer and have had to modify some procedures, whereas LUNR and MLMIS found it necessary to augment some procedures with more computer assistance. Gradually a clearer understanding of man-machine roles is emerging.

Examination of both early and current experience with geographic information systems documents the importance of the classic criteria for information systems--timeliness, accuracy, flexibility, reliability, etc. In addition to these, spatial and attribute resolution are important determinants of utility of geographic information systems. Resolution that is too fine usually results in violating one or more of the classic criteria, usually cost or time overruns. Too coarse spatial or temporal resolutions do not allow addressing questions with enough precision to be telling.

In sum, what appear to be pragmatic and straightforward decisions concerning spatial and attribute resolution turn out to be the most crucial decisions concerning system utility. Spatial and attribute resolution decisions are usually a result of data volume or system constraints, rather than a direct result of application requirements. Yet there is little guidance or experience in relating system resolution to application requirements. This relationship needs further development.

References

Dangermond, J., (1976), "Integrated Terrain Unit Mapping (ITUM)--An Approach for Automation of Polygon Natural Resource Information," paper presented at Fourteenth Annual Conference of Urban and Regional Information Systems Association, August 29-September 2, 1976, Atlanta, GA.

Dangermond, J., (1978), "A Case Study of the Zulia Regional Planning Study, Describing Work Completed," a paper prepared for Harvard Computer Graphics Week, Environmental Research Systems Institute, Redlands, CA.

Dueker, K. J., (1976), "Cartographic Data Structures: Alternatives for Geographic Information Systems, Computer Graphics, Vol. 10, No. 2, Summer, 1976, pp. 167-172.

Dueker, K. J., (1978), "Land Resource Information Systems: A Review of Fifteen Years Experience," Technical Report No. 110, Institute of Urban and Regional Research, the University of Iowa, Iowa City, IA.

Dutton, G. J. and W. G. Nisen, (1978), "The Expanding Realm of Computer Cartography," Datamation, June 1978, pp. 234-142.

Tomlinson, R. F., H. W. Calkins, and D. G. Marble, (1976), Computer Handling of Geographic Data, The UNESCO Press, Paris.