Ottawa Street Address Conversion System

HUGH W. CALKINS, Research Planner, National Capital Commission, Ottawa

•THE Ottawa Street Address Conversion System is a geocoding system being developed for the Ottawa-Hull Metropolitan Area in Ontario. This report describing the system will be presented in two parts. The first part contains some general remarks relating to the function of a geocoding system and the second part describes the current project in Ottawa, a geocoding system based on street addresses.

GEOCODING SYSTEMS

Geocoding is the process of assigning a spatially significant identifier to a data observation. A spatially significant identifier is defined as one from which the relative spatial location of a data observation is evident from the value of the identifier. For example, if two properties are identified by parcel numbers (No. 202-013-001 and No. 202-016-024) or street addresses (1091 3rd Ave. and 820 Main St.), these identifications do not describe the relative spatial locations of the parcels. However, if the two parcels are identified in terms of x-y coordinate locations (2510, 4400 and 2730, 3782), the values of the identifier represent relative spatial locations. Coordinates are the most flexible spatial identification system. Coordinates, however, do not replace nonspatial forms of data identification, but rather, the coordinates are a generalization which represent all spatial reference units. Coordinates can replace all geographic coding for areal units, i.e., traffic zones, census tracts, etc. Computer techniques can then be used to assemble the data by any set of areal units using the coordinate identifiers of the data observations.

For example, the first use of the Ottawa street address geocoding system will be to determine coordinates for each household in the Ottawa Metropolitan Area as enumerated in the 1966 census. These coordinates will be placed on the household data files maintained by the census department. Population characteristics can then be summarized and reported by any arbitrary set of areal units, i.e., traffic zones, planning districts, etc.

A geocoding system is only one part of an overall information system capability. The major steps in an information system are (a) collecting data; (b) organizing data; (c) storing data; and (d) retrieving data. A geocoding system is part of the process of organizing the data and it establishes the identifiers needed to enable maximum flexibility for retrieval on spatial criteria.

In order for data with spatial significance to have the capability of providing meaningful information for a specified problem, it must have an identifier which is spatially significant. That is, all data must be referenced to a point in space or areal unit. In the past, data have usually been referenced to a set of more or less single purpose areal units, e.g., census tracts, and traffic zones. Two major flaws to this system have long been recognized.

1. All anticipated sets of areal units must be specified before the time the data are collected, a severe restraint at the beginning of the study; and
2. Data collected for different sets of areal units are often non-comparable.

Data referenced to the larger areal units are in fact summarized to the level of that areal unit, and retrieval on spatial criteria is limited to the set of areal units as defined before the data were collected. As long as data are processed by manual methods,

Paper sponsored by Committee on Urban Measurements and presented at the 46th Annual Meeting.

it is virtually impossible to break away from this system due to limitations in manpower resources. However, the speed offered by computers overcomes this restriction.

Spatial units can be generalized as follows: coordinate region, query regions, and uniform data regions. The coordinate region is the area represented by a single coordinate point on a grid system and has a size and shape equal to a square whose sides are equal in length to the distance between points on the grid system. The query region is an arbitrarily delineated areal unit, e.g., traffic zone or census tract. Query regions are represented as polygons, each vertex of the polygon being represented by a coordinate region. Uniform data regions indicate the areal extent of some data characteristic, and are described in the same manner as query regions. The common point of reference to all three regions is the coordinate region.

A coordinate system is therefore used to establish and retain flexibility in the definition of query regions and in the description of uniform data regions. Individual data observations must be identified by unique coordinates. These coordinates should be based on a standard grid system such as the State Plane-Coordinate System.

Having established the need for data to be identified by coordinate values, it only remains to develop the techniques of assigning coordinates to the data observations. The magnitude of this task should not be underestimated because it involves assigning a coordinate value to every individual data observation, e.g., property, household, etc. However, automated techniques are being developed to substantially reduce the time and manpower resources needed to do this job. The street address conversion system being developed in Ottawa is such a technique.

OTTAWA GEOCODING SYSTEM

The system described here is one which derives the coordinates from street addresses. There are three coordinate sets (x, y) assigned to each address: (a) the unique proxy; (b) the midpoint of the block face; and (c) the block centroid (Fig. 1). The unique proxy coordinate value is a unique identification for each address; however, its spatial location may not always be exact. The other two coordinate values represent small areal units for data summarization—the block face and the block.

The sequence of steps in this system is shown in Figure 2. The data observations, identified by street addresses, are converted to machine records (punch cards or magnetic tape). These records are processed by a computer program which assigns any combination of the three possible coordinate sets to each data record. The program

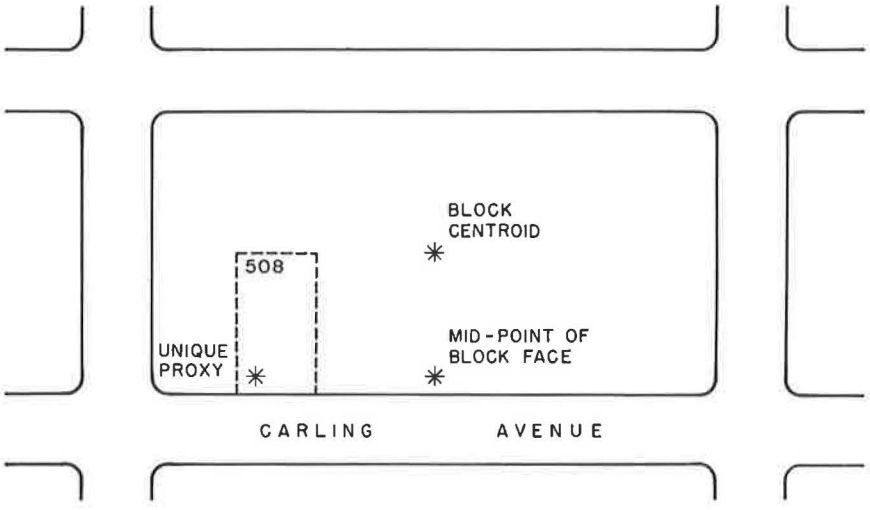

Figure 1. Coordinates.

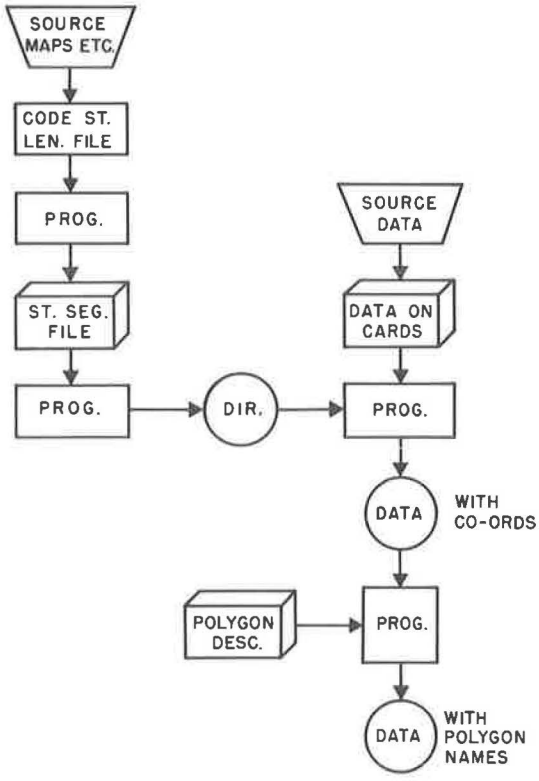

Figure 2. Flow chart.

relies on a directory for the information needed to assign the coordinates. The product of this first step, data with coordinate identifiers, is then processed by another computer program which assigns to each data record an identification code representing the appropriate areal unit (query region). The set of areal units is described as polygons.

The final result is data identified by coordinates and by specified areal units. At this point the data can be summarized, analyzed, and reported on the basis of any of the identifiers, coordinate sets or areal units. The capability to retrieve data by arbitrary polygon sets is the query capability built into the current geocoding system. Although this is far from a general query capability, it obviates the need to code geographic identifications at the time the data are assembled.

In the flow chart of the system (Fig. 2), the directory is shown as one of the major inputs to the step in which coordinate sets are assigned. The building of this directory is a computer-assisted process.

The steps required to build the directory are (a) code the street length file from the source documents; (b) run the computer program to create the street segment file from the street length file; (c) make corrections to the street segment file (manual process); and (d) run the computer program to build the directory from the street segment file.

The basic unit of the directory is the street segment record (Fig. 3). The street segment record is a straight-line segment between the centers of two consecutive intersections or abrupt changes in a street's direction. The following information makes up a street segment record:

1. Street name;
2. Range of street addresses;
3. Coordinate values (x, y) at each end;
4. Block centroids (x, y) on each side; and
5. An indicator for determining which side of the street relates to even and odd addresses.

The address range represents both sides of the street. In order for the system to function properly, the addresses on each side of the street segment must fall within the same range.

The street segment file is prepared from the street length file by a computer program. A street length is an assumed straight length of street over which addresses are always increasing. A street length represents a series of street segments by the following (Fig. 4):

1. Street name;
2. Range of street addresses;
3. Coordinate values (x, y) at each end; and
4. An indicator for determining which side of the street relates to even and odd addresses.

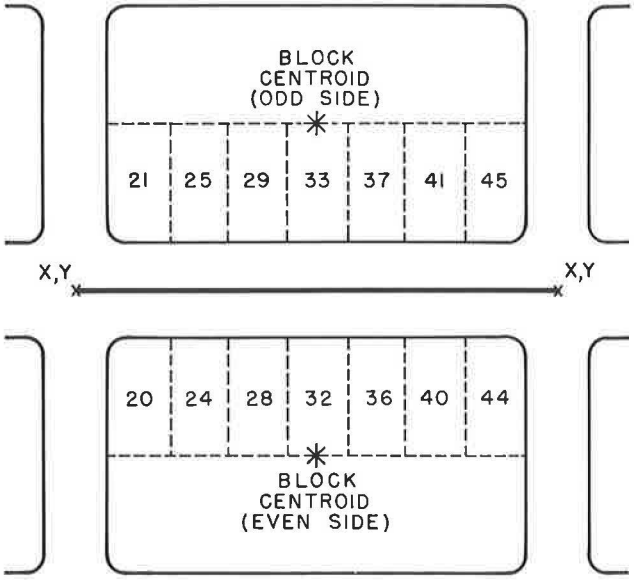

Figure 3. Street segment.

The program divides each street length into its appropriate number of segments and calculates the address range, coordinate values for the end points and the block centroids for each segment. Thus, for any street pattern the initial coding process need only record the end points of all street lengths. The hypothetical street pattern (Fig. 5) would require 19 street length records. Processing these records by the computer would yield 131 street segment records. The street segment file as produced by the computer is not complete. In areas of very regular street patterns, up to 90 percent of the street segment file is created by the computer, the remainder must be done by hand. However, as the street pattern becomes less regular the amount of work done by the computer drops, thereby increasing the manual work.

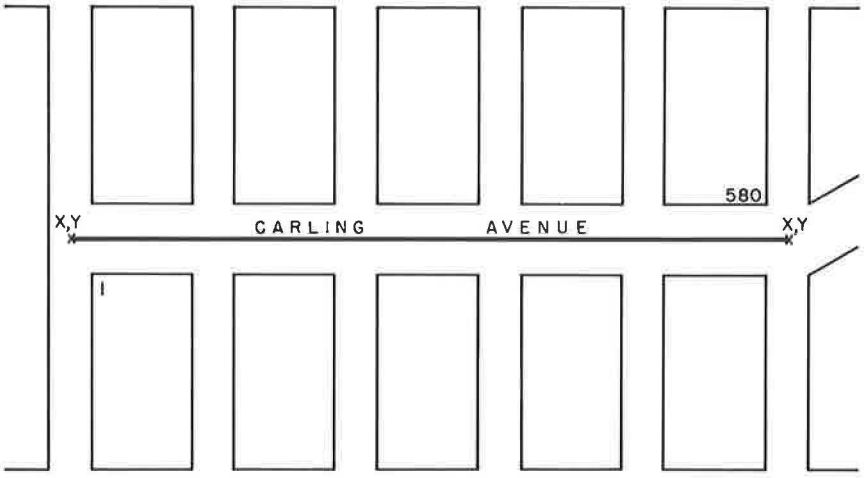

Figure 4. Street length.

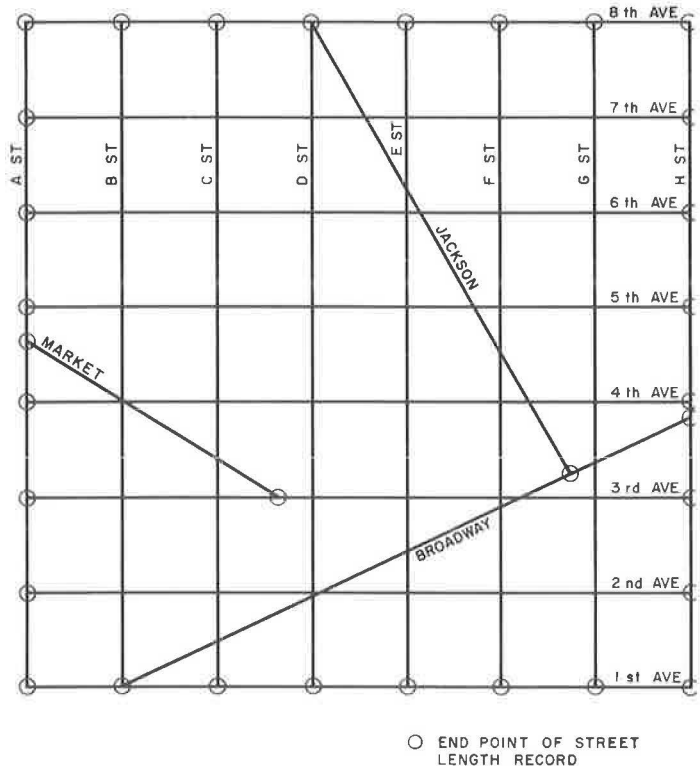

Figure 5. Hypothetical street pattern.

Operational experience in the application of the street address geocoding system to the Ottawa Metropolitan Area disclosed certain problems. The first problem related to the regularity of street addresses.

As was noted earlier, the system requires that the address range for a street segment represent both sides. Addresses must always be increasing by more or less constant increments to successfully meet this requirement. Many cities have adopted a system of assigning a range of 100 addresses to each block, thus producing a very regular address pattern. However, when any system of a less regular nature is used, then it is possible for address ranges to have characteristics which are not compatible with the system. Briefly, some of the situations which can develop are (a) addresses increasing faster for one side of a street than the other; (b) even and odd addresses on the same side of a street segment, such as on a circle or crescent; and (c) addresses out of sequence.

Another major problem which can limit the use of this geocoding system is the areal extent of urban street addresses. If the area of interest is completely urbanized, then this is not a problem. However, if rural areas are to be included, the present geocoding system cannot be used over the entire area. In the Ottawa Metropolitan Area, urban and rural areas are intermixed (Fig. 6) so that the street address geocoding system can only represent a discontinuous area.

The original system has required only slight modifications to deal with the situations described. To solve the problem of irregular patterns of street addressing, new techniques have been developed:

1. Block face records are substituted for street segment records in the directory when one address range cannot represent both sides of the street without creating an ambiguous situation with respect to other street segments of the same street.

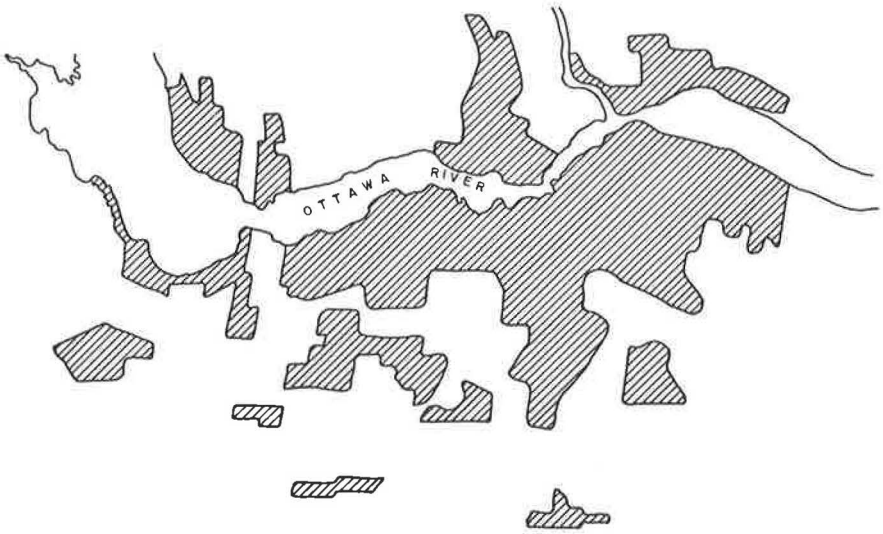

Figure 6. Extent of urban postal addresses.

2. Block face records are used to record segments where both even and odd numbers are on the same side of the street. Generally, in this case only one side of the street is developed.

3. Out-of-sequence addresses are coded as individual entries, the coordinates being derived manually.

Although it is possible to accommodate almost any situation, the amount of manual effort required to build the directory increases substantially as the street addressing pattern becomes less regular.

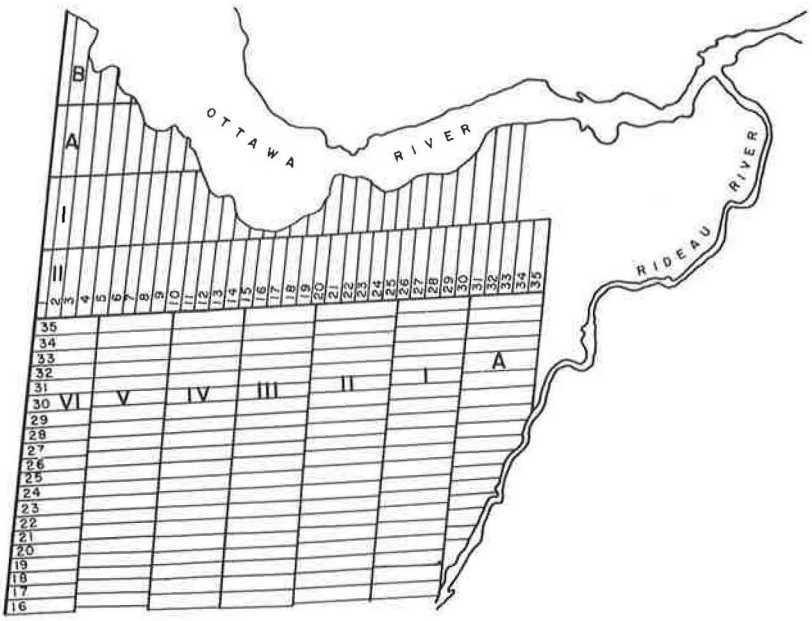

Figure 7. Lots and concessions.

In order to include the rural fringe areas in the geocoding process, the methods of rural addressing and land identification were examined. Three methods were identified: (a) rural mail delivery routes; (b) P.O. Box numbers; and (c) rural land survey (concessions and lots, Fig. 7).

The concession/lot system was found to have spatial characteristics similar to the concept of the street segment. The significant characteristic for the geocoding system is that there is a linear relationship between all lots within one concession. Therefore, if the lot numbers are used in place of the address range and the concession name is substituted for the street name, a pseudo street segment can be created to represent a whole concession. If this pseudo segment is located through the middle of the concession, each lot can be split into two pieces (north half, south half or east half, west half) for geocoding, each piece being 100 acres in size.

The changes introduced to the system have broadened its scope and increased its potential application. For example, the Canadian Dominion Bureau of Statistics is currently investigating geocoding applications through participation with the National Capital Commission in developing the Ottawa street address system. The Dominion Bureau of Statistics (DBS) plans to use geocoding techniques for census data as well as other nationally collected statistics, thereby giving DBS the capability to summarize any of their data to any set of areal units as a service to users. The ability to automatically geocode data on the basis of various identifiers substantially increases usefulness of the data in all studies relating to urban areas.

REFERENCES

1. Dial, Robert B. Street Address Conversion System. Res. Rept. No. 1, Urban Data Center, Univ. of Washington, 1964.
2. Calkins, Hugh W. Operations Manual for Street Address Conversion System. Res. Rept. No. 2, Urban Data Center, Univ. of Washington, 1965.