THE INSTITUTE OF URBAN AND REGIONAL RESEARCH

A Review of:

URBAN GEOCODING

by

Kenneth J. Dueker

102 CHURCH STREET
THE UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

A Review of:

URBAN GEOCODING

by

Kenneth J. Dueker

February 1973

(revised July, 1973)

Institute of Urban and Regional Research
University of Iowa

Dr. Dueker is Director, Institute of Urban and Regional Research and Professor of Urban and Regional Planning and Geography, University of Iowa.

URBAN GEOCODING

Introduction

Geographic research, focusing within urban areas or otherwise, requires location-specific data. Place names and street addresses are common means of referencing data, which either cognitively or with the aid of a map,convey locational relationships. An occasional ambiguity of such intuitive and simple referencing of places can be tolerated when relying on manual processing of data, but computer processing of large data files requires precise location referencing in an unambiguous manner and routines to detect error conditions. These two requirements are becoming increasingly important as researchers emphasize analytic methods, mathematical models, and computerized spatial analysis for studying real world processes. However, a conceptual framework for location referencing or coding is not well developed despite the logical need for the conceptual development to precede the technical developments.

The discipline of geography is beginning to assume a greater degree of leadership in developing concepts and applications of spatial referencing of data. Although geographers have participated in planning and management applications requiring the development of spatial referencing schemes, research geographers generally have been reluctant to divert attention from substantive areas of inquiry to develop coding concepts for data assembly. An individual's data problems appear too immediate for an investment of time and energy toward a general solution. Non-recurring problems inherent in developing a substantive line of inquiry and small sample sizes do not encourage the investment necessary for computer-assisted data preparation. Consequently, few research geographers are making inputs to a methodological development which could be quite useful to their own work, and to the work of others.

The above situation has often been reinforced by actual experience. In attempting to use existing geocoding systems in an urban area, say for street address to census area conversion, a health planner or academic researcher not involved in setting up the system for the local area will most often find it difficult and time consuming to make use of the computerized procedures. Too frequently, the potential user is a one-time user of such systems, or a considerable investment is called for on the researcher's part to support or alter the system or develop an interface to adapt data to requirements of the system. Such experience may lead researchers to believe that such computerized systems are "more trouble than they are worth", and rather than demanding better systems, they can become disenchanted and return to manual methods.

Developments are emerging as a result of spatial analysis needs of urban planners, demographers, and urban administrators. These applications are stimulating conceptual developments and are slowly forcing the development of consistent terminology. However, the plethora of acronyms, e.g., ACG, DIME, SACS, ADMATCH and inadequate technical definitions of common terms such as place, point, code, block, address, etc., have caused confusion and delays in the development of geocoding technology. In general, concepts have followed, rather than led, developments in application. This might be overcome if the field were more systematically structured as an area of geographic research.

The purpose of this paper is to describe and evaluate recent efforts in spatial referencing problems and to assess the utility of the developments for urban research particularly, and to speculate on future developments in the field. This is not a well-defined area of research within geography or within the applied fields associated with urban problems. This review article attempts to structure the issues and review literature and directions of what has become known as "geocoding" with a view toward encouraging more vigorous investigation of the many problems facing the field.

Definitions

Geocoding is a concatenated term deriving from <u>geo</u>graphy, that is spatial, and the concept of systematizing such information as an aid in analysis or communication, that is <u>coding</u>. The term, therefore, implies the purpose of the process, namely, to provide a methodology for maintaining and using information about the spatial relationships among units used for data collection purposes.

Tobler[1] refers to geocoding as place naming of which there are two types -- 1) a name of a place that tells something about the place or 2) a place name such as coordinates that describe the spatial relationships of that place. The first type requires reference to outside sources or a map to infer situation from the name of place. For example, the name of a place such as a city is insufficient without a map or knowledge of the city's juxtaposition to other cities. A coordinate system makes the spatial relationships explicit. Generally the thrust of geocoding is to move from a type (1), place name geocode to a more explicit geocoding of spatial relationships, i.e., type (2). Consequently, an important feature of geocoding systems should be a capability to translate data geocoded by place names to geocodes describing the spatial relationships of places, such as hierarchial areal unit codes, or coordinates defining places as points, lines or areas.

The geocoding process captures some portion of the geometric structure of a map in a machine-readable data file. Depending on the application, geographic phenomena (or structures) are abstracted. In one application census tracts comprise the structure, whereas vegetation cover patterns might comprise the structure for another application. In addition different aspects of that structure are captured -- points, line segments, and areas. These elements may be described by coordinate or name of place identifiers. In order to encode the geometric structure of points, lines and areas fully, the relationship and incidence among the elements -- points, line segments, and areas -- must be encoded, and the point locations making up these lines and areas must be specified in terms of coordinates.

Essentially, geocoding consists of choosing a geographic unit, and encoding and identifying the geographic units. Geography units are functional -- political, administrative, economic, statistical, natural etc. -- or arbitrary units based on a coordinate system. These units can be encoded as an intuitive representation of "place" or represented in a geometric, coordinate or topological framework.

In urban geocoding, the current emphasis is on capturing the structure of the street network -- the coordinate and name descriptions of the points, lines, and areas comprising the network. Such a machine-readable network comprises a powerful tool for manipulating and relating data describing urban phenomena.

In sum, a geocode can be considered a nominal, ordinal, or cardinal spatial index code describing uniquely identified points, lines and areas. Nominal indexes, such as city names, street names, and building names, and ordinal indexes, such as postal zip codes, census area codes and numbered street names, and cardinal indexes, such as coordinate systems, provide increasing powerful systems for location identification.

Types of Application Leading to Developments in Geocoding

Computer technology has been applied to a variety of problems that have lead to developments in geocoding. Initially, these problems have been pursued independently and separate systems have been developed for: 1) converting areal unit codes from one scheme to another, 2) computer mapping, and 3) allocation/location problems. Subsequently, the commonalities of these applications have been recognized and geocoding systems are being developed which are capable of handling all these functions.

The geographic code conversion problem was initially dealt with by constructing cross-reference directories which enabled coverting from one areal unit code to another, such as census blocks to traffic zones or school enrollment zones to facilitate aggregation of data for analysis of facilities and service areas. A second type of conversion problem is to translate commonly used locational identifiers, such as street address, to areal unit codes or coordinates that are more easily manipulated for spatial aggregation and analysis.

Computer cartography developments explicitly require geocoded data. Thematic cartography requires cartographic symbols to be positioned according to the relative location of the places being mapped. Thematic cartography enables analysis of patterns and correlation of data in a spatial context and requires geocoded data to be employed for spatial classification. Geocoding requirements for thematic cartography are substantially different than geocoding requirements for automatic map compilation. Both applications, however, have geocoding requirements. Developments in both areas have been important in developing geocoding systems.

The third type of applications that have lead to development of geocoding technology are allocation/location problems. These problems are typified by a spatial representation of demand related by time/distance to spatially specific points or areas of supply. Legislative redistricting, assignment of people to service facilities, such as schools and health care centers, are examples of these kinds of applications which require geocoded data.

As will be discussed in a subsequent section the pursuit of these applications independently has evolved into efforts to develop a generalized geographic base file (GBF) that can handle all these types of applications.

Geocoding Systems

Schumacker[2] distinguishes between geo-defining systems and geocoding systems. Geo-defining systems are used for precise geometric definition of boundaries of functional areas, whereas geocoding systems deal with "codes for places that are not geometrically defined." However, use of the term geocoding has come to encompass what Schumacker refers to as geo-defining systems, because current efforts strive for both capabilities within the same system. Strictly, geocoding consists of assigning geocodes to records in a data file (without spatial definition of places)[3], but often times the Geographic Base File (GBF), which serves as a directory for the translation of "name of place" to geocodes, is also geo-defined, i.e., contains geometric definition of spatial structure of the areas as well as geocodes for the areas.

Geocoding systems (as differentiated from geo-defining systems) utilize machine-readable directories or indexes such as MEDList, ACG, and ANSI Place Code[4] to convert name of place to geocodes or from one geocode to another. These directories do not geometrically determine the places, but merely provide a correspondence table of different codes that are used for that place. Coordinates of places can be added to these indexes where applications allow consideration of the places as points. In this way the basic directory can be made more useful for determining spatial relationships, such as distance and orientation of data observations.

An additional refinement to geocoding systems has been the geographic definition of a place by defining perimeters as polygons and the geographic definition of street systems as graphs and metric networks. In addition to serving geocoding needs, these spatially augmented indexes or geographic base files are in effect machine-readable maps which can be manipulated and related to thematic data. Geographic base files that provide for both geocoding and geo-defining will likely become more prevalent, because geocoded socio-economic data constitutes the demand side for services and facilities and can be related to the geo-defined data that describe the location and/or service or jurisdiction areas for facilities that supply those needs. Travel demand is served by a spatially defined network, and primary education is served by spatially defined facilities and service areas. Management, planning, and research are all served by geographic base files that enable the bringing of demand and supply data together in a spatial context.

Urban level geocoding developments are presently concentrating on operationalizing street network geographic base files and exploring the feasibility of parcel level geo-defined areas.[5] At the national level, geocoding is dealing with compatible geocodes and conversion files for places. Compatibility problems between different geocoding systems developed for different uses pose considerable problems in proposing and adopting a single system or system of conversion files at the national level. (A companion review article on National Geocoding describes these efforts).

Development of land resource information systems at the statewide or regional level has provided considerable impetus for developing geo-defining systems. Early systems coded information to fairly large grid cells; whereas more recently, extremely small cells are used to capture patterns from images, or patterns are encoded as polygons. A recent IGU Symposium on Geographic Data Sensing and Handling[6] dealt with the capture, processing, and display of data from images. This more general problem of capturing data and spatial relationships from imagery emphasizes the importance of a framework for the capture or encoding of spatial data.

Encoding Spatial Data

Development of geocoding systems requires a conceptual framework for encoding geographic data, i.e., points, lines, and areas, and then evaluating alternative ways in which the encoded geographic data can be input, stored, retrieved, and output. For example, data can be abstracted from maps in terms of coordinates for points, and areas encoded as a sequence of points. These data might be stored as coordinate values for points making up areas, and the data might be displayed as line segments making up a system of areas. Clearly, the input, storage, and output encodings are not independent. Translation from one stage to another must be thought out in advance.

There are a number of alternative ways of encoding spatial data.[7] Depending upon ultimate needs, the storage media, and file structure environment, not all possible encoding schemes need be utilized. However, more than one is usually needed to provide a redundant coding for editing purposes to detect errors for quality control and completeness. Some encodings can also be generated from others which provides the means for edit. Each encoding procedure has a variety of advantages and disadvantages when viewed in light of the storage, comparison, retrieval, and output requirements of a geocoding system.

The choice of an encoding method must be made considering: 1) usefulness in terms of purpose, 2) ease of data capturing or encoding, 3) ability to generate other encodings for error detection, 4) ambiguities, caused by non-unique identification, 5) ease of synchronizing graphic data to descriptive data at input, and 6) storage media and file structure available. When comparing these criteria with the possible encodings, some encodings are more appropriate than others for typical use of image data in machine-readable form. For example, areas encoded as polygons are highly useful as is, can be used to generate other encodings, or can be encoded with ease. Conversely, areas encoded as being contiguous to other areas are less useful and may not uniquely identify an area when two separate areas are wholly contained within the same larger area. Also, point connectivity matrices are difficult to encode directly, although the encoding can be generated from line segment encodings. For error detection, it is important to have two independent encodings, one of which can generate the other for comparison.

A conceptual framework for encoding spatial data is essential to evaluate urban geocoding and to identify research needs. This first section provides a limited perspective from which urban geocoding developments can be viewed.

Review of Urban Geocoding Developments

Cooke describes development in geographic base files and geocoding as evolving from largely manual assignment of location codes to fully automated storage, retrieval and processing of geographic data at the individual parcel level.[8] He identifies four generations of development as follows:

1. 1961-1964, the first major geocoding systems: Automatic Location Table (AULT) and Street Address Conversion Systems (SACS),[9]

2. 1966-1969, the first nationwide systems: Address Coding Guides (ACG),[10]

3. 1967-1971, the second nationwide system: Dual Independent Map Encoding (DIME),[11]

4. 1968-?, future systems: parcel files.

First Generation Files

The first generation files were in response to needs in transportation planning to code large files of origin-destination data to traffic zones. These early directories for translating street addresses to traffic zones quickly gave way to more ambitious efforts to relate data to smaller areas, i.e., city blocks and to geographically define the areas.

> "The Tri-State Transportation Commission (New Jersey, New York, and Connecticut) developed AULT (Automatic Location Table) to help reduce vast amounts of travel survey and land use data from the New York City metropolitan area . . . Tri-State personnel measured block corner coordinates for most city blocks within 75 miles of New York City. They digitized street features in New York City with accuracy sufficient for calculation of street and block areas.
>
> Tri-State and the Paul Rosenberg Associates consulting firm designed and built the AULT system specifically for the Commission's requirements. The system consisted of a back or front lighted, vertically mounted digitizer, with joy-stick cursor controls and a keyboard for entry of alpha-numeric data. The operator fed digitizer output

directly into the AULT plotter to verify his work immediately."[12]

Edgar M. Horwood has directed the implementation of SACS (Street Address Conversion System) at the Urban Data Center, the University of Washington. Mr. Robert Dial was responsible for initial conceptualization of the system and performed the initial programming. The SACS Geographic Base File consists of street segment records containing ranges of street addresses and end-point coordinates. The system contained address matching routines which would augment data records with a coordinate pair corresponding to the address. One could then retrieve data geographically using point-in-polygon techniques. In addition, the system was designed to generate street network maps for purposes of display of data and for editing the geographic base file visually.

The dependence upon point-in-polygon retrieval is both the strength and the weakness of SACS. Point-in-polygon is completely flexible so that one is not tied to a fixed definition of districts, a drawback of systems which code data to census tracts only. However, one needs a digitizer to define retrieval polygons in any quantity, and retrieval algorithms are time consuming compared to testing standard areal unit codes.

Both the digitizing efforts at Tri-State and the development of SACS have proved to be significant in initiating urban geocoding, which has been greatly extended by the U.S. Bureau of the Census.

Geocoding the 1970 Census

The U.S. Bureau of the Census enumerated urban residents (approximately 60 percent of the households in the country) by mail questionnaires. Before the Bureau generated mailing labels they had the addresses coded to 1970 tract and block using Address Coding Guides (ACG's) for 147 SMSA's (Standard Metropolitan Statistical Areas). The unit record of an ACG described a block-face or side of a block. It related a census block number to a range of addresses along one side of the street.

The Bureau relied to a great extent on local assistance in developing the ACG's on the assumption that ACG's could be used for coding local data to census blocks. They also implied the possibility of ACG's augmented by coordinates.

In late 1966, the New Haven (Connecticut) Census Use Study was undertaken. The Census Use Study proposed to work in five areas: computer mapping, address matching, special user-defined tabulations of census data, analysis of merging census data with health and transportation surveys, and analysis of data used by local agencies. A dress-rehearsal pre-test of 1970 census techniques was also conducted in the New Haven SMSA in 1967, which served as the first full-scale test of the Address Coding Guide. The Use Study's research in computer mapping identified problems in adding grid coordinates to the Address Coding Guides.

An attempt to metrically describe the ACG required adding block-face endpoint coordinates to each ACG block-face. With eight block-face endpoints at a typical intersection, each intersection required digitizing as many as eight times. These attempts to map the street network illustrated that the logical structure of the ACG -- block-face records -- were not amenable for computer mapping. Also, ACG's proved difficult to edit and detect errors without considerable clerical assistance.

The Development of DIME

A new technique was developed for generating geographic base files at the New Haven Census Use Study. The technique -- called DIME (Dual Independent Map Encoding) -- was based on the street segment as the basic record (See Figure 1) and treats an urban street network as a mathematical linear graph.[13]

A linear graph -- a network made up of points, lines and areas -- can be described by any of three incidence matrices which define the relationships between 1) lines and points, 2) lines and areas, or 3) points and areas. Whereas the SACS system recorded information that can generate a lines and points matrix, the DIME system went one step further and recorded data that enables generating two of the matrices. This redundant (or dual) encoding has a number of advantages. The powerful aspect of the DIME approach is that computer editing can be employed to eliminate coding errors in the relationship between the point (or node) numbers and the block codes. The edit attempts to 'bound' a block by linking its boundary segments together on the node numbers. Any failure to bound a block results in a message that identifies coding errors.

The block chaining edit serves as a check on the accuracy of the census tract code, since the records are sorted prior to computer processing. If the nodes chain and the first "from" node is the same as the last "to" node, the block is considered topologically correct. If any segments remain or if the block cannot be chained, the block records are rejected as potential errors. If the node numbers or block numbers are reversed, the block would not be chained properly and would be rejected. A node chaining edit can also be employed which chains blocks around the node. If the blocks chain and the first left block is the same as the last right block, the node is considered to be topologically correct. If any segment remains or if the node cannot be chained, the node records are rejected as potential errors.[14]

The DIME file technology ultimately was implemented nationally, after further experimentation in New Haven. The U.S. Department of Housing and Urban Development authorized 701 funds for DIME file construction in 79 cities not covered by ACG's and for conversion of ACG's to DIME files in other cities. During 1970 local agencies again launched efforts to make new DIME files and 'add DIME features to the ACG's'. In late 1970, the completed fieldwork was returned to the Census Bureau for the long series of edits, corrections, and digitizing.[15]

DIME's primal network consists of street segment links on which traffic flows and the dual network interprets those links as city block boundaries. Consequently, DIME is both a flow network and a boundary network. This is the essence of its power, for editing, computer mapping, and allocation/location analysis.

The Present Generation

DIME technology has provided geographic base files in metropolitan areas that offer considerable potential for urban administrators, planners, and researchers. However, there are several unresolved problems with respect to their utility. The immediate program is one of updating the geographic base files which represent the street networks in 1969. As changes occur in our urban areas, these files are becoming obsolete. Although the Census Bureau has an interest in updating the files for use in the next census, it does not have the resources or funds to take over all maintenance activities.[16] On the other hand, local users of the files often have best access to information on changes, but they lack experience and funding to carry out a maintenance program. The second problem as identified by Cooke is a combination of lack of software and user experience in geographical data processing, overselling of DIME capabilities, and bad user experiences with ACG's.[17] User-oriented software packages to make use of geographic base files are limited. ADMATCH (Address Matching) for converting street addresses to census tract block codes and coordinates for computer mapping have been developed through the efforts of the Census Use Study. These packages are not generally operational in most metropolitan areas, however, and efforts to do so meet with considerable frustration for local users. Geographic base files have considerable potential for merging census and local data, computer mapping, generating tabulations of user data (including survey data) to census areas, calculation of areas of zones and density statistics, and redistricting of schools and routing of vehicles. These applications, however, require software packages that are not generally available or operational. The utility or support program environment for GBF's are not readily available, and _ad hoc_ efforts are leading to confusion and incompatibilities.

The user environment for geographic base files is being improved, however , and the U.S. Bureau of the Census is providing significant leadership in coordinating applications of DIME as the geographic base file for urban information system development. DIME is proving to be an effective mechanism for data linkage, display, and analysis for local government planners and decision-makers. The DIME Workshops are providing their coordinating initiative[18] and considerable effort is underway to make DIME a nationwide geographic base file on a continuing basis.[19]

The efforts of USAC to develop Integrated Municipal Information Systems hold promise of providing geocoding procedures and software.[20] On one hand, USAC can be viewed as providing geocoding capabilities within a broad perspective of municipal information systems for urban administration, which emphasizes the use of geocoding to deal with urban problems rather than geocoding per se. On the other hand, it can be argued that the USAC Integrated Municipal Information Systems Project has extremely broad and ambitious goals that will not enable enough emphasis on geocoding to realize a nationwide and sophisticated urban geocoding system.

Urban geocoding systems still are not operable in most areas, and considerably more effort and resources are necessary. The development or urban geocoding systems is severely constrained by lack of concerted research and development, poor system maintenance opportunities, lack of guidelines and standards, and management and direction which is largely aspatial in orientation.

In spite of Census Use Study activities and USAC, there are no federal programs that explicitly require the use of geographic base files and directly support their development for planning or management activities. Until use of GBF's is required by federal agencies, it is doubtful that universal adoption will take place. Many federal agencies seemingly have a stake in seeing that up-to-date geographic base files exist in metropolitan areas. The Department of Justice, in terms of low enforcement, is concerned about the spatial distribution criminal activity. Similarly, the Department of Housing and Urban Development is concerned with the spatial distribution of housing changes; the Department of Health, Education and Welfare supports many programs in metropolitan areas where the spatial distribution of students, welfare recipients, and persons needing health services is of importance. The Department of Transportation continues to have need for small area data primarily by traffic zones for the continuing phase of urban transportation planning. Jointly or through the Bureau of Census these agencies have a stake in funding consistent updating of geographic base files and development of support programs for their use.

In an analogous situation, the Bureau of Public Roads (now Federal Highway Administration) developed a package of transportation planning programs for use by the various area transportation studies in the 1960's and funded the planning process. This battery of programs proved indispensable in conducting area transportation studies called for by the 1962 Highway Act. A similar federal effort is warranted to maintain geographic base files and to develop the support programs for the use of the geographic base files. Similarly, federal agencies may in the future require metropolitan planning agencies to perform certain functions that require the use of geographic base files. For example, HEW and HUD could well require vital statistics and building permits to be geocoded to census tracts. Federal requirements for geocoded data will probably be the biggest incentive to developing a usable nationwide set of urban geographic base files.

The existence of geographic base files is necessary for and supportive of more advanced technological developments, for example, interactive graphics to handle complex problems, such as service area delineation, network analysis and planning, and other forms of spatial analysis. The street segments of a GBF become the basic element to which facility data, e.g., street width, pavement type and condition, accident data, utility characteristics, and size and type of shopping, school, and job opportunities, can be related. Similarly, demand data, e.g., volumes can be assigned to these same street segments to determine adequacy in a spatial context. Several such applications of GBF's are reported in Geocoding - 71[21] and Geocoding -72[22]. These applications are: measuring accessibility, transit planning, refuse collection routing, student to school assignment, computer mapping, traffic engineering. More generally, GBF's provide a basis for analysis of space-serving facilities.

Future Potential

Largely, the future will bring future attempts to operationalize the many potential applications of geocoding systems that were mentioned in the prior section. Since very few of these efforts can be considered widely operational at this time considerable effort will and should be extended in these directions.

The next generation of urban geographic base files will probably be more detailed than the existing street segment network. DIME editing features are also applicable to generating files that describe individual parcels of land and street rights-of-way. The basic DIME procedures of node numbering, segment coding, editing, and digitizing are valid for parcel-level files. The essential difference between a street segment geographic base file and a parcel boundary file is in size. Where a city of 200,000 may have 6,000 street segments, it is likely to have 50,000 - 100,000 parcel boundary segments. The problems of creating and processing files of this size are considerably different, e.g., line follower digitizing might be more efficient than manual encoding of parcel segments. Potentially, it would seem possible to produce a machine-readable detailed land use map where users could specify a proposed freeway right-of-way and generate lists of parcels affected and property values for alternative rights-of-way. Ultimately, computer mapping of parcel-level geographic base files could replace many manual mapping applications in metropolitan areas. The geographic base file could be used to create maps at various scales for various sections of the metropolitan area with user selection of the data and background detail to be mapped.

This transition from a street segment network to a parcel boundary system is a significant shift, in that the coordinate and scale aspects must be translated from one of insuring positional uniqueness of street segment nodes (with an accuracy of plus or minus 50 feet) to a coordinate system with sufficient accuracy to integrate with land survey data. This kind of accuracy is not possible given the present methods of coordinate location. The scale of the maps from which coordinates are taken and the method of placement of nodes on these maps places severe limitations on accuracy. The typical scale for the Metropolitan Map Series from which the Bureau of the Census coordinates are derived was 1" - 800' (1:9600), and those maps were created for use in the address coding guide operation where digitizing was not a prime consideration. In any event, it is recognized that the coordinates established by these methods are too inaccurate for engineering and legal purposes, and suitable only for planning and spatial analysis. Consequently, parcel-level files will require complete redigitizing. Even then they will be inadequate for legal descriptions. A two-tier system will be necessary where the geographic base file might be used to index more specific locational coordinates of parcels for legal needs.

Additional applications of urban GBF's are emerging. One is to attach elevation as an attribute of nodes which enables three-dimensional analysis of link and node data and facilitates integrating utility line data into GBF's. Another application is on-line street address conversion for real-time needs of dispatching police, fire and dial-a-ride vehicles. This application requires storage of the GBF in direct access mode in lieu of the more conventional sequential storage that is used for most research, planning, and management applications.

Research Areas

Geocoding is at a stage where the operationalization problems are paramount. However, there are several areas in which new research has begun. Geocoding research and development is moving to greater detail, on one hand, and toward greater generality or larger grain systems on the other. These are exemplified by: 1) composite networks, e.g., parcel-level geographic base files and 2) application of geocoding to land resource information systems. Both cases require an overlay capability (or polygon intersection) to compare metric based networks. Efficient comparison of separate networks is an important research area that is emerging.

Composite Networks

For street address translation, geographic base files need only contain nodes and edges comprising the street system. Urban geographic base files also include non-street features such as rivers, railroads, and major jurisdictional boundaries. In general, a composite network exists when several kinds of features are encoded as nodes and edges. As long as each area that the composite network creates is uniquely identified, the DIME edits can be employed.

For example, Becker and Hayes created a composite network of census enumeration districts and precinct boundaries in their work in redistricting California.[23] DIME edits were employed to purge the composite network of errors. This is an example of a more general utility of encoding patterns, jurisdictions, or nets as graphs.

Spatial features, such as patterns (soil type, vegetation cover type), jurisdictions (counties, city boundaries, townships), and nets (roads, rail lines, streams), are amenable to encoding as individual metric based graphs or as a composite network.

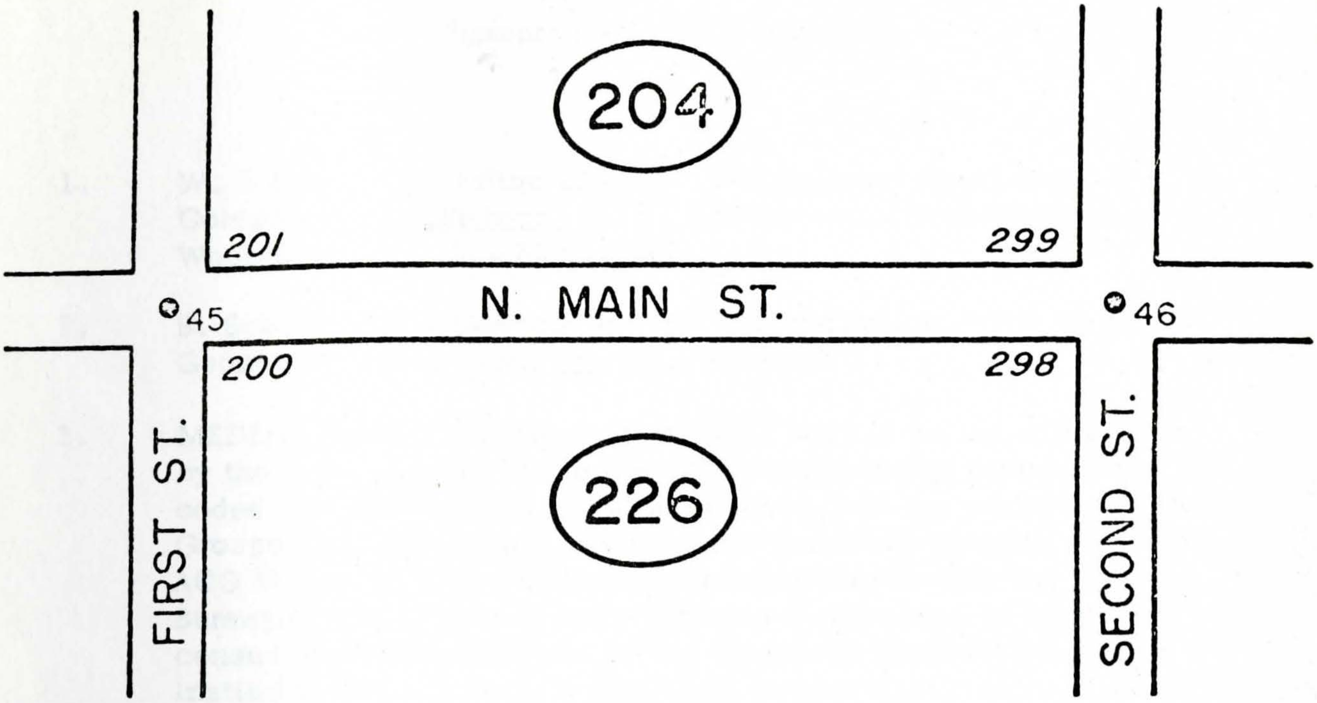
Land Resource Information Systems

The composite network concepts have considerable application at the statewide or regional level where it is desirable to encode many different kinds of features into a single network. A composite network might contain highways, rail lines, streams, minor civil division boundaries, etc. This composite network then could be compared to patterns of vegetation cover, land ownership, land use, and soil characteristics. These patterns might be represented as extremely small grid cell values or as polygons. If the patterns are complex, small grid cells are preferable. Nevertheless, these kinds of overlays or phenomena could be related to the composite networks representing the minor civil divisions, analysis areas, and linear features.

Considerable research activity on problems and systems for handling geographic data is occurring. This research concentrates an analysis of hardware and software for input, processing, retrieval, and display of geographic data. Consequently, this research is linked to remote sensing, pattern recognition, data encoding and storage, and computer mapping. A literature in this field is beginning to emerge.[24]

Closing Note

Urban geocoding has progressed rapidly within a relatively short amount of time. This advancement, however, has not occurred without difficulty as terminology and concepts are largely following rather than leading applications. Experience had indicated that dual encoding of geographic phenomena is advantageous so as to employ logical edits. However, considerable manual input is necessary. Although DIME-like methods are suitable for more detailed parcel level geographic data and for land resource inventory efforts, research may show that different encoding methods are more suitable and more amenable to machine-assistance. The discipline of geography has not provided persons in applied geocoding with a well-structured conceptual basis. The applications have both directed and forced the development of geocoding concepts. The extension of geocoding concepts and applications will probably continue to be a cooperative effort between geographers and others concerned with applications.



STREET NAME	MAIN
PREFIX DIRECTION	N
STREET TYPE	ST
FROM NODE	45
TO NODE	46
LEFT LOW ADDRESS	201
LEFT HIGH ADDRESS	299
RIGHT LOW ADDRESS	200
RIGHT HIGH ADDRESS	298
LEFT BLOCK	204
RIGHT BLOCK	226

Figure 1. The Basic DIME Record Structure

Source: System Development Corporation, A Geographic Base File for Urban Data Systems, Santa Monica, May, 1969.

Footnotes - Urban Geocoding

1. W. Tobler, "Geocoding Theory," The National Geocoding Conference <u>Proceedings,</u> U. S. Department of Transportation, Washington, D. C., May, 1972.

2. B. Schumacker, "Geo-coding and Geo-definition," The National Geocoding Conference, <u>op. cit.</u>, footnote 1.

3. MEDList (Master Enumeration District List) is an index prepared by the U. S. Bureau of the Census containing the geographic codes and place names for which census data are tabulated. Block Groups and Enumeration Districts are the smallest units on the MEDList. ACG (Address Coding Guide) is an index prepared by the U. S. Bureau of the Census to code individual addresses to specific census area for tabulation. ANSI (American National Standards Institute) Place Code is a fine-digit number assigned in alphabetic sequence to each named place with a state. See P. Werner, <u>A Survey of National Geo-Coding Systems,</u> U. S. Department of Transportation, Washington, D. C., February, 1972, for a description of those and other geocoding systems.

5. The Urban and Regional Information Systems Association (URISA), through a Special Interest Group on Geographic Base Files, reports activities in urban geocoding and has published two volumes, <u>Geocoding - 71,</u> Papers from the Information Systems Conference, September, 1971 (Available from NTIS, PB 211-743, micro fiche only) and <u>Geocoding - 72,</u> Papers presented in Geographic Base File SIB (Special Interest Group) Session, 1972 URISA Conference, September, 1972 (Available from NTIS, PB 211-744, micro fiche only).

6. R. F. Tomlinson (ed.), <u>Geographic Data Handling</u>, Symposium Edition, A publication of the International Geographical Union Commission on Geographical Data Sensing and Processing for the UNESCO/IGU Second Symposium on Geographical Information Systems, Ottawa, August, 1972, 2 volumes.

7. K. J. Dueker, "A Framework for Encoding Spatial Data," <u>Geographical Analysis,</u> January, 1972.

8. D. F. Cooke, "Geocoding and Geographic Base Files: The First Four Generations," Presented at the American Institute of Planners "Confer-in West," San Francisco, California, October 26, 1971, (Available from Urban Data Processing, Inc., 675 Massachusetts Avenue, Cambridge, Massachusetts, 02139).

9. R. C. Barraclough and P. Rosenberg, "The Ault System," Photogrammetric Engineering, September, 1966; R. B. Dial, "Street Address Conversion System," Urban Data Center, University of Washington, Seattle, 1964, (Available from NTIS, PB 197-348).

10. W. T. Fay, "The Geography of the 1970 Census: A Cooperation Effort," Planning 1966, American Society of Planning Officials, Chicago, 1966.

11. D. F. Cooke and W. H. Maxfield, "The Development of a Geographic Base File and Its Uses for Mapping," Urban and Regional Information Systems for Social Programs, Urban and Regional Information Systems (URISA) Conference Proceedings, Kent State University, 1967; U. S. Bureau of the Census, Census Use Study: The DIME Geocoding System, Report No. 4, Washington, D. C., 1970; J. Corbett and G. Farnsworth, "Theoretical Basis of Dual Independent Map Encoding," In Geocoding-71, op. cit., footnote 5, (reprinted in U. S. Bureau of the Census, Census Use Study: The First International DIME Colloquium: Conference Proceedings, Washington, D. C., 1973).

12. Cooke, op. cit., footnote 8, pp. 1-2.

13. U. S. Bureau of the Census, op. cit., footnote 11.

14. Ibid., pp. 27-29.

15. Cooke, op. cit., footnote 8, p. 6.

16. U. S. Bureau of the Census, Southern California Regional Information Study: ACG/DIME Updating System: The Long Beach California Experience, Report No. 8, Washington, D. C., 1971.

17. Cooke, op. cit., footnote 8.

18. U. S. Bureau of the Census, Census Use Study: DIME Workshop: An Interim Report, Washington, D. C., 1973.

19. U. S. Bureau of the Census, Geographic Base File System -- Establishing A Continuing Program, Report GE 60 No. 4, Washington, D. C., 1973.

20. K. L. Kraemer, "USAC: An Evolving Mechanism for Urban Information Systems Development," In: J. E. Rickert and S. L. Hale (eds.), Urban and Regional Information Systems: Past, Present and Future, Papers from the 1970 URISA Conference, Center for Urban Regionalism, Kent State University, Kent, Ohio, 1970, pp. 66-79. See various papers in J. E. Rickert, et al., (eds.), Urban and Regional Information Systems: Information Systems and Political Systems, Papers from the 1971 URISA Conference, Stockton State College, Pomona, New Jersey, 1972.

21. Geocoding - 71, op. cit., footnote 5.

22. Geocoding - 72, op. cit., footnote 5.

23. "Becker and Hayes Geo-Base Coding," Becker and Hayes, Inc., 10835 Santa Monica Blvd., Los Angeles, California, 90025, September, 1971.

24. R. F. Tomlinson, (ed.), Geographic Data Handling, Symposium Edition, A publication of the International Geographical Union Commission on Geographical Data Sensing and Processing for the UNESCO/IGU Second Symposium on Geographical Information Systems, Ottawa, August, 1971; C. Runge, et al., (eds.), Symposium: A Survey of Programs for Statewide Land Resource Inventories, Working Paper 8D, The Institute for Environmental Studies, University of Wisconsin - Madison, December, 1972; and K. J. Dueker and J. S. Drake, "Information Systems for Land Resource Planning," Technical Report No. 16, Institute of Urban and Regional Research, University of Iowa, Iowa City, Iowa, January, 1973.

25. System Development Corporation, A Geographic Base File for Urban Data Systems, Santa Monica, May, 1969.